Modelización analítica y técnicas avanzadas para el AML/CFT

"Un modelo resulta siempre parcial, pero ofrece recursos para progresar en el conocimiento"
Jean-Pierre Changeux[73]

En esta sección se describen algunas de las tendencias y prácticas más innovadoras del sector basadas en la modelización analítica y en técnicas avanzadas para la identificación, gestión, control y supervisión del blanqueo de capitales.

El contexto para el enfoque analítico de la evaluación del AML/CFT

Con la aparición de una normativa más restrictiva, que busca una mejor y más rápida identificación del riesgo, y las nuevas tecnologías disponibles, las entidades financieras están avanzando en un nuevo viaje de transformación en lo que respecta a la implementación de la adopción de análisis avanzados de AML/CFT[74]. Las tres herramientas principales que se utilizan para detectar el blanqueo de capitales son la evaluación del riesgo del cliente, la monitorización de transacciones y el escaneo de sanciones.

Evaluación de riesgo del cliente

La evaluación del riesgo del cliente es un modelo basado en los factores de riesgo asociados a la identificación del blanqueo de capitales, como el país del cliente, la ocupación y el salario, los productos bancarios, etc.

Los modelos estadísticos se han convertido en la práctica habitual para la evaluación del riesgo de los clientes, mediante la aplicación de diferentes técnicas para resolver el problema de detección de anomalías. Sin embargo, este problema es complejo de identificar o reproducir, y produce muestras desbalanceadas.

La aplicación de métodos avanzados de datos permite superar estas limitaciones, mejora la precisión de la evaluación del riesgo del cliente y fomenta su relevancia a lo largo del programa de AML/CFT. La evaluación del riesgo del cliente evoluciona progresivamente hacia una evaluación del riesgo del cliente basada en el comportamiento, en la que se actualizan continuamente los datos y se enriquece el proceso de identificación del riesgo[75]. Además, los propios modelos están incorporando la ventaja de utilizar técnicas de aprendizaje automático. Los métodos de aprendizaje supervisado, como el *random forest*, son los primeros en aplicarse para desvelar las relaciones ocultas entre los factores de riesgo en un conjunto aumentado de factores.

A medida que aumenta la potencia de cálculo y la riqueza y profundidad de los datos, estos modelos de comportamiento también pueden incorporar desencadenantes de una posible estructuración de las transacciones, es decir, estrategias colectivas de blanqueo de dinero por parte de múltiples individuos a través de pequeñas cantidades, para evitar la detección por parte de las estrategias clásicas de detección estática. La capacidad de construir algoritmos y estrategias que se ejecutan no en base a un cliente individual o un cliente más una transacción, sino en conjuntos de clientes, permite la identificación de la agrupación de transacciones de una manera más proactiva y eficaz. Los llamados algoritmos gráficos[76,77] aprovechan las conexiones potenciales procedentes de diferentes fuentes de información[78]. Además, la capacidad de construir una representación de red completa de todos los clientes aporta el valor adicional de agilizar el proceso de investigación de alertas, entre otros.

[73] Jean-Pierre Changeux (b.1936) es un neurocientífico francés conocido por sus investigaciones en varios campos de la biología, desde la estructura y función de las proteínas, al desarrollo temprano del sistema nervioso hasta las funciones cognitivas.

[74] Sin embargo, no hay uniformidad en el grado de adopción de estas técnicas de análisis avanzadas. Mientras que algunas entidades financieras están experimentando con soluciones innovadoras, las aplicaciones simples son más habituales en el sector, y la dependencia del soporte analítico está en sus inicios para otras. No obstante, el presente y el futuro de los programas de AML/CFT no pueden entenderse sin examinar las nuevas tecnologías y metodologías disponibles.

[75] Por ejemplo, incorporando información procedente del seguimiento de las transacciones, el escaneo de pagos o el análisis de valores atípicos en torno a los canales, los volúmenes, la geolocalización, etc.

