



Detección de Fraude Interno en Colocaciones Financieras mediante la Ley de Benford y Análisis de Redes Transaccionales

Víctor Armando Echavarría Hernández

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor

Carlos Andrés Mera Banguero, Ph.D.

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia
2024

Cita	(Echavarría-Hernández, 2024)
Referencia	Echavarría-Hernández V. A. (2024). <i>Detección de Fraude Interno en Colocaciones Financieras mediante la Ley de Benford y Análisis de Redes Transaccionales</i> . Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Especialización en Analítica y Ciencia de Datos, Cohorte IV.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Danny Alejandro Munera Ramírez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**MONOGRAFIA DE GRADO
ESPECIALIZACIÓN EN ANALÍTICA Y CIENCIA DE DATOS**

Dedicatoria

A mis padres y mi hermana por ser mi soporte incondicional y la razón que me impulsa a ser mejor cada día.

A Karime, Andrés C, Ximena, JuanMa, Lina, Anita, Alessa, a mi ADC y amigos, por su conocimiento experto, su acompañamiento constante y por ayudarme a crecer tanto personal como profesionalmente. Gracias por ser un ejemplo de dedicación y motivación diaria.

Agradecimientos

A la Universidad de Antioquia, por brindarme las herramientas académicas y el entorno propicio para mi formación, contribuyendo significativamente a mi crecimiento profesional y personal.

A mi asesor, por sus valiosos aportes, orientación y compromiso en el desarrollo de esta propuesta metodológica. Su guía fue esencial para la concreción de este trabajo.

A Luce y Chucho, por enseñarme a ver el mundo desde una perspectiva diferente, inspirándome a crecer y a valorar cada aprendizaje.

A mis compañeros, quienes me acompañaron y apoyaron a lo largo de mis estudios. Su amistad y colaboración fueron fundamentales, convirtiendo este proceso en una experiencia enriquecedora y memorable.

A todos aquellos que, de una u otra manera, contribuyeron a la realización de este proyecto, mi más sincero agradecimiento.

Tabla de contenido

Resumen	9
Abstract	10
1.	Descripción del problema 11
2.	Objetivos 12
2.1.	Objetivo general 12
2.2.	Objetivos específicos 12
3.	Metodología 13
3.1.	Fuentes Crudas 15
3.2.	Preprocesamiento de los Datos 16
3.3.	Aplicación de la ley de Benford 19
3.3.	Análisis de Grafos 20
4.	Resultados y discusión 21
7.	Conclusiones 27
Referencias	28
Anexos	29

Lista de tablas

Tabla 1 Análisis Estadístico Monto de los Créditos	32
Tabla 2 Análisis Estadístico Días de Mora	32
Tabla 3 Resultados del análisis de la Ley de Benford aplicados a los montos de colocaciones financieras	34
Tabla 4 Métricas clave del grafo transaccional Algoritmo PageRank	35

Lista de figuras

Figura 1	Diagrama de flujo de metodología propuesta	14
Figura 2	Forma Teórica de la ley de Benford	20
Figura 3	Distribución Primer Dígito Monto Crédito Ley de Benford	22
Figura 4	Distribución Riesgo Total Cuantitativo Jurisdicción	23
Figura 5	Subgrafo Nodo Con Mayor Score PageRank	24
Figura 6	Distribución Riesgo Total Cuantitativo Jurisdicción Por Cada Grado de Transacción	25
Figura 7	Variación monto transaccional hacia el nodo concentrador	26
Figura 8	Salidas Nodo Concentrador & Entradas Nodos De Interés Interno	26
Figura 9	Cantidad de tipos de crédito asignados	30
Figura 10	Estado de los créditos asignados	31
Figura 11	Tipo de plan crédito vs Monto Contratado	33
Figura 12	Leyenda de los grafos	37

Siglas, acrónimos y abreviaturas

LAFT	Lavado de Activos, Financiación del Terrorismo y Corrupción
ETL	Extracción, Transformación y Cargue
HDFS	Sistema Distribuido de Ficheros de Hadoop en español
UdeA	Universidad de Antioquia

Resumen

El presente proyecto tiene como objetivo identificar patrones y relaciones que podrían estar asociados a fraudes internos en colocaciones financieras, utilizando herramientas como la Ley de Benford, análisis de grafos y el algoritmo PageRank. Los datos analizados incluyen información sobre colocaciones de crédito, relaciones familiares de empleados y transacciones financieras de primer y segundo grado, obtenidos de bases históricas de una entidad financiera.

La metodología inició con la aplicación de la Ley de Benford, que permitió identificar anomalías significativas en los montos de crédito, destacando subconjuntos de datos para análisis posterior. Este enfoque reveló que ciertos montos se desviaban notablemente de la distribución esperada, vinculándose principalmente a sucursales con más del 40% de clientes en categorías de alto riesgo. A partir de estas anomalías, se construyeron grafos transaccionales para modelar las relaciones entre empleados, clientes y relacionados, aplicando el algoritmo PageRank para identificar nodos con alta centralidad.

