



Evasión tributaria: Análisis de redes

Centro de Ciencias de la Complejidad (C3)

Instituto de Física (IF)

Facultad de Contaduría y Administración (FCA)

Universidad Nacional Autónoma de México (UNAM)

en colaboración con:

Department of Network and Data Science (DNDS)

Central European University (CEU)

Índice

1. Objetivo	3
2. Resumen ejecutivo	3
3. Revisión del marco legal	5
3.1. La Evasión Tributaria en el Impuesto sobre la Renta	5
3.2. Ley del Impuesto sobre la Renta	6
3.2.1. La Ley del Impuesto sobre la Renta, su objeto	6
3.2.2. Sujetos del Impuesto sobre la Renta en México	7
3.2.3. El residente en México y el residente en el extranjero como determinante de la acumulación del ingreso en el Impuesto sobre la Renta	7
3.2.4. Tributación en ISR de las Personas Morales	8
3.2.5. Las pérdidas fiscales en la determinación del Resultado Fiscal	9
3.3. Código Fiscal de la Federación	9
3.3.1. Artículo 58, determinación presuntiva de la utilidad fiscal	10
3.3.2. Artículo 27, apartado A, inciso III y apartado B, incisos V y VI, Registro Federal de Contribuyentes, obligaciones, libro de socios y accionistas y aviso al RFC por cambio de socios y accionistas	10
3.4. Ley Federal para la Prevención e Identificación de Operaciones con Recursos de Procedencia Ilícita (LFPIORPI)	11
3.4.1. Artículo 17, fracción V, prestación habitual o profesional de servicios de construcción y en general transmisión de propiedad y constitución de derechos sobre bienes	11
3.4.2. Prestación de servicios profesionales, artículo 17, fracción XI, incisos a) a e)	12
4. Experiencias internacionales de investigación	12
5. Descripción de datos	13
6. Metodología de investigación	14
6.1. Ciencia de redes	15
6.1.1. Concentración de EFOS en actividades económicas	15
6.1.2. Redes de emisiones de CFDI entre actividades económicas	22
6.1.3. Comparación de montos en emisiones de CFDI	24
6.1.4. Redes bipartitas de RFCA de personas morales y sus asociados	26
6.2. Aprendizaje de máquinas	30
6.2.1. Modelo <i>Gradient Boosting Trees</i> (GBT)	30
6.2.2. Resultados del modelo de identificación de EFOS	31
6.2.3. Modelo <i>Redes Neuronales de Grafos</i> (DGCNN)	34
6.2.4. Resultados del modelo <i>DGNN</i>	35
7. Integración de los métodos	37

8. Cálculos de evasión	39
9. Conclusiones y limitaciones	41
9.1. Recomendaciones	42
9.2. Trabajo a futuro	43
A. Anexo	46
A.1. Los CFDI en la determinación de la base de cálculo del ISR de las Personas Morales	46
A.1.1. El CFDI y los ingresos acumulables de las Personas Morales del Grupo I	48
A.1.2. El CFDI y las deducciones autorizadas	48
A.2. Tipos de Ingresos que pueden percibir las Personas Físicas y las Personas Morales	48
A.2.1. Ingresos totales	49
A.2.2. Ingresos gravables	49
A.2.3. Ingresos acumulables	49
A.2.4. Ingresos no acumulables	50
A.2.5. Ingresos Exentos o no sujetos al pago del ISR	50
A.2.6. Ingresos no objeto	50
A.3. Grupos de Personas Morales	51
A.4. Artículo 58 del CFF	51
A.5. Tasas de utilidad por sector definidas en este estudio	53
A.6. Agregación de información y construcción de enlaces	53
A.7. Distribución de RFCA sospechosos en Sectores y Actividades Económicas	54

1. Objetivo

Describir el comportamiento de las redes entre actores que comercializan Comprobantes Fiscales Digitales por Internet (CFDI) alrededor de las Empresas que Facturan Operaciones Simuladas (EFOS) ya detectadas por el Servicio de Administración Tributaria (SAT), mediante herramientas de teoría de redes y ciencia de datos. Esto con el fin de estimar un presunto monto de evasión potencial del Impuesto sobre la Renta (ISR) y diseñar estrategias de combate a este fenómeno.

2. Resumen ejecutivo

La evasión fiscal, siguiendo la definición de Oldman y Holland, puede presentarse desde el pago extemporáneo de impuestos hasta la evasión fraudulenta y deliberada. En México, este segundo aspecto ha llevado a la comercialización de facturas apócrifas. Con los cambios al Código Fiscal de la Federación (CFF) (artículo 29: nuevas formas de expedición de los comprobantes fiscales) y la introducción del Comprobante Fiscal Digital por Internet (CFDI), los defraudadores fiscales han creado nuevas formas de evasión: las llamadas Empresas que Facturan Operaciones Simuladas (EFOS) y su contraparte, las Empresas que Deducen Operaciones Simuladas (EDOS).

Para detectar estas EFOS, a partir del 1 de enero de 2014 se incorporó en el CFF el artículo 69B, en el que se establecen condicionantes generales y necesarias para la operación de un negocio o desarrollo de una profesión, relacionados con la expedición de los CFDI, que tienen la obligación de emitir, como es el caso de los activos, el personal, la infraestructura y la capacidad material y/o física; o bien, cuando el contribuyente no se encuentre localizado y a falta de ellas, la autoridad fiscal presumirá como inexistentes las operaciones amparadas en ellos. De esta forma inicia el proceso en que la autoridad le dará a conocer al contribuyente, a través del buzón tributario y mediante publicación en el Diario Oficial de la Federación (DOF), para que éste desvirtúe la presunción, manifestando lo que a su derecho convenga y presente la documentación e información respectiva.

Así mismo, considerando el CFF en su artículo 58, dentro de las facultades de las autoridades fiscales, contempla la determinación presuntiva de la utilidad fiscal de los contribuyentes para efectos de la Ley del ISR; por medio de la aplicación de tasas establecidas a los montos totales percibidos en un ejercicio, de acuerdo al tipo de actividad que realizan. Por lo que, considerando dicha metodología y tomando las bases de datos proporcionadas para la investigación, tendremos una estimación de la utilidad fiscal de los contribuyentes, identificados por su Registros Federales de Contribuyentes Anonimizados (RFCA), conforme se establece en dicho artículo, permitiéndonos realizar una estimación de la evasión de ISR partiendo de un conjunto de RFCA sospechosos identificados por los distintos métodos considerados en el estudio.

Sin embargo, reconocemos que los RFCA sospechosos pudieran tener el total de sus operaciones simuladas, lo que nos llevaría a considerar otra forma de determinación de la evasión en ISR directamente de los ingresos declarados, sin dejar de lado la implicación que pudiera tener sobre cualquier otro ingreso no declarado y percibido de forma adicional.

Con los datos que nos fueron proporcionados realizamos distintas exploraciones para la caracterización del comportamiento de EFOS, en específico, hemos realizado una exploración de las variables categóricas asociadas a las EFOS y RFCA, patrones de emisión de CFDI entre actividades económicas y características de la constitución legal de personas morales y sus socios, asociados y representantes legales.

En términos de la metodología, utilizamos técnicas computacionales e interdisciplinarias derivadas de la llamada ciencia de redes. En primer lugar, analizamos las emisiones de CFDI entre sectores de actividad económica. En segundo, estudiamos la conformación legal de personas morales. Las redes bipartitas resultantes (es decir, relaciones de nodos con grupos de nodos) describen a la persona moral y su relación con otras personas morales, representantes legales o socios y asociados. De forma complementaria utilizamos el enfoque de ciencia de datos, a partir de la implementación de dos modelos de aprendizaje de máquinas. En el primero, llamado potenciación del gradiente (GBT, por sus siglas en inglés), se utiliza como insumo la información categórica. En el segundo, el modelo de clasificación llamado Modelo de Redes Neuronales de Grafos, se toma como insumo las redes bipartitas.

Partiendo de la información categórica de RFCA y el listado de EFOS, calculamos la distribución de la presencia de EFOS en los distintos sectores, ramas y actividades. Hemos identificado una sobre-representación de EFOS en los sectores económicos de Comercio al por mayor, Construcción y Minería. Esta sobre-representación sugiere una preferencia de las EFOS de pertenecer a ciertos sectores o actividades que puedan ser favorables para la simulación de operaciones u otras actividades ilícitas.

Posteriormente, construimos redes de emisión de CFDI entre actividades económicas basadas en los registros de las EFOS, y las comparamos con redes construidas por muestras aleatorias de RFCA. Encontramos ciertas parejas de actividades (enlaces en la red de emisión) para las cuales los montos promedio asociados a los CFDI entre ellas corresponden a montos mucho más altos que los observados en el resto de la población. De igual manera, resultado del cálculo de medidas de centralidad de los nodos (actividades) en las redes, identificamos ciertas actividades que juegan papeles importantes para la emisión, recepción y mediación de las operaciones asociadas a las EFOS. Estos resultados sugieren que las EFOS realizan emisiones de montos altos con mayor preferencia entre ciertas actividades, las cuales están asociadas con mayor frecuencia a la prestación de servicios o la comercialización de intangibles.

Es importante notar que no podemos asegurar que la mayor presencia de EFOS en ciertos sectores económicos corresponda a un comportamiento real de las EFOS, o a un sesgo en los métodos y mecanismos utilizados por el SAT para su identificación. Es necesario relacionar el comportamiento observado con características legales o fiscales de los sectores económicos para tener mayor certeza en esta aseveración.

De manera adicional a las emisiones de CFDI, construimos redes bipartitas de Personas Morales (PM) y sus socios, asociados y representantes legales, las cuales representan la constitución legal de las empresas. Dentro de este listado hay un conjunto de EFOS, el cual nos permite identificar su relación con el resto de las PM en el ecosistema empresarial mexicano. Hemos observado redes bipartitas en las que las PM corresponden a EFOS, las cuales pueden ser consideradas como redes organizadas de RFCA especializadas en la conformación y operación de empresas asociadas a la emisión de CFDI de operaciones simuladas.

Motivados por la observación de una sobre-representación de EFOS en ciertos sectores, implementamos un modelo de aprendizaje de máquinas (Machine Learning) basado en la información categórica de las EFOS, con el objetivo de extrapolar posibles patrones existentes para la identificación de RFCA sospechosos. El modelo presenta una exactitud de 86% en la detección de las EFOS en el listado proporcionado y, resultado de la aplicación del modelo al total de RFCA en las bases de datos, obtuvimos un listado de RFCA sospechosos de ser EFOS.

Dado que cada RFCA tiene una actividad económica, podemos pasar de emisiones de CFDI entre contribuyentes a emisiones de CFDI entre actividades económicas. La motivación principal de esta exploración es la suposición de que, además de que las EFOS tienen preferencia por pertenecer a ciertas actividades y sectores económicos, también presentan una preferencia por realizar sus operaciones (emisiones y recepciones de CFDI) entre ciertas actividades.

Haciendo uso de las redes bipartitas asociadas a EFOS, finalmente implementamos una Red Neuronal Convolutiva Profunda (DGCNN por sus siglas en inglés) capaz de usar no sólo la información categórica de los RFCA, sino también la topología de las redes bipartitas. Este modelo nos permite identificar redes bipartitas sospechosas y, en consecuencia, obtener un listado adicional de RFCA sospechosos conformado por los RFCA que forman parte de dichas redes.

Resultado de la aplicación de la metodología descrita, obtuvimos un listado de 4,026 RFCA sospechosos. A partir de éste, y considerando las tasas de utilidad del art. 58 del CFF, calculamos un estimado de evasión potencial anual de ISR. Cabe hacer notar que los estimados obtenidos pueden verse modificados resultado de un análisis fiscal detallado, por ejemplo, al incorporar el impacto de las pérdidas fiscales generadas artificialmente en ejercicios anteriores.

La metodología utilizada en este estudio se basa en el comportamiento y características observadas en las EFOS del listado proporcionado. Por lo tanto, el refinamiento y la ampliación de los métodos utilizados dependen de la calidad e integridad de las bases de datos con las que cuenta el SAT. En este sentido, recomendamos hacer una revisión de la información disponible y realizar monitoreos periódicos de las mismas, con el objetivo de tener la posibilidad de generar perfiles integrales y que caractericen el comportamiento de los contribuyentes al paso del tiempo.

Hasta el momento hemos estudiado solamente el comportamiento de las EFOS. Sin embargo, las EDOS juegan un papel fundamental en los mecanismos de evasión que aún no ha sido explorado, ya que no contamos con la información necesaria. De contar con un listado análogo al de EFOS, en un futuro podremos investigar y caracterizar el comportamiento de EDOS, tanto en las redes de emisión de CFDI como en las redes bipartitas, y así obtendremos una visión más completa del fenómeno y las prácticas de evasión tributaria.

Los métodos utilizados en este estudio no buscan reemplazar los esfuerzos realizados dentro del SAT, sino complementarlos para realizar una identificación de RFCA sospechosos de ser EFOS de forma más eficiente y expedita. Tanto la ciencia de redes como el aprendizaje de máquinas son líneas de investigación prometedoras tanto para automatizar la identificación de patrones en actividades ilícitas, como para tener una visión más holística del importante fenómeno de la evasión tributaria.

3. Revisión del marco legal

3.1. La Evasión Tributaria en el Impuesto sobre la Renta

El término evasión tributaria, aunque muchos autores lo abordan y profundizan, no todos se detienen en su conceptualización, por lo que tomaremos los diferentes planos en los que Cosulich [1] lo hace, desde una perspectiva general, hasta llegar a una particular y específica. En el plano general considera a la evasión *como la falta de cumplimiento de sus obligaciones fiscales por el contribuyente*, o bien, -parte específica-, considera, existe *evasión cuando se violan las normas jurídicas establecidas*, sin embargo, esta falta de cumplimiento o de violación a las normas jurídicas, puede llevar consigo, tanto la falta de cumplimiento de las obligaciones formales, como de las obligaciones sustanciales, que incluye la falta de pago del impuesto; por lo que para precisar acudimos al documento de Oldman y Holland [2] que consideran como evasión, *a todas aquellas actividades por las cuales se responsabiliza la persona que no paga el impuesto que la legislación vigente cobra sobre su ingreso, < ... > el contribuyente comete una falta tanto desde la hipótesis de que su delito sea por declaración intencionalmente fraudulenta como hasta la de una simple falla al hacer su declaración o al pagar el impuesto fuera del plazo determinado*, que pese a los años transcurridos, consideramos

precisa y vigente en este momento¹.

Derivado de la definición de evasión tributaria podemos darnos cuenta que la misma corre en vertientes que van desde una simple falla en las declaraciones que llevan al pago del impuesto, de manera particular en el Impuesto sobre la Renta (ISR), correspondientes al ejercicio, como el pago extemporáneo de las mismas, hasta los motivos de evasión fraudulenta y deliberada.

Con el fin de contrarrestar la evasión, la autoridad fiscal ha implementado herramientas que permiten dar seguimiento en tiempo real a las operaciones entre contribuyentes, tal es el caso de la incorporación del Comprobante Fiscal Digital por internet (CFDI) a las disposiciones fiscales –Código Fiscal de la Federación (CFF), artículo 29-. Sin embargo, a pesar de dicha incorporación los defraudadores fiscales han creado nuevas formas de evasión, que han tenido como resultado las llamadas Empresas que Facturan Operaciones Simuladas (EFOS) y su contra parte, las Empresas que Deducen Operaciones Simuladas (EDOS).

En este sentido, para detectar a estas EFOS, a partir del 1 de enero de 2014 se incorpora en el CFF el artículo 69B, en el que se establecen condicionantes generales y necesarias para la operación de un negocio o desarrollo de una profesión, relacionados con la expedición de los CFDI, que tienen la obligación de emitir, como es el caso de los activos, el personal, la infraestructura y la capacidad material y/o física; o bien, cuando el contribuyente no se encuentre localizado y a falta de ellas, la autoridad fiscal presumirá como inexistentes las operaciones amparadas en dichos CFDI. De esta forma inicia el proceso en que la autoridad le dará a conocer al contribuyente, a través del buzón tributario y mediante publicación en el Diario Oficial de la Federación (DOF), para que éste desvirtúe la presunción, manifestando lo que a su derecho convenga y presente la documentación e información respectiva.

Transcurrido el plazo sin que el contribuyente haya desvirtuado la presunción establecida por la autoridad, se le notificará en este sentido y la autoridad publicará a los treinta días siguientes de haber emitido su resolución el listado correspondiente en el DOF y en la página del SAT el listado de quienes no desvirtuaron la presunción y por lo tanto se considera que realizan operaciones inexistentes².

3.2. Ley del Impuesto sobre la Renta

En la presente investigación nos enfocamos en analizar la evasión en ISR generada por las posibles EFOS que se detecten a través de los modelos de ciencia de redes, por lo que resulta importante conocer algunos aspectos relevantes de este impuesto.

3.2.1. La Ley del Impuesto sobre la Renta, su objeto

La recaudación tributaria en México descansa mayoritariamente en el Impuesto Sobre la Renta (ISR), acudimos para establecer su objeto, en virtud de su vigencia, claridad y precisión al enunciado de la Ley de ISR publicada en el DOF del 31 de diciembre de 1953 en el que el artículo 1 establecía, *El impuesto sobre la Renta grava los ingresos provenientes del capital, del trabajo o de la combinación de ambos en los términos de este ordenamiento* [3]. La Ley de ISR vigente, aunque en términos generales y de manera indirecta también los considera de esa forma, es más

¹Esta definición es la aceptada usualmente y es tomada de la Comisión Real de Tasación de Ganancias y Rentas del Reino Unido en 1955.

²La autoridad cuenta con un plazo que no excederá de 50 días hábiles para resolver al respecto; sin embargo, dentro de los primeros 20 días podrá solicitar al contribuyente documentación adicional y éste deberá aportarla dentro de un plazo de 10 días posteriores a su notificación mediante el buzón tributario e interrumpirá el plazo de 50 días, hasta en tanto no venzan los 10 referidos.

específica al determinar el alcance de su objeto considerando las condicionantes de residentes en México, residentes en el extranjero con establecimiento permanente en el país y residentes en el extranjero con fuente de riqueza en territorio nacional.

3.2.2. Sujetos del Impuesto sobre la Renta en México

La Ley del Impuesto sobre la Renta (LISR) en México grava los ingresos de las personas morales y las personas físicas; dentro de las primeras, se consideran entre otras, *las sociedades mercantiles, los organismos descentralizados que realicen preponderantemente actividades empresariales, las instituciones de crédito, las sociedades y asociaciones civiles y la asociación en participación cuando a través de ella se realicen actividades empresariales en México* –artículo 7 LISR-, incluyendo también, dentro de entre otras, las establecidas en el artículo 25 del Código Civil Federal (CCF), y son: la Nación, los Estados y los Municipios, las demás corporaciones de carácter público reconocidas por la ley, los sindicatos, las asociaciones profesionales, incluyendo las integradas por obreros y empresarios para la defensa de sus intereses, las sociedades cooperativas y mutualistas, las asociaciones distintas de las enumeradas que se propongan fines políticos, científicos, artísticos, de recreo o cualquiera otro fin lícito, siempre que no fueren desconocidas por la ley y las personas morales extranjeras de naturaleza privada, que se registrarán por el derecho de su constitución.

3.2.3. El residente en México y el residente en el extranjero como determinante de la acumulación del ingreso en el Impuesto sobre la Renta

Un concepto que determina la acumulación total de los ingresos en México, y por lo tanto, sobre ellos el pago del ISR, sin importar la fuente de dónde procedan, es el de residencia en territorio nacional para efectos fiscales, entendiendo como tal, en el caso de personas físicas, cuando tengan su casa-habitación en el país, o cuando no teniéndola, obtengan más del 50 % de sus ingresos de fuente de riqueza ubicada en México, o tengan en el país el centro principal de sus actividades profesionales (intereses vitales) –artículo 9 del CFF-.

Dos excepciones a lo indicado en el párrafo anterior, lo tienen los funcionarios de Estado o trabajadores del mismo de nacionalidad mexicana, aun cuando obtengan más del 50 % de sus ingresos del extranjero o tengan en él su centro principal de actividades profesionales, se consideran residentes en territorio nacional. La segunda excepción se presenta cuando una persona física de nacionalidad mexicana acredita su nueva residencia en un país o territorio en donde sus ingresos se encuentren sujetos a un Régimen Fiscal Preferente (REFIPRE), considerando como tales, aquellos en los que los ingresos no son gravados para ISR, o bien, son gravados a una tasa inferior al 75 % de la que corresponde en México –artículo 176 tercer párrafo LISR-; esta persona física seguirá considerándose residente en territorio nacional durante el año en que presentó el aviso ³ de cambio de residencia y los tres siguientes, salvo que el país o jurisdicción de la nueva residencia sea un país que tenga celebrado un acuerdo amplio de intercambio de información con México. Las personas físicas de nacionalidad mexicana se consideran residentes en territorio nacional, salvo prueba en contrario.

Tratándose de personas morales, son residentes en territorio nacional, cuando hayan establecido en él la administración principal del negocio, o su sede de dirección efectiva; *contario sensu* las personas físicas y morales que no se ubiquen

³Las personas físicas y personas morales están obligadas a presentar al Registro Federal de Contribuyentes (RFC) un aviso de cambio de residencia fiscal a más tardar dentro de los 15 días inmediatos anteriores a dicho cambio –artículo 30, fracción XIV del Reglamento de CFF-. La persona moral debe considerar, en este caso, como si efectuara una liquidación, con el propósito de que pague el ISR por los ingresos y/o utilidades que se generaron en México –artículo 12, segundo párrafo LISR-.

bajo estos supuestos, se consideran residentes en el extranjero y podrán tener ingresos en México, ya sea a través de un establecimiento permanente⁴ (EP) -sucursal, agencia, representación, etc.-, de forma constante, o bien, ingresos de fuente de riqueza en territorio nacional; en el primer caso, tributarán bajo el régimen de Personas Morales del Título II, o Personas Físicas del Título IV de la LISR, capítulo II –actividades empresariales y profesionales-, según se trate, conjuntamente con los residentes en México.

Cuando los residentes en el extranjero no tengan EP en el país, o teniéndolo, sus ingresos no correspondan al mismo, se considerarán residentes en el extranjero con fuente de riqueza en territorio nacional y pagarán el ISR tomando en cuenta dos supuestos, si el país de donde procedan tiene celebrado con México un Tratado para evitar la Doble Tributación, se acogerán a las disposiciones establecidas en él, o en su defecto, lo harán bajo el título correspondiente de la LISR (V), residentes en el extranjero; en ambos casos, el ISR lo cubren mediante retención, que le hará la persona física o la persona moral que le efectúe el pago, por lo que, en este caso, los ingresos que perciba el residente en el extranjero, ya pagaron impuesto en nuestro país, y la erogación formará parte de las deducciones de dichas personas.

En este estudio, no consideramos el cálculo de evasión por estos ingresos, ya que no contamos con los CFDI correspondientes. Sin embargo, se podrían incluir en estudios futuros, para cumplir con las acciones BEPS de la OCDE, las cuales buscan evitar la erosión y traslado de utilidades a otros países con baja o nula regulación fiscal, algunas de las cuales ya son aplicadas por México.

3.2.4. Tributación en ISR de las Personas Morales

Para la determinación y cálculo del ISR del ejercicio es necesario tomar en cuenta que la Ley contempla diferentes tipos de ingresos. que pueden tener, tanto las personas físicas como las personas morales, dentro de los cuales se encuentran, los ingresos totales, gravables, acumulables, no acumulables, exentos y no objeto (véase la Sección A.2 del Anexo).

En el caso de las personas morales, la mecánica de tributación de las PM en la LISR cambia dependiendo su tipo y características, que pudieran clasificarse –las indicamos por separado, ya que así tenemos dividida la información que nos proporcionaron y sobre la que enfocamos nuestra investigación, aunque algunas pertenecen al régimen general- en personas morales del régimen general, sector primario y de autotransporte, con fines no lucrativos, régimen de consolidación y grupos de sociedades y cooperativas de producción (véase la Sección A.3 del Anexo). En el presente estudio nos enfocamos en el régimen general, derivado de que los reportes que integran una gran parte de la base gravable para el cálculo del ISR lo constituyen los importes incluidos en el CFDI –artículo 9 LISR-, que como indicamos, son de diferentes tipos (véase la Sección A.1 del Anexo); sin embargo nos concentramos en los CFDI de ingreso por la alternativa que tomamos para estimar la evasión fiscal.

Cabe aclarar que los CFDI de ingreso, en términos generales, constituyen parte de los ingresos acumulables que integran la base para el cálculo del ISR causado, tanto de PM como de PF, o bien, de la determinación del remanente distribuible en algunas PM con fines no lucrativos. Se debe tomar en cuenta, sin embargo, algunas consideraciones sobre esta acumulación de importes del CFDI para la determinación de los ingresos acumulables, tal es el caso, de la enajenación de bienes de activo fijo –inversiones-, partes representativas del capital –títulos valor, acciones, partes sociales, certificados de participación- o terrenos, en los que la acumulación del ingreso no es el total del mismo, sino únicamente la ganancia acumulable - artículos 18 fracción IV, 19, 22 y 24, 31 sexto párrafo LISR- y no todas

⁴Cfr artículo 2 y 3 LISR

ellas están amparadas con un CFDI, como es el caso de las enajenaciones de un inmueble, un terreno o las partes representativas del capital.

Existen otros ingresos acumulables que no están amparados con CFDI, como son: el ajuste anual por inflación, determinado por el excedente de las deudas sobre los créditos –artículo 44 fracción II, 45 y 46 de la LISR– o como deducción en el caso contrario; los ingresos determinados presuntivamente por las autoridades fiscales en los casos en que procedan –artículo 18 fracción I, 91 tercer párrafo, 101 segundo párrafo y último párrafo aplicable a extranjeros–; la recuperación de un crédito deducido por incobrable –artículo 18 fracción V; las condonaciones, quitas y remisiones –artículo 101 fracción I, 197 párrafos tercero y cuarto LISR–.

Las PF con ingresos por actividades empresariales no consideran como ingreso acumulable el ajuste anual por inflación, ya que el momento de acumulación de sus ingresos es al cobro –flujo de efectivo–, salvo, en este último caso, cuando se trata de ingresos por la venta de bienes que se exporten y no se reciba el pago, deberán acumularlos transcurridos doce meses de la exportación –artículo 102, último párrafo LISR–.

3.2.5. Las pérdidas fiscales en la determinación del Resultado Fiscal

Por lo que se refiere a las pérdidas fiscales de ejercicios anteriores, se tiene que acudir al seguimiento de las mismas en la contabilidad electrónica y en las declaraciones anuales. Una vez con todo esto estaríamos en posibilidad de determinar el resultado fiscal, al que se aplicará la tasa del 30 % para determinar el ISR del ejercicio. Sin embargo, ya que no determinamos el resultado fiscal por no contar con la información para hacerlo, optamos por la aplicación del art. 58 del CFF sin considerarlas.

Aunque mensualmente se tiene la obligación de realizar enteros a cuenta del ISR del ejercicio –pagos provisionales de ISR artículo 14 LISR– se pueden tener por el contribuyente, en algunos casos, cantidades enteradas en importes menores que generan recursos con los que se deja de contar, que estarán sujetos a las actualizaciones, recargos y cualquier otro accesorio; aunque en el ejercicio fiscal se debe efectuar el pago de ISR de la cantidad que corresponda, pudiendo, inclusive ser inferior al enterado mensualmente; en caso contrario, se cubrirá el ISR del ejercicio por los enteros mensuales no realizados.

3.3. Código Fiscal de la Federación

Durante la investigación, buscamos diferentes alternativas que nos pudieran llevar a alcanzar el objetivo planteado, mediante la información con que contábamos; con base en ésta, decidimos tomar la alternativa establecida en el artículo 58 del CFF. A la par, indagamos mediante el uso de redes bipartitas, las relaciones que se presentan en la estructura constitutiva de la persona moral: representantes legales y socios y accionistas -considerando dentro de éstos últimos, a los asociados y cualquier otro de naturaleza análoga- con hallazgos importantes que sugerimos profundizar en investigaciones posteriores.

Los siguientes artículos del CFF, además del 69B comentado al inicio, como indicamos, los consideramos para el cálculo de la estimación de evasión de ISR –artículo 58– y para apoyar y fundamentar las redes bipartitas de socios y asociados y representantes legales –artículo 27, Apartado B, fracción VI–.

3.3.1. Artículo 58, determinación presuntiva de la utilidad fiscal

En el Código Fiscal de la Federación –artículo 58-, dentro de las facultades de las autoridades fiscales, contempla que para determinar presuntivamente la utilidad fiscal de los contribuyentes para efectos de la Ley del ISR, podrán aplicar a los ingresos brutos declarados o determinados presuntivamente, el coeficiente de 20 % o el que corresponda, tratándose de alguna de las actividades señaladas –las tasas varían dependiendo éstas- y pueden ser del 6 %, 12 %, 15 %, 22 %, 23 %, 25 %, 27 %, 39 % y/o 50 %, esta última aplicable a servicios profesionales. Este artículo se encuentra descrito con mayor detalle en la Sección A.4 del Anexo.

3.3.2. Artículo 27, apartado A, inciso III y apartado B, incisos V y VI, Registro Federal de Contribuyentes, obligaciones, libro de socios y accionistas y aviso al RFC por cambio de socios y accionistas

A partir del 1 de enero de 2020 el artículo 27 del CFF, que contempla disposiciones en materia del RFC, se reestructuró, incluyendo dos Apartados, el A, que contempla los sujetos y sus obligaciones y el Apartado B, el catálogo general de obligaciones que deben cumplir, de acuerdo a cada caso, estos sujetos.

En la fracción III del Apartado A del artículo en comento, se establecen las obligaciones que deben cumplir los representantes legales, socios y accionistas de la persona moral listados en el Apartado B (fracciones I, II, III y IV), además de la recién incorporada fracción VI, presentar un aviso al RFC *a través del cual informen el nombre y la clave del RFC de los socios y accionistas, cada vez que se realice alguna modificación o incorporación respecto a éstos*

Conforme a la exposición de motivos del paquete económico para 2020 [4], que indica: *Toda vez que actualmente la Ley no obliga a los representantes legales, socios y accionistas de las personas morales con fines no lucrativos a que soliciten la inscripción ante el RFC, ello impide al SAT conocer la información de las personas morales con fines no lucrativos, respecto dichos sujetos, lo cual es fundamental cuando se realizan acciones de detección de operaciones inexistentes, toda vez que las personas morales con fines no lucrativos no están exentas de involucrarse en conductas tendientes a la defraudación fiscal o desvío de recursos, o bien, alguno de sus socios o representantes legales.*

Derivado de lo anterior se elimina para el año 2020 la salvedad que se tenía para las personas morales con fines no lucrativos. Esta eliminación originó que dichas personas morales “tuvieran dudas” para cumplir con la obligación por considerar que no contaban con socios y accionistas, por *lo que, para el 2021 se cambia el contenido de la fracción VI e incorporan a asociados y demás personas, cualquiera que sea el nombre con el que se les designe, que por su naturaleza formen parte de la estructura orgánica y que ostenten dicho carácter conforme a los estatutos o legislación bajo la cual se constituyen* [5].

También a partir del 1 de enero de 2020 se incorpora en la fracción VI del apartado B del artículo 27, la obligación de presentar aviso al RFC cuando haya modificación de socios o accionistas –para 2021 se incorporan también a los indicados en el párrafo anterior-; la exposición de motivos indicaba: *De igual manera se sugiere establecer que las personas morales tendrán la obligación de presentar un aviso en el RFC cada vez que sus socios o accionistas sean modificados, atendiendo a la necesidad de que se combata la creación de empresas que se constituyen únicamente con la finalidad de facturar o deducir operaciones inexistentes, las cuales regularmente están integradas con los mismos socios o accionistas personas físicas que se dedican a esta práctica indebida.*

Por su parte la regla 2.4.19 de la Resolución Miscelánea Fiscal (RMF) para 2020, publicada en el DOF del 28 de

diciembre de 2019, a este respecto establece que, deberán presentar ante el RFC un aviso en el que informarán el nombre y la clave del RFC de los socios y accionistas cada vez que se realice una modificación o incorporación conforme a la ficha de trámite 295/CFE “Aviso de actualización de socios o accionistas” contenida en el Anexo 1-A dentro de los treinta días hábiles siguientes al que se realice el supuesto ante la Administración Desconcentrada de Servicios al Contribuyente (ADSC). El artículo cuadragésimo sexto transitorio de la RMF en comento, establece que el aviso al RFC con la información de la estructura con la que se encuentren las PM en ese momento, por única ocasión se deberá presentar a más tardar el 30 de junio de 2020.

La ampliación del plazo para presentar el aviso con la información de socios y accionistas que conforman la estructura actualizada de la persona moral –que se publicó al cierre de esta investigación-, indica, al igual que en 2020, por única ocasión, para quienes no lo hayan presentado en dicho año, lo podrán hacer a más tardar el 31 de marzo de 2021 [6]. Así como, para quienes a partir del 1 de enero de 2021 y durante el primer semestre tengan la obligación de presentar el aviso por modificación o incorporación de socios y accionistas, asociados u otras personas de naturaleza análoga, lo podrán hacer a más tardar el 30 de septiembre de 2021 -Artículo Quincuagésimo Tercero Transitorio para 2021-.

3.4. Ley Federal para la Prevención e Identificación de Operaciones con Recursos de Procedencia Ilícita (LFPIORPI)

Por lo que corresponde a la LFPIORPI, disposición de carácter administrativo, llamada comúnmente Ley Anti lavado, cuyo objeto es proteger al sistema financiero y a la economía nacional de las operaciones con recursos de procedencia ilícita –artículo 2-, tomamos el artículo 17, De las actividades vulnerables, fracciones V y XI para apoyar y fundamentar los hallazgos mostrados en las Secciones 6.1.1 y 7, en donde identificamos sectores económicos sobre representados de EFOS, en específico los sectores de construcción y prestación de servicios. Las actividades vulnerables de esta Ley, corresponden, como su nombre lo indica, a aquellas en las que se tiene mayor riesgo de ser utilizadas para incorporar a la economía nacional, recursos obtenidos de forma ilícita.

3.4.1. Artículo 17, fracción V, prestación habitual o profesional de servicios de construcción y en general transmisión de propiedad y constitución de derechos sobre bienes

La Ley en comento establece la obligación, a quienes realizan estas actividades vulnerables, de inscribirse en el padrón de sujetos obligados a cumplir con la LFPIORPI, así como, entre otras, identificar al cliente, usuario y en su caso, beneficiario final y presentar avisos mensuales, con los datos de las personas que recibieron estas operaciones y que fueron previamente identificadas, cuando el monto llegue o supere el umbral de 8,025 Unidades de Medida y Actualización (UMA) -\$678,032.25 para 2020-. Estas operaciones de servicios de construcción y en general transmisión de propiedad y constitución de derechos sobre bienes, a su vez, están relacionadas con las obligaciones, que en materia de esta ley tienen los notarios, en cuanto a identificación –siempre-, presentación de avisos -16,000UMA, \$1,351,840.-, prohibición de recibir o pagar operaciones con dinero en efectivo, metales preciosos o divisas, a partir de ciertos montos – 8,025 UMA \$678,032.25 para estas operaciones-, y en su caso, anotar en la escritura o documento en el que otorgan fe pública, monto, fecha y forma de pago y moneda o divisa con que fue cubierta la operación –artículos 17, fracción XII, apartado A, incisos a) a e); 32; 33 de la LFPIORPI y 42 a 45 del Reglamento de dicha ley-.

3.4.2. Prestación de servicios profesionales, artículo 17, fracción XI, incisos a) a e)

Por cuanto a la prestación de servicios profesionales, sin que medie una relación laboral, establecida en la fracción XI, incisos a) a e), tienen las mismas obligaciones que los de la fracción V -salvo en el umbral para presentar aviso-, en cuanto a inscripción al padrón, identificación del cliente, usuario, o en su caso, beneficiario final y presentación de avisos cuando lleven a cabo operaciones por cuenta del cliente, en cuanto a compra de bienes inmuebles o cesión de derechos sobre éstos; administración y manejo de recursos, valores y manejo de cualquier otro activo; manejo de cuentas bancarias, de ahorro o de valores; organización de aportaciones de capital o cualquier otro tipo de recursos para la constitución, operación y administración de sociedades mercantiles; o bien, la constitución, escisión, o fusión, operación y administración de personas morales o vehículos corporativos, incluido el fideicomiso y la compra o venta de entidades mercantiles. La presentación de avisos, con los datos de las personas a quienes se les prestó el servicio, previamente identificadas, es siempre, cuando lleven a cabo a su nombre y representación operaciones financieras relacionadas con las antes indicadas.

4. Experiencias internacionales de investigación

El fenómeno de la evasión fiscal ha sido estudiado desde distintos puntos de vista, entre los que destaca en tiempos recientes el enfoque de la Ciencia de Redes, el cual ha mostrado su utilidad para la descripción de sistemas y fenómenos de distinta naturaleza [7].

Además del fenómeno específico de la evasión fiscal, la ciencia de redes, mediante la construcción y análisis de las redes identificadas, ha sido utilizada para la detección de cárteles de manipulación de precios [8], o para analizar el riesgo de posibles prácticas corruptas en licitaciones públicas [9]. Así mismo, se ha utilizado para describir redes de corrupción política, por medio de la identificación de nodos estructurales en la topología de redes de figuras públicas [10], o como en el estudio presentado en la referencia [11], en donde se estudia la dinámica de actores en distintos eventos de corrupción, identificando conjuntos de actores que han participado de forma recurrente en este tipo de prácticas y sus mecanismos de colusión con el resto de los actores por medio de la construcción y el análisis de las redes de corrupción identificadas.

Este tipo de esfuerzos también han sido útiles para la descripción de redes de empresas pantalla (coloquialmente conocidas como empresas fantasma) y asociadas a la desviación de recursos públicos, tomando como ejemplo el caso conocido como “la estafa maestra” [12], en donde, haciendo uso de información abierta relacionada con la asignación de contratos públicos y la constitución legal de empresas, se construyen redes entre empresas y dependencias públicas, así como redes bipartitas de las empresas a las que se les asignaron los contratos, que permiten explorar distintos aspectos de estas empresas pantalla, tales como patrones en las fechas de constitución de las mismas así como identificar que estas empresas operan de manera organizada en redes de corrupción, operadas por un conjunto de actores que coordinan la operación de las mismas.

Otro ejemplo de la aplicación de técnicas de la ciencia de redes para el estudio de actividades ilícitas se presenta en el estudio mostrado en la referencia [13], en donde se trata el fenómeno del lavado de dinero. En dicho trabajo se realiza una caracterización de actores, representados por nodos en las redes en las que, mediante distintas medidas de centralidad, los autores asignan distintos roles a los actores (nodos), que permitirían identificar conjuntos de actores clave en estas redes para la implementación de acciones de intervención efectivas.

Este conjunto de ejemplos ofrece un panorama de la aplicación de técnicas similares a las empleadas en este estudio,

y muestran como la ciencia de redes hace una importante aportación para el entendimiento y desarrollo de posibles mecanismos de intervención para mitigar distintos tipos de actividades ilícitas como la evasión fiscal, la corrupción o el lavado de dinero, entre otros. Esta es un área de investigación abierta la cual, gracias a un enfoque multidisciplinario, posibilita desde la caracterización de los mecanismos legales sobre los cuales estas prácticas ocurren, hasta la incorporación de modelos de aprendizaje de máquinas que permiten la generalización de comportamientos para la identificación de conductas sospechosas.

En las siguientes secciones desarrollamos las metodologías utilizadas y los resultados obtenidos de la aplicación de las mismas en este estudio, lo cual nos ha permitido realizar un estimado de la evasión de ISR asociado a RFCA sospechosos de ser posibles EFOS.

5. Descripción de datos

Los datos que nos fueron proporcionados, en distintas bases de datos con información previamente anonimizada por parte del SAT, incluyen:

- Agregados mensuales de las emisiones de CFDI entre RFCA, en el periodo 2015–2019, conformados por los RFCA de emisores, receptores y los montos asociados.
- Base de datos con información categórica de cerca de 80 millones de RFCA, que incluyen:
 - Fecha de creación
 - Fecha inicio de operaciones
 - Entidad federativa
 - Clave numérica del Sector Económico
 - Descripción del Sector Económico
 - Clave numérica de la Rama Económica
 - Descripción de la Rama Económica
 - Clave numérica de la Actividad Económica
 - Descripción de la Actividad Económica
- Declaraciones anuales de ISR de Personas Físicas y Morales en el periodo 2015–2019, en las que se reportan los ingresos e impuestos que los contribuyentes autodeterminan y declaran.
- Listado de relaciones entre Personas Morales y sus Representantes Legales y Socios y Asociados identificados por medio de su RFCA, así como información categórica asociada a cada uno de ellos.
- Listado de EFOS definitivas y presuntas identificadas por medio de su RFCA, previamente procesadas por el SAT.

Del total de los RFCA en el padrón, 10,678 RFCA están clasificados como EFOS por el SAT, las cuales pueden ser definitivas o presuntas. En la Tabla 1 se muestran los números exactos de RFCA en cada una de las categorías. A los contribuyentes o RFCA que no pertenecen al padrón de EFOS proporcionado por el SAT los denominamos no clasificados, pues no se conoce si se clasifican como EFOS o no. Cabe hacer notar que este listado de EFOS es

esencial, ya que la identificación y caracterización de su comportamiento en las distintas bases de datos, es lo que permite implementar los métodos utilizados en este estudio, que generalizan su comportamiento para la identificación de RFCA sospechosos de ser EFOS.

Número de EFOS definitivas y presuntas en el listado	
EFOS Definitivas	EFOS Presuntas
9,663	1,015

Tabla 1: Número de EFOS Definitivas y Presuntas proporcionadas por el SAT

El otro conjunto de datos que fue utilizado fue el listado de personas morales y sus socios, asociados y representantes legales, en el cual consta de 2,012,442 RFCA, de los cuales 1,161 están clasificados como EFOS definitivas o presuntas.

Partiendo de estos conjuntos de datos construimos redes de emisión de CFDI entre actividades económicas, redes bipartitas entre personas morales y sus asociados, así como implementamos modelos de Aprendizaje de Máquinas para la identificación de RFCA sospechosos. Los detalles del procesamiento y análisis de los datos se describe en las siguientes secciones.

6. Metodología de investigación

Debido a que los datos que nos fueron proporcionados permiten describir tanto a los contribuyentes como a las relaciones entre ellos desde distintos puntos de vista, utilizamos diferentes técnicas para la exploración de su comportamiento.

En primer lugar, haciendo uso de la base de datos de información categórica de contribuyentes, calculamos la distribución de las EFOS dentro de los distintos sectores, ramas y actividades económicas (véase la Sección 6.1.1). Posteriormente, agrupamos a los contribuyentes por actividades económicas y construimos redes de emisión de CFDI entre ellas (véase la Sección 6.1.2), en las cuales caracterizamos a los nodos en las redes por medio del cálculo de medidas de centralidad, y comparamos los montos promedio de las operaciones de EFOS (los cuales asignamos a pesos de los enlaces en las redes) con los de las operaciones realizadas por el resto de la población (véase la Sección 6.1.3).

A la luz de los resultados obtenidos en las secciones 6.1.1 y 6.1.2, partiendo de la información categórica de las EFOS definitivas y presuntas, implementamos un modelo de clasificación de Aprendizaje de Máquinas que nos permite realizar una clasificación del resto de los contribuyentes como sospechosos de ser EFOS (véase la Sección 6.2.1). La evaluación del modelo presenta una muy buena precisión en tanto a que clasifica correctamente a las EFOS proporcionadas por el SAT con una alta probabilidad.

En segundo lugar, a partir del listado de relaciones entre personas morales y sus representantes legales y socios o asociados construimos redes bipartitas, las cuales representan la composición legal de personas morales (véase la Sección 6.1.4). Considerando las redes compuestas exclusivamente por las personas morales en el listado de las EFOS y sus asociados, observamos una variedad de topologías (patrones de interacción entre nodos). Este conjunto de redes es utilizado para el entrenamiento de un segundo modelo de clasificación basado en Aprendizaje de Máquinas (véase la Sección 6.2.3) el cual, a diferencia del basado en la información categórica de los contribuyentes y que asigna una probabilidad de ser sospechoso de ser EFOS a un RFCA en particular, realiza una clasificación de redes bipartitas, asignándoles ahora a las redes completas una probabilidad de ser sospechosas, generalizando las propiedades topológicas de las redes bipartitas observadas alrededor de las EFOS.

En las siguientes secciones se describe a detalle la metodología y resultados obtenidos de cada una de las exploraciones realizadas.

6.1. Ciencia de redes

En esta sección describimos la aplicación de distintas técnicas de la ciencia de redes para describir algunos patrones de emisión de CFDI, y de relaciones de la composición legal de personas morales por medio de sus representantes legales, socios y asociados, observados en las EFOS.

En cada una de estas aplicaciones el papel de los nodos y los enlaces en la red es distinto. Por un lado, los nodos corresponden a los sectores, ramas o actividades económicas de los contribuyentes, y los enlaces representan la emisión y recepción de CFDI. Por otro lado, en la aplicación en la que estudiamos la conformación legal de personas morales, los nodos, aunque en este caso también representan a contribuyentes, pueden ser de tres tipos distintos: personas morales, representantes legales o socios y asociados. En este caso los enlaces en la red representan una relación entre RFCA mediante la composición legal de las personas morales.

6.1.1. Concentración de EFOS en actividades económicas

Hicimos uso de la base de datos de información categórica en la que se incluye el sector, rama y actividad económica⁵ para identificar si las EFOS prefieren pertenecer a alguna de las categorías en las que está dividida la economía, es decir, si buscan actividades que les permita con mayor facilidad simular operaciones y por ende causar erosión de la base gravable, defraudación o evasión fiscal. Este análisis motivó la exploración de los patrones de emisión de CFDI entre actividades económicas de la Sección 6.1.2 y se convirtió en el punto de partida para uno de los procedimientos de aprendizaje de máquinas: potenciación del gradiente, presentado en la Sección 6.2.1.

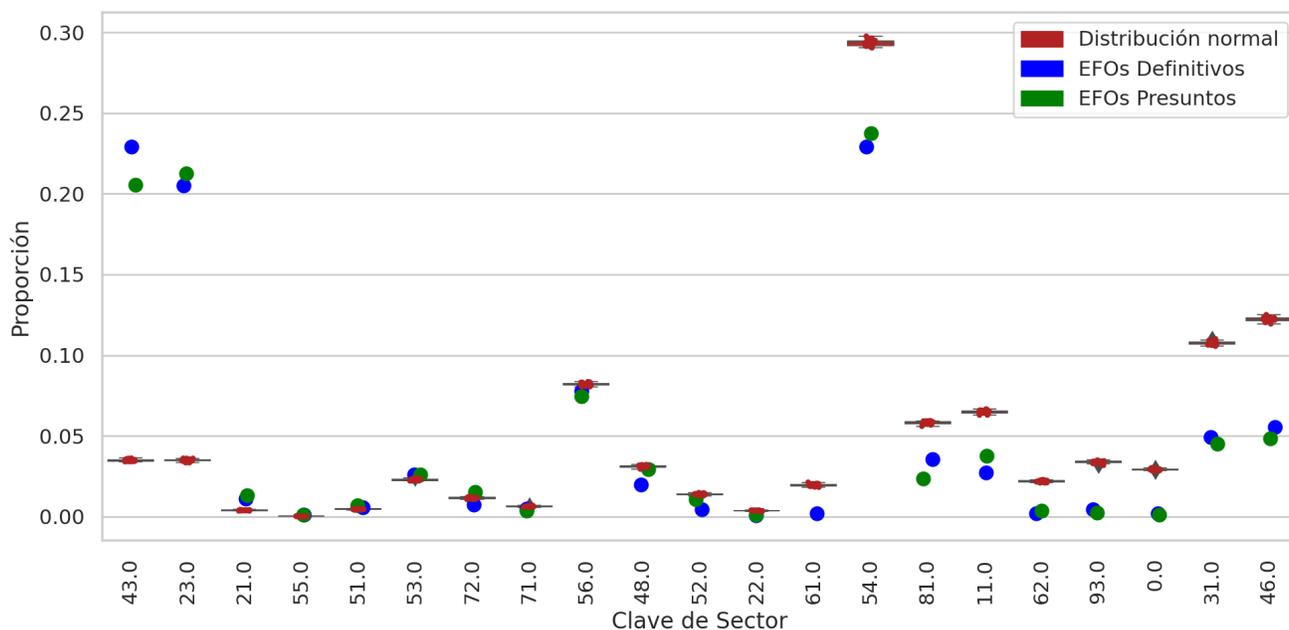
Para identificar si existe tal preferencia, se podría comparar la distribución en sectores y actividades económicas que presenta la base de EFOS con el total de la población. Sin embargo, para robustecer esta comparativa y tener un mejor nivel de confianza se tomaron 100 muestras de datos, cada una conteniendo 10 mil registros aleatorios de la tabla de RFCA no clasificados (no sabemos si son posibles EFOS o no) para construir las distribuciones de las distintas actividades y sectores económicos y así compararlos con la distribución que presentan las EFOS.

Como resultado de lo anterior, observamos una sobre representación de las EFOS en los sectores del comercio al por mayor, la industria de la construcción y la minería (véase la Figura 1). Por ejemplo, la proporción media de la población general dedicada al sector del comercio al por mayor es del 3.5 %, con una desviación estándar del 0.06 %, mientras que en la población de EFOS definitivas y presuntas, corresponde al 18.3 %. Esto indica una sobre representación, pues claramente difiere de lo que podría atribuirse al azar (297 desviaciones estándar).

Las figuras 3 y 4 exploran aún más los sectores sobre representados de comercio al por mayor y la construcción, y observamos que dentro de estos sectores existen ramas y actividades con participación significativamente mayor de las EFOS presuntas y definitivas, por ejemplo, “otras construcciones de Ingeniería Civil u obra pesada”, en el sector de la construcción (véase la Fig. 3) o la actividad de “otros intermediarios de comercio al por mayor” en el sector del comercio al por mayor (véase la Fig. 4).

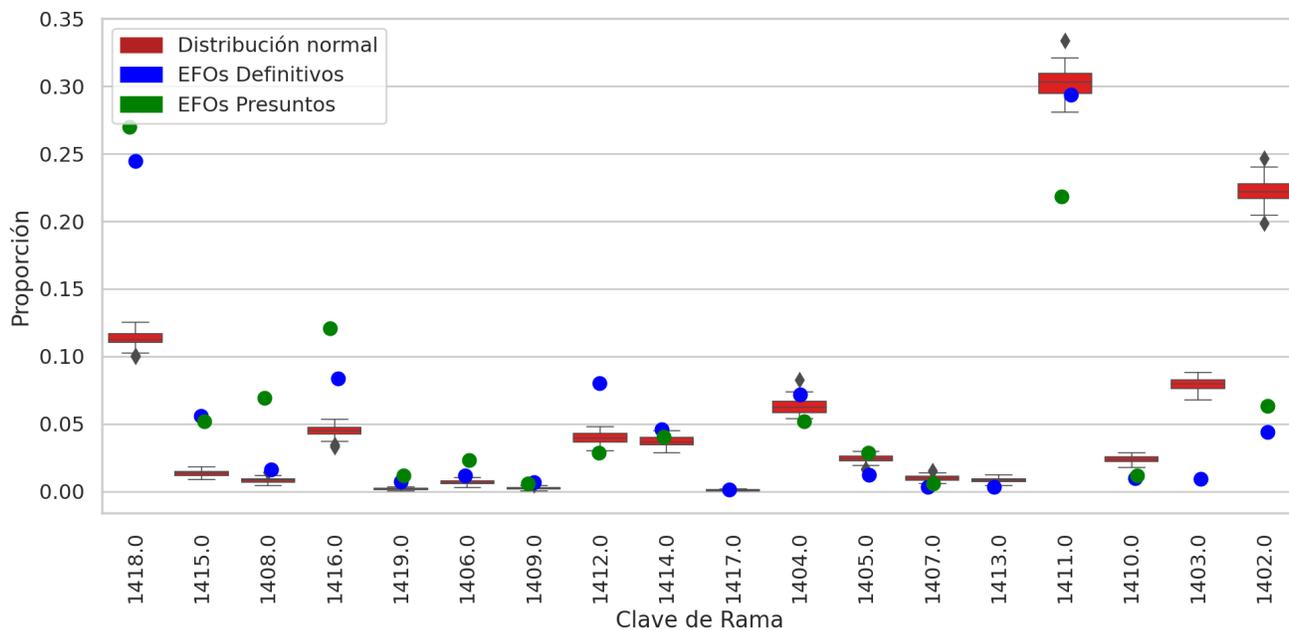
⁵Los sectores, ramas y actividades económicas se organizan de manera jerárquica, siendo los sectores la categoría más general, seguido de las ramas, para finalizar con las actividades económicas, la cual corresponde a la categoría más específica y numerosa de las tres. Cada uno de estos niveles nos permite observar el comportamiento de los contribuyentes a diferentes escalas.

A pesar de la fuerte evidencia de que las EFOS pudieran inclinarse por ciertas actividades de ciertos sectores económicos no podemos concluir que este hecho sea por un comportamiento natural de las EFOS o porque los criterios de fiscalización y detección de EFOS por parte del SAT sesga la investigación en estas áreas económicas. En cualquiera de los dos casos, surge la pregunta de si es posible mejorar este perfil EFOS utilizando únicamente la información categórica de RFCA.



Sectores Económicos	
Clave	Descripción
43.0	COMERCIO AL POR MAYOR
23.0	CONSTRUCCIÓN
21.0	MINERÍA
55.0	DIRECCIÓN DE CORPORATIVOS Y EMPRESAS
51.0	INFORMACIÓN EN MEDIOS MASIVOS
53.0	SERVICIOS INMOBILIARIOS Y DE ALQUILER DE BIENES MUEBLES E INTANGIBLES
72.0	SERVICIOS DE ALOJAMIENTO TEMPORAL Y DE PREPARACIÓN DE ALIMENTOS Y BEBIDAS
71.0	SERVICIOS DE ESPARCIMIENTO CULTURALES Y DEPORTIVOS, Y OTROS SERVICIOS RECREATIVOS
56.0	SERVICIOS DE APOYO A LOS NEGOCIOS Y MANEJO DE DESECHOS Y SERVICIOS DE REMEDIACIÓN
48.0	TRANSPORTES, CORREOS Y ALMACENAMIENTO
52.0	SERVICIOS FINANCIEROS Y DE SEGUROS
22.0	ELECTRICIDAD, AGUA Y SUMINISTRO DE GAS POR DUCTOS AL CONSUMIDOR FINAL
61.0	SERVICIOS EDUCATIVOS
54.0	SERVICIOS PROFESIONALES, CIENTÍFICOS Y TÉCNICOS
81.0	OTROS SERVICIOS EXCEPTO ACTIVIDADES DEL GOBIERNO
11.0	AGRICULTURA, GANADERÍA, APROVECHAMIENTO FORESTAL, PESCA Y CAZA
62.0	SERVICIOS DE SALUD Y DE ASISTENCIA SOCIAL
93.0	ACTIVIDADES DEL GOBIERNO Y DE ORGANISMOS INTERNACIONALES Y EXTRATERRITORIALES
0.0	ACTIVIDAD PENDIENTE DE ACLARACIÓN
31.0	INDUSTRIAS MANUFACTURERAS
46.0	COMERCIO AL POR MENOR

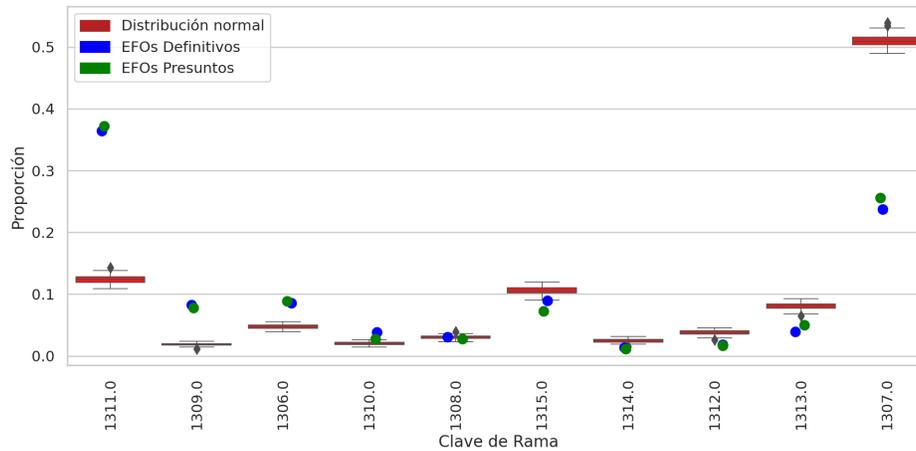
Figura 1: Proporción de EFOS definitivas y presuntas (puntos azules y verdes, respectivamente), que pertenecen a los diferentes sectores económicos, en comparación con la distribución de proporciones de 100 diferentes muestras de RFCA representativas de la población en general (diagramas de caja en rojo). Los sectores con más sobre representación en la población de EFOS son el 43, el 23 y el 21, los cuales corresponden a los sectores de comercio al por mayor, construcción y minería respectivamente.



Ramas Económicas dentro del Sector de comercio al por mayor

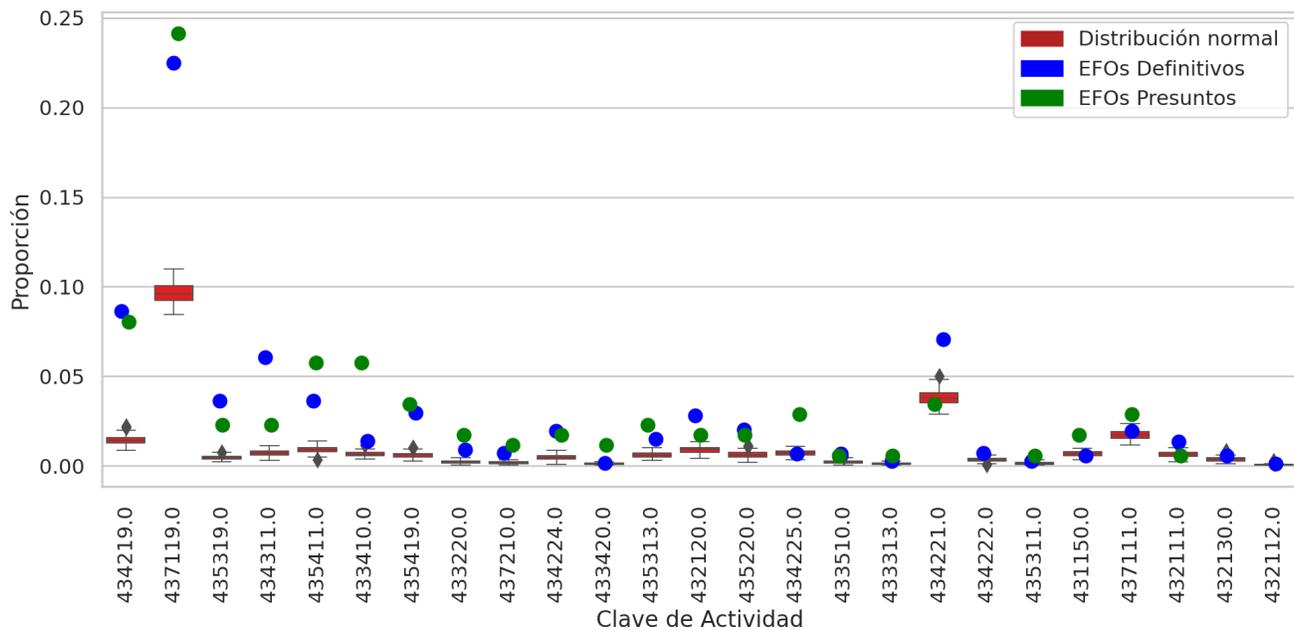
Clave	Descripción
1418.0	INTERMEDIACIÓN AL POR MAYOR.
1415.0	COMERCIO AL POR MAYOR DE MAQUINARIA Y EQUIPO PARA LOS SERVICIOS Y PARA ACTIVIDADES COMERCIALES.
1408.0	COMERCIO AL POR MAYOR DE ARTÍCULOS DE PAPELERÍA, LIBROS, REVISTAS Y PERIÓDICOS.
1416.0	COMERCIO AL POR MAYOR DE MAQUINARIA, MOBILIARIO Y EQUIPO DE USO GENERAL.
1419.0	COMERCIO AL POR MAYOR POR MEDIOS MASIVOS DE COMUNICACIÓN Y OTROS MEDIOS.
1406.0	COMERCIO AL POR MAYOR DE ARTÍCULOS DE PERFUMERÍA, JOYERÍA Y OTROS ACCESORIOS DE VESTIR.
1409.0	COMERCIO AL POR MAYOR DE ELECTRODOMÉSTICOS MENORES Y APARATOS DE LÍNEA BLANCA.
1412.0	COMERCIO AL POR MAYOR DE MATERIALES DE DESECHO
1414.0	COMERCIO AL POR MAYOR DE MAQUINARIA Y EQUIPO PARA LA INDUSTRIA.
1417.0	COMERCIO AL POR MAYOR DE CAMIONES.
1404.0	COMERCIO AL POR MAYOR DE PRODUCTOS TEXTILES Y CALZADO.
1405.0	COMERCIO AL POR MAYOR DE PRODUCTOS FARMACÉUTICOS.
1407.0	COMERCIO AL POR MAYOR DE DISCOS, JUGUETES Y ARTÍCULOS DEPORTIVOS.
1413.0	COMERCIO AL POR MAYOR DE MAQUINARIA Y EQUIPO AGROPECUARIO, FORESTAL Y PARA LA PESCA.
1411.0	COMERCIO AL POR MAYOR DE MATERIAS PRIMAS PARA LA INDUSTRIA.
1410.0	COMERCIO AL POR MAYOR DE MATERIAS PRIMAS AGROPECUARIAS.
1403.0	COMERCIO AL POR MAYOR DE BEBIDAS Y TABACO.
1402.0	COMERCIO AL POR MAYOR DE ALIMENTOS Y ABARROTES.

Figura 2: Proporción de EFOS definitivas y presuntas que pertenecen a las diferentes ramas dentro del sector 43 (comercio al por mayor), en comparación con la distribución de proporciones de 100 diferentes muestras de RFCA representativas de la población general perteneciente a este sector. Las ramas más sobre representadas en la población de EFOS son la 1418, 1415 y 1408, que corresponden a las ramas de: Intermediación al por mayor, Comercio al por mayor de maquinaria y equipo para los servicios y para actividades comerciales y Comercio al por mayor de artículos de papelería, libros, revistas y periódicos, respectivamente.



Ramas Económicas dentro del Sector de construcción	
Clave	Descripción
1311.0	OTRAS CONSTRUCCIONES DE INGENIERÍA CIVIL U OBRA PESADA
1309.0	DIVISIÓN DE TERRENOS Y CONSTRUCCIÓN DE OBRAS DE URBANIZACIÓN
1306.0	EDIFICACIÓN RESIDENCIAL
1310.0	CONSTRUCCIÓN DE VÍAS DE COMUNICACIÓN
1308.0	CONSTRUCCIÓN DE OBRAS PARA EL ABASTECIMIENTO DE AGUA, PETRÓLEO, GAS, ELECTRICIDAD Y TELECOMUNICACIONES
1315.0	OTROS TRABAJOS ESPECIALIZADOS PARA LA CONSTRUCCIÓN
1314.0	TRABAJOS DE ACABADOS EN EDIFICACIONES
1312.0	CIMENTACIONES, MONTAJE DE ESTRUCTURAS PREFABRICADAS Y TRABAJOS EN EXTERIORES
1313.0	INSTALACIONES Y EQUIPAMIENTO EN CONSTRUCCIONES
1307.0	EDIFICACIÓN NO RESIDENCIAL

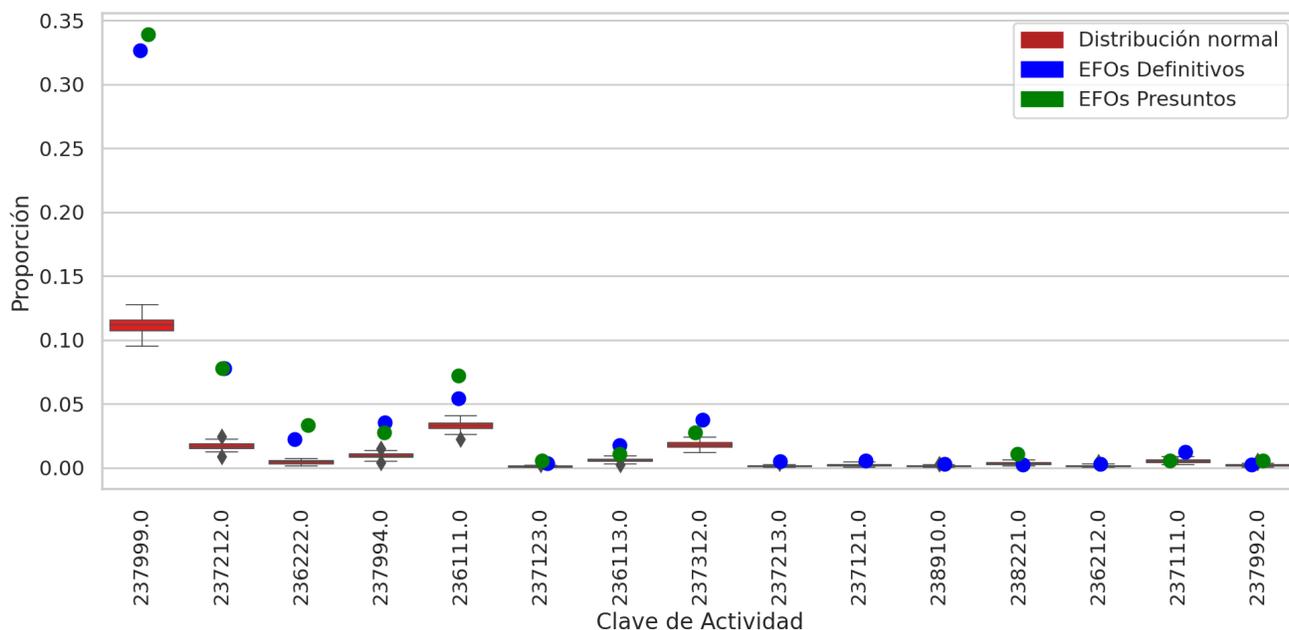
Figura 3: Proporción de EFOS definitivas y presuntas que pertenecen a las diferentes ramas dentro del sector 23 (construcción), en comparación con la distribución de proporciones de 100 diferentes muestras de RFCA representativas de la población general perteneciente a este sector. Las ramas más sobre representadas en la población de EFOS son la 1311, la 1309 y la 1306 correspondientes a las ramas de: Otras construcciones de Ingeniería Civil u obra pesada, División de terrenos y construcción de obras de urbanización y Edificación residencial, respectivamente.



Actividades Económicas dentro del Sector de comercio al por mayor

Clave	Descripción
434219.0	Comercio al por mayor de otros materiales para la construcción, excepto de madera
437119.0	Otros intermediarios de comercio al por mayor
435319.0	Comercio al por mayor de maquinaria y equipo para otros servicios y para actividades comerciales
434311.0	Comercio al por mayor de desechos metálicos
435411.0	Comercio al por mayor de equipo y accesorios de cómputo
433410.0	Comercio al por mayor de artículos de papelería para uso escolar y de oficina
435419.0	Comercio al por mayor de otra maquinaria y equipo de uso general
433220.0	Comercio al por mayor de artículos de joyería y otros accesorios de vestir
437210.0	Comercio al por mayor por medios masivos de comunicación (como correo e internet) y otros medios
434224.0	Comercio al por mayor de madera
433420.0	Comercio al por mayor de libros
435313.0	Comercio al por mayor de mobiliario, equipo e instrumental médico y de laboratorio
432120.0	Comercio al por mayor de ropa
435220.0	Comercio al por mayor de maquinaria y equipo para la industria manufacturera
434225.0	Comercio al por mayor de equipo y material eléctrico
433510.0	Comercio al por mayor de electrodomésticos menores y aparatos de línea blanca
433313.0	Comercio al por mayor de artículos y aparatos deportivos
434221.0	Comercio al por mayor de materiales metálicos
434222.0	Comercio al por mayor de productos químicos para uso industrial
435311.0	Comercio al por mayor de equipo de telecomunicaciones, fotografía y cinematografía
431150.0	Comercio al por mayor de semillas y granos alimenticios, frutas secas, chiles secos y especias (clavos, pimienta, azafrán, comino, nuez moscada, canela)
437111.0	Venta al por mayor por comisión y consignación
432111.0	Comercio al por mayor de fibras, hilos y telas
432130.0	Comercio al por mayor de calzado
432112.0	Comercio al por mayor de blancos

Figura 4: Proporción de EFOS definitivas y presuntas que pertenecen a las diferentes actividades dentro del sector 43 (comercio al por mayor), en comparación con la distribución de proporciones de 100 diferentes muestras de RFCA representativas de la población general perteneciente a este sector. Se observa una mayor sobre representación de EFOS en las actividades 437119, 434219, 434311, 435319, 435411 y 435419, las cuales corresponden a: Otros intermediarios de comercio al por mayor, Comercio al por mayor de otros materiales para la construcción, excepto de madera, Comercio al por mayor de desechos metálicos, Comercio al por mayor de maquinaria y equipo para otros servicios y para actividades comerciales, Comercio al por mayor de equipo y accesorios de cómputo, Comercio al por mayor de otra maquinaria y equipo de uso general, respectivamente.



Actividades Económicas dentro del Sector de construcción

Clave	Descripción
237999.0	Otras construcciones de ingeniería civil u obra pesada
237212.0	Construcción de obras de urbanización
236222.0	Administración y supervisión de Construcción de inmuebles comerciales, institucionales y de servicios
237994.0	Administración y supervisión de construcción de otras obras de ingeniería civil u obra pesada
236111.0	Construcción de vivienda unifamiliar
237123.0	Administración y supervisión de construcción de obras para petróleo y gas
236113.0	Administración y supervisión de construcción de vivienda
237312.0	Construcción de carreteras, autopistas, terracerías, puentes, pasos a desnivel y aeropistas
237213.0	Administración y supervisión de división de terrenos y de construcción de obras de urbanización
237121.0	Construcción de obras relacionadas con la distribución de petróleo y gas
238910.0	Preparación de terrenos para la construcción
238221.0	Instalaciones hidrosanitarias y de gas en construcciones
236212.0	Administración y supervisión de Construcción de naves y plantas industriales
237111.0	Construcción de obras para el tratamiento, distribución y suministro de agua y drenaje
237992.0	Construcción de obras marítimas, fluviales y subacuáticas

Figura 5: Proporción de EFOS definitivas y presuntas que pertenecen a las diferentes actividades dentro del sector 23 (construcción), en comparación con la distribución de proporciones de 100 diferentes muestras de RFCA representativas de la población general perteneciente a este sector. Se observa una mayor sobre representación de EFOS en las actividades 237999, 237212, 236222, 237994 y 236111, que corresponden a: Otras construcciones de ingeniería civil u obra pesada, Construcción de obras de urbanización, Administración y supervisión de Construcción de inmuebles comerciales, institucionales y de servicios, Administración y supervisión de construcción de otras obras de ingeniería civil u obra pesada y Construcción de vivienda unifamiliar, respectivamente.

6.1.2. Redes de emisiones de CFDI entre actividades económicas

El análisis de las redes de emisiones de CFDI entre actividades económicas, puede ser útil para la identificación de actividades susceptibles de participar con mayor frecuencia en la evasión de impuestos y ser útil para la focalización de recursos y la detección temprana de posibles evasores. Con el objetivo de comparar las emisiones de CFDI entre actividades económicas que realizan tanto las EFOS como los RFCA no clasificados, construimos redes de emisiones para cada una de las clasificaciones de las actividades de los contribuyentes (sector, rama, actividad) en las que los enlaces corresponden al agregado de las operaciones entre EFOS, ⁶ ya sean definitivas o presuntas, y redes en las que los enlaces corresponden a las emisiones del resto de los contribuyentes (La manera en la que agregamos la información y que nos permite pasar de interacciones entre RFCA a interacciones entre actividades económicas se describe en la Sección A.6 del Anexo).

En la Fig. 6 se muestran ejemplos de las redes de emisiones de CFDI entre sectores, ramas y actividades obtenidas de los RFCA previamente identificados como EFOS. Se puede observar como el número de elementos se incrementa conforme se consideran las más específicas en la jerarquía de la clasificación de sectores, ramas y actividades.

Dado que el nivel de actividad económica se compone de un mayor número de elementos y nos permite realizar una mejor descripción y caracterización de las propiedades de dichas redes de emisión de CFDI, en lo sucesivo analizamos este nivel y prescindiremos de los niveles de sector y rama económicas.

A continuación, calculamos medidas de centralidad de los nodos en la red, para cuantificar su importancia respecto a distintos roles que desarrollan en ella, lo que nos permite identificar a las actividades económicas más utilizadas por las EFOS para la realización de sus operaciones.

Se utilizaron las siguientes medidas de centralidad:

- I) Centralidad de grado de entrada: la cual identifica a los nodos (o actividades) que reciben emisiones de CFDI desde un mayor número de actividades distintas
- II) Centralidad de grado de salida: que identifica actividades económicas que realizan emisiones de CFDI hacia un mayor número de actividades distintas
- III) Centralidad de *betweenness*: la cual nos permite identificar actividades que fungen como intermediarias de las emisiones de CFDI entre un mayor número de parejas de actividades económicas.

En la Tabla 2 se listan las actividades con los valores más altos de las medidas de centralidad de grado de entrada, de salida y *betweenness*. Se puede observar como la actividad “Servicios de consultoría en administración” es la que consistentemente presenta los valores más altos, seguida de otras actividades como “Otros intermediarios de comercio al por mayor”, “Otros servicios profesionales, científicos y técnicos”, “Servicios especializados de apoyo a las empresas”, entre otros. Siendo el común denominador de estas actividades el que están relacionadas con la prestación de servicios con generalidades o ambigüedad de la descripción de la actividad que se desarrolla. Por otro lado, otro conjunto de actividades que también presentan valores altos de las medidas de centralidad son las relacionadas con el sector de la construcción, el cual como hemos mencionado anteriormente, presenta una sobre representación de EFOS.

El que este tipo de actividades sean utilizadas con mayor frecuencia como emisoras, receptoras o intermediarias de los CFDI asociados a las operaciones de EFOS, puede indicarnos, ya sea que las EFOS que emiten los CFDI tienen

⁶Solamente tomaremos en cuenta el número de facturas activas y el monto subtotal activo de cada registro.

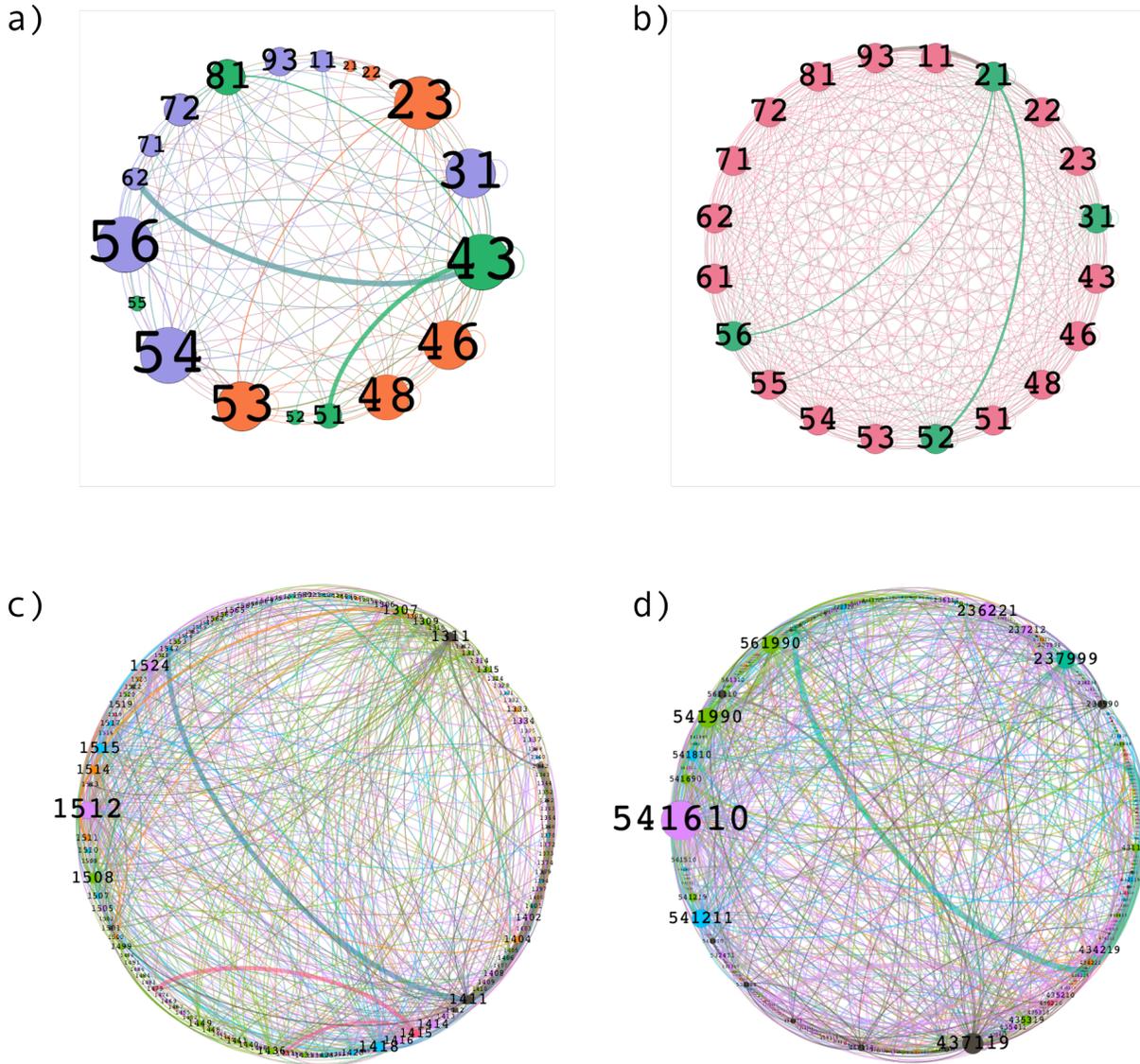


Figura 6: Redes de emisiones de CFDI entre a) sectores, c) ramas y d) actividades económicas asociadas a EFOS y al resto de la población. Las redes entre b) sectores económicos asociadas a las emisiones de RFCA no clasificados son más homogéneas que las asociadas a EFOS. Los números que etiquetan a los nodos en las figuras corresponden a las claves numéricas que identifican a los sectores, ramas y actividades económicas en la base de datos que nos fue proporcionada, mientras que el tamaño de los nodos es proporcional a su grado (número de enlaces que parten de o llegan a cada nodo). En a) destacan los sectores, 23, 43, 46, 48, 53, 54 y 56, que corresponden a los sectores de: Construcción, Comercio al por mayor, Comercio al por menor, Transportes, correos y almacenamiento, Servicios inmobiliarios y de alquiler de bienes muebles e intangibles, Servicios profesionales, científicos y técnicos, y Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación. En c) destacan las ramas 1508, 1512, 1514, 1515, 1524, asociadas a: Servicios de Contabilidad, auditoría y servicios relacionados, Servicios de consultoría administrativa, científica y técnica, Otros servicios profesionales, científicos y técnicos, y Otros servicios de apoyo a los negocios. Finalmente en d) destacan las actividades, 237999, 437119, 541211, 541610, 541990 asociadas a: Otras construcciones de ingeniería civil u obra pesada, Otros intermediarios de comercio al por mayor, Servicios de contabilidad y auditoría, Servicios de consultoría en administración, y otros servicios profesionales, científicos y técnicos.

Actividades Económicas con mayor centralidad		
Grado de Entrada	Grado de Salida	<i>Betweenness</i>
Servicios de consultoría en administración	Servicios de consultoría en administración	Servicios de consultoría en administración
Otros intermediarios de comercio al por mayor	Otros servicios profesionales, científicos y técnicos	Administración y supervisión de construcción de otras obras de ingeniería civil u obra pesada
Otros servicios de apoyo a los negocios	Otros intermediarios de comercio al por mayor	Otras construcciones de ingeniería civil u obra pesada
Otros servicios profesionales, científicos y técnicos	Servicios de contabilidad y auditoría	Construcción de inmuebles comerciales, institucionales y de servicios
Construcción de inmuebles comerciales, institucionales y de servicios	Otras construcciones de ingeniería civil u obra pesada	Comercio al por mayor de equipo y accesorios de cómputo

Tabla 2: Actividades económicas que presentan los valores más altos de las medidas de centralidad consideradas. Se puede observar como dichas actividades están relacionadas en su mayoría con la prestación de servicios y su descripción es ambigua o muy general.

dentro de su objeto social o como actividad preponderante la prestación de servicios, o bien la preferencia por utilizar este tipo de actividades responde a que, debido a su ambigüedad en su descripción y al no estar relacionadas con la producción o venta de bienes tangibles, pudiera permitir con mayor holgura la posibilidad de no tener que comprobar, por ejemplo, activos en una auditoría.⁷ Sin embargo, es necesario realizar una investigación más detallada de las regulaciones y normativas para tener un mejor entendimiento de este comportamiento.

6.1.3. Comparación de montos en emisiones de CFDI

Adicional a la caracterización de las actividades económicas en las redes de emisión de CFDI por medio de las medidas de centralidad, también se estudiaron los montos promedio observados para sustentar evidencia de que las EFOS realizan transacciones con montos significativamente mayores a los del resto de la población. Esto se realizó comparando los montos promedio por operación de las emisiones de las EFOS (los cuales asignamos como peso a los enlaces dentro de las redes de emisión de CFDI⁸), con los observados en 100 muestras aleatorias de RFCA no clasificados con tantos elementos como el número de EFOS proporcionadas (10,678 RFCA entre presuntas y definitivas).

Este conjunto de redes, que interpretamos como “el resto de la población”, es nuestro punto de comparación con respecto al cual identificamos posibles desviaciones, las cuales asociamos a un comportamiento característico de EFOS y que podrían ser utilizadas para el monitoreo y detección de CFDI con montos sospechosos, como un elemento dentro de una estrategia eficiente de detección de evasores o transacciones potencialmente anómalas.

Cada una de las redes de emisión de CFDI resultantes del proceso descrito anteriormente tiene asociado un conjunto de pesos (que representan los montos promedio por operación entre pares de actividades económicas) que nos permite, considerando al mismo par de nodos i, j y el peso del enlace entre ellos en cada una de las 100 redes, calcular el valor promedio del peso de dicho enlace, μ_{ij} , definido como:

⁷Cabe mencionar que a partir de 2014, pueden estar sujetos a la aplicación del artículo 69 B, al no contar con los activos, infraestructura o capacidad material, directa o indirecta para proporcionar los servicios que amparen los CFDI; o bien, al encontrarse como no localizado.

⁸A cada enlace entre cualquier pareja de nodos i y j se le asocia el peso, ω_{ij} , el cual definimos como el cociente del agregado total de los montos subtotaes de las emisiones mensuales de CFDI, m_{ij} , entre el total de facturas activas emitidas, f_{ij} , el cual corresponde al monto promedio por factura, es decir: $\omega_{ij} = m_{ij}/f_{ij}$.

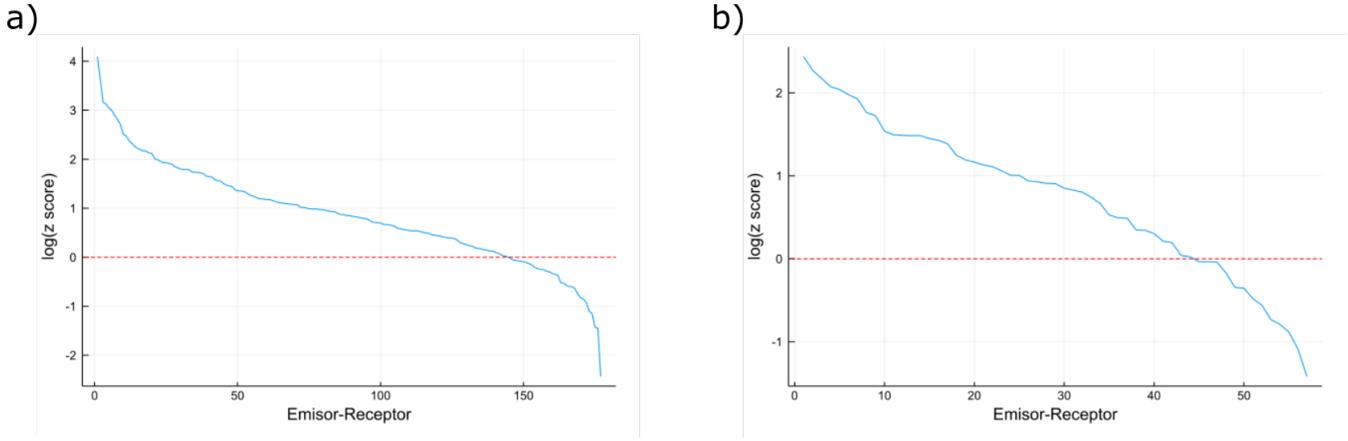


Figura 7: Comparación de características de los enlaces en las redes de actividades económicas asociadas a EFOS y al resto de la población. El eje horizontal corresponde a parejas emisor - receptor, en este caso, correspondientes a actividades económicas de cualquier sector. Los datos se encuentran ordenados de forma descendente. El eje vertical corresponde al logaritmo del z -score para cada una de las parejas de actividades. Se puede observar como un conjunto de actividades (aquellas por encima de la línea horizontal punteada) realiza operaciones con montos mayores comparados con el resto de la población. La figura a) corresponde a Junio de 2017 mientras que la figura b) corresponde a Noviembre de 2018, sin embargo este comportamiento es consistente para todos los meses.

$$\mu_{ij} = \frac{1}{N_{ij}} \sum_{k=1}^{100} \omega_{ij}^k, \quad (1)$$

donde ω_{ij}^k es el peso del enlace entre los nodos i y j de la red de emisión entre actividades económicas de la k -ésima red de referencia, y N_{ij} es el número total de redes en las que $\omega_{ij} \neq 0$. Así mismo calculamos la desviación estándar, σ_{ij} , de este mismo conjunto de valores.

Para comparar los pesos de los enlaces entre actividades económicas observadas en las redes basadas en las emisiones de las EFOS con los de las redes del resto de la población, hacemos uso del z -score, el cual es una forma de medir la desviación de un elemento con respecto al valor promedio de un conjunto de elementos en unidades de la desviación estándar de los valores en el conjunto, el cual calculamos para cada enlace entre cualquier par de nodos i, j , representamos como z_{ij} , y está definido por:

$$z_{ij} = \frac{\omega_{ij} - \mu_{ij}}{\sigma_{ij}}, \quad (2)$$

de tal forma que valores $z_{ij} > 1$ nos indican que el peso ω_{ij} en las redes basadas en las emisiones de EFOS, corresponden a montos mayores a una desviación estándar del valor promedio observado en el resto de la población.

Como se muestra en la Fig. 7, observamos un conjunto de enlaces (parejas de actividades económicas) para los cuales el monto promedio por operación corresponde a valores mucho mayores que los observados para el resto de la población. Lo cual sugiere que las EFOS, además de realizar sus operaciones desde un conjunto de actividades económicas que puedan ser favorables para sus intereses, también realizan operaciones por montos más elevados que el promedio de los contribuyentes. En la Tabla 3 listamos las parejas de actividades económicas con los valores más altos de su z -score, z_{ij} , y que por lo tanto se desvían más del comportamiento del resto de la población.

En conjunto, este par de análisis, la caracterización de nodos asociados a actividades económicas por medio de

Actividades Económicas con mayor desviación del monto promedio	
Actividad Emisora	Actividad Receptora
Comercio al por mayor de madera	Otros servicios de orientación y trabajo social prestados por el sector público
Telefonía tradicional	Comercio al por mayor de equipo de telecomunicaciones, fotografía y cinematografía
Fabricación de bolsos de mano, maletas y similares	Comercio al por mayor de maquinaria y equipo para la industria manufacturera
Fabricación de otros productos de papel y cartón	Comercio al por mayor de madera
Confección de otros accesorios de vestir	Otros servicios de consultoría científica y técnica
Confección de otros accesorios de vestir	Administración y supervisión de construcción de vivienda
Otros servicios de apoyo a los negocios	Otros servicios profesionales, científicos y técnicos
Servicios de recepción de llamadas telefónicas y promoción por teléfono	Otros servicios relacionados con la contabilidad
Servicios de Arquitectura	Reparación y mantenimiento de otro equipo electrónico y de equipo de precisión

Tabla 3: Parejas de actividades económicas con una mayor desviación en su valor de z-score, z_{ij} , lo cual representa el uso de montos mucho mayores en los montos promedio de las operaciones entre estas actividades cuando están asociadas a las emisiones de EFOS, en comparación con los montos observados en el resto de la población.

sus medidas de centralidad, y la desviación de los pesos de los enlaces, que representan los montos promedio de las operaciones entre actividades económicas asociadas a las operaciones de las EFOS, con respecto al resto de la población, muestran un comportamiento característico de las EFOS que, si bien es necesario investigar a mayor profundidad, podría ser útil para la identificación de operaciones sospechosas.

6.1.4. Redes bipartitas de RFCA de personas morales y sus asociados

Como parte de la información que nos fue proporcionada se incluye un listado de RFCA que representan tanto a personas morales como a sus socios y asociados o representantes legales. Esta base de datos consta de un total de 2,012,442 RFCA, dentro de las cuales 1,161 están identificadas como EFOS (11 % del listado de EFOS).

Considerando a los socios, asociados y representantes legales como miembros de la misma categoría, la cual denominamos simplemente como **asociados**, y a las personas morales como una categoría aparte, construimos redes bipartitas en donde los nodos representan a RFCA dentro de alguna de estas dos categorías, y los enlaces representan los roles que desarrollan los RFCA en relación a la conformación de las personas morales. A diferencia de las redes basadas en los registros de emisiones de CFDI entre contribuyentes, estas redes no representan la actividad económica de los contribuyentes sino la composición y organización legal de las personas morales, por lo que consideramos que estas relaciones son más estrechas al representar un acuerdo expreso entre ambas partes con un efecto legal.

En primera instancia, nos centramos en las redes bipartitas compuestas por las personas morales que están clasificadas como EFOS, ya sean definitivas o presuntas. Como se puede observar en la Fig. 8, estas redes bipartitas se organizan en componentes de distintos tamaños y tipos, en las que las personas morales pueden presentarse de manera aislada al centro de una topología de tipo estrella, o de manera conjunta en componentes más grandes donde un conjunto de personas morales comparten algunos de sus asociados.

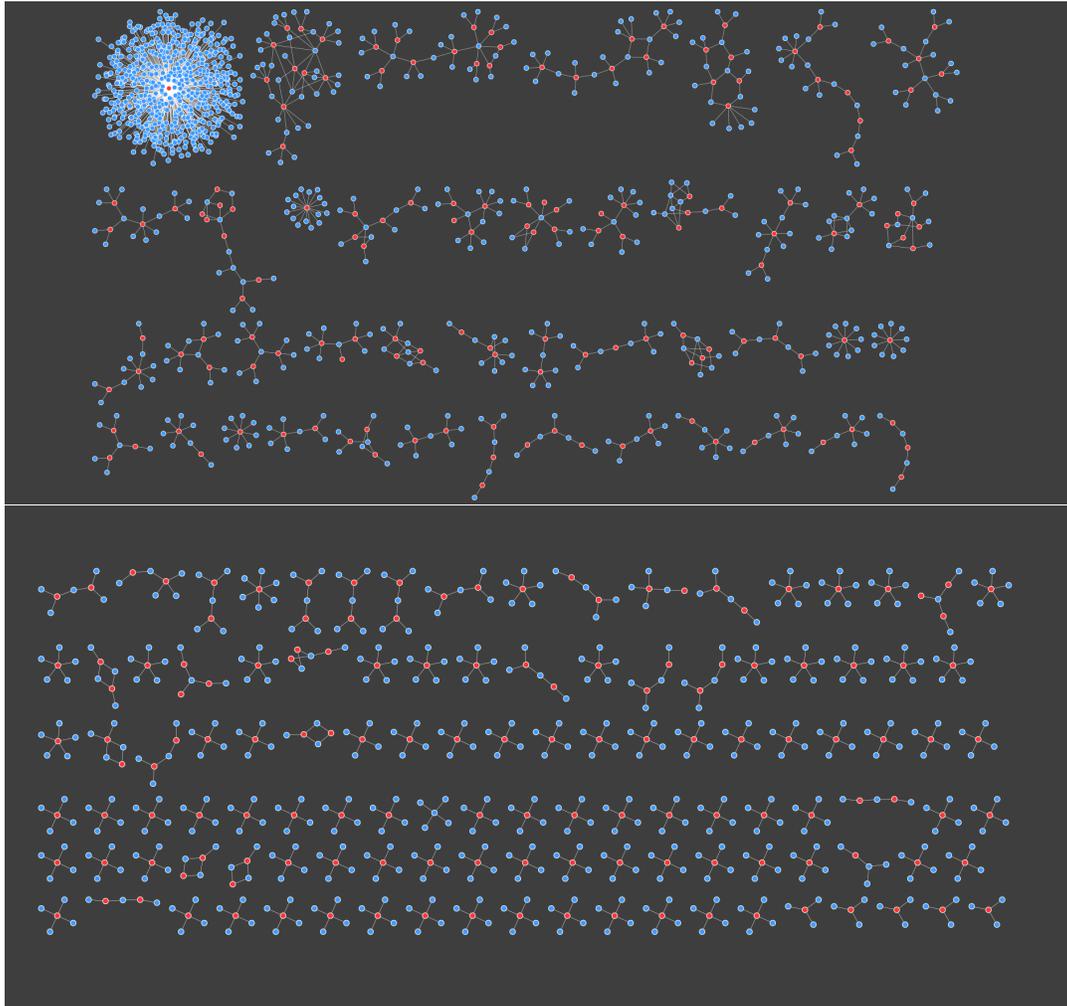


Figura 8: Ejemplos de las redes bipartitas observadas partiendo de las personas morales clasificadas como EFOS (nodos rojos) y sus asociados (nodos azules). Se pueden observar distintos tipos de topologías (patrones de conexión entre nodos) en las que hay casos en donde una sola persona moral tiene su propio conjunto de asociados y otros casos donde distintas personas morales comparten algunos de sus asociados formando redes más grandes.

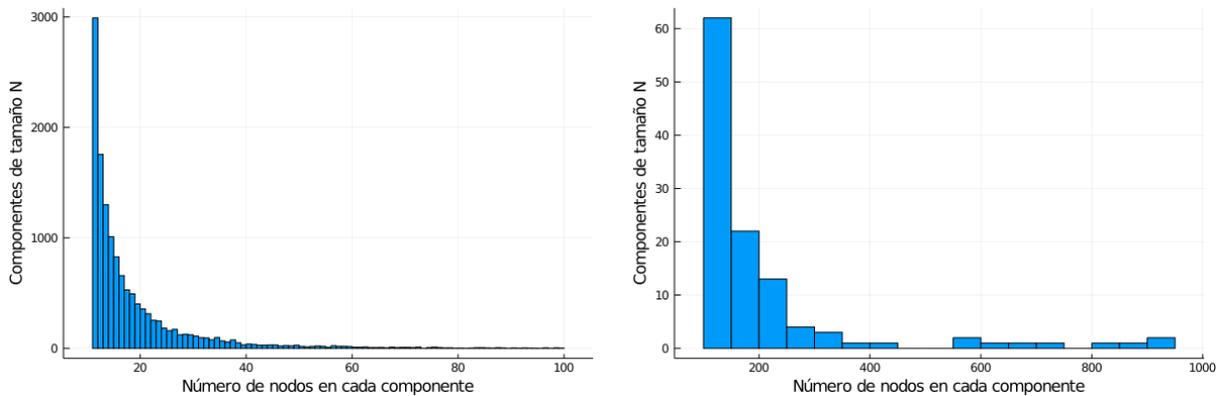


Figura 9: Distribución de tamaños de componentes observadas con el total de las relaciones entre personas morales y sus asociados en el intervalo $10 < N < 100$ (izquierda) y de tamaños $N > 100$ (derecha).

Es importante enfatizar que todas las personas morales que forman parte de las componentes mostradas en la Fig. 8, corresponden a RFCA ya identificados como EFOS, y que las componentes en las que se observa un conjunto de personas morales que comparten asociados, se pueden asociar a redes organizadas de RFCA que se han especializado en la constitución y operación de personas morales que se dedican a la emisión de facturas de operaciones simuladas.

Estas redes bipartitas centradas en personas morales clasificadas como EFOS son utilizadas en la Sección 6.2.3 como conjunto de entrenamiento de un modelo de Aprendizaje de Máquinas para la clasificación de redes bipartitas compuestas por RFCA no clasificados como sospechosas de presentar un comportamiento similar a los de las EFOS.

Si ahora, en vez de considerar solamente a las personas morales clasificadas como EFOS en el listado, tomamos en cuenta a la totalidad de ellas y construimos las redes bipartitas formadas por las personas morales y sus asociados, podemos observar que se organizan en componentes de distintos tamaños, cuya distribución se muestra en la Fig. 9 y se pueden observar ejemplos de las mismas en la Fig. 10. Cada una de estas componentes corresponde a un conjunto de RFCA que representan tanto a personas morales como asociados, los cuales se encuentran conectados entre ellos mediante un conjunto de personas morales, por medio de las relaciones entre sus asociados.

Aunque previamente consideramos a las personas morales clasificadas como EFOS de forma aislada para su caracterización inicial (véase la Fig. 8), estas no operan de forma independiente, sino que también participan en las redes bipartitas formadas por el total de los RFCA en el listado de personas morales y asociados⁹. En la Tabla 4 reportamos el número de componentes observadas y el número de EFOS presentes en cada una de ellas. Como se puede observar, debido principalmente a la gran diferencia entre el número total de RFCA (2,012,442) y el número de EFOS (1,161) en el listado, la gran mayoría de las componentes observadas no contienen EFOS, sin embargo pudimos identificar una componente en la que están presentes cerca del 40% de las EFOS en el listado. En la Fig. 10 mostramos algunos ejemplos de las componentes de las redes bipartitas observadas con presencia de EFOS al considerar el total de los RFCA en el listado de personas morales y asociados.

⁹Haciendo uso de las redes obtenidas considerando a la totalidad de los RFCA en el listado, realizamos el experimento de entrenar un modelo de Aprendizaje de Máquinas para realizar una clasificación de nodos individuales dentro de las redes bipartitas como sospechosos de ser EFOS, sin embargo, debido a la baja densidad de EFOS en este listado, no obtuvimos resultados favorables.

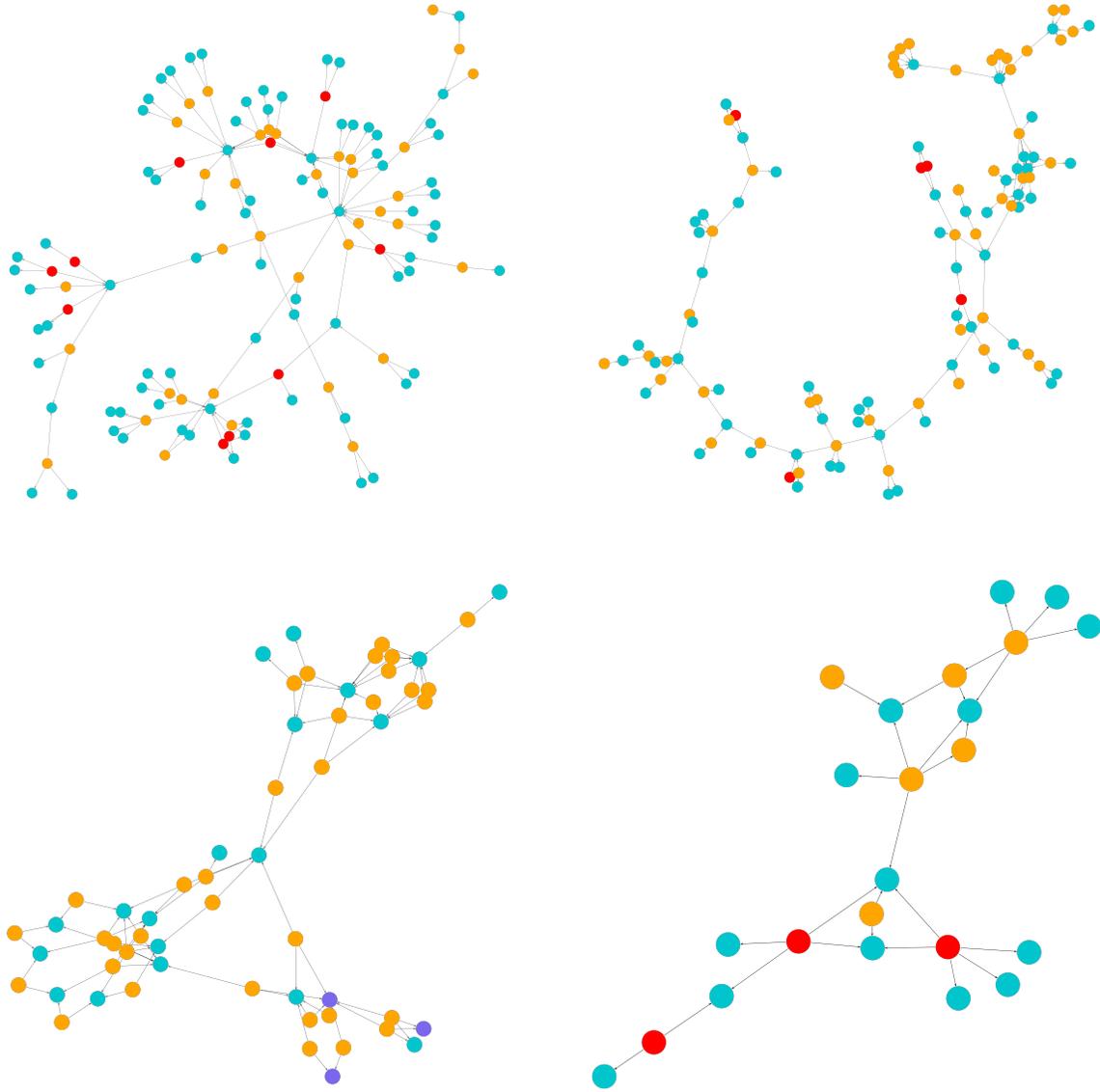


Figura 10: Ejemplos de las redes bipartitas observadas al considerar el conjunto completo de los RFCA en el listado de personas morales y asociados. Los nodos amarillos corresponden a personas morales no etiquetadas como EFOS, mientras que los rojos corresponden a personas morales EFOS, de forma similar los nodos azules corresponden a RFCA asociados no etiquetados como EFOS mientras que los morados representan RFCA asociados presentes en el listado de EFOS que nos fue proporcionado.

Presencia de EFOS en las componentes de las redes bipartitas	
Número de Componentes	Número de EFOS en la componente
409,880	0
578	1
36	2
9	3
1	4
1	5
1	10
1	465

Tabla 4: Número de EFOS observadas en las componentes bipartitas al considerar la totalidad de personas morales y asociados en el listado proporcionado por el SAT.

6.2. Aprendizaje de máquinas

Las técnicas de aprendizaje de máquina construyen modelos basados en un conjunto de datos de ejemplo (datos de entrenamiento) que se conforman de una serie de variables de entrada o *features* y el valor de salida para cada uno de los casos. Los modelos resultantes “aprenden” de este conjunto de datos de entrenamiento para hacer predicciones consistentes con los valores de salida sin necesidad de tener las reglas de conversión explícitamente programadas, puede decirse por ende, que estos algoritmos mejoran automáticamente a través de la experiencia [14]. En nuestro caso, consideramos las características de EFOS publicadas por el SAT y las comparamos con las de RFCA no clasificados, para entrenar estos modelos con ejemplos de ambos casos.

En este proyecto se utilizaron dos técnicas de aprendizaje de máquina para detectar patrones específicos relacionados con EFOS. La primer técnica utilizada fue un modelo tipo *Gradient Boosting Trees* [15] [16] para detección de EFOS individuales y entrenado en sus características no fiscales, y el segundo es un modelo de *Redes Neuronales de Grafos* [17] [18] para la detección de redes de socios, asociados y representantes legales relacionados con la creación de EFOS personas morales.

6.2.1. Modelo *Gradient Boosting Trees* (GBT)

Gradient Boosting o *Potenciación del gradiente* es una técnica de aprendizaje de máquina utilizada para problemas de regresión y clasificación, que produce un modelo de predicción en la forma de un ensamble o suma de modelos más débiles, típicamente árboles de decisión [15]. El algoritmo empieza entrenando un primer modelo simple a los datos de entrenamiento, después se construye un segundo modelo que se enfoque en predecir el error del primer modelo. La suma de ambos modelos nos da una predicción más acertada del valor objetivo real. Este proceso se repite muchas veces, en cada paso entrenando un modelo que prediga el error de la suma de todos los anteriores, hasta que el error del modelo combinado no se reduzca más en el conjunto de datos de prueba.¹⁰ Este algoritmo es considerado uno

¹⁰Específicamente consideremos un algoritmo de potenciación de gradiente en un problema de regresión con M etapas entrenado sobre un conjunto de n observaciones x que usamos para predecir un conjunto de respuestas y . Durante cada etapa m ($1 \leq m \leq M$) supondremos un modelo imperfecto F_m (por ejemplo, para un valor de m pequeño, este modelo podría simplemente regresar el valor medio de los valores observados y : $\hat{y}_i = \bar{y}$). Para mejorar F_m , el algoritmo debe añadir un nuevo estimador $h_m(x)$. Por lo tanto,

$$F_{m+1}(x) = F_m(x) + h_m(x) = y$$

o, de forma equivalente,

$$h_m(x) = y - F_m(x)$$

. Es decir, el método de potenciación del gradiente calcula h_m como el *residual* o error de la etapa m , $y - F_m(x)$. Esto significa que cada nuevo modelo F_{m+1} intenta corregir el error de su predecesor F_m

de los estándares de la industria para el entrenamiento con datos no homogéneos en formato tabular.

Existen una gran variedad de implementaciones de esta técnica en librerías de código abierto. En este proyecto utilizaremos la implementación del algoritmo de la librería *Catboost*, desarrollada inicialmente por la empresa *Yandex* y publicada con librería de código abierto [16].

Se entrenó el modelo con todos los RFCA en la categoría de EFOS presuntas o definitivas (un total de 10,573 muestras) a los cuales se catalogó con la etiqueta "1" y con una muestra aleatoria de 50,000 RFCA no clasificados, a los cuales se les dio la etiqueta de "0". Se entrenó un modelo de clasificación binaria usando la clase *CatBoostClassifier* de la librería *CatBoost* que utiliza las siguientes características provenientes de la información categórica para cada RFCA:

- Situación fiscal
- Tipo de Persona (física o moral)
- Edad de operación (Calculada como el número de días transcurridos desde la fecha de inicio de operaciones a la actualidad)
- Entidad Federativa
- Municipio
- Clave de Actividad
- Clave de Rama
- Clave de Sector
- Si pertenece al sector automotriz
- Si pertenece al sector público

6.2.2. Resultados del modelo de identificación de EFOS

La figura 11 muestra lo que se conoce como la curva *ROC* (acrónimo de **Receiver Operating Characteristic**, o *Característica Operativa del Receptor*) del modelo, la cuál es una gráfica de la proporción de verdaderos positivos (PVP) frente a la proporción de falsos positivos (PFP) según se varía el umbral de discriminación del modelo (valor a partir del cuál decidimos que un caso es positivo).¹¹ Observamos que la curva generada por nuestro modelo en el conjunto de datos de prueba tiene un área bajo la curva (AUC) de 0.97, muy cercano al valor ideal de 1.0 que tendría un modelo de clasificación perfecta. Para poder tener un clasificador se necesita un punto de corte en la salida del modelo que se tomará como el umbral de discriminación, es decir, el valor a partir del cuál se clasificará a un RFCA como sospechoso. La figura 12 muestra la matriz de confusión del modelo con un valor de corte de 0.85 evaluada en el conjunto de datos de prueba. En esta matriz, las columnas representan los RFCA que el modelo clasificó en cada una de las categorías (0 para no sospechosos, 1 para sospechosos), mientras que las filas son los valores reales de estos RFCA. De esta manera, los valores de la diagonal corresponden a los casos clasificados de forma correcta, mientras que los errores se encuentran fuera de la diagonal. Se puede observar que del total de 3,793 RFCA en el

¹¹Esta curva ilustra la habilidad de diagnóstico de un sistema de clasificación binaria, con un clasificador perfecto ubicándose en la esquina superior izquierda (PVP igual a 1, PFP igual a 0) y un modelo aleatorio representado por una línea recta de pendiente 1, puesto que en este caso la proporción de verdaderos positivos sería igual a la de falsos positivos para todos los valores de discriminación.

conjunto de prueba, el modelo clasifica 518 como sospechosos de ser EFOS, siendo 446 de estos verdaderos EFOS y 72 falsos positivos. Esto da un porcentaje de precisión del 86 %. El reporte completo de clasificación se puede ver en la tabla 5.

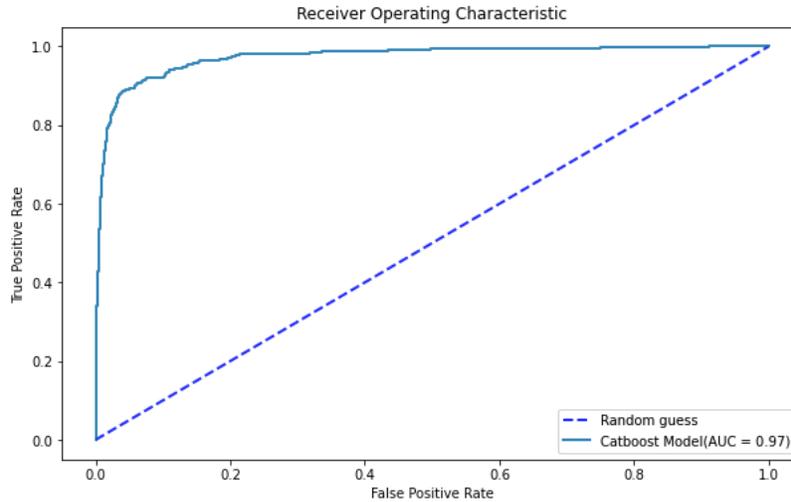


Figura 11: Curva ROC (Receiver Operating Characteristic) del modelo Catboost. Se grafica la proporción de falsos positivos (eje horizontal) contra la proporción de verdaderos positivos para el modelo a diferentes umbrales de discriminación. El modelo aleatorio (donde clasificamos un RFCA como EFOS o no EFOS puramente basados en el azar), se muestra como la línea punteada. Se observa un poder discriminatorio marcadamente superior al azar y cercano al punto ideal (esquina superior izquierda).

		Predicción del modelo		Total
		0	1	
Valores reales	0	3171	72	3243
	1	104	446	550
Total		3275	518	3793

Figura 12: Matriz de confusión del modelo Catboost. Las columnas representan los RFCA que el modelo clasificó como sospechosos (1) o no sospechosos (0), mientras que las filas son los valores reales de los RFCA. Se observa que de los 3,243 RFCA provenientes de la población de no sospechosos, el modelo solo clasificó 72 (2%) como sospechosos, mientras que de los 550 provenientes de la población de EFOS definitivas y presuntas, el modelo pudo identificar 446 (81%) como sospechosas. Puesto que sólo 72 de los que clasificó como sospechosos eran en realidad de la población de RFCA no clasificados, la precisión total del modelo es de 86 %.

Las figuras 13 y 14 muestran los valores de importancia relativa para las diferentes variables categóricas en el resultado final del modelo, producidos por el proceso **SHAP** (SHapley Additive exPlanations) [19]. La figura 13 muestra el valor absoluto de estos valores, mientras que la figura 14 extiende este análisis mostrando el efecto que cada variable tiene en la salida del modelo conforme sube o baja su valor.

Por ejemplo, observamos que la característica que el modelo considera más importante es la edad de operación (definida como el número de días desde la fecha de inicio de operaciones del RFCA). Al examinar con más detalle el efecto que tiene esta variable en el modelo, podemos ver que valores más bajos de la edad (RFCAs de reciente creación, color azul claro en la figura 14) tienen un efecto muy negativo en la salida del modelo, lo que indica que mueven el resultado más hacia una clasificación de no EFOS o, lo que es lo mismo, que el modelo no encontró muchos ejemplos de EFOS definitivas o presuntas de reciente creación. En el caso de los RFCAs más antiguos (edad de operación alta, color rojo en el diagrama) podemos ver que se concentran alrededor de la región central, indicando que esa característica

Reporte de Clasificación				
	precisión	recall	score f1	soporte
0	0.97	0.98	0.97	3243
1	0.86	0.81	0.84	550

Tabla 5: Reporte de Clasificación para el modelo de Catboost. Se muestra la precisión, recall, score f1 y soporte tanto para las clasificaciones de **0** (no sospechosos) como las de **1** (sospechosos) en el conjunto de prueba. La precisión es el radio de verdaderos positivos contra la suma de verdaderos positivos y falsos positivos. El recall o exhaustividad es el radio de verdaderos positivos contra la suma de verdaderos positivos y falsos negativos. El score f1 es la media armónica de la precisión y el recall, pesados de tal forma que el mejor score es 1.0 y el peor es 0.0. El soporte es el número de muestras de cada clase presentes en el conjunto de prueba (3242 RFCA no sospechosos y 550 de la población de EFOS definitivas y presuntas).

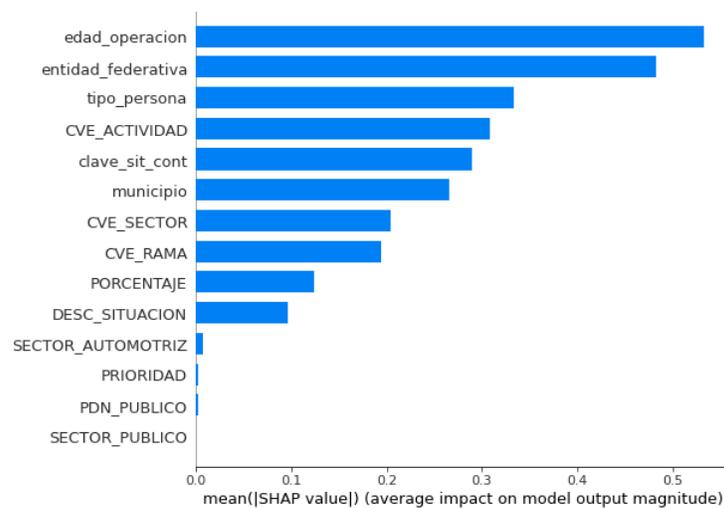


Figura 13: Valores SHAP de importancia absoluta para las variables en el modelo de Catboost. Observamos que la edad de operación del RFCA es la variable que el modelo considera como la más importante, seguida por la entidad federativa, el tipo de persona (física o moral) y finalmente, la clave de actividad.

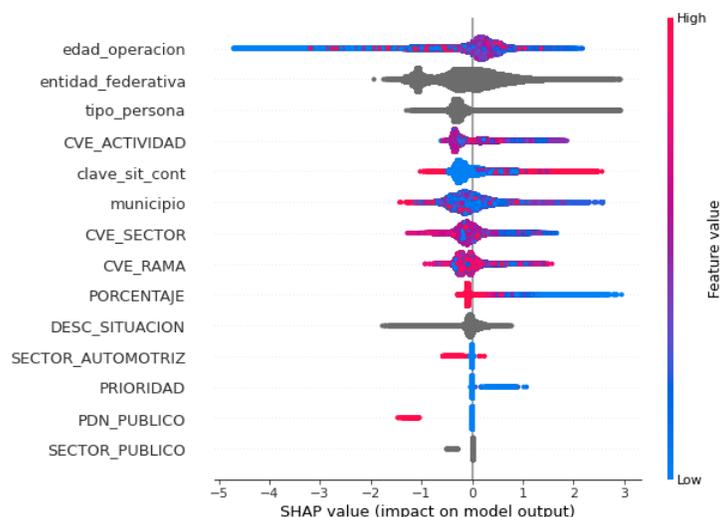


Figura 14: Impacto de las variables como función de sus valores correspondientes en el mismo orden de importancia que en la figura 13. La distribución a lo largo del eje horizontal muestra el efecto que diferentes valores de la variable (representados por el color) tienen sobre el resultado final del modelo, con valores hacia la derecha apuntando a una clasificación más sospechosa, mientras que los valores a la izquierda mueven el resultado hacia una clasificación menos sospechosa. Podemos ver que valores más bajos de la edad (RFCA de reciente creación) tienen un efecto muy negativo en la salida del modelo, lo que indica que mueven el resultado más hacia una clasificación de no sospechosos, mientras que valores de la edad de operación en el período 2008-2013 mueven el resultado final hacia una clasificación de sospechosos.

no ejerce un efecto muy significativo en la salida del modelo. Finalmente observamos un conjunto de valores de la edad de operación con coloración similar que mueven el resultado final hacia una clasificación de sospechosos. Estos valores se concentran principalmente en los RFCA que iniciaron operaciones en el período 2008-2013. Cabe hacer notar que la incorporación del artículo 69-B al Código Fiscal de la Federación, el cual se publicó en el Diario Oficial de la Federación el 9 de diciembre de 2013 e inició su vigencia a partir del 1° de enero de 2014, puede condicionar este supuesto.

6.2.3. Modelo *Redes Neuronales de Grafos* (DGCNN)

Las *Redes Neuronales Artificiales* (RNA) son un modelo de aprendizaje automático inspirado en la arquitectura de un cerebro. Formadas por una colección de unidades interconectadas, de manera semejante a como se conectan las neuronas en un cerebro, por lo que comúnmente se les conocen a estas unidades como *neuronas*. Cada neurona alterna su estado entre activa e inactiva de acuerdo a la información que recibe de las neuronas con las que está conectada. Alterando el peso de la interacción entre neuronas se cambia la manera en que se procesa la información. Es justamente mediante la modificación de dichos pesos como una red neuronal aprende a identificar patrones.

Para la clasificación de redes de constitución de personas morales se utilizó el algoritmo conocido como *Deep Graph Convolutional Neural Network*, o de redes neuronales convolucionales profundas de grafos, llamado así porque, a diferencia de una red neuronal convencional, que requeriría la inserción de nodos individuales de la red y el cálculo e inclusión de características de red agregadas para cada uno de ellos, permite la inclusión directa de redes completas como observaciones individuales y es capaz de aprender su arquitectura interna y de usar sus características de red para la clasificación de una etiqueta de red global.

La DGNN tiene tres etapas de procesamiento secuenciales: En la primer etapa, se introduce la red o grafo de entrada a una serie de *capas convolucionales de grafos*, propuestas por primera ocasión en [18]. Estas capas extraen las características de los vértices ¹² y las propagan entre sus vecinos, de forma que se pueda extraer información de subestructuras locales en la red de forma automática. Estos descriptores extraídos por las capas convolucionales se envían a una *capa SortPooling* cuyo objetivo es ordenarlos de forma consistente antes de procesarlos por una red neuronal convolucional clásica ¹³. Esta red convolucional tradicional lee la representación de los grafos ordenados y hace las predicciones finales ¹⁴.

Para el entrenamiento y la definición de etiquetas se utilizaron solamente redes bipartitas de más de 3 nodos, donde cada nodo corresponde a un RFCA y las conexiones se definen por relaciones con sus socios, asociados y representantes legales. Se les dio un peso mayor a las conexiones con representantes legales, lo que significa que el modelo vería esas conexiones como más importantes que las que se observan con socios y asociados. Se entrenó la red neuronal con un conjunto de 480 redes bipartitas centradas alrededor de EFOS definitivas o presuntas, a las que se les dio etiqueta 1; y con una muestra aleatoria de 480 redes típicas (cuyos RFCA miembros tenían clasificación desconocida) a las que se les dio la etiqueta 0. Incluimos también la misma información categórica para cada nodo que utilizamos para el entrenamiento del modelo GBT en la sección 6.2.1, a saber:

- Situación fiscal
- Tipo de Persona
- Edad de operación
- Entidad Federativa
- Municipio
- Clave de Actividad
- Clave de Rama
- Clave de Sector
- Si pertenece al sector automotriz
- Si pertenece al sector público

6.2.4. Resultados del modelo *DGNN*

La figura 15 muestra la curva ROC del modelo DGCNN para la clasificación de redes bipartitas. Observamos que el modelo demuestra una capacidad de predicción marcadamente mayor al azar, con un área bajo la curva (AUC) de 0.87.

En la figura 16 se muestra la matriz de confusión para un valor de corte de 0.65 (el valor a partir del cual se clasifica la red como sospechosa). Esta matriz está calculada sobre un conjunto de datos de prueba separados de los datos de

¹²En el caso del presente estudio, cada vértice corresponde a un RFCA y su información categórica

¹³La pregunta es, en qué orden debemos poner los vértices. En el caso de clasificación de imágenes, por ejemplo, los píxeles están ordenados de forma natural por su ubicación espacial. En clasificación de textos podemos ordenar las palabras por orden alfabético. Para redes, los vértices pueden ser ordenados de acuerdo a sus *roles estructurales* dentro del grafo (niveles de conectividad, centralidad, etc., véase la sección 6.1)

¹⁴Utilizamos la arquitectura propuesta en [17] y [18] usando la implementación de código abierto de la librería *StellarGraph* [20].

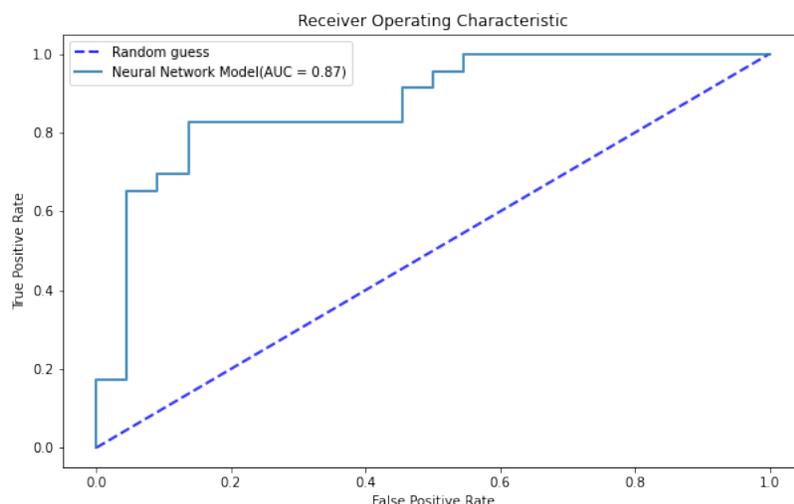


Figura 15: Curva ROC del modelo de Redes Neuronales. Se grafica la proporción de falsos positivos (eje horizontal) contra la proporción de verdaderos positivos para el modelo a diferentes umbrales de discriminación. El modelo aleatorio (donde clasificamos un RFCA como EFOS o no EFOS puramente basados en el azar), se muestra como la línea punteada. Se observa un poder discriminatorio marcadamente superior al azar con score de AUC (área bajo la curva) de 0.87.

		Predicción del modelo		Total
		0	1	
Valores reales	0	21	1	22
	1	14	9	23
Total		35	10	45

Figura 16: Matriz de confusión del modelo de Redes Neuronales. Las columnas representan las redes bipartitas que el modelo clasificó como sospechosas (1) o no sospechosas (0), mientras que las filas son los valores reales de estas redes. Se observa que de las 22 redes provenientes de la población de no sospechosas, el modelo solo clasificó 1 (5%) como sospechosa, mientras que de los 23 provenientes de la población de redes centradas en EFOS definitivas o presuntas, el modelo pudo identificar 9 (30%) como sospechosas. Puesto que sólo 1 de las que clasificó como sospechosas eran en realidad de la población de redes centradas en RFCA no clasificados, la precisión total del modelo es de 90%.

entrenamiento del modelo. Se puede observar que, del total de 45 redes en el conjunto de prueba, el modelo clasifica 9 como sospechosas, siendo 8 de estas verdaderas redes centradas alrededor de EFOS y 1 falso positivo. Esto da un porcentaje de precisión del 90%. El reporte completo de clasificación se puede ver en la tabla 6.

Reporte de Clasificación				
	precisión	recall	score f1	soporte
0	0.60	0.95	0.74	22
1	0.90	0.39	0.55	23

Tabla 6: Reporte de Clasificación para el modelo de redes neuronales de grafos. Se muestra la precisión, *recall*, score f1 y soporte tanto para las clasificaciones de **0** (no sospechosas) como las de **1** (sospechosas) en el conjunto de redes de prueba. La precisión es el radio de verdaderos positivos contra la suma de verdaderos positivos y falsos positivos. El *recall* o exhaustividad es el radio de verdaderos positivos contra la suma de verdaderos positivos y falsos negativos. El score f1 es la media armónica de la precisión y el *recall*, pesados de tal forma que el mejor score es 1.0 y el peor es 0.0. El soporte es el número de muestras de cada clase presentes en el conjunto de prueba (22 redes centradas en RFCA no clasificados y 23 provenientes de la población de redes centradas en EFOS definitivas y presuntas)

Podemos observar que aunque el *recall* (la proporción de redes verdaderas centradas en EFOS que el modelo logra identificar) es relativamente bajo (39%) esto se compensa con el hecho de que la precisión del modelo para las redes que clasifica como sospechosas es del 90% en este conjunto de datos de prueba. Es decir, que aunque el modelo no logre identificar a todas las redes centradas alrededor de EFOS que existen, se puede tener un mayor nivel de confianza en las que sí se identifican como sospechosas. Se puede obtener un mayor valor de *recall* ajustando el valor de corte de clasificación para este modelo, pero esto forzosamente vendría a expensas de reducir la precisión del modelo.

7. Integración de los métodos

Las metodologías utilizadas en este estudio caen dentro de dos grandes categorías: métodos de análisis de la ciencia de redes, y modelos de aprendizaje de máquinas (Machine Learning).

Por medio de la primera, realizamos una descripción de las redes de emisión de CFDI entre RFCA o entre actividades económicas mediante una agregación de los registros entre contribuyentes por medio de su información categórica, y de las redes entre PM y sus asociados, que representan la constitución legal de empresas en México; y por medio de la segunda, implementamos algoritmos de clasificación de RFCA que generalizan el comportamiento de EFOS, y que nos permiten identificar sospechosos dentro del total de RFCA que forman parte de las bases de datos.

Estas aproximaciones, aunque operan en las bases de datos de manera independiente, están vinculadas por medio de las observaciones y resultados obtenidos por cada uno de ellos, ya que motivan la exploración e implementación de algoritmos que complementan la visión del fenómeno desde cada enfoque.

De forma específica, la observación de una sobre representación de EFOS en ciertos sectores económicos motivó, tanto a la implementación del modelo de *Gradient Boosting Trees* para la identificación de RFCA sospechosos haciendo uso exclusivamente de la información categórica de los contribuyentes (véase la Sección 6.2.1), como a la construcción de redes de emisión de CFDI entre actividades económicas (véase la Sección 6.1.2). Lo cual nos ha permitido identificar no solamente un conjunto de sectores económicos en los que se observan EFOS con mayor frecuencia, que además corresponden a sectores identificados como vulnerables asociados a operaciones con recursos de procedencia ilícita (véase la Sección 3.4), sino también a las principales actividades económicas que utilizan para realizar operaciones,

por medio del cálculo de medidas de centralidad de los nodos y la caracterización de los enlaces en dichas redes, que representan a las actividades económicas y las operaciones entre ellas respectivamente.

Así mismo, el análisis de las redes bipartitas (véase la Sección 6.1.4), el cual pudo realizarse gracias a la base de datos de PM y sus socios y asociados y representantes legales que nos fue proporcionada, nos permitió en primer lugar, identificar como las PM dentro del listado proporcionado se organizan en componentes de distintos tamaños, i.e. estas no se observan de manera aislada sino que existen grupos de PM relacionadas entre ellas por medio de sus socios y asociados y representantes legales, dentro de las cuales pudimos observar la distribución de las EFOS previamente identificadas por el SAT. En segundo lugar, este análisis también nos permitió identificar redes bipartitas compuestas en su totalidad por PM ya clasificadas como EFOS, las cuales es posible asociar con redes organizadas de RFCA dedicados tanto a la constitución como a la operación de PM asociadas a la simulación de operaciones. Estos hallazgos resuenan con lo expresado en el Marco Legal de este estudio (véase la Sección 3), en específico lo relacionado con el art. 27 del CFF, apartado B fracción VI (véase la Sección 3.3.2), en el cual se expresan reformas con el objetivo de contar con mayor información respecto a la constitución de personas morales, debido a que la ausencia de la misma dificulta la detección de operaciones inexistentes. Como mostramos en este estudio, las herramientas de la ciencia de redes son útiles para la descripción de este tipo de relaciones las cuales, de la mano de acciones legales, pueden eficientizar la detección de evasores mediante su relación con EFOS ya identificados. Así mismo, en el mismo artículo se menciona como las personas morales a partir de 2020, tienen la obligación de notificar cambios en sus socios y accionistas. Un registro eficiente y monitoreo de esta información con las técnicas utilizadas en este estudio o similares, permitiría identificar conductas anómalas o sospechosas y estudiar la dinámica de la composición legal de las empresas al paso del tiempo.

La identificación de estas componentes alrededor de PM EFOS, motivó la implementación de un modelo de aprendizaje de máquinas capaz de no solamente realizar una clasificación de RFCA individuales, sino de redes bipartitas completas, que generaliza las propiedades topológicas (patrones de conexión entre nodos) observadas en las redes bipartitas asociadas a EFOS, el cual toma como conjunto de entrenamiento las redes bipartitas asociadas a EFOS identificadas por medio del análisis basado en ciencia de redes. El modelo utilizado para realizar esta clasificación (véase la Sección 6.2.3) es una técnica muy reciente con la que hemos obtenidos buenos resultados y que consideramos merece la pena seguir explorando.

Resultado tanto de los análisis basados en ciencia de redes como de los modelos de clasificación de RFCA sospechosos de ser EFOS basados en Machine Learning, hemos podido identificar sectores y actividades económicas a los que pertenecen los RFCA sospechosos. Una característica recurrente de estos sectores y actividades económicas, los cuales se listan en las Tablas 11 y 12 del Anexo, es que están relacionadas con mayor frecuencia a la prestación de servicios y la comercialización de intangibles, esta observación resuena con lo expresado en las páginas XCII a XCIII del paquete económico para 2020, entregado el 8 de septiembre de 2019, en el cual haciendo referencia a las empresas que emiten, comercializan y usan comprobantes fiscales de operaciones inexistentes, parafraseando el texto en dicho documento se expresa que, estas empresas cuentan con un objeto social amplio para poder ofrecer CFDI acordes a una amplia gama de actividades de los contribuyentes que adquieren estos comprobantes, así como que las principales actividades de estas empresas están relacionadas con: la prestación de servicios, subcontratación laboral, consultoría, capacitación, investigación, entre otros; en concordancia con los resultados obtenidos en nuestra investigación.

Este estudio nos ha permitido obtener una mejor descripción del comportamiento de las EFOS y obtener un listado de RFCA sospechosos de ser EFOS, el cual utilizamos en la siguiente sección para obtener estimados de evasión de ISR en distintos ejercicios fiscales.

8. Cálculos de evasión

Como resultado de la aplicación de los métodos de clasificación basados en aprendizaje de máquinas, obtenemos un listado de RFCA sospechosos de ser EFOS por cada uno de ellos¹⁵. Cada método operó de manera independiente sobre las bases de datos, basados en conjuntos independientes de la información de los RFCA: información categórica e información de la constitución legal de PM. En la Tabla 7 se listan los números de RFCA sospechosos identificados por cada modelo, así como los que fueron clasificados como sospechosos por ambos.

Número de RFCA sospechosos identificados	
Modelo de Catboost basado en información categórica	100,000
Modelo de Clasificación de Redes Bipartitas	19,653
RFCA en la intersección de los listados	4,026

Tabla 7: Número de RFCA sospechosos identificados por cada método de clasificación basado en aprendizaje de máquinas y en la intersección de los listados.

Del total de los RFCA clasificados como sospechosos por ambos modelos, 3,784 corresponden a RFCA asociados a Personas Morales, mientras que 242 corresponden a Personas Físicas. Este conjunto de RFCA es el que utilizamos como punto de partida para el cálculo de los estimados de evasión de ISR que presentamos en este documento.

Tomando en cuenta la información disponible de las emisiones de CFDI entre contribuyentes, en un inicio propusimos un acercamiento a la estimación de la evasión en ISR, considerar la tasa del 30 % a los importes totales del CFDI de los RFCA sospechosos PM y aplicar una tasa promedio del 20 % a los RFCA sospechosos PF, ya que la tarifa aplicable tiene un excedente máximo del 35 %.

La estimación de la evasión en ISR considerando los CFDI se aplicaría en el Grupo 1 de PM (Régimen General de Ley de Personas Morales), conforme al artículo 9 de la Ley de ISR para llegar al Resultado Fiscal y aplicarle la tasa del 30 %. Sin embargo, en esta alternativa no contábamos con los diversos CFDI para llegar a una estimación del ISR, ya que como indicamos anteriormente, los ingresos acumulables del ejercicio no están dados por los importes de los CFDI expedidos, hay partidas en las que se acumula, no el total, sino sólo la ganancia, es el caso de enajenaciones de activos fijos, terrenos y partes sociales, así como, el ajuste anual por inflación acumulable que no está representado en un CFDI. De igual manera, la identificación de inconsistencias entre los montos de ingreso obtenido registros de CFDI y los reportados en las bases de datos de declaraciones anuales, no nos permitieron explorar esta alternativa.

Como segunda alternativa para determinar de manera estimada la evasión en ISR, partimos de los EFOS para llegar a los presuntos EDOS conforme a los CFDI expedidos por los primeros, la hipótesis planteada fue que los EDOS ya tienen sus operaciones efectuadas durante el ejercicio, pero al ver que el resultado fiscal, o bien la base sobre la cual se determina el ISR es alta, acuden a los EFOS para adquirir CFDI para reducirla; con lo que, la base para la estimación de la evasión del ISR será la determinada sobre la base de los importes de esos CFDI recibidos por los EDOS. Esta alternativa tuvo que ser descartada debido a que no contamos con un listado de EDOS, de la forma con la que contamos con un listado de EFOS, aunado a que el ejercicio que realizamos para la posible identificación de EDOS, por medio de la identificación de enlaces recurrentes alrededor de EFOS, no fue concluyente. Sin embargo, cabe hacer notar que un estudio a mayor detalle de la formación y características de dichos enlaces apoyado, ya

¹⁵El modelo de *Catboost* (Sección 6.2.1) identificó en primera instancia a 1,141,880 RFCA como sospechosos, de los cuales solamente 502,814 RFCA realizaron emisiones de CFDI en el periodo 2015–2019. Para realizar un filtrado adicional de este conjunto, consideramos a los que fueron clasificados como posibles EFOS por el modelo con una probabilidad mayor a 0.99, resultando en un total de 100,000 RFCA sospechosos.

sea con un listado de EDOS o un conjunto de características de ellos validado por el SAT, permitiría incorporar el comportamiento de EDOS y complementar la exploración realizada en este estudio.

Como tercera alternativa y como hemos discutido en la sección referente al marco legal en el que se basa este estudio (véase la Sección 3) el ISR se calcula tomando como base la utilidad fiscal o gravable obtenida por cada contribuyente resultado de sus actividades comerciales, a la cual se le aplica la tasa estatutaria vigente durante el ejercicio. Dado que la determinación de la base para el pago de ISR de cada contribuyente sería un procedimiento complicado y para el cual no contamos con la información suficiente para realizar un cálculo detallado, hemos tomado la decisión de realizar una aproximación de la utilidad generada por cada uno de los RFCA sospechosos, partiendo de los ingresos reportados por cada uno de ellos para cada ejercicio en las bases de datos de declaraciones anuales que nos fueron proporcionadas, aplicando las tasas de utilidad expresadas en el art. 58 del CFF (véanse la Sección A.5 y la Tabla 10 del Anexo). Si bien dichas tasas de utilidad en el art. 58 del CFF hacen referencia a las actividades económicas de los contribuyentes, hemos realizado una clasificación de estas actividades en sectores económicos, identificando especialmente aquellos a los que pertenecen los RFCA sospechosos identificados, en las Tablas 11 y 12 listamos la distribución de RFCA sospechosos en los distintos sectores y actividades económicas, de tal forma que en vez de aplicar las tasas por actividad, lo hacemos por sectores económicos. Esta decisión es meramente operativa, ya que nos facilita tanto la clasificación como el procesamiento de los datos.

Una vez que hemos obtenido la utilidad estimada de cada RFCA sospechoso para cada ejercicio por medio de la información reportada en las bases de datos de declaraciones anuales y la aplicación de las tasas de utilidad por sector del art. 58 del CFF, aproximamos la recaudación potencial del ISR de cada uno de los RFCA sospechosos aplicando una tasa del 30%, conforme lo establece el artículo 9 de la Ley del Impuesto sobre la Renta, en el caso de los RFCA sospechosos Personas Morales, y una tasa del 20%, como estimación promedio, con base en la tarifa del artículo 152 de la Ley del Impuesto sobre la Renta, para Personas Físicas.

Ya que obtuvimos la recaudación potencial, comparamos dicho monto con el ISR efectivamente pagado reportado en las declaraciones anuales de los contribuyentes, lo cual nos permite definir el estimado de evasión individual de cada uno de los RFCA sospechosos como la diferencia entre el ISR potencial menos el ISR reportado en su declaración anual. Este cálculo nos permite obtener los estimados de evasión de ISR resultado del análisis realizado en este estudio, los cuales reportamos en la Tabla 8 para los RFCA sospechosos PF y en la Tabla 9 para los sospechosos PM.

Estimados de evasión de ISR de RFCA sospechosos PF					
Ejercicio	Ingreso (MDP)	ISR CFF (MDP)	ISR Declarado (MDP)	Evasión ISR (MDP)	Evasión Promedio ISR (MDP, RFCA)
2016	242	10	8	2	0.04 (43)
2017	420	15	14	2	0.04 (46)
2018	433	18	11	7	0.15 (49)
2019	171	12	1	11	0.35 (33)

Tabla 8: Estimados de evasión de ISR en millones de pesos (MDP) para cada ejercicio, asociados a los RFCA sospechosos Personas Físicas identificados en el estudio. No se muestra el ejercicio 2015 en la tabla debido a la ausencia de información para los RFCA sospechosos en la base de datos de declaraciones anuales. Los números entre paréntesis en la última columna corresponden al número de RFCA a los que corresponden los datos en cada ejercicio.

Cabe hacer notar que en la Tabla 9 no listamos el monto de ISR declarado ya que esta información no se encuentra reportada en las declaraciones anuales de estos RFCA, a diferencia de los RFCA PF sospechosos (véase la Tabla 8) en donde si encontramos montos de ISR efectivamente pagado reportado en las declaraciones anuales. Así mismo, no reportamos estimado de evasión para el ejercicio 2015 en el caso de personas físicas, debido a la ausencia de

Estimados de evasión de ISR de RFCA sospechosos PM			
Ejercicio	Ingreso (MDP)	Evasión ISR (MDP)	Evasión Promedio ISR (MDP, RFCA)
2015	20,787	1,993	1.0 (2261)
2016	32,816	3,085	1.1 (2868)
2017	50,167	4,669	1.6 (3010)
2018	70,501	6,491	2.2 (2957)
2019	1,166	113	1.5 (78)

Tabla 9: Estimados de evasión de ISR asociados a los RFCA sospechosos Personas Morales identificados en el estudio. En este caso no mostramos el ISR declarado debido a la ausencia de esta información en la base de datos de declaraciones anuales. Los números entre paréntesis en la última columna corresponden al número de RFCA a los que corresponden los datos en cada ejercicio. La disminución que se observa en 2019 se debe a que, para dicho ejercicio, solamente se cuenta con registros de emisiones de CFDI de 78 de los 4,026 RFCA sospechosos, a diferencia de los años anteriores para los cuales se cuenta con aproximadamente 2,500 RFCA sospechosos por ejercicio.

información de declaraciones anuales en dicho ejercicio para los RFCA PF sospechosos.

Los estimados de evasión mostrados en las Tablas 8 y 9 corresponden a cotas inferiores de la posible evasión de ISR realizada por los RFCA sospechosos que hemos identificado. Esto debido a que hemos considerado solamente una fracción de sus ingresos como la base gravable para la determinación del impuesto por medio de las tasas de utilidad del art. 58 del CFF. Sin embargo, si hacemos la suposición de que el total de sus ingresos acumulables, autodeterminados por cada RFCA en sus declaraciones anuales, corresponde a operaciones inexistentes o que este total fue utilizado por EDOS para la evasión de impuestos, y lo tomamos como la base gravable para el cálculo del ISR, obtenemos un estimado de evasión anual promedio por RFCA sospechoso de 1.4 MDP para el caso de RFCA PF y de 4.6 MDP para los RFCA PM, lo cual corresponde a una evasión anual promedio global de 60 MDP para RFCA PF y de 12,760 MDP para el caso de RFCA PM. Estos montos pueden considerarse como cotas superiores de evasión de acuerdo a la información que tenemos disponible, los cuales podrían incrementarse al tomar en cuenta el efecto de las EDOS en el fenómeno de la evasión, y la disponibilidad de bases de datos más amplias y detalladas.

9. Conclusiones y limitaciones

En este estudio hemos utilizado métodos de análisis de ciencia de redes para la descripción de la topología (patrones de conexión) tanto de redes de emisiones de CFDI entre contribuyentes, como de las relacionadas a la composición legal de empresas en México por medio de redes bipartitas entre personas morales y sus socios, asociados y representantes legales.

La exploración a distintos niveles de agregación de los datos (desde la escala de contribuyentes individuales hasta la de sectores y actividades económicas) nos ha permitido identificar la sobre-representación de EFOS en varios sectores económicos. Nuestro estudio ha mostrado el uso sistemático de un conjunto de actividades económicas para la emisión, recepción y mediación de operaciones entre EFOS (asociadas principalmente a la prestación de servicios o la comercialización de intangibles que dificultan la materialización de las operaciones), así como la emisión de CFDI con montos mucho mayores a los del resto de la población entre ciertas actividades económicas.

Un aspecto explorado en este estudio son las redes bipartitas entre PM y sus socios, asociados y representantes legales, con el objetivo de analizar la evasión fiscal más allá de las emisiones de CFDI entre contribuyentes. El estudio de estas redes bipartitas, con apoyo en las recientes reformas del art. 27 del CFF (en lo que se refiere a la obligación de los

contribuyentes de notificar cambios en sus socios y accionistas), junto con los esfuerzos del SAT en identificación de EFOS, nos permite detectar prácticas y mecanismos relacionados con la constitución de empresas utilizadas para la simulación de operaciones, así como su evolución en el tiempo. Este tipo de análisis ofrece una línea de investigación relevante para la identificación eficiente y pragmática de evasores fiscales, profundiza el entendimiento del fenómeno.

Los métodos utilizados en este estudio (análisis basados en ciencia de redes y aprendizaje de máquinas) permiten realizar una descripción cuantitativa del fenómeno de la evasión fiscal en México y se complementan al permitir un diálogo entre ambos enfoques. Así mismo, han permitido obtener visos de una mejor colaboración entre este tipo de aproximaciones científicas y otros enfoques más cercanos al tratamiento legal y fiscal del comportamiento de los evasores de impuestos, en específico de los que simulan operaciones. De esta forma, esperamos que nuestro estudio forme la base de futuras líneas de investigación interdisciplinarias que se complementen de aún mejor manera con el tratamiento legal y fiscal del fenómeno de evasión en México.

9.1. Recomendaciones

Como hemos mostrado en este estudio, el contar con un listado de EFOS identificados por el SAT nos ha permitido identificar y generalizar su comportamiento para así identificar RFCA sospechosos. Recomendamos generar un listado similar de EDOS, o al menos crear un perfil general relacionado con la información en las bases de datos disponibles, para poder caracterizar y generalizar su comportamiento de manera similar a como lo hemos hecho con los EFOS. Esto permitiría obtener una visión más completa del fenómeno de evasión. Sin embargo, el refinamiento y la ampliación de los métodos utilizados dependen de la calidad e integridad de las bases de datos con las que cuenta el SAT. En este sentido, recomendamos hacer una revisión de la información disponible y realizar monitoreos periódicos de las mismas, con el objetivo de tener la posibilidad de generar perfiles integrales y que caractericen el comportamiento de los contribuyentes al paso del tiempo.

Así mismo, proponemos crear un listado de RFCA cumplidos y pensar en generar incentivos, para tener una referencia de la contraparte del fenómeno y observar el contraste con EFOS y EDOS. De igual manera, dado que hemos identificado una preferencia de pertenecer y realizar operaciones entre cierto conjunto de actividades económicas, proponemos incorporar datos en la inscripción al RFC que ligue con la realización de actividades vulnerables.

Recomendamos integrar un sistema automático de monitoreo y detección de EFOS y EDOS sospechosas, basado en los métodos de este estudio o similares, a las herramientas tecnológicas con las que ya cuenta el SAT. Al analizar la actividad cotidiana de todos los contribuyentes, se podrían identificar presuntos evasores de manera ágil e informar a las instancias correspondientes.

Hasta la finalización de este estudio (diciembre 2020), no todas las operaciones que realiza la persona moral y/o, en algunas partidas, la persona física, están amparadas por un CFDI, lo que dificulta el cálculo total de los ingresos acumulables y las deducciones autorizadas. Al contar con un CFDI de todas las operaciones que realice o reciba el contribuyente durante el ejercicio, se simplificaría el cálculo de ISR de manera directa para la autoridad fiscal y por ende las alertas en el caso de que haya diferencias. Adicionalmente tenemos las siguientes recomendaciones:

1. Por el Ajuste Anual por Inflación Acumulable (AAIA), la persona moral expida un CFDI de ingreso con una clave especial que identifique esta partida que deba emitirse a más tardar dentro de los tres meses siguientes al cierre del ejercicio, y/o al momento de presentar la declaración anual y por el Ajuste Anual por Inflación Deducible (AAID) la persona moral expida un CFDI de egreso.

2. Evaluar la posibilidad de expedir el CFDI de ingreso por los ingresos acumulables que correspondan a una aportación de capital –aportación mediante entrega de un bien, o un pago en especie-, o por la adjudicación de un bien, ya sea o no, adjudicación judicial, o tratándose de los ingresos por accesión que se dan al dejar el arrendatario, al término del contrato de arrendamiento, en favor del arrendador las mejoras a los bienes arrendados.
3. Por la enajenación de activos fijos, terrenos o acciones, se expida el CFDI de ingreso que corresponda, por el valor de la enajenación y los impuestos a que haya lugar, en su caso, en el que se especifique, mediante una clave, previamente establecida, el saldo pendiente por deducir, el costo comprobado de adquisición y el costo promedio por acción por el total de acciones respectivamente; para indicar también mediante una clave especial, la ganancia acumulable en enajenación de activos, en enajenación de terrenos o en enajenación de acciones o partes sociales.
4. Por la deducción de la inversión -activos fijos, cargos o gastos diferidos- se establezca la emisión de un CFDI de egreso por el importe de la deducción en el ejercicio, que haga referencia al tipo de bien, la fecha, monto y folio del CFDI de adquisición, así como el porcentaje de deducción que corresponda en el año, incluyendo el factor de actualización.

9.2. Trabajo a futuro

Hasta el momento no ha sido posible cuantificar el impacto de las EDOS en la evasión fiscal, así como su papel en las redes de emisión de CFDI entre contribuyentes y en las redes bipartitas. La razón principal es la falta de un listado de RFCA asociados a EDOS similar al que existe para las EFOS, o una caracterización de acuerdo a su información categórica, comportamiento contable, etc. Debido a la ausencia de esta información, una exploración sobre la formación de enlaces recurrentes en redes de emisión de CFDI (con el objetivo de identificar EDOS) no fue concluyente y, por lo tanto, no fue considerada en el estudio. De contar con un listado de EDOS de esta naturaleza, podríamos generalizar su comportamiento por medio de métodos de aprendizaje de máquinas para la identificación de EDOS sospechosos, y obtener una descripción de su comportamiento dentro de las redes de emisión de CFDI.

Además de un análisis más profundo de EDOS, a continuación listamos una serie de líneas de investigación prometedoras para el futuro cercano:

- Comparar la lista entregada al SAT de RFCA con alta probabilidad de ser EFOS, con los resultados de las investigaciones internas del SAT, a fin de mejorar nuestros métodos de clasificación y detección automática de EFOS.
- Ampliar la descripción de las redes bipartitas de PM y sus asociados tomando en cuenta el papel de los notarios en la constitución de las empresas.
- Ahondar en la caracterización del comportamiento contable y fiscal de EFOS para refinar los métodos de identificación de RFCA sospechosos por medio de la identificación de comportamientos anómalos.
- Explorar la formación y recurrencia temporal de emisiones de CFDI en la localidad de EFOS, tanto en las redes de emisión de CFDI, como geográficamente, con el objetivo de obtener una mejor descripción tanto del comportamiento de EFOS como de EDOS.

- Investigar posibles métodos para la caracterización e identificación de EDOS con el objetivo de incorporarlos en la descripción del fenómeno de la evasión fiscal en México y así obtener mejores estimados de montos de evasión.
- Analizar las distribuciones de motifs (patrones locales de redes) alrededor de EFOS y EDOS, para detectar posibles anomalías en las distribuciones. Estas anomalías podrían ser útiles para detectar rápidamente EFOS y EDOS sospechosos (sin que estén relacionados con nodos ya conocidos).
- Hacer un análisis más profundo de las propiedades temporales de las redes bipartitas entre EFOS y sus asociados, así como de su estructura local (evolución de motifs, etc.), a fin de aumentar la precisión de los estimados de evasión y detectar estrategias de defraudación más complejas.

Créditos

Por orden alfabético.

Directores del proyecto

Dr. Carlos Gershenson, Dr. Gerardo Iñiguez, Dr. Carlos Pineda.

Investigadores

M. en C. Nephtalí Garrido, Dra. Martha Gómez, Dr.(c) Martín Zumaya.

Agradecimientos

Romel Calero, Humberto Carrillo Calvet, Adriana Cruz Cortés, Alejandro Frank Hoefflich, Javier García García, José Luis Gordillo, William Lee Alardín, José Luis Mateos Trigos, Patricia Peña, Ana Leonor Rivera López, Aleida Rueda, José Saniger Blesa, Laura Vargas Parada.

Referencias

- [1] Jorge Cosulich Ayala. La evasión tributaria. In *Serie Política Fiscal*, volume 39, page 9. Proyecto Regional de Política Fiscal Naciones Unidas CEPAL-PNUD, Santiago de Chile, 1993.
- [2] Oliver Oldman and Daniel M. Holland. Medición de la evasión fiscal, 1971.
- [3] 31 de diciembre de 1953 Diario Oficial México. Sección cuarta, ley del impuesto sobre la renta. http://www.dof.gob.mx/nota_to_imagen_fs.php?codnota=4645807&fecha=31/12/1953&cod_diario=199946. Recuperado: 25 de octubre de 2020.
- [4] Iniciativas del Ejecutivo Federal Gaceta parlamentaria. Que reforma, adiciona y deroga diversas disposiciones de la ley del impuesto sobre la renta, de la ley del impuesto al valor agregado de la ley del impuesto especial

- sobre producción y servicios y del código fiscal de la federación. <http://gaceta.diputados.gob.mx/PDF/64/2019/sep/20190908-D.pdf>.
- [5] Diario Oficial de la Federación del 8 de diciembre de 2020. Decreto por el que se reforman, adicionan y derogan diversas disposiciones de la ley del impuesto sobre la renta, de la ley del impuesto al valor agregado y del código fiscal de la federación. http://www.dof.gob.mx/nota_detalle.php?codigo=5606951&fecha=08/12/2020.
- [6] DOF del 29 de diciembre de 2020. Resolución miscelánea fiscal para 2021. http://www.dof.gob.mx/2020/SHCP/Resolucion_Miscelanea_Fiscal_2021.pdf. Recuperado: 25 de octubre de 2020.
- [7] Maria R. D'Orsogna and Matjaž Perc. Statistical physics of crime: A review. *Physics of Life Reviews*, 12:1–21, mar 2015.
- [8] Johannes Wachs and János Kertész. A network approach to cartel detection in public auction markets. *Scientific Reports*, 9:10818, 2019.
- [9] Johannes Wachs, Mihály Fazekas, and János Kertész. Corruption risk in contracting markets: a network science perspective. *International Journal of Data Science and Analytics*, jan 2020.
- [10] Philip C. Solimine. Political corruption and the congestion of controllability in social networks. *Applied Network Science*, 5(1), may 2020.
- [11] Haroldo V Ribeiro, Luiz G A Alves, Alvaro F Martins, Ervin K Lenzi, and Matjaž Perc. The dynamical structure of political corruption networks. *Journal of Complex Networks*, 6(6):989–1003, jan 2018.
- [12] Issa Luna-Pla and José R. Nicolás-Carlock. Corruption and complexity: a scientific framework for the analysis of corruption networks. *Applied Network Science*, 5(1), feb 2020.
- [13] Rafał Dreżewski, Jan Sepielak, and Wojciech Filipkowski. The application of social network analysis algorithms in a system supporting money laundering detection. *Information Sciences*, 295:18–32, feb 2015.
- [14] T.M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997.
- [15] Leo Breiman. Arcing the edge. *Statistics*, 4, 1997.
- [16] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: Unbiased boosting with categorical features, 2018.
- [17] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An end-to-end deep learning architecture for graph classification. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4438–4445. AAAI Press, 2018.
- [18] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.
- [19] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *CoRR*, abs/1705.07874, 2017.
- [20] CSIRO's Data61. Stellargraph machine learning library. <https://github.com/stellargraph/stellargraph>, 2018.
- [21] Tipos de CFDI. <https://www.cfdi.org.mx/tipos-de-cfdi/>. Recuperado: 8 de julio de 2020.

A. Anexo

A.1. Los CFDI en la determinación de la base de cálculo del ISR de las Personas Morales

Como indicamos, nos enfocamos, en un inicio, en el Grupo 1: 601. Régimen General de Ley Personas Morales para, de forma global determinar el ISR del ejercicio y con ello, estimar la evasión en ISR de los presuntos EFOS, apoyados en la información con que contábamos.

Los importes que integran gran parte de la base para el cálculo del ISR, en este caso, de las PM del Grupo 1, lo constituyen los incluidos en los Comprobantes Fiscales Digitales por Internet (CFDI) –artículo 29 CFF-, que están obligados a expedir, tanto las personas físicas como las personas morales, con algunas excepciones, como es el caso de los Sindicatos y organismos que los agrupen –artículo 86 segundo párrafo LISR-, CFDI que pueden ser de diferentes tipos y, por lo tanto, sus importes pueden formar parte de las partidas que integran los ingresos acumulables, o bien, de las deducciones autorizadas, y son los siguientes -Anexo 20 de la Guía de llenado de los Comprobantes Fiscales Digitales por Internet en el Apéndice 2:

- CFDI de ingreso
- CFDI de egreso
- CFDI de traslado
- CFDI de recepción de pagos
- CFDI de nómina
- CFDI de retenciones e información de pagos

CFDI de ingreso. –

El CFDI de ingreso, como su nombre lo indica, lo emiten los contribuyentes por las contraprestaciones que reciben o van a recibir como pago de la venta de sus productos, de la prestación de sus servicios, del arrendamiento de sus bienes, de los donativos que reciben o de cualquier otra actividad, mercantil o civil que realicen. Representan los importes que incluyen los CFDI, los ingresos que obtienen los contribuyentes por realizar las operaciones recurrentes propias de su actividad, o de aquellas que realizan de forma esporádica, como es el caso de la enajenación de activos fijos o de terrenos.

Los CFDI de ingreso, en términos generales, constituyen parte de los ingresos acumulables que integran la base para el cálculo del ISR causado, tanto de PM como de PF, o bien, de la determinación del remanente distribuible en algunas PM con fines no lucrativos. Se debe tomar en cuenta, sin embargo, algunas consideraciones sobre esta acumulación de importes del CFDI para la determinación de los ingresos acumulables, tal es el caso, como mencionamos, de la enajenación de bienes de activo fijo –inversiones-, partes representativas del capital –títulos valor, acciones, partes sociales, certificados de participación- o terrenos, en los que la acumulación del ingreso no es el total del mismo, sino únicamente la ganancia acumulable –artículos 18 fracción IV, 19, 22 y 24, 31 sexto párrafo LISR- y, no todas ellas están amparadas con un CFDI, como es el caso de las enajenaciones de un inmueble, un terreno o las partes representativas del capital.

Existen otros ingresos acumulables que no están amparados con CFDI, como son: el ajuste anual por inflación, determinado por el excedente de las deudas sobre los créditos –artículo 44 fracción II, 45 y 46 de la LISR- o como deducción en el caso contrario; los ingresos determinados presuntivamente por las autoridades fiscales en los casos en que procedan –artículo 18 fracción I, 91 tercer párrafo, 101 segundo párrafo y último párrafo aplicable a extranjeros-; la recuperación de un crédito deducido por incobrable –artículo 18 fracción V; las condonaciones, quitas y remisiones –artículo 101 fracción I, 197 párrafos tercero y cuarto LISR-.

Las PF no consideran como ingreso acumulable el ajuste anual por inflación, ya que el momento de acumulación de sus ingresos es al cobro –flujo de efectivo-, salvo, en este último caso, cuando se trata de ingresos por la venta de bienes que se exporten y no se reciba el pago, deberán acumularlos transcurridos doce meses de la exportación –artículo 102, último párrafo LISR-.

Algunas PM tienen también, el momento de acumulación de sus ingresos al cobro, en el caso de los obtenidos por la prestación de servicios independientes de las sociedades y asociaciones civiles, los obtenidos por los servicios de suministro de agua potable o recolección de basura doméstica prestados por organismos descentralizados y por otros autorizados –artículo 17 segundo párrafo-; de igual forma, las PM que identificamos para efectos de este documento dentro del Grupo 1, constituidas por PF que en el ejercicio anterior sus ingresos totales no hayan superado cinco millones de pesos, podrán ejercer la opción de pagar el ISR con base en flujo de efectivo y también, al igual que las PF el ingreso por la enajenación bienes que exporten, lo acumularán a los 12 meses de la exportación, si antes no recibieron el pago –artículo 196 primer párrafo y 197 último párrafo-. Los coordinados, PM que administren activos fijos, o activos fijos y terrenos relacionados directamente con actividades de autotransporte terrestre de carga y de pasajeros, pagarán el ISR con base en flujo de efectivo –artículo 72 cuarto párrafo LISR-; de igual forma las PM con actividades del sector primario –artículo 74 párrafo siete, fracción I LISR-.

CFDI de egreso.-

El CFDI de egreso corresponde a aquél que se emite para disminuir los ingresos, como es el caso de las devoluciones descuentos y bonificaciones sobre ventas; cancelación de un CFDI de ingreso, previo acuerdo con la persona a favor de quien se emitió originalmente. Se utiliza también para corregir los montos del CFDI de ingresos cuando se aplican anticipos, y corresponden a las llamadas notas de crédito.

CFDI de traslado.-

El CFDI de traslado es emitido para amparar la propiedad de las mercancías que se transportan durante su trayecto.

CFDI de recepción de pagos.-

El CFDI de recepción de pagos se emite en aquellos casos en los que no se recibe el pago, ya sea porque será cubierto en parcialidades o de forma diferida (PPD), o bien cuando no está definida la forma en que será cubierto (PPD 099). Este CFDI de recepción de pagos permite seguir la secuencia de los pagos y a su vez las PF y las PM que tributan en el ISR con base en flujo de efectivo, cuando reciben cada uno de los pagos tendrán el ingreso acumulable para el pago del ISR.

CFDI de nómina.-

El CFDI de nómina representa un comprobante fiscal que forma parte de los gastos por el pago de sueldos, salarios y asimilables; normalmente quien recibe el ingreso debe emitir un CFDI, ya sea de ingreso, o bien de recepción de pagos; sin embargo el CFDI de nómina lo emite el empleador, ya que debe llevar a cabo el timbrado de la nómina

para cumplir con requisitos fiscales y formar parte de las deducciones, tanto de las PF como de las PM, cuando requieran en el desarrollo de su actividad, contar con personas sujetas a una relación laboral, o bien, realicen pagos por asimilables contemplados en el artículo 94 fracciones II a VI de la LISR.

CFDI de retenciones e información de pagos.–

Este tipo de CFDI que propiamente no pertenece al catálogo de tipos de comprobante porque se genera con el estándar contenido en el rubro II del Anexo 20 de la Resolución Miscelánea Fiscal 2020, vigente a partir del 19 de julio de 2017 [21] se expide cuando se realizan retenciones de impuestos, incluyendo los pagos realizados a residentes en el extranjero para efectos fiscales y las retenciones que se les realicen.

A.1.1. El CFDI y los ingresos acumulables de las Personas Morales del Grupo I

Por lo indicado anteriormente podemos concluir que, aunque los importes de los diferentes tipos de CFDI forman parte de la base para el cálculo del ISR del ejercicio de las PM del Grupo I, el CFDI del emisor forma parte de sus ingresos acumulables, con ajustes, si es que se trata de ingresos por los que se acumula solamente la ganancia y no el total de la operación, con la adición del ajuste anual por inflación cuando así corresponda y los ingresos presuntos determinados por la autoridad fiscal cuando así proceda.

A.1.2. El CFDI y las deducciones autorizadas

Por cuanto a las deducciones autorizadas corresponde y considerando todos los requisitos que deben cumplir, estarían constituidas por el importe de los CFDI de egreso, los CFDI de ingreso recibidos de otro emisor, los CFDI de nómina, que incluyen la PTU del ejercicio pagada, tomando en cuenta, no considerar las partidas que no son deducibles –artículo 28 fracción XXX LISR-, faltando en su caso, la deducción del ajuste anual por inflación cuando los créditos excedan a las deudas, así como, la correspondiente deducción de inversiones por las que se recibió el CFDI de otros emisores en el ejercicio en que se determina el impuesto, o en ejercicios anteriores, por los que dicha deducción es parcial hasta los porcentajes máximos establecidos –artículos 33 y 34 LISR-.

A.2. Tipos de Ingresos que pueden percibir las Personas Físicas y las Personas Morales

Para la determinación y cálculo del ISR del ejercicio es necesario tomar en cuenta que la Ley del ISR contempla diferentes tipos de ingresos que pueden tener las personas físicas y, en su caso, las personas morales, estos son:

- Ingresos totales
- Ingresos gravables
- Ingresos acumulables
- Ingresos no acumulables
- Ingresos exentos o no sujetos al pago del ISR e
- Ingresos no objeto

A.2.1. Ingresos totales

Los ingresos totales, como su nombre lo indica, resultan de la suma de los ingresos que perciben durante el ejercicio, ya que el ISR se paga por ejercicios fiscales que son iguales a los del año calendario, salvo en los casos de:

- a) Inicio y reanudación de actividades o inicio de fusión y escisión de sociedades, que comprenden desde el día que el contribuyente inicia actividades, las reanuda, inicia la fusión o la escisión, hasta el 31 de diciembre del mismo año;
- b) Suspensión de actividades, fusión y escisión de sociedades e inicio de liquidación, que va del 1 de enero al día anterior a la suspensión, fusión, escisión o liquidación de sociedades; y
- c) Liquidación de sociedades, o bien, cambio de residencia fiscal al extranjero; en la liquidación por disolución o por acuerdo de socios –artículo 234 de la Ley General de Sociedades Mercantiles (LGSM)-, el ejercicio comprende desde el día que inicia la liquidación hasta que ésta concluye –durante plazos menores o mayores a un año-, caso en el que el ISR se causa por todo el ejercicio mientras dure la liquidación, con enteros mensuales y ajustes anuales; la liquidación por cambio de residencia de personas morales se da en el ejercicio en que se decide hacer el cambio, a partir del 1 de enero de dicho ejercicio, hasta el día previo al cambio dentro del ejercicio calendario.

A.2.2. Ingresos gravables

Los ingresos gravables son aquéllos sobre los que se causa el ISR y forman la base para su cálculo, en algunos casos, puede ser a los que se les aplica la tasa de ISR, tratándose de PM -30 % en 2020, artículo 9 LISR-, o bien en PF la tarifa, desde un porcentaje sobre el excedente de 1.92 % hasta el 35 % -artículo 152 LISR-. En el caso de PM con fines no lucrativos, en términos generales no pagan ISR, quienes lo pagan son sus integrantes a través del remanente distribuible; si se trata de donatarias autorizadas, éstas no reparten remanentes, pero llegan a pagar ISR cuando perciben ingresos por actividades no propias de su objeto que superen el 10 % -artículo 80, último párrafo LISR-, sobre dicho excedente, considerándose, en este caso, ISR definitivo.

Dentro de los ingresos gravables se pueden tener también los llamados Ingresos presuntos o presuntivos (que a su vez son ingresos acumulables en algunos casos) -artículo 18 fracción I, LISR-, y son los determinados por la autoridad fiscal considerando algunas conductas u omisiones de los contribuyentes.

A.2.3. Ingresos acumulables

Los ingresos acumulables corresponden a los montos que forman parte de la base para el pago del ISR a los que se les disminuye las deducciones autorizadas, la participación a los trabajadores en las utilidades –PTU- pagada en el ejercicio y, en su caso, las pérdidas fiscales de ejercicios anteriores, para llegar al resultado fiscal, si se trata de personas morales del régimen general de Ley, o bien, a la utilidad gravable en el caso de personas físicas con actividades empresariales y profesionales.

En algunos casos para el cálculo de ISR del ejercicio de las personas físicas, como es cuando perciben salarios y reciben durante el ejercicio pagos por compensación por separación, jubilaciones en un solo pago, o bien, realizan una enajenación de bienes, el total de ingresos gravables (o ganancia acumulable o pérdida acumulable en enajenación de

bienes), se divide en dos partes, una de ellas es el ingreso acumulable (ganancia acumulable o pérdida acumulable) y la otra parte la constituye el ingreso no acumulable (ganancia no acumulable) –artículos 95 y 120 fracción II, LISR-.

A.2.4. Ingresos no acumulables

Los ingresos no acumulables son aquellos que, por disposición de Ley, se determina el ISR aplicando la llamada “tasa efectiva”, que resulta de dividir el ISR determinado sobre los ingresos acumulables entre el total de los mismos ingresos, expresando el resultado en por ciento; no quiere decir que sobre ellos no se paga ISR, sino que se permite, dado que la tarifa de PF es progresiva, que estos ingresos causen un ISR limitado al que resultó de los ingresos acumulables; el ISR causado del ejercicio será entonces la suma del ISR de ingresos acumulables y el ISR de los no acumulables. Como ejemplos de su aplicación, tenemos el caso de PF que perciben ingresos en el ejercicio por compensación por separación –artículo 95 LISR-, o cuando reciben ingresos por enajenación de bienes –artículo 120 fracción II LISR-.

A.2.5. Ingresos Exentos o no sujetos al pago del ISR

Los ingresos exentos son aquellos que se encuentran en el supuesto de la obligación fiscal de pago de ISR, pero son señalados, por disposición de Ley, de forma específica como tales; así tenemos los contemplados en el artículo 93 de la LISR; algunas percepciones son exentas total o parcialmente, si cumplen con los requisitos establecidos –prestaciones de salario mínimo, horas extras, jubilaciones, compensaciones por separación, partidas de previsión social, enajenación de bienes, herencias y legados, donaciones etc.- algunas de las cuales se deben considerar en la declaración anual de ISR mediante disminución de los ingresos totales para llegar a los ingresos gravados, como es el caso de prestaciones por salarios y otros que son totalmente exentos –donativos en línea recta ascendente o descendente y entre cónyuges; viáticos; enajenación de casa habitación; herencias y legados-, se deben incluir en los datos informativos de la declaración anual, a partir del año 2002, junto con otros que, no son ingresos –préstamos- o bien, están gravados y pagan el ISR mediante retención con carácter de impuesto definitivo –premios- (artículos 90 segundo párrafo y 150 tercer párrafo LISR).

A.2.6. Ingresos no objeto

En cuanto a los ingresos no objeto, corresponden a aquellos que de forma enunciativa por disposición de Ley, se consideran como tal y, por lo tanto, ni siquiera se incluyen dentro de los ingresos totales, ya que, como su nombre lo indica, no son objeto de la LISR; podemos mencionar entre ellos, en el caso de las personas morales -artículo 16 segundo párrafo LISR-, los ingresos por aumentos de capital, por pago de pérdidas por los accionistas, por primas por la colocación de acciones que emita la propia sociedad, o por utilizar para valuar sus acciones el método de participación, ni tampoco los que obtengan por la revaluación de sus activos y de su capital (son ingresos no objeto, pero tienen una repercusión fiscal al formar parte, ya sea de la Cuenta de Capital de Aportación (CUCA), o bien, de la integración del capital contable de la PM, con efectos en las reducciones y reembolsos de capital) -artículo 78 LISR-. En PF podemos señalar, entre otros, tratándose de ingresos por salarios, los servicios de comedor y comida y el uso de bienes o herramientas de trabajo -artículo 94 último párrafo LISR-.

A.3. Grupos de Personas Morales

Los grupos de PM que formamos son:

Grupo 1, 601. Régimen General de Ley Personas Morales.

Grupo 2, Personas Morales del Sector primario, de autotransporte y otros. En este grupo incluimos: 602. Régimen Simplificado de Ley Personas Morales; 622. Régimen de Actividades Agrícolas, Ganaderas, Silvícolas y Pesqueras (PF y PM).

Grupo 3, 603. Personas Morales con fines no lucrativos.

Grupo 4, Personas Morales que consolidan, 623. Régimen de Opcional para Grupos de Sociedades y 609. Régimen de Consolidación.

Grupo 5, Personas Morales cooperativas de producción con integrantes personas físicas, 620. Sociedades Cooperativas de Producción que optan por diferir sus ingresos.

A.4. Artículo 58 del CFF

A continuación se cita el artículo, en comento, del CFF:

Artículo 58.

Las autoridades fiscales, para determinar presuntivamente la utilidad fiscal de los contribuyentes a que se refiere la Ley del Impuesto sobre la Renta, podrán aplicar a los ingresos brutos declarados o determinados presuntivamente, el coeficiente de 20 % o el que corresponda tratándose de alguna de las actividades que a continuación se indican:

I. Se aplicará 6 % a los siguientes giros:

Comerciales: Gasolina, petróleo y otros combustibles de origen mineral.

II. Se aplicará 12 % en los siguientes casos:

Industriales: Sombreros de palma y paja.

Comerciales: Abarrotes con venta de granos, semillas y chiles secos, azúcar, carnes en estado natural; cereales y granos en general; leches naturales, masa para tortillas de maíz, pan; billetes de lotería y teatros.

Agrícolas: Cereales y granos en general.

Ganaderas: Producción de leches naturales.

III. Se aplicará 15 % a los giros siguientes:

Comerciales: Abarrotes con venta de vinos y licores de producción nacional; salchichonería, café para consumo nacional; dulces, confites, bombones y chocolates; legumbres, nieves y helados, galletas y pastas alimenticias, cerveza y refrescos embotellados, hielo, jabones y detergentes, libros, papeles y artículos de escritorio, confecciones, telas y artículos de algodón, artículos para deportes; pieles y cueros, productos obtenidos del mar, lagos y ríos, substancias y productos químicos o farmacéuticos, velas y veladoras; cemento, cal y arena, explosivos; ferreterías y tlapalerías; fierro y acero, pinturas y barnices, vidrio y otros materiales para construcción, llantas y cámaras, automóviles, camiones, piezas de repuesto y otros artículos del ramo, con excepción de accesorios.

Agrícolas: Café para consumo nacional y legumbres.

Pesca: Productos obtenidos del mar, lagos, lagunas y ríos.

IV. Se aplicará 22 % a los siguientes rubros:

Industriales: Masa para tortillas de maíz y pan de precio popular.

Comerciales: Espectáculos en arenas, cines y campos deportivos.

V. Se aplicará 23 % a los siguientes giros:

Industriales: Azúcar, leches naturales; aceites vegetales; café para consumo nacional; maquila en molienda de nixtamal, molienda de trigo y arroz; galletas y pastas alimenticias; jabones y detergentes; confecciones, telas y artículos de algodón; artículos para deportes; pieles y cueros; calzado de todas clases; explosivos, armas y municiones; fierro y acero; construcción de inmuebles; pintura y barnices, vidrio y otros materiales para construcción; muebles de madera corriente; extracción de gomas y resinas; velas y veladoras; imprenta; litografía y encuadernación.

VI. Se aplicará 25 % a los siguientes rubros:

Industriales: Explotación y refinación de sal, extracción de maderas finas, metales y plantas minero-metalúrgicas.

Comerciales: Restaurantes y agencias funerarias.

VII. Se aplicará 27 % a los siguientes giros:

Industriales: Dulces, bombones, confites y chocolates, cerveza, alcohol, perfumes, esencias, cosméticos y otros productos de tocador; instrumentos musicales, discos y artículos del ramo; joyería y relojería; papel y artículos de papel; artefactos de polietileno, de hule natural o sintético; llantas y cámaras; automóviles, camiones, piezas de repuesto y otros artículos del ramo.

VIII. Se aplicará 39 % a los siguientes giros:

Industriales: Fraccionamiento y fábricas de cemento.

Comerciales: Comisionistas y otorgamiento del uso o goce temporal de inmuebles.

IX. Se aplicará 50 % en el caso de prestación de servicios personales independientes.

A.5. Tasas de utilidad por sector definidas en este estudio

En la Tabla 10 listamos las tasas de utilidad por sector económico utilizadas para la determinación de la utilidad de cada ejercicio, asociada a los ingresos declarados por los RFCA sospechosos. Estas tasas son una definición operativa y corresponden a una clasificación de actividades en sectores económicos, de acuerdo a las tasas por actividad expresadas en el art. 58 del CFF.

Tasas de Utilidad de los Sectores Económicos asociados a RFCA sospechosos	
Sector Económico	Tasa de Utilidad
Comercio al por menor	0.12
Construcción	0.39
Transportes, correos y almacenamiento	0.25
Servicios profesionales, científicos y técnicos	0.50
Comercio al por mayor	0.15
Industrias manufactureras	0.23
Agricultura, ganadería, aprovechamiento forestal, pesca y caza	0.15
Otros servicios excepto actividades del gobierno	0.50
Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	0.39
Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	0.50
Servicios de alojamiento temporal y de preparación de alimentos y bebidas	0.39
Servicios financieros y de seguros	0.50
Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos	0.22
Información en medios masivos	0.39
Servicios de salud y de asistencia social	0.25
Actividades del gobierno y de organismos internacionales y extraterritoriales	0.25
Dirección de corporativos y empresas	0.50
Minería	0.25
Servicios educativos	0.50
Electricidad, agua y suministro de gas por ductos al consumidor final	0.25

Tabla 10: Tasas de utilidad por Sector Económico basadas en las tasas de utilidad por actividad económica definidas en el art. 58 del CFF. Esta agrupación es meramente operativa, que nos facilita tanto la clasificación como el procesamiento de los datos.

A.6. Agregación de información y construcción de enlaces

Dado que uno de los objetivos es el de continuar con la caracterización del comportamiento de las EFOS, comenzamos por explorar las interacciones entre actividades económicas partiendo de las emisiones de CFDI entre contribuyentes.

Haciendo uso de la información categórica con la que contamos, representamos a cada contribuyente, C_i , por medio de su RFCA, su estatus, (EFOS definitiva, presunta o RFCA desconocido) y su conjunto de variables categóricas asociadas, lo cual lo podemos expresar como:

$$C_i = (\text{RFCA}_i, \theta_i, s_i, r_i, a_i) \quad (3)$$

donde RFCA_i es la cadena de caracteres que identifica a cada contribuyente y θ_i, s_i, r_i, a_i corresponden a su estatus y el sector, rama y actividad económica a las que pertenece.

Como hemos visto, dado que es posible asociarle un conjunto de variables categóricas a cada contribuyente, podemos “transformar” cada emisión de CFDI entre pares de contribuyentes en emisiones de CFDI entre pares de sectores, ramas y actividades económicas, de tal forma que podamos investigar si existe una preferencia de las EFOS tanto para pertenecer a alguna como para realizar operaciones entre un conjunto de ellas.

Con este objetivo, partiendo de los registros de emisiones de CFDI mensuales entre contribuyentes, construimos registros de operaciones entre sectores, ramas y actividades económicas, tomando en cuenta el estatus (Definitivo, Presunto, Sospechoso o Desconocido) tanto del emisor como del receptor de los CFDI. Dada la experiencia obtenida del estudio anterior, solamente tomaremos en cuenta el número de facturas activas y el monto subtotal activo de cada registro, de tal forma que partiendo de un registro entre contribuyentes de la forma:

$$(\text{RFCA}_R, \text{RFCA}_E, f_{RE}, m_{RE}) \quad (4)$$

donde el subíndice E corresponde al contribuyente emisor y el subíndice R al contribuyente receptor, y f_{RE} es el número de facturas activas y m_{RE} es el monto subtotal activo de las emisiones del mes, pasamos a un registro de la forma:

$$(\theta_E, \theta_R, s_E, s_R, r_E, r_R, a_E, a_R, f_{RE}, m_{RE}), \quad (5)$$

lo cual nos permite obtener el número de facturas y los montos asociados a las operaciones entre cada una de las actividades económicas etiquetadas por el estatus del emisor y el receptor. Esto nos será útil para poder comparar las operaciones entre actividades económicas partiendo de las operaciones que se realizan entre EFOS con las que se ocurren entre RFCA desconocidos.

El estatus de un contribuyente θ_i puede tomar los valores $\theta_i = D, P, U$, los cuales corresponden a EFOS definitiva, presunta o RFCA desconocido, respectivamente. Considerando el estatus del emisor y el receptor podemos construir redes entre actividades en las cuales los enlaces están asociados solamente a emisiones entre EFOS o entre RFCA desconocidos o a combinaciones entre ellos.

A.7. Distribución de RFCA sospechosos en Sectores y Actividades Económicas

En las Tablas 11 y 12, listamos los principales sectores y actividades económicas a las que pertenecen los RFCA sospechosos identificados por los métodos de clasificación basados en Aprendizaje de Máquinas. Es posible observar como estas corresponden en su mayoría a la prestación de servicios o a la comercialización de intangibles.

Sectores Económicos de los RFCA sospechosos	
Sector Económico	RFCA Sospechosos
Servicios profesionales, científicos y técnicos	967
Comercio al por mayor	839
Construcción	814
Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	252
Comercio al por menor	244
Industrias manufactureras	234
Transportes, correos y almacenamiento	134
Otros servicios excepto actividades del gobierno	124
Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	117
Agricultura, ganadería, aprovechamiento forestal, pesca y caza	104
Servicios de alojamiento temporal y de preparación de alimentos y bebidas	66
Información en medios masivos	50
Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos	27
Servicios financieros y de seguros	24
Servicios de salud y de asistencia social	9
Dirección de corporativos y empresas	8
Minería	5
Actividades del gobierno y de organismos internacionales y extraterritoriales	4
Servicios educativos	3
Electricidad, agua y suministro de gas por ductos al consumidor final	1

Tabla 11: Distribución de RFCA sospechosos identificados por los métodos de clasificación en los distintos sectores económicos.

Actividades Económicas de los RFCA sospechosos

Actividad Económica	RFCA Sospechosos
Otras construcciones de ingeniería civil u obra pesada	277
Servicios de consultoría en administración	185
Otros intermediarios de comercio al por mayor	178
Servicios de consultoría en computación	137
Construcción de inmuebles comerciales, institucionales y de servicios	122
Otros servicios profesionales, científicos y técnicos	122
Servicios de contabilidad y auditoría	87
Construcción de obras de urbanización	74
Agencias de publicidad	73
Otros servicios de consultoría científica y técnica	64
Otros servicios de apoyo a los negocios	60
Otro autotransporte foráneo de carga general	58
Construcción de vivienda unifamiliar	51
Comercio al por mayor de mobiliario, equipo e instrumental médico y de laboratorio	49
Bufetes jurídicos	47
Reparación y mantenimiento de maquinaria y equipo industrial	47
Comercio al por mayor de frutas y verduras frescas	45
Comercio al por mayor de otros materiales para la construcción, excepto de madera	45
Comercio al por mayor de equipo y accesorios de cómputo	44
Socio o accionista	44
Comercio al por mayor de materiales metálicos	38
Construcción de carreteras, autopistas, terracerías, puentes, pasos a desnivel y aeropistas	37
Servicios de protección y custodia mediante el monitoreo de sistemas de seguridad	36
Comercio al por menor en ferreterías y tlapalerías	33
Siembra, cultivo y cosecha de aguacate	33
Comercio al por mayor de equipo y material eléctrico	32
Inmobiliarias y corredores de bienes raíces	32
Otros trabajos especializados para la construcción	32
Administración y supervisión de construcción de otras obras de ingeniería civil u obra pesada	31
Otros intermediarios del comercio al por menor	31
Servicios de ingeniería	31
Instalaciones eléctricas en construcciones	28
Otros servicios de publicidad	28
Restaurantes-bar con servicio de meseros	28
Comercio al por mayor de maquinaria y equipo para la industria manufacturera	26
Servicios de investigación y de protección y custodia, excepto mediante monitoreo	26
Servicios de administración de negocios	25
Comercio al por mayor de cemento, tabique y grava	24
Comercio al por mayor de fertilizantes, plaguicidas y semillas para siembra	23
Comercio al por mayor de otra maquinaria y equipo de uso general	23
Comercio al por mayor de otras materias primas para otras industrias	23
Construcción de vivienda multifamiliar	23
Construcción de naves y plantas industriales	22
Otros servicios relacionados con la contabilidad	22

Tabla 12: Distribución de RFCA sospechosos identificados por los métodos de clasificación en las distintas actividades económicas.