Más Allá de los Deep Fakes: Detección de Explotación Masiva de Identidades Reales para Originación de Cuentas Mula en Economías en Desarrollo

Carlos Gómez[1]

[1]Backbone Systems, México
Correspondencia: carlos@backbonesystems.io

Resumen—Los marcos actuales de prevención de lavado de dinero (PLD) para la detección de fraude en procesos de Conoce a tu Cliente (KYC) se enfocan predominantemente en la detección de identidades sintéticas y reconocimiento de deep fakes durante la originación de cuentas. Sin embargo, en economías en desarrollo con altas tasas de pobreza, emerge un modelo de amenaza fundamentalmente diferente: la explotación masiva de identidades *reales* mediante el reclutamiento organizado de personas económicamente vulnerables para la originación fraudulenta de cuentas. Este artículo examina el fenómeno de las "granjas de identidad" en México, donde redes organizadas reclutan personas de comunidades rurales empobrecidas—de manera análoga a la práctica política de los "acarreados"—para abrir cuentas bancarias legítimas que posteriormente sirven como vehículos de cuentas mula. Argumentamos que los controles KYC tradicionales son estructuralmente incapaces de detectar esta amenaza porque cada elemento de verificación es genuino. Proponemos un marco de detección basado en clusters que explota las restricciones operacionales inherentes a la explotación masiva de identidades: concentración geográfica de identidades reclutadas (sección electoral y código postal de la credencial de elector), compresión temporal de eventos de originación, anomalías de geolocalización durante procesos de KYC remoto, y perfilamiento transaccional post-originación contra líneas base socioeconómicas derivadas de datos censales. El marco introduce tres escenarios de detección—clusters de originación en sucursal, granjas de originación con KYC remoto, y detección de spoofing de geolocalización—cada uno produciendo señales de riesgo que individualmente pueden ser insuficientes pero colectivamente permiten un monitoreo efectivo. Discutimos la integración con indicadores socioeconómicos del INEGI y la alineación regulatoria con las disposiciones mexicanas de PLD (LFPIORPI) y los requisitos de BANXICO para la recolección de datos de geolocalización durante la apertura remota.

Palabras clave—cuentas mula, explotación de identidad, fraude KYC, detección por clusters, originación de cuentas, prevención de lavado de dinero, detección de anomalías geoespaciales, economías en desarrollo

1. INTRODUCCIÓN

La industria de prevención de crímenes financieros ha invertido significativamente en tecnologías de detección de deep fakes para procesos de Conoce a tu Cliente (KYC), impulsada por la creciente sofisticación del fraude de identidad sintética (Gao et al., 2024). Los proveedores de verificación de identidad promueven la detección de vida biométrica, la autenticación de documentos y algoritmos de comparación facial como la defensa primaria contra la originación fraudulenta de cuentas. Si bien estas tecnologías abordan un vector de amenaza genuino, operan bajo una suposición implícita de que las cuentas fraudulentas se abren usando identidades *falsas*—sintéticas, robadas o digitalmente manipuladas.

En economías en desarrollo caracterizadas por desigualdad significativa y pobreza concentrada, un modelo de amenaza fundamentalmente diferente domina: la explotación masiva de identidades *reales* de poblaciones económicamente vulnerables. En este modelo, cada elemento de la verificación KYC es genuino. La persona está físicamente presente. El documento de identidad es auténtico. La verificación biométrica es exitosa. Ningún sistema de detección de deep fakes generará una alerta porque no hay nada sintético que detectar (Schuetz, 2021).

Este artículo examina el fenómeno tal como se observa en el sistema financiero mexicano, donde redes organizadas reclutan personas de comunidades rurales empobrecidas para abrir cuentas bancarias que posteriormente sirven como vehículos de cuentas mula. La práctica es paralela al sistema políticamente motivado de "acarreados"—donde se paga a individuos para asistir a mítines políticos o votar—adaptado aquí para la originación de cuentas financieras. Datos recientes indican que casi 2 millones de cuentas de lavado de dinero fueron reportadas globalmente en 2024 (BioCatch, 2024), con economías en desarrollo contribuyendo desproporcionadamente debido a la disponibilidad de poblaciones explotables (U.S. Department of the Treasury, 2024).