Los resultados mostraron que los nodos con mayor importancia en la red correspondían a entidades con alta concentración de transacciones entrantes, el 79% de las cuales provenían de jurisdicciones de alto riesgo. Además, se detectaron patrones de flujo irregular, incluyendo triangulación y picos transaccionales significativos, lo que refuerza la sospecha de actividades anómalas. Como producto de esta metodología, se generaron alertas con una cobertura del 70% sobre la red total, las cuales fueron enviadas al equipo de Investigaciones de Fraude y Lavado de Activos para revisión y validación. Estos hallazgos demuestran la eficacia de combinar herramientas computacionales y enfoques analíticos para la detección de riesgos financieros.

Palabras clave: Ley de Benford, análisis de grafos, PageRank, Fraude Interno, Lavado de Activos, Redes Transaccionales.

Abstract

The present project aims to identify patterns and relationships potentially associated with internal fraud in financial placements, using tools such as Benford's Law, graph analysis, and the PageRank algorithm. The analyzed data includes information on credit placements, employees' family relationships, and financial transactions of the first and second degree, obtained from the historical databases of a financial entity.

The methodology began with the application of Benford's Law, which identified significant anomalies in credit amounts, highlighting subsets of data for further analysis. This approach revealed that certain amounts deviated significantly from the expected distribution, primarily linked to branches where over 40% of customers were in high-risk categories. Based on these anomalies, transactional graphs were constructed to model the relationships between employees, customers, and their connections, applying the PageRank algorithm to identify nodes with high centrality.

The results showed that the most important nodes in the network corresponded to entities with a high concentration of incoming transactions, 79% of which originated from high-risk jurisdictions. Additionally, irregular flow patterns were detected, including triangulation and significant transactional peaks, reinforcing suspicions of anomalous activities. As an outcome of this methodology, alerts covering 70% of the total network were generated and forwarded to the Fraud and Money Laundering Investigation Team for review and validation. These findings demonstrate the effectiveness of combining computational tools and analytical approaches for detecting financial risks.

Keywords: Benford's Law, graph analysis, PageRank, Internal Fraud, Money Laundering, Transactional Networks

1. Descripción del problema

En general, las entidades financieras enfrentan el constante desafío de identificar a tiempo empleados que puedan estar incurriendo en prácticas fraudulentas, especialmente en la asignación de préstamos crediticios. Este tipo de fraude interno puede incluir acciones como la manipulación de datos, la aprobación de créditos con fines no legítimos, o el uso indebido de información transaccional para beneficio personal.

El impacto de estas prácticas fraudulentas se ha evidenciado en casos como el de Wells Fargo en 2016, donde empleados crearon cuentas falsas para cumplir con objetivos comerciales (Corkery, 2016). Este escándalo no solo generó pérdidas económicas significativas, sino que también afectó gravemente la reputación de la entidad, disminuyendo la confianza de sus clientes y generando repercusiones legales y regulatorias. Estas experiencias resaltan la importancia de identificar comportamientos anómalos a tiempo, ya que tales situaciones comprometen la estabilidad financiera, la confianza de los clientes y la integridad institucional.

Debido a la naturaleza confidencial de la información manejada por las entidades financieras y a las estrictas normativas de reserva bancaria, no se conocen trabajos previos publicados que aborden específicamente la detección de fraude interno en el sistema financiero. Esta falta de documentación limita el acceso a referencias o investigaciones comparables, lo que resalta la novedad y relevancia de este trabajo en el contexto de análisis de datos para la mitigación de riesgos internos.

Por lo tanto, el problema que se busca abordar en este trabajo es la dificultad para identificar, de manera efectiva y escalable, patrones de fraude interno en colocaciones financieras. Este problema radica en la complejidad de los datos, la falta de herramientas específicas y la necesidad de mecanismos analíticos que permitan emitir alertas tempranas basadas en anomalías detectadas en redes transaccionales y distribuciones numéricas.

2. Objetivos

2.1 Objetivo general

Detectar patrones de fraude interno en colocaciones financieras utilizando ciencia de datos mediante análisis de redes transaccionales y distribuciones numéricas. Este estudio busca aplicar la Ley de Benford para identificar anomalías en los montos de crédito y emplear métricas de centralidad en grafos transaccionales para reconocer conexiones sospechosas y tramitadores de alto riesgo, con el fin de establecer alertas de prevención para la institución financiera.

2.2 Objetivos específicos

- Comprender el contexto del negocio y la dinámica de la asignación de préstamos crediticios en la organización.
- Entender el conjunto de datos a partir de técnicas estadísticas descriptiva y visualización de datos, implementando técnicas de limpieza e imputación que garanticen la calidad y consistencia de la información
- Detectar posibles anomalías a partir de la Ley de Benford en las operaciones de colocaciones financieras, identificando desviaciones significativas en los montos de crédito.
- Evaluar la efectividad de la ley de Benford en la identificación de patrones inusuales o inconsistentes en los datos.
- Identificar concentradores transaccionales, utilizando métricas de centralidad para reconocer nodos clave que puedan indicar conexiones de alto riesgo dentro de la red de relaciones financieras.

3. Metodología

En todas las entidades financieras parte importante del riesgo de fraude se encuentra asociado al originado por el sesgo de conocimiento de los diferentes procesos por parte de sus colaboradores, ya que estos por acción u omisión pueden incidir en alguna práctica que vaya en detrimento del propósito de dichas organizaciones.