[76] Soltani, Reza & Nguyen, Uyen & Yang, Yang & Faghani, Mohammad & Yagoub, Alaa & An, Aijun. (2016. 1-7. 10.1109/UEMCON.2016.7777919

[77] Aprendizaje gráfico escalable para la lucha contra el blanqueo de dinero: Un primer vistazo; Weber, M; Chen, J.; Suzumura, T.; Pareja, A.; Ma, T.; Kanezashi, H., Kaler, T.; Leisersen C.E.; Schardl, Tao B

[78] Por ejemplo, circuitos cerrados de transaccionalidad -transferencias periódicas-, a la titularidad de cuentas conjuntas, a la dirección única, a la sucursal elegida o a las sucursales o cajeros más visitados, al geoposicionamiento a través de la app móvil, a la coincidencia de comercios, etc.

Monitorización de transacciones

El enfoque más común para la monitorización de transacciones consiste en un sistema basado en reglas, al estilo de un árbol de decisiones. Cada regla está configurada para identificar un comportamiento definido que enmascara las posibles actividades de blanqueo de los clientes y las entidades implicadas en la transacción[79]. Estas reglas se identifican generalmente como "escenarios". Las reglas y los escenarios más complejos tratan de abordar la identificación de cuentas anidadas y de relaciones más sofisticadas entre las partes, pero la base de la identificación de valores atípicos sigue siendo, en general, el nivel de la transacción individual, examinando los datos recibidos durante el proceso transaccional. Cuando se identifica un valor atípico, se activa una alerta, que posteriormente requiere la evaluación de un experto[80].

En este proceso, el conjunto inicial de reglas se desglosa en una segmentación más profunda de los comportamientos en la que la línea de negocio, el nivel de actividad transaccional y la evaluación del riesgo del cliente determinan los valores atípicos finales del comportamiento, es decir, las alertas que se dispararían.

Los métodos de análisis de datos pueden aprovecharse para detectar más alertas de calidad, aumentando los verdaderos positivos y reduciendo los falsos negativos, es decir, se identifican más alertas verdaderas sin aumentar el ruido en la identificación. Las técnicas de análisis de datos y aprendizaje automático se implementan para optimizar la segmentación proporcionando una identificación más precisa de los patrones gracias a la exploración de los datos históricos[81].

No obstante, las entidades financieras que estén estudiando activamente la incorporación de métodos avanzados en su programa de AML/CFT podrían decidir centrarse en la priorización de las alertas. El enfoque de las reglas genera grandes cantidades de alertas incluso cuando se aplica un ajuste adecuado de los umbrales del escenario y se ha optimizado la segmentación. Para solucionar esto, muchos bancos implementan métodos de aprendizaje supervisado para clasificar las alertas en términos de productividad[82]. El aspecto clave que determina el éxito de este enfoque es la utilización de métricas diferenciales, más allá de las variables esperadas e inamovibles disponibles a nivel de transacción.

El enfoque más disruptivo para la identificación de los riesgos de AML/CFT consiste en abandonar el enfoque tradicional de las reglas individuales para desvelar la relación oculta con la analítica avanzada. Sin embargo, pocas entidades financieras están explorando la utilización de metodologías alternativas. Algunas de ellas son:

- La analítica de grafos, que está ocupando su espacio en la identificación de las relaciones de la red y es cada vez más determinante para las actividades de blanqueo en el mundo financiero interconectado.

- Técnicas de clustering, que ayudan a identificar los valores atípicos sin asumir comportamientos específicos; por lo tanto, capturando con mayor frecuencia nuevas actividades ilícitas potenciales.

Avanzar hacia un enfoque no basado en reglas no implica automáticamente abandonar las buenas prácticas de optimización previamente identificadas. De hecho, la utilización de análisis avanzados para mejorar la segmentación de clientes, combinada con la detección de redes y de valores atípicos, y la utilización de la priorización de alertas podría considerarse una solución integral.

[79] Este comportamiento sospechoso se basará muy probablemente en los valores atípicos de la ubicación, el recuento de transacciones o los importes de las mismas.