Realizamos tres contribuciones. Primero, caracterizamos la mecánica operacional de la explotación masiva de identidades para originación de cuentas, distinguiendo tres escenarios de originación con firmas de detección distintas. Segundo, proponemos un marco de detección basado en clusters que explota las restricciones operacionales inherentes al reclutamiento masivo—concentración geográfica, compresión temporal y anomalías de geolocalización—para identificar granjas de originación. Tercero, describimos un enfoque de monitoreo post-originación que combina la pertenencia a un cluster con perfilamiento socioeconómico para detectar comportamiento transaccional anómalo indicativo de activación de cuentas mula.

2. ANTECEDENTES Y TRABAJO RELACIONADO

2.1. Reclutamiento de cuentas mula en economías en desarrollo

Las cuentas mula son individuos que, consciente o inconscientemente, permiten que sus cuentas bancarias sean utilizadas para transferir fondos ilícitos (Schuetz, 2021). Los métodos de reclutamiento tradicionales incluyen ingeniería social, estafas de empleo en línea y fraude romántico, dirigidos predominantemente a adultos jóvenes y estudiantes (HM Government, 2024). Sin embargo, en economías con pobreza extrema, los incentivos económicos directos crean un canal de reclutamiento mucho más eficiente.

El panorama de pobreza de México proporciona las condiciones estructurales para la explotación masiva de identidades. Según CONEVAL, aproximadamente 36.3 millones de mexicanos (28.9% de la población) vivían en pobreza moderada y 7.1 millones (5.7%) en pobreza extrema en 2022 (CONEVAL, 2022). En comunidades rurales donde los salarios diarios pueden estar por debajo de 200 MXN, una oferta de 500–1,000 MXN por "un día de trabajo" representa una proposición económica convincente. Patrones similares han sido documentados en Filipinas, donde la pobreza sistémica permitió el reclutamiento de cuentas mula a gran escala en poblaciones rurales (Flores and Santos, 2024).

La distinción crítica respecto al reclutamiento convencional de mulas es la escala y organización. En lugar del reclutamiento individual mediante ingeniería social, las redes organizadas realizan operaciones por lotes: viajando a comunidades rurales, reclutando docenas de individuos simultáneamente y procesándolos a través de la originación de cuentas en una ventana de tiempo comprimida.

2.2. Limitaciones de los enfoques de verificación de identidad

Los sistemas actuales de detección de fraude KYC están diseñados alrededor de tres modelos de amenaza primarios: (a) fraude de identidad sintética, donde se construyen identidades ficticias a partir de elementos de datos fabricados o combinados; (b) fraude documental, donde los documentos de identidad son falsificados, alterados u obtenidos mediante corrupción; y (c) suplantación de identidad, donde una persona presenta las credenciales de otra, cada vez más facilitado por tecnología de deep fakes (Gao et al., 2024).

La explotación masiva de identidades reales cae fuera de las tres categorías. La identidad no es sintética—pertenece a una persona real y viva. El documento no es falsificado—es una credencial genuina emitida por el Instituto Nacional Electoral (INE). La persona no está suplantando a nadie—está presente y consintiendo (aunque no comprenda completamente las consecuencias). Esta incompatibilidad categórica significa que las inversiones en detección de deep fakes, análisis forense de documentos y pruebas de vida biométrica, aunque valiosas para otros vectores de amenaza, proporcionan cero capacidad de detección para este ataque específico.

2.3. Análisis de clusters para detección de fraude

Los métodos de clustering basados en densidad, particularmente DBSCAN (Ester et al., 1996), han sido aplicados exitosamente a la detección de fraude financiero. Pourhabibi et al. (2020) proporcionan una revisión exhaustiva de la detección de anomalías basada en grafos en sistemas financieros, demostrando que los patrones a nivel de red frecuentemente revelan fraude que el análisis a nivel de transacción no detecta. Chandola et al. (2009) establecen los fundamentos teóricos para la detección de anomalías en general, señalando que las anomalías contextuales—puntos de datos que son anómalos solo dentro de un contexto específico—son particularmente relevantes para la detección de fraude donde registros individuales pueden parecer legítimos en aislamiento.

El clustering geoespacial ha emergido como una herramienta poderosa para la detección de crímenes financieros, con el sistema de indexación hexagonal H3 de Uber (Brodsky, 2019) permitiendo la agrupación espacial eficiente de eventos. La aplicación de clustering espacio-temporal a eventos de originación de cuentas, sin embargo, ha recibido atención limitada en la literatura. Nuestro trabajo aborda esta brecha proponiendo un marco específicamente diseñado para detectar patrones de originación masiva.