La materialización de fraude al interior de las organizaciones se ha convertido en un elemento fundamental que permite mitigar el riesgo residual, su detección es por tanto un reto a fin de prevenir eventos masivos o réplicas de estos. Esta propuesta para la detección de posibles patrones de riesgo interno se aborda desde una perspectiva fundamentada en analítica de datos, considerando desde el análisis exploratorio hasta el uso de herramientas avanzadas para el análisis de datos que permitan una óptima identificación efectiva de alertas tempranas. Entre las técnicas aplicadas, se destaca la visualización de datos, considerada una herramienta fundamental dentro del ecosistema de Big Data, que permite identificar patrones, tendencias y anomalías de manera intuitiva en grandes volúmenes de información y se usa como un elemento crucial para comprender las conexiones y relaciones transaccionales que serían difíciles de detectar en datos tabulares.

En este mismo sentido, se integra el análisis de grafos como una técnica avanzada de ciencia de datos, ampliamente utilizada para modelar relaciones y descubrir patrones que no son evidentes en datos tradicionales. Según Neo4j (n.d.), esta técnica permite a los científicos de datos analizar y procesar relaciones como estructuras relacionales complejas, lo que facilita la detección de conexiones significativas y la priorización de entidades relevantes dentro de las mismas. En este trabajo, los nodos representan empleados, clientes y entidades relacionadas, mientras que las aristas capturan las interacciones transaccionales entre ellos. Usando el algoritmo PageRank, se priorizan las conexiones más relevantes, identificando individuos o relaciones que puedan representar un mayor riesgo. Este enfoque es eficiente en el manejo de datos relacionales y amplifica las capacidades de las herramientas de visualización y análisis exploratorio.

Una característica clave de este problema es la cantidad de datos disponibles para el análisis, el peso de los archivos y la complejidad en su manejo, por lo cual el procesamiento de la información

requiere infraestructuras de Big Data, el uso de formatos optimizados para el almacenamiento y análisis masivo de datos, tales como *Parquet* y sistemas distribuidos como *HDFS*.

La Figura 1 presenta de manera sucinta la metodología usada para el procesamiento y solución basado en algoritmos analíticos. Esta metodología inicia con un proceso de extracción de datos en el que fue necesario recolectar la información de los préstamos, información sociodemográfica de los actores, relacionados directos de los colaboradores y los movimientos transaccionales en una entidad bancaria del País. La segunda etapa corresponde al preprocesamiento de los datos. Una vez los datos tienen la calidad deseada se realiza la aplicación de la Ley de Benford para la selección de los créditos cuyo monto resulta importante para el alertamiento, estos se convierten en la nueva unidad de información sobre la cual vamos a centrar el análisis. Una vez se identifican los créditos riesgosos, se realiza la extracción de la información transaccional y se validan las relaciones complejas entre las partes usando un grafo dirigido para identificar los concentradores (colaboradores) que pueden estar vinculados en la materialización de fraude. Esta selección se itera de manera secuencial a través de un algoritmo PageRank (Page, 1999) que permite identificar aquellos nodos con una mayor relevancia a nivel de conexiones.

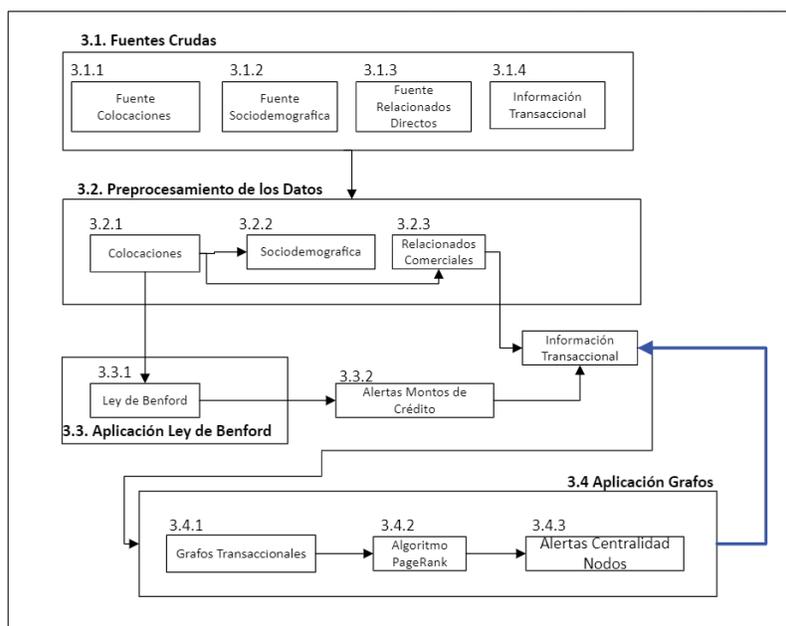


Figura 1. Diagrama de flujo de metodología propuesta

Fuente: Elaboración propia

3.1 Fuentes Crudas

La información se origina en los sistemas bancarios de la organización, se almacena en el lago de datos en formato *parquet* (Cloudera). Estos datos son gestionados y consultados mediante *Hadoop Impala* como herramienta de procesamiento de datos. A través de una serie de transformaciones y con el aporte del conocimiento del negocio en áreas clave como Cartera, Conocimiento de Clientes y Canales, los datos son estructurados y enriquecidos, lo que permite que estén disponibles para el análisis de manera óptima y alineada a las necesidades del proyecto.