[80] Véase Scalable Graph Learning for Anti-Money Laundering: A First Look; Weber, Chen, Suzumura, Pareja, Ma, Kanezashi, Kaler, Leisersen Schardl, Tao.

[81] El ajuste de umbrales basado en datos permite optimizar los cubos de productividad creciente a lo largo de las variables utilizadas en los escenarios (más verdaderos positivos), al tiempo que proporciona medidas del riesgo potencial no identificado (limitando los falsos negativos). Estos enfoques comunes se basan en los motores existentes basados en reglas.

[82] Este enfoque puede considerarse como una imitación de la revisión de las alertas por parte de los analistas de nivel 1; sin embargo, podría ser una identificación más compleja de abordar y no todas las entidades tienen éxito en este esfuerzo.

Un ejemplo de evaluación nacional de riesgos

El gobierno del Reino Unido publica periódicamente una evaluación nacional de riesgos[1], que informa sobre los riesgos de delito financiero a los que se enfrenta a nivel nacional. A través de esta evaluación nacional de riesgos se aportan referencias sobre las técnicas más habituales utilizadas en el ML/FT y su nivel de implantación en el país y son una referencia importante para las propias entidades en su evaluación del riesgo.

Una compañía debe realizar una evaluación del riesgo de delito financiero y utilizarla para diseñar sus controles de AML/CFT. La evaluación del riesgo nacional sirve, por tanto, como una base sólida sobre la que construir esta evaluación, en la que la compañía toma medidas adicionales para comprender, de forma más específica, los riesgos a los que se enfrenta.

Esto tendría en cuenta, entre otras cosas, su cartera de clientes y los productos que tienen: las cuentas corrientes personales sirven como medio de evasión fiscal para muchas pequeñas empresas, además de introducir la exposición a muchas otras técnicas de blanqueo de capitales debido a su capacidad para realizar transferencias rápidas de fondos y aceptar transacciones en efectivo. Además, una revisión de la actividad delictiva histórica puede ayudar a comprender cualquier tipología adicional a la que se enfrente el banco.

Las transacciones en efectivo, que entran y salen de las cuentas, son una forma fácil para los blanqueadores de dinero de romper los rastros de las transacciones. Aunque el uso de efectivo en el blanqueo de capitales está muy extendido y se incluye en muchas de las estrategias utilizadas, los controles en torno a los riesgos del efectivo suelen ser los más sencillos, en gran medida debido a la poca información disponible sobre las transacciones en efectivo.

Las mulas de dinero son terceras partes que se utilizan, consciente o inconscientemente, para realizar transacciones adicionales en efectivo y transferencias de fondos que enmascaran los rastros de las transacciones. Esto puede utilizarse junto con otras estrategias, por ejemplo, la compra de activos de alto valor y revendibles, para eliminar casi por completo las sospechas sobre el origen de los fondos, cuando las cuentas temporales podrían ser las de una red de mulas. Esto es difícil de detectar utilizando los métodos tradicionales, ya que ninguna cuenta, ni ningún cliente, puede ser utilizado para grandes volúmenes de las transacciones utilizadas en cualquier etapa de este proceso.

Del mismo modo, los negocios con gran cantidad de dinero en efectivo suponen otro reto para los métodos de detección tradicionales. Negocios como los salones de belleza, los quioscos de prensa y los lavaderos de coches son utilizados por los blanqueadores de capitales para documentar el dinero en efectivo procedente de actividades delictivas como ingresos comerciales legítimos, de modo que grandes volúmenes de los fondos ilícitos de las redes delictivas puedan centralizarse en una sola cuenta. Esto resulta difícil de detectar, ya que los ingresos en efectivo del negocio pueden parecer coherentes con su propio historial, así como con los ingresos de sus compañeros, y por lo tanto es posible que las transacciones en efectivo del negocio no levanten sospechas. Sin embargo, estas empresas suelen estar también vinculadas a la trata de personas y a la esclavitud moderna, que incluyen sus propios comportamientos transaccionales que pueden ser más fáciles de detectar. Al igual que con el uso de mulas de dinero, estas tipologías suelen implicar una red de terceros aparentemente no relacionados. Estos terceros pueden ser los facilitadores o incluso las víctimas de estos delitos y, por lo tanto, hay comportamientos específicos que uno esperaría ver. Las transacciones en varias ciudades diferentes, especialmente en ciudades con centros de transporte, el uso intensivo de restaurantes de comida rápida, las transacciones múltiples en el mismo hotel en el mismo día, los pagos múltiples a proveedores de telefonía móvil, las transferencias de fondos entre cuentas con comportamientos similares y las transacciones internacionales, especialmente las transferencias de efectivo y de fondos, son fuertes indicadores de estas tipologías. Si se puede vincular a estas partes con el negocio de uso intensivo de efectivo, se podría descubrir la red completa.

Las transacciones internacionales son otra operación de alto riesgo identificada en la evaluación nacional de riesgos. Se observan en una variedad de técnicas de blanqueo de capitales, además de presentar un riesgo en otros aspectos de la delincuencia financiera. Esto se ve en el tráfico de personas, que se estima que es uno de los mayores generadores de ganancias criminales a nivel mundial. El tráfico de personas requiere el envío al extranjero de los miembros de la banda de delincuencia organizada asociada en los países relacionados con el tráfico. Esto puede ser en forma de dinero en efectivo retirado en el Reino Unido y trasladado físicamente al extranjero o a través de mulas de dinero de manera similar al comportamiento asociado con los depósitos en efectivo descritos anteriormente.

La financiación del terrorismo está identificada como una tipología de alto riesgo en el Reino Unido. La recaudación y el movimiento de fondos no se consideran un objetivo primordial de los terroristas, especialmente porque la mayoría de los recientes atentados terroristas han sido de bajo presupuesto y poca sofisticación, y con frecuencia han sido planificados, financiados y realizados por un individuo. La financiación del terrorismo se utiliza habitualmente para trasladar fondos al extranjero a través de métodos relativamente sencillos, como el traslado físico de dinero en efectivo al extranjero o el empleo de empresas de servicios monetarios (MSB). Por lo tanto, la detección de la financiación del terrorismo requiere una recopilación de indicadores clave de la misma manera que se requiere para el uso de empresas con gran cantidad de efectivo en el blanqueo de capitales.

El riesgo asociado a los criptoactivos crece año tras año a medida que los criptoactivos se vuelven más comunes y de fácil acceso, pero los controles en torno a ellos siguen siendo relativamente nuevos, ya que el Reino Unido no introdujo normas sobre el uso de criptoactivos para el blanqueo de capitales hasta enero de 2020. Las bandas criminales organizadas utilizan los criptoactivos para el blanqueo de dinero comprando primero los criptoactivos con sus fondos ilícitos, potencialmente después de una etapa inicial de estratificación, antes de vender los activos para proporcionar una fuente legal de sus fondos. Además, los criptoactivos pueden moverse fácilmente a través de las fronteras, lo que permite a los delincuentes mover importantes fondos a nivel internacional con gran facilidad en comparación con las monedas fiduciarias.

Este es un ejemplo de los nuevos riesgos que surgen en el ámbito del delito financiero y que suponen un nuevo reto para las entidades, que deben desarrollar y poner en práctica nuevos controles de forma regular para mantenerse al día con los cambios y la evolución de los blanqueadores de capitales.

[1]HM Treasury: National risk assessment of money laundering and terrorist financing 2020. December 2020.

Escaneo de sanciones

Los motores de escaneo de sanciones comparan a las personas o empresas con la lista de sanciones designada utilizando técnicas de coincidencia difusa. Los enfoques más sencillos se basan en una amplia gama de transformaciones aplicadas a los "nombres" (cambio de orden del nombre, iniciales, transliteración, errores vocales o consonánticos comunes, etc.). Los nombres transformados se normalizan como cadenas y se comparan con los nombres de la lista de sanciones, también normalizados siguiendo las mismas reglas. Las reglas o lógicas de comparación miden el grado de separación entre las dos cadenas. El motor puede devolver una puntuación de la coincidencia, o una alerta basada en una regla de coincidencia predefinida, sin embargo, la lógica subyacente es la misma, es decir, las dos cadenas son lo suficientemente similares como para conceder una revisión de expertos.

Como en el caso de la monitorización de transacciones, estas reglas producen un gran número de falsos positivos[83]. Además, el potencial de optimización basado en el ajuste es menor que en el caso de la monitorización de transacciones.

Por ello, las entidades están explorando métodos alternativos para mejorar la calidad de la identificación basados en tecnologías de traducción y transliteración, y en la aplicación de tecnologías de procesamiento de lenguaje natural (NLP) para mejorar la coincidencia de nombres. La mejora de los métodos analíticos para el escaneo de sanciones va en paralelo a la exploración de estas técnicas en la identificación de noticias negativas.

Los próximos pasos en los enfoques analíticos de la evaluación de AML/CFT

La aplicación de métodos y tecnologías innovadoras no se detiene en las destacadas anteriormente. El procesamiento extendido del lenguaje natural y el aprendizaje profundo, las aplicaciones de blockchain, la verificación electrónica de la identidad, el reconocimiento de voz y del habla, la biometría o la geolocalización son otras tecnologías que pueden contribuir a la identificación de actividades ilícitas.

Detrás de todos estos posibles enfoques, se encuentran varias tendencias en el análisis de AML/CFT:

- Se implementa un análisis más profundo de los datos tanto de la transacción, como del cliente y sus relaciones. Algunas de las opciones analíticas señaladas anteriormente se vuelven impotentes si no se dispone de datos diferenciales y se incorporan al análisis.

- Se requieren datos complementarios de las fuentes internas y de las diferentes dimensiones del programa de AML/CFT (es decir, calificación del riesgo del cliente, diligencia debida, escaneo de sanciones, transacciones) y fuentes externas (datos públicos sobre PEP, relaciones de propiedad, fuentes de reputación, búsquedas abiertas) para crear un enfoque holístico de la identificación del riesgo de ML/FT.

- Las tecnologías y los métodos pueden ser tan complejos como lo permita la innovación, pero dimensionar los más adecuados a la naturaleza del negocio y a la evaluación de riesgos de la entidad es fundamental para optimizar el uso de los recursos tecnológicos y humanos, al tiempo que se garantiza el cumplimiento de la normativa.

Los supervisores y reguladores son en general reacios a los cambios repentinos y favorecen las metodologías bien

[83] Los motores pueden ser más o menos complejos en la incorporación de transformaciones innovadoras aplicadas a los nombres, o incorporar más fuentes de sanción de calidad mejoradas con información PEP, sin embargo, todos presentan las mismas debilidades.

establecidas antes de adoptar plenamente los cambios revolucionarios. Sin embargo, para aquellas instituciones que están dispuestas a embarcarse en un programa de transformación total de la analítica de AML/CFT, se han producido una serie de avances en los últimos años[84]: desde desarrollos específicos de aplicaciones de coincidencia difusa o de detección de PEP en colaboraciones conjuntas, hasta la constitución de centros de innovación y *sandboxes*.

En el camino hacia una identificación de riesgos más sofisticada, la interpretabilidad y el control adecuado de los riesgos siguen siendo el centro de las preocupaciones del regulador (y de las instituciones).

El uso de la analítica avanzada en el programa de AML/CFT está vinculado a que las reglas implementadas se consideren modelos y estén por tanto, sujetos a las prácticas de identificación, monitorización y control que las entidades han desplegado bajo la función de Gestión del riesgo de modelo (MRM). Mientras que la distinción para la calificación del riesgo de cliente es clara, ya que cumple todas las condiciones típicamente establecidas en el marco de la gestión del riesgo de modelo (MRM) para ser un modelo o, al menos, una herramienta de usuario que debe ser supervisada, los motores de AML/CFT no han sido considerados como modelos inicialmente. La asimilación de los motores de reglas de AML/CFT en la disciplina de la gestión del riesgo de modelo no se ha producido de manera uniforme en todas las jurisdicciones y los principales actores quieren evitar la carga de un escrutinio incremental de los programas de AML/CFT[85].

Sin embargo, las tecnologías de *machine learning* para mejorar la identificación de riesgos están ampliando la concepción de lo que se entiende como modelo sujeto a MRM. A pesar de su voluntad de fomentar su aplicación a los programas de AML/CFT, los supervisores dejan clara la necesidad de garantizar un grado adecuado de comprensión e interpretabilidad de las metodologías aplicadas y los resultados obtenidos. Hay que evitar los modelos de caja negra. Los modelos de *machine learning* pueden adolecer de falta de transparencia en la selección y explicación de las características, la evaluación del rendimiento del modelo, etc. Una documentación adecuada, la comprobación del modelo, los módulos de interpretabilidad; los principios básicos de un marco robusto de MRM apoyarán la adecuación de estos modelos para el uso de AML/CFT.

Caso práctico: mejorar la detección de patrones sospechosos mediante el análisis de redes

Una de las técnicas aplicadas con éxito para detectar el fraude es el denominado análisis de redes. Esta técnica puede ayudar a identificar, detectar y caracterizar comportamientos sospechosos utilizando métricas, técnicas de aprendizaje automático y algoritmos difusos.

Para desarrollar el análisis de redes, hay que dar tres pasos relevantes: (i) recopilar datos relevantes y construir un gráfico que represente las relaciones entre las entidades; (ii) decidir la estrategia de identificación que permita identificar el cluster de entidades y relaciones sospechosas; y (iii) caracterizar esos clusters mediante métricas apropiadas que se utilizarán como características de los modelos de detección (véase figura 4).

Etapa 1. Representación de la red

Una red permite examinar relaciones complejas entre entidades relacionadas, ya sea mediante vínculos de datos internos, como transacciones, o externos, como direcciones y titularidades

Figura 4. Etapas para la detección mediante el análisis de redes.

[84] En palabras del reciente documento publicado por el GAFI, "las nuevas tecnologías tienen el potencial de hacer que las medidas de AML/CFT sean más rápidas, baratas y eficaces". Además, el GAFI enumera las múltiples iniciativas de supervisores y entidades de todo el mundo que constituyen la vanguardia de la evolución del sector Ver: Opportunities and challenges of new technologies for AML/CFT, disponible en https://www.fatf-gafi.org/media/fatf/documents/reports/Opportunities-Challenges-of-New-Technologies-for-AML-CTF.pdf

[85] Una declaración conjunta de la FRS, la FDIC y la OCC abordó las preguntas del sector sobre cómo deben aplicarse las directrices del MRM a los modelos de cumplimiento de la BSA/AML. Los supervisores consideran que no se requiere que todos los sistemas se clasifiquen como modelos, y que el propio banco puede categorizar los modelos como considere oportuno. Y lo que es más importante, afirmaron que los bancos no están obligados a tener procesos duplicados ni a llevar a cabo actividades de prueba duplicadas para cumplir con la normativa BSA. Aunque proporciona cierto grado de maniobra a las instituciones financieras, la declaración refuerza la opinión de que el banco debe abordar los riesgos asociados a los sistemas de AML (con modelos o sin ellos).

(véase la figura 5). La construcción de las redes requiere computar suficientes puntos de datos granulares que puedan conectar las entidades con diferentes objetos, como empresas, direcciones, usos digitales, etc., y considerar la fuerza de estas relaciones (por ejemplo, conexión transaccional). Esta red puede estructurarse como un grafo (tanto dirigido como no dirigido, y ponderado o no ponderado). La red construida y la información contenida en ella determinarán la idoneidad de determinadas técnicas (por ejemplo, un grafo no dirigido ponderado podría tratarse en los pasos siguientes mediante técnicas de agrupación, como la agrupación espectral).

Etapa 2. Estrategia de identificación

Se necesita una estrategia de identificación para descubrir posibles pautas de blanqueo de capitales u otras actividades ilícitas dentro de la red identificada. Existen diferentes estrategias que pueden utilizarse, por ejemplo:

- Enfoques heurísticos basados en la proximidad a casos o entidades sospechosos confirmados.
- Enfoques probabilísticos y reconocimiento de patrones.
- Enfoque de detección de comunidades basado en técnicas de aprendizaje automático.

Al aplicar el enfoque de detección de comunidades, es necesario descubrir las distintas comunidades. Una comunidad es un subgrafo de la red con un mayor número y una relación más intensa entre los miembros de la comunidad en comparación con subgrafos aleatorios y poco informativos (véase la figura 6). La detección de comunidades es un enfoque útil para detectar y caracterizar las estructuras objetivo, que puede requerir el uso de algoritmos como *k-means*, *clustering* jerárquico, *clustering* espectral, algoritmos evolutivos u optimización de la modularidad[86].

Para encontrar las comunidades óptimas, se optimiza una función específica: la función de modularidad. Dada una red representada como un grafo ponderado y particionada en comunidades o módulos, esta fórmula depende de la estructura específica de la representación gráfica, y expresa la definición matemática de modularidad en términos de pesos:

$$Q = \frac{1}{2w}\sum_i\sum_j\left(w_{ij} - \frac{w_i w_j}{2w}\right)\delta(C_i, C_j)$$

Donde Ci es la comunidad a la que está asignado el nodo i, w_{ij} representa el valor del peso en el enlace entre los nodos i y j (0 si no existe enlace), $w_i = \sum_j w_{ij}$, y $w=\sum_i w_i$. Por último, la función δ corresponde a la función delta de kronecker: δ(i,j) toma el valor 1 si los nodos i y j están en el mismo módulo y 0 en caso contrario.

Etapa 3. Uso de funciones

Una vez identificadas las comunidades objetivo dentro de la red, pueden definirse métricas o características específicas para evaluar la profundidad e importancia de las relaciones o el riesgo de las conexiones entre entidades. Estas características pueden utilizarse en reglas o algoritmos de aprendizaje automático para mejorar la capacidad predictiva de los modelos reduciendo los falsos positivos e identificando mejor los patrones sospechosos. El enfoque basado en reglas que incorporan características "enriquecidas" puede ser útil para producir alertas cualitativas, ya que incorporan nueva información aparte de la base transaccional tradicional

[86] Varios autores han desarrollado algoritmos óptimos para la detección de patrones. Véase L. Alsedà, A. Awasthi, Jörg Lässig (2012).

Figura 5. Representación del algoritmo de detección de comunidades (pasos (a) a (d)) utilizando la optimización de la modularidad.
La comunidad superior contiene un comportamiento sospechoso (objetivo) que corresponde en este caso de uso a la red del Cliente A de la Etapa 1 (representado en la figura 7 en la página siguiente).

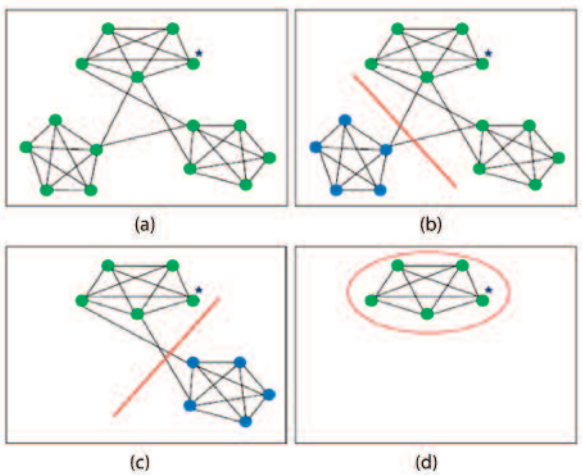
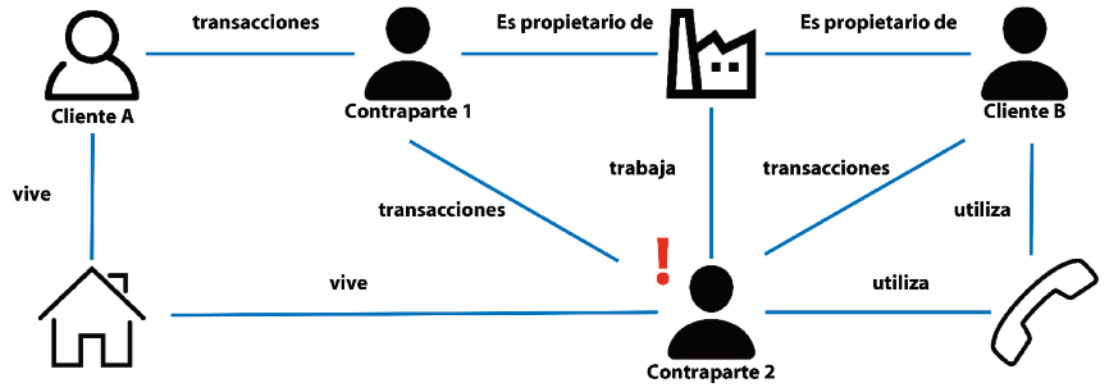


Figura 6. Lógica simplificada para un escenario basado en características obtenidas mediante el análisis de la red. Otras características, como los importes de las transacciones, pueden incorporarse en función del patrón o la combinación de patrones que se desee detectar. Este ejemplo es meramente ilustrativo.

Lógica y parametrización de reglas de detección con características de red

- Si el tipo de entidad = **Individual**
- Si la comunidad sospechosa identificada: conexiones sospechosas identificadas = **Sí**
- Si el número de rutas al nodo sospechoso >= **num_rutas_a_nodo_sospechoso**
- Si la distancia al nodo sospechoso es <= **distancia_máxima_a_nodo_sospechoso**
- Si las conexiones primarias al nodo sospechoso = **tipo_conexión_principal**

Figura 7. Representación de una red con las relaciones del cliente A, donde se incluye un nodo sospechoso (la contraparte 2 está en una lista negra) y posibles entidades sintéticas (nodos relacionados con la empresa)



relacionada con el cliente (véase la figura 8). Sin embargo, las técnicas de aprendizaje automático pueden desvelar relaciones más sólidas que permitan separar las alertas positivas verdaderas de las falsas.

En el ejemplo de la figura 7, cuya información de redes se presenta en la figura 8, el cliente A y el cliente B pertenecen al mismo *cluster* sospechoso con conexiones a la entidad sospechosa (Contraparte 2), pero el cliente B tiene la relación más fuerte, tanto personal como profesionalmente con la Contraparte 2. Basándonos en este escenario, si los umbrales se calibran para que sean *num_rutas_a_nodo_sospechoso* = 1, *distancia_máxima_a_nodo_sospechoso* = 5 y *tipo_conexión_principal* = "all" (ya sea transaccional, personal o de cualquier tipo), entonces tanto el Cliente A como el B serán marcados como entidades sospechosas (o sus transacciones relacionadas, etc.). Sin embargo, considerando un enfoque más tradicional, sin el análisis de redes, sólo el cliente B sería marcado como tal; el cliente A no tiene conexiones transaccionales con la Contraparte 2.

Se pueden evaluar características complejas y entrenar distintos tipos de algoritmos de aprendizaje automático, lo que permite asignar un mayor riesgo al cliente B y a las transacciones asociadas. La incorporación de nuevas características a los modelos permite también aumentar la precisión y detectar más comportamientos potencialmente arriesgados (reduciendo las falsas alertas negativas), al tiempo que se discrimina mejor el riesgo entre esos comportamientos identificados (reduciendo las falsas alertas positivas).

Figura 8. Información sobre los clientes para la identificación de conexiones sospechosas

Empresa	Distancia mínima al nodo sospechoso	Conexión fundamental al nodo sospechoso	Conexión de datos personales	Número de rutas hacia el nodo sospechoso	Cluster identificado	Conexiones sospechosas identificadas
Cliente A	2	Transaccional	Sí	2	1	Sí
Cliente B	1	Transaccional	Sí	4	1	Sí