3. MODELO DE AMENAZA: EXPLOTACIÓN MASIVA DE IDENTIDADES PARA ORIGINACIÓN DE CUENTAS

3.1. Mecánica operacional

El proceso operacional sigue un patrón consistente a través de las instancias observadas. Una red organizada despliega operadores en comunidades rurales caracterizadas por pobreza extrema e inclusión financiera limitada. Los operadores ofrecen pagos en efectivo a los residentes a cambio de su participación en procesos de apertura de cuentas. Se observan tres escenarios de originación:

(a) **Originación en sucursal**: los individuos reclutados son transportados en autobús o camioneta a sucursales bancarias, típicamente en centros urbanos cercanos, donde abren cuentas a través de procesos KYC presenciales estándar;

(b) **Granja de originación con KYC remoto**: los individuos reclutados son transportados a una ubicación fija (bodega, oficina rentada o propiedad residencial) donde múltiples dispositivos se utilizan para completar procesos de KYC remoto simultáneamente;

(c) **Originación en sitio**: el proceso de KYC se realiza directamente en la comunidad rural utilizando dispositivos móviles, eliminando la necesidad de transporte.

En los tres escenarios, los individuos reclutados proporcionan su credencial de elector (credencial para votar) genuina del INE, se someten a verificación biométrica y completan el proceso de apertura de cuenta. Las cuentas son posteriormente controladas por la red organizadora, que las utiliza para lavado de dinero, dispersión de productos de fraude u otros flujos financieros ilícitos.

3.2. Restricciones operacionales explotables

A pesar del desafío fundamental de las identidades genuinas, el modelo de originación masiva introduce restric-

ciones operacionales que crean patrones detectables. Identificamos cuatro propiedades invariantes:

Concentración geográfica de identidades. Dado que los operadores reclutan dentro de una sola comunidad o área geográfica pequeña, los individuos reclutados comparten datos demográficos codificados en sus credenciales del INE: código postal, sección electoral y municipio. Estos campos no pueden ser alterados por los atacantes porque están impresos en el documento de identidad genuino.

Compresión temporal. El proceso de originación debe ocurrir dentro de una ventana de tiempo comprimida (típicamente horas a pocos días) porque los operadores no pueden recolectar credenciales para uso posterior—la persona debe estar físicamente presente para la verificación KYC. Esto genera un patrón de ráfaga en las marcas temporales de originación de cuentas.

Clustering de geolocalización. Para procesos de KYC remoto, las disposiciones de BANXICO requieren la recolección de coordenadas de geolocalización del dispositivo (Banco de México, 2021). Cuando la originación ocurre en una granja, múltiples dispositivos en la misma ubicación física procesan cuentas simultáneamente, creando un cluster geoespacial denso.

Desajuste socioeconómico-transaccional. Se espera que las cuentas abiertas por individuos de zonas de pobreza extrema exhiban volúmenes transaccionales bajos consistentes con su contexto socioeconómico. Cuando estas cuentas comienzan a procesar transacciones de alto valor o alta frecuencia, la desviación del perfil esperado es un indicador fuerte de activación de cuenta mula.

4. MARCO DE DETECCIÓN PROPUESTO

4.1. Escenario 1: Detección de clusters de originación en sucursal

Para cuentas abiertas en sucursales físicas, proponemos monitorear la distribución de identificadores geográficos derivados del INE en las nuevas originaciones de cuentas. El mecanismo de detección opera de la siguiente manera:

Para cada sucursal b y ventana de tiempo w (p. ej., 7 días), computamos el conjunto de nuevas cuentas $A_{b,w}$ y extraémos la sección electoral s_i del INE de cada titular. Luego computamos la distribución de frecuencia $f(s)$ de secciones electorales y marcamos cualquier sección s^* donde:

$$f(s^*) > \mu_{b,w} + k \cdot \sigma_{b,w} \tag{1}$$

donde $\mu_{b,w}$ y $\sigma_{b,w}$ son la media histórica y desviación estándar de los conteos de originación por sección para la sucursal b durante ventanas comparables, y k es un parámetro de sensibilidad (recomendado $k = 3$ para despliegue inicial). Una sucursal que repentinamente registra 30 cuentas de una sola sección electoral rural dentro de una semana, cuando la línea base histórica es 0–2, produce una señal de anomalía fuerte.

Alternativamente, un enfoque no paramétrico usando el método del rango intercuartílico (IQR) puede aplicarse cuando las distribuciones históricas son asimétricas:

$$f(s^*) > Q_3 + 1.5 \cdot \text{IQR} \tag{2}$$

4.2. Escenario 2: Detección de granjas de originación con KYC remoto

Para la originación remota de cuentas, los datos de geolocalización proporcionan una dimensión adicional de detección. Aplicamos DBSCAN (Ester et al., 1996) sobre el conjunto de eventos de originación $E = \{(lat_i, lon_i, t_i, s_i)\}$ donde lat_i, lon_i son coordenadas del dispositivo, t_i es la marca temporal de originación, y s_i es la sección electoral del INE del solicitante.

El clustering opera en dos etapas. Primero, DBSCAN espacio-temporal agrupa eventos de originación que son geográficamente próximos ($\epsilon_{geo} \leq 500$ m) y temporalmente cercanos ($\epsilon_{time} \leq 48$ h). Segundo, dentro de cada cluster identificado, evaluamos la homogeneidad de secciones electorales. Un cluster donde $> 70\%$ de los solicitantes comparten la misma sección electoral o código postal, y esa sección está geográficamente distante de las coordenadas de originación, constituye una señal fuerte de granja de originación.

La condición de detección se puede formalizar como:

$$\text{score}(C) = \frac{|\{e_i \in C : s_i = s_{mode}\}|}{|C|} \times \mathbb{1}[d(s_{mode}, \bar{g}_C) > \delta] \tag{3}$$

donde C es un cluster detectado, s_{mode} es la sección electoral más frecuente dentro del cluster, $\bar{g}_C$ es el centroide geográfico del cluster, $d(\cdot)$ mide la distancia entre el municipio registrado de la sección electoral y el centroide del cluster, y δ es un umbral de distancia mínima (p. ej., 50 km).

4.3. Escenario 3: Detección de spoofing de geolocalización

Los atacantes sofisticados pueden falsificar coordenadas GPS para dispersar los eventos de originación geográficamente, intentando evadir la detección de granjas de originación del Escenario 2. Sin embargo, esta contramedida paradójicamente crea un patrón más anómalo: un gran número de individuos de la misma sección electoral rural iniciando originación de cuentas simultáneamente desde ubicaciones geográficamente dispersas.

La detección aprovecha la inverosimilitud del patrón resultante. Para una sección electoral s y ventana de tiempo w dados, computamos el conjunto de eventos de originación $E_{s,w}$ y medimos la dispersión geográfica $\sigma_{geo}(E_{s,w})$ como la desviación estándar de las distancias entre pares de coordenadas de originación. La condición de detección es:

$$|E_{s,w}| > \theta_{count} \ \wedge \ \sigma_{geo}(E_{s,w}) > \theta_{disp} \tag{4}$$

donde θ_{count} es un tamaño mínimo de cluster (p. ej., 10 cuentas) y θ_{disp} es un umbral de dispersión. Esto captura el escenario donde muchas personas de la misma pequeña comunidad simultáneamente deciden abrir cuentas desde ubicaciones dispersas por todo el país—un evento estadísticamente inverosimíl para el perfil demográfico de comunidades rurales con índices de pobreza extrema.

5. MONITOREO POST-ORIGINACIÓN

5.1. Monitoreo de comportamiento basado en clusters

La detección de un cluster de originación no constituye, por sí misma, evidencia suficiente para clasificar cuentas como fraudulentas o vehículos mula. La señal de cluster sirve dos propósitos operacionales.

Primero, todas las cuentas dentro de un cluster marcado son etiquetadas para monitoreo de comportamiento coordinado. Si las cuentas del cluster comienzan a recibir y dispersar fondos en patrones sincronizados—una característica distintiva de la activación de redes mula (Weber et al., 2019)—la evidencia de uso ilícito se fortalece significativamente. Las métricas de coordinación incluyen sincronización temporal de transferencias entrantes, similitud en montos de transacciones, contrapartes comunes y dispersión rápida de fondos (típicamente dentro de 24–48 horas de recibidos).

Segundo, las cuentas individuales dentro de clusters marcados están sujetas a perfilamiento transaccional mejorado, como se describe en la siguiente sección.

5.2. Perfilamiento transaccional socioeconómico

Proponemos construir perfiles transaccionales esperados basados en las características socioeconómicas de la dirección registrada del titular de la cuenta. El INEGI proporciona datos socioeconómicos granulares a nivel de área geoestadística básica (AGEB) y localidad, incluyendo índices de marginación, ingreso promedio por hogar, actividad económica y métricas de inclusión financiera (INEGI, 2020).

Para cada unidad geográfica g, se construye un perfil transaccional esperado P_g a partir del comportamiento observado de cuentas existentes dentro de la misma unidad:

$$P_g = \{\mu^g_{vol}, \sigma^g_{vol}, \mu^g_{amt}, \sigma^g_{amt}, \mu^g_{freq}, \sigma^g_{freq}\} \tag{5}$$

donde vol, amt y $freq$ representan volumen mensual de transacciones, monto promedio de transacción y frecuencia de transacciones, respectivamente. Una cuenta a con dirección registrada en la unidad geográfica g es marcada cuando su comportamiento transaccional observado se desvía significativamente de P_g:

$$z_a = \max\left(\frac{vol_a - \mu^g_{vol}}{\sigma^g_{vol}}, \frac{amt_a - \mu^g_{amt}}{\sigma^g_{amt}}\right) > z_{thresh} \tag{6}$$

Una persona registrada en una dirección de una localidad clasificada como de alta marginación por CONEVAL no debería presentar patrones transaccionales consistentes con actividad comercial de alto ingreso. Cuando una cuenta marcada mediante detección de clusters de originación también exhibe anomalías transaccionales relativas a su línea base socioeconómica, la evidencia combinada proporciona una base sólida para el reporte de actividades sospechosas.

6. DISCUSIÓN

6.1. Consideraciones prácticas de implementación

El marco propuesto requiere acceso a elementos de datos que típicamente están disponibles dentro de las instituciones financieras mexicanas: identificadores geográficos derivados del INE (recolectados durante el KYC), marcas temporales de originación de cuentas, geolocalización del dispositivo para apertura remota (requerida por BANXICO), y registros transaccionales. Los requisitos computacionales para la detección de clusters son modestos—DBSCAN opera en tiempo $O(n \log n)$ con indexación espacial—y pueden integrarse en infraestructura de monitoreo existente como procesos por lotes o analítica de streaming en tiempo casi real.

Las instituciones con bases de clientes suficientes en regiones geográficas específicas pueden construir perfiles transaccionales robustos a partir de sus propios datos. Las instituciones más pequeñas pueden necesitar depender de indicadores socioeconómicos externos del INEGI y CONEVAL como aproximaciones hasta que se acumulen datos internos suficientes.

6.2. Alineación regulatoria

El marco se alinea con las regulaciones mexicanas de PLD existentes. La LFPIORPI (Gobierno de México, 2012) establece que las instituciones financieras deben implementar enfoques basados en riesgo para la debida diligencia del cliente, y la identificación de patrones anómalos de originación constituye una medida legítima de debida diligencia reforzada. El requisito de BANXICO para la recolección de geolocalización durante la apertura remota (Banco de México, 2021) proporciona la base de datos para los Escenarios 2 y 3.

Es importante señalar que el marco no propone el cierre automático de cuentas basado en la pertenencia a un cluster. Más bien, genera señales de riesgo que activan monitoreo mejorado y, cuando son corroboradas por anomalías transaccionales, respaldan la presentación de reportes de operaciones inusuales (ROI) ante la Unidad de Inteligencia Financiera (UIF).

6.3. Consideraciones éticas

El marco inherentemente involucra perfilamiento basado en origen geográfico e indicadores socioeconómicos, lo cual plantea preocupaciones legítimas sobre discriminación contra individuos de zonas empobrecidas. Enfatizamos que las señales de detección de clusters por sí solas son insuficientes para acciones adversas contra cualquier cuenta individual. El marco está diseñado para producir marcas de monitoreo ponderadas por riesgo, no determinaciones. Los individuos de regiones empobrecidas que legítimamente abren cuentas y exhiben comportamiento transaccional esperado no serán afectados adversamente.

Además, el objetivo último del marco es proteger a las mismas poblaciones que están siendo explotadas. Los individuos cuyas identidades son usadas para cuentas mula enfrentan responsabilidad penal potencial por lavado de dinero, a pesar de ser víctimas de explotación económica. La detección temprana de operaciones de reclutamiento masivo puede habilitar intervenciones que protejan a comunidades vulnerables de ser participantes involuntarios en crímenes financieros.

6.4. Limitaciones

Se deben señalar varias limitaciones. Primero, el marco no ha sido validado en datos de producción a gran escala con etiquetas confirmadas de verdad fundamental para cuentas mula originadas mediante reclutamiento masivo. Segundo, la efectividad de la detección basada en geolocalización depende de la integridad de los datos GPS, que pueden estar comprometidos en escenarios más allá del spoofing simple (p. ej., desplazamiento de ubicación basado en VPN, emulación de dispositivo). Tercero, el enfoque de perfilamiento socioeconómico puede producir falsos positivos para individuos que han migrado legítimamente de áreas empobrecidas o cuyas circunstancias económicas han cambiado. Cuarto, los umbrales específicos propuestos ($k = 3$, $\epsilon_{geo} = 500$ m, $\delta = 50$ km) están basados en razonamiento operacional y requieren calibración empírica con datos institucionales.

7. CONCLUSIÓN

Mientras la industria de prevención de crímenes financieros se enfoca en la detección de deep fakes y fraude de identidad sintética, la explotación masiva de identidades reales de poblaciones económicamente vulnerables representa una amenaza significativa y subatendida a la integridad del KYC en economías en desarrollo. Este artículo ha caracterizado la mecánica operacional de la explotación masiva de identidades para originación de cuentas mula en México, identificado las propiedades invariantes que crean patrones detectables a pesar del uso de identidades genuinas, y propuesto un marco de detección basado en clusters que abarca tres escenarios de originación.

El insight fundamental es que el cambio del análisis individual al colectivo revela patrones que la verificación a nivel individual no puede detectar. Ninguna cuenta individual en una operación de reclutamiento masivo parece anómala. La anomalía emerge solo a nivel agregado: demasiadas personas, del mismo lugar, abriendo cuentas al mismo tiempo, en la misma ubicación, con comportamiento transaccional posterior inconsistente con su contexto socioeconómico.

El trabajo futuro debería enfocarse en (a) validación empírica con datos de producción con etiquetas confirmadas de cuentas mula, (b) desarrollo de métodos adaptativos de calibración de umbrales, (c) integración con análisis de redes basado en grafos para rastrear el flujo descendente de fondos a través de clusters de mulas identificados, y (d) mecanismos de intercambio de información inter-institucional que permitan la detección de campañas de originación distribuidas entre múltiples instituciones financieras.

REFERENCIAS

Banco de México (2021). Disposiciones de carácter general aplicables a las instituciones de crédito. Technical report, Banco de México.

BioCatch (2024). Nearly 2 million money laundering accounts reported in 2024. *BioCatch Annual Report*.

Brodsky, I. (2019). H3: Uber's hexagonal hierarchical spatial index. In *Uber Engineering Blog*.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58.

CONEVAL (2022). Medición de la pobreza en méxico 2022. *Consejo Nacional de Evaluación de la Política de Desarrollo Social*.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 226–231.

Flores, R. and Santos, M. (2024). Systemic money mule recruitment in developing economies: Evidence from the Philippines. *Journal of Financial Crime*, 31(5):1102–1118.

Gao, Y., Liu, Y., and Zhang, W. (2024). Deepfake detection in identity verification: A systematic review. *Computers & Security*, 138:103672.

Gobierno de México (2012). Ley federal para la prevención e identificación de operaciones con recursos de procedencia ilícita. Technical report, Diario Oficial de la Federación.

HM Government (2024). Money mule and financial exploitation action plan 2024. Technical report, UK Government.

INEGI (2020). Censo de población y vivienda 2020. *Instituto Nacional de Estadística y Geografía*.

Pourhabibi, T., Ong, K.-L., Kam, B. H., and Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133:113303.

Schuetz, P. (2021). The role of money mules in money laundering schemes. *Journal of Financial Crime*, 28(3):823–839.

U.S. Department of the Treasury (2024). 2024 national money laundering risk assessment. Technical report, U.S. Department of the Treasury.

Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., and Leiserson, C. E. (2019). Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics. *KDD Workshop on Anomaly Detection in Finance*.