La información utilizada en este proyecto se compone de cuatro conjuntos de datos principales, cada uno con características específicas que permiten una visión integral para el análisis:

Colocaciones financieras: este conjunto de datos en general abarca todas las colocaciones financieras realizadas desde 1990 hasta 2024, para el objeto del presente estudio por vigencia de dicha información se considera solo las colocaciones realizadas de 2018 al 2024. Contiene un total aproximado de 4 millones de registros y ocupa aproximadamente 1GB. Este recurso representa la información histórica de préstamos y asignaciones financieras, siendo la fuente primaria del análisis. La fuente cuenta con los campos necesarios definidos en el protocolo para la extracción de la información.

Información demográfica: este conjunto de datos incluye información sociodemográfica de clientes y empleados de la organización, recopilados históricamente desde 2018 hasta 2024. Tiene un aproximado de 293 millones de registros y 303 atributos, ocupando un total de 8TB. Este insumo aporta un contexto detallado sobre los perfiles de los individuos analizados. La fuente cuenta con los campos necesarios definidos en el protocolo para la extracción de la información.

Relaciones directas: este conjunto de datos abarca la información sobre las personas (naturales y jurídicas) que son relacionados directos de los empleados y los clientes. Esa población incluye familiares y empresas donde sean representantes legales o accionistas. Estos datos cubren el periodo de 2018 a 2024. Con un aproximado de 278 millones de registros con 16 atributos y un tamaño de almacenamiento en disco de 13GB, esta fuente permite entender las conexiones de

primer grado de los involucrados. La fuente cuenta con los campos necesarios definidos en el protocolo para la extracción de la información.

Transacciones consolidadas: es un conjunto de datos sobre transacciones consolidadas de clientes y empleados desde 2018 hasta 2024. Este conjunto contiene aproximadamente mil millones de registros con 14 atributos y un tamaño de almacenamiento en disco de 91GB. Este recurso ofrece una visión completa de las relaciones transaccionales y permite analizar patrones de comportamiento financiero. La fuente cuenta con los campos necesarios definidos en el protocolo para la extracción de la información.

Dado que los datos tienen información sensible, estos se anonimizan ocultando campos que identifican individual o colectivamente a las personas, además se eliminan ciertas variables que son confidenciales, atendiendo a la política de *habeas data* y la reserva bancaria de la organización financiera.

3.2 Preprocesamiento de los Datos

Primero, se realiza una selección inicial de los atributos relevantes con ayuda de expertos del negocio, reduciendo el conjunto de 250 a 29 atributos clave. Los datos tienen como insumo principal la base de colocaciones financieras, que contiene los detalles de asignación de créditos a clientes. Algunos de los atributos que se mantuvieron en el conjunto de datos son: la ciudad del desembolso, tipo de crédito, monto, fecha del desembolso, total de pagos realizados y días de mora, entre otras. También, sólo se consideraron aquellos créditos cuyo interés por naturaleza hace parte de los objetivos propios del negocio. Esta exclusión optimiza el uso de recursos y asegura que solo se analicen colocaciones relevantes para el estudio de riesgos.

La información sobre la sucursal de donde se originó el crédito se complementa con datos sobre la región, la ciudad y el nivel de riesgo asociado a la ubicación, montos de dinero en efectivo que están manejando y sus tipos de clientes. Esta información ayuda a generar un perfil de riesgo macro que contribuye al análisis general de riesgos de las colocaciones financieras. Además, se incorpora información adicional sobre los actores relacionados en las colocaciones, tales como clientes y empleados, para mejorar el perfil de riesgo de cada caso.

En cuanto a la ventana de observación de los datos, con relación a la información sociodemográfica utilizada, esta se limita al último registro obtenido para cada titular del crédito aún vigente en la organización. En este caso, se considera el estado del cliente al momento de recibir la colocación financiera, es decir, la información del cliente en el momento específico de la asignación del crédito. Estos datos se integran a la base de colocaciones financieras y se añaden como insumo adicional para análisis posteriores.

Respecto a los relacionados directos, estos se consideran como aquellos con vínculos cercanos con los empleados comerciales de la entidad financiera, e incluyen familiares, y empresas de las cuales son accionistas o representantes legales. Esta información se almacena en una base unificada para facilitar su análisis.

Por último, se incluye la información de relacionados transaccionales, que abarca los movimientos de dinero asociados a los actores de interés. Se definen las transacciones de primer grado, que involucran recursos que salen de las cuentas de los actores clave (comerciales, relacionados y clientes). A partir de estos datos, se genera una segunda base de transacciones de segundo grado, que representa a las personas o entidades destino de las transacciones de primer grado. Esta información se complementa con variables de riesgo internas de la entidad y con la marcación de otros actores relevantes como ex empleados, proveedores, ex proveedores, y si se encuentra localizado en una jurisdicción de riesgo.

Con estos pasos, se generan tres conjuntos de datos estructurados: Colocaciones, Relacionados Comerciales y Relacionados Transaccionales. Cada conjunto de datos está preprocesado y organizado para facilitar el análisis y optimizar el uso de recursos en el clúster de *Hadoop/Impala*. La exclusión de información irrelevante, como tipos de créditos de bajo riesgo, evita el procesamiento de datos no necesarios, lo que previene un uso excesivo de *CPU* y memoria en los clústeres. Esta estructuración asegura una integración coherente en los próximos análisis y proporciona una base sólida para validar los resultados obtenidos con precisión y consistencia, manteniendo la eficiencia en el uso de recursos computacionales.

En la base de Colocaciones Financieras, que contiene un total de 115,473 registros sin valores nulos, se identificaron 19 variables categóricas y 17 variables numéricas tales como el estado del crédito, la identificación del empleado y cliente, estado del crédito. Relacionados directos 1 variable categórica y 2 numéricas, siendo la identificación e identificación del relacionado junto con el tipo de relación. Relación transaccional junto con 5 variables categóricas y 4 numéricas, siendo aquellos campos como el número de identificación de la transacción, monto transaccional en el mes, cantidad de transacciones. En el Anexo A se presenta un análisis descriptivo de la información que permite evidenciar la relevancia de las técnicas seleccionadas y el uso de la ley de Benford y al análisis de grafos.

Como parte de la metodología utilizada se considera un proceso iterativo entre las diferentes etapas, de manera que se hace una exploración, un refinamiento, un filtrado y el respectivo modelamiento hasta obtener un resultado que cumpla los parámetros requeridos para ser llevado a investigación. Este proceso en primer lugar considera una exploración inicial con Cytoscape para realizar un análisis gráfico de transacciones, colocaciones y relaciones familiares de empleados. Esta herramienta proporciona una representación visual y métricas propias, pero el proceso falló debido al gran volumen de datos (más de 5 millones de registros), lo que pone en evidencia la necesidad de optimizar la carga y selección de datos. Para superar esta limitación y como segunda etapa iterativa, se modificó el proceso de ETL (Extracción, Transformación y Carga), incorporando variables internas adicionales que enriquecieran en términos de riesgo y posibles casos de eventos de riesgo, tanto internos como externos. También se añadieron datos de listas internacionales, como la lista OFAC y la Lista Clinton, para incluir información relevante de sanciones y riesgos internacionales. En este proceso, se seleccionaron únicamente las variables consideradas relevantes para la entidad financiera y en las que se habían identificado riesgos previamente materializados. Sin embargo, al generar el grafo con estas variables, los datos no mostraron patrones anómalos específicos; más bien, reflejaban un comportamiento transaccional cotidiano en el cual, por ejemplo, un cliente A transfiere dinero a un cliente B por un bien o servicio (Financial Crime Academy, 2024). Sin embargo, se detectaron comportamientos de cuentas o entidades con una alta captación de dinero que, por su naturaleza, es esperable y normal debido a que están destinadas a recibir pagos de servicios públicos o impuestos. Estas transacciones regulares y legítimas, propias

de la función de estas entidades, llevaron a excluirlas del análisis para evitar la generación de falsos positivos que podrían desviar la atención de casos de riesgo real.

Como parte de este proceso iterativo en la tercera etapa se realiza el filtrado de patrones normales y depuración de nodos: En iteraciones posteriores, el análisis se centró en sucursales con niveles elevados de cartera castigada, es decir, créditos considerados incobrables o difíciles de recuperar (El Universo, 2023). Se ajustó el grafo eliminando nodos que representaban relaciones normales, como transacciones entre comerciales y empleados de la misma sucursal que, en algunos casos, actuaban como clientes dentro de la entidad. Este ajuste fue necesario para depurar los datos y centrar el análisis en relaciones que pudieran representar un riesgo real.

3.3 Aplicación de la ley de Benford

La Ley de Benford, ver **Figura 2**, describe la frecuencia esperada de los dígitos en un conjunto de datos numéricos, donde los dígitos más bajos, como el "1", son significativamente más comunes como primer dígito. Esta distribución, que sigue una fórmula logarítmica, se utiliza ampliamente en la detección de fraudes porque los datos generados de manera natural suelen ajustarse a esta ley, mientras que las manipulaciones o irregularidades presentan desviaciones significativas (M, 2023). Para esta propuesta el uso de la Ley de Benford, es fundamental para analizar los montos asociados a colocaciones financieras, permitiendo identificar patrones numéricos inusuales que podrían indicar comportamientos anómalos en el sistema financiero.

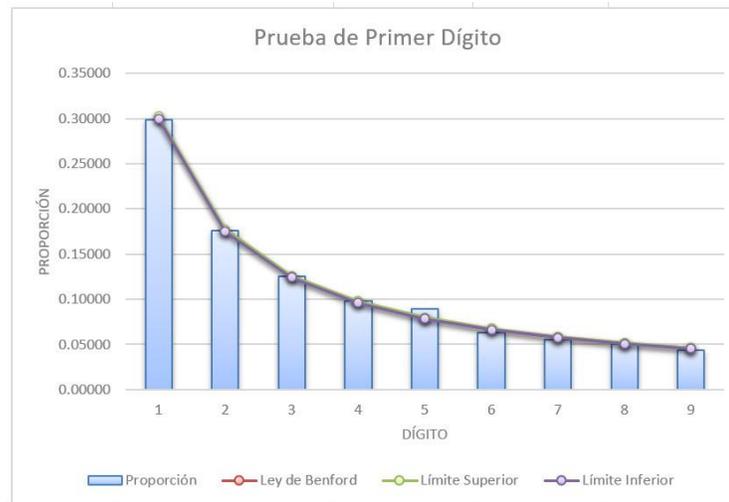


Figura 2. Forma Teórica de la ley de Benford

Fuente: CPA Informatics

La Ley de Benford, se aplica según esta propuesta metodológica para la selección de los montos de los créditos asignados en un periodo específico, extrayendo su primer dígito significativo y comparando su frecuencia relativa con la distribución esperada, logrando identificar desviaciones importantes en ciertos rangos de montos, donde el dígito "5" apareció más frecuentemente de lo esperado.

3.4 Análisis de Grafos

Los datos transaccionales y de colocaciones financieras pueden ser modelados como un grafo porque representan relaciones explícitas entre entidades, como empleados, clientes y transacciones. En este modelo, los nodos del grafo corresponden a las entidades involucradas (por ejemplo, comerciales, clientes y relacionados), mientras que las aristas representan las relaciones o interacciones entre ellos, como la asignación de un crédito o transferencias de dinero. Este enfoque permite estructurar las conexiones en forma de red, facilitando la identificación de patrones anómalos, concentradores transaccionales y posibles vínculos sospechosos en las relaciones financieras de las partes vinculadas en la red. Además, el tamaño de los nodos se configuró para reflejar el número de transacciones asociadas a cada entidad, destacando aquellos actores que interactúan con mayor frecuencia en el sistema financiero.

Una vez construido el grafo, se aplica el algoritmo PageRank, que calcula la importancia relativa de cada nodo en función de las conexiones que tiene y su relevancia en la red (Page, 1999). Este análisis permite priorizar nodos clave que actúan como concentradores, es decir, actores con alta centralidad que representan puntos críticos en la red y que resultan en alertas proactivas de fraude al interior de la organización. Por ejemplo, nodos con valores altos de PageRank incluyen empleados que gestionan múltiples créditos para clientes relacionados, así como clientes que transferían montos significativos a un número reducido de destinatarios, lo que generó alertas para investigación.

El modelamiento por grafos y la aplicación de PageRank aportan una perspectiva novedosa en la mitigación del riesgo de fraude interno a través de modelos analíticos, permitiendo identificar actores y relaciones de alto riesgo de manera eficiente. Los resultados ayudaron a enfocar las investigaciones en conexiones críticas dentro de la red, complementando otras herramientas como la Ley de Benford para emitir alertas más completas y precisas.

Para la implementación de algoritmos avanzados (camino cortos y PageRank) se utiliza la librería NetworkX en Python y se implementa el algoritmo de camino cortos para identificar rutas de menor costo en el grafo (Graph everywhere, n.d.). Sin embargo, este método resultó ineficaz debido a la estructura cíclica de la red, donde las relaciones bidireccionales entre relacionados comerciales generaban bucles sin fin, lo que impedía el análisis efectivo de rutas únicas. Finalmente, se utiliza el algoritmo PageRank (Page, 1999) para identificar nodos de alta centralidad en el grafo, lo que permite la detección de un nodo específico con un rol centralizador en la red, lo cual permite identificar una alerta significativa de riesgo interno. Este hallazgo resulta clave para el análisis, ya que ayuda a identificar un patrón de comportamiento potencialmente riesgoso dentro de la red de relaciones.

4 Resultados y discusión

Se aplica la Ley de Benford para analizar la frecuencia de los dígitos iniciales en los montos de los créditos otorgados. Los resultados revelaron que ciertos dígitos superaron significativamente las frecuencias esperadas según la distribución de Benford, la cual se presenta en la **Figura 3**.

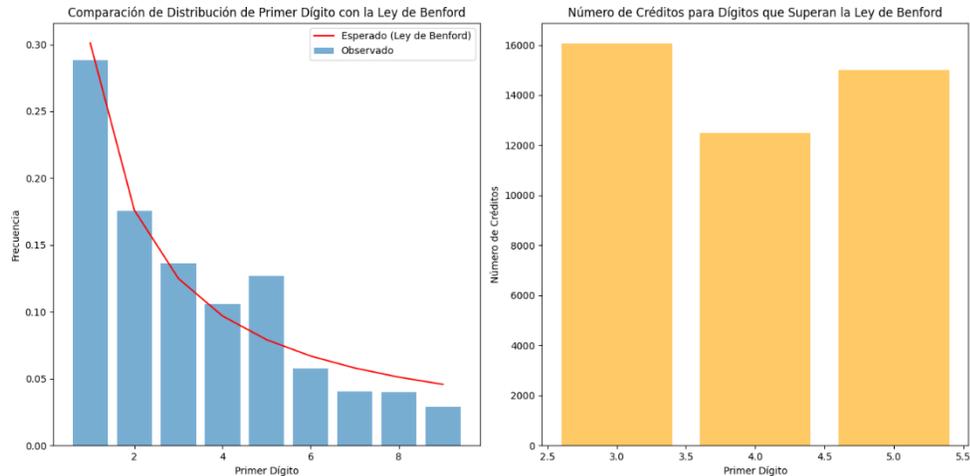


Figura 3. Distribución Primer Dígito Monto Crédito Ley de Benford

Fuente: Elaboración Propia

Adicionalmente, al analizar el riesgo de las sucursales junto con la información disponible al interior de la organización, frente a otros riesgos críticos que de manera conjunta pueden afectar el desempeño, ver **Figura 4**, se puede identificar que más del 40% de los clientes tiene un riesgo alto y solo el 27% es de bajo riesgo, lo cual refuerza la sospecha de montos de crédito anómalos en las sucursales.

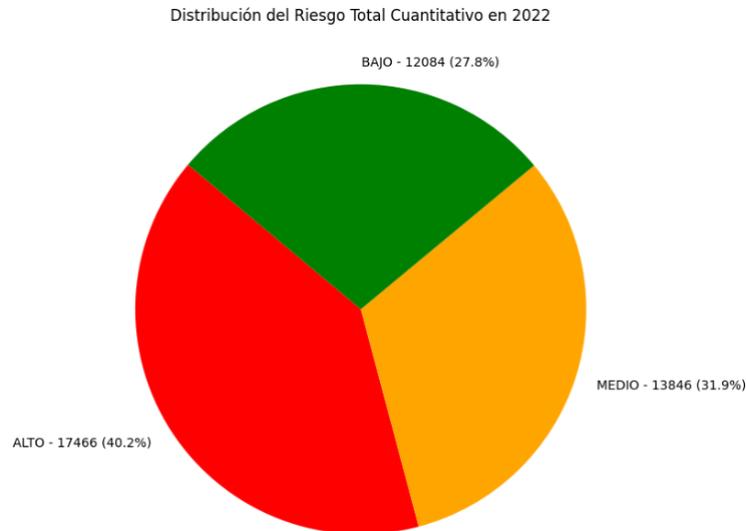


Figura 4. *Distribución Riesgo Total Cuantitativo Jurisdicción*

Fuente: Elaboración Propia

En la **Figura 3** como resultado del análisis se obtiene un subconjunto de créditos que son aquellos detectados por la Ley de Benford como anómalos, es decir aquellos dígitos que se desvían considerablemente de la distribución teórica asociada y a partir de este, se construyen los grafos transaccionales basados en las relaciones entre comerciales, clientes y relacionados de los comerciales.

La aplicación de la Ley de Benford resulta en una herramienta eficaz en el proceso de detección de fraudes al proporcionar una herramienta estadística para priorizar investigaciones en áreas de mayor riesgo. Estas alertas iniciales sirvieron como base para explorar relaciones transaccionales más complejas mediante el análisis de grafos, contribuyendo significativamente al diseño de un sistema reactivo para identificar posibles casos de fraude interno, el análisis detallado de los resultados de la ley de Benford aplicados sobre el conjunto de datos reales.

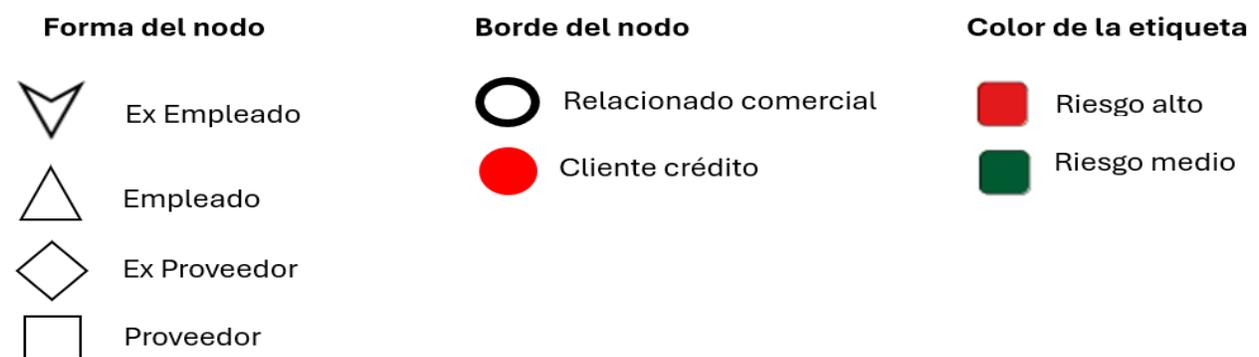
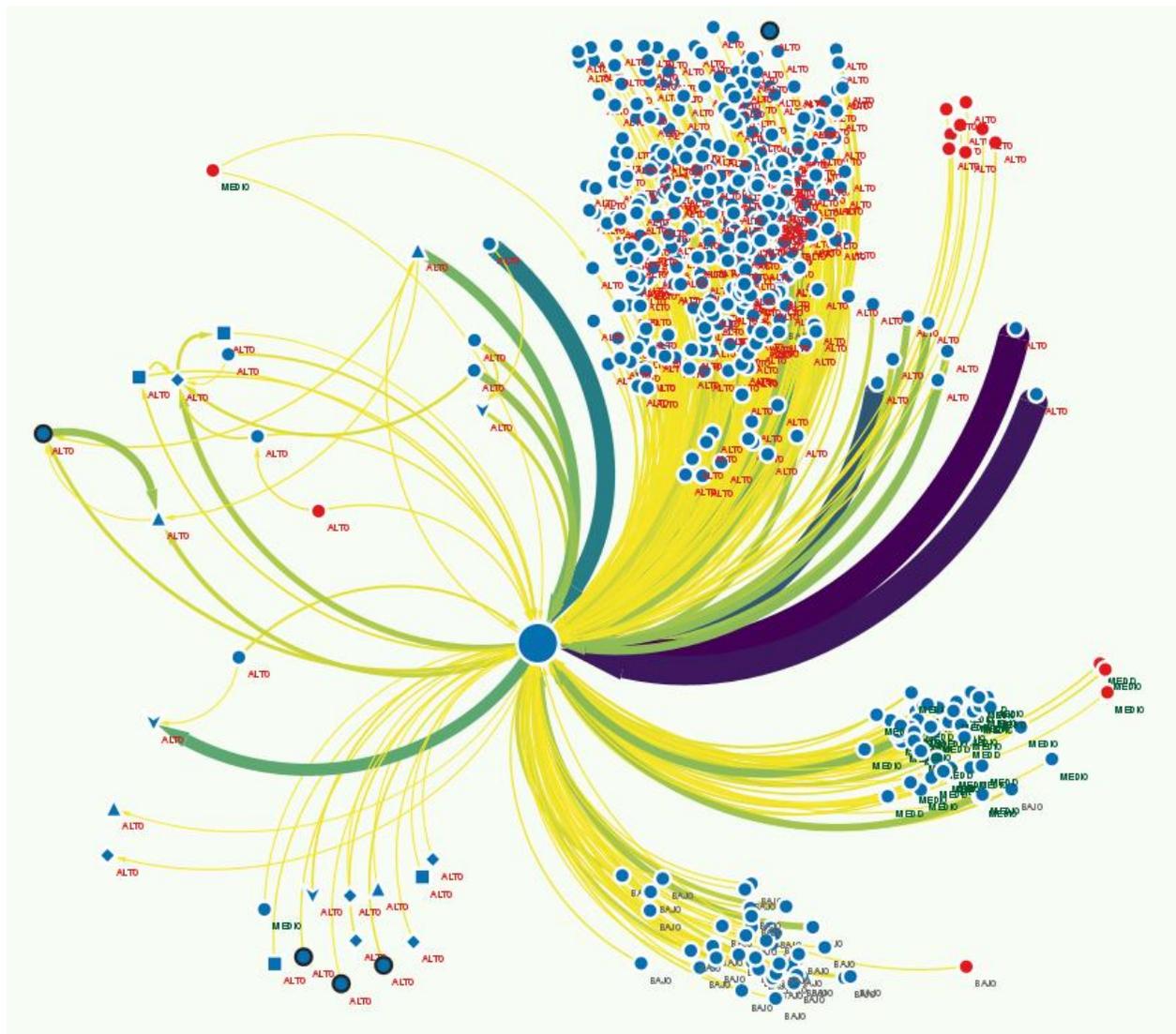


Figura 5. Subgrafo Nodo Con Mayor Score PageRank

Fuente: *Elaboración Propia*

La **Figura 5** muestra el nodo con la mayor centralidad de importancia, identificado mediante el algoritmo de PageRank aplicado sobre el grafo transaccional de las partes vinculadas para aquellos créditos seleccionados a través de la Ley de Benford. Las aristas con un mayor grosor y color más oscuro representan montos transaccionales altos donde el rango transaccional es de 500 mil pesos a 97 millones de pesos. El nodo con mayor tamaño representa el centralizador de las transacciones. En este subgrafo, se observa una alta concentración de conexiones entrantes, donde el 98% de las transacciones provienen de conexiones de segundo grado. Además, el 79% de estas conexiones se originan en jurisdicciones catalogadas como de alto riesgo, distribuidas en un total de 123 ubicaciones diferentes, ver **Figura 6**.

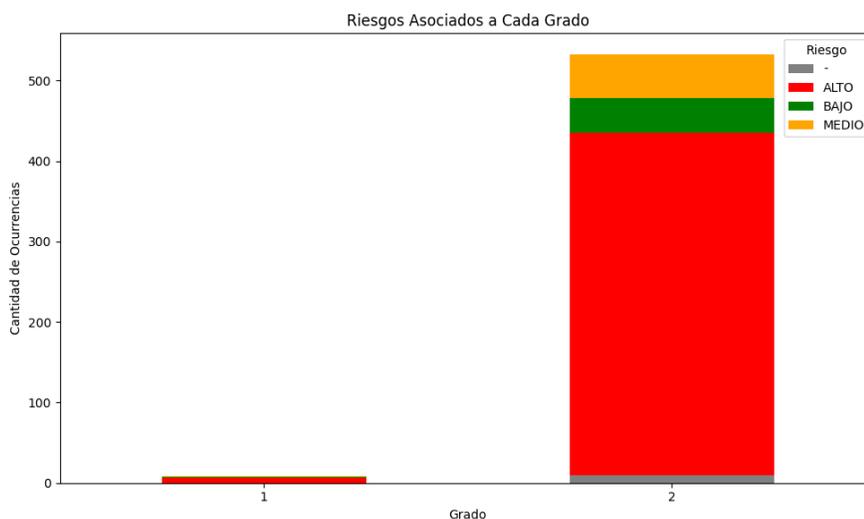


Figura 6. *Distribución Riesgo Total Cuantitativo Jurisdicción Por Cada Grado de Transacción.*

Fuente: *Elaboración Propia*

Con base en la selección realizada, de acuerdo con los resultados obtenidos para la **Figura 6**, se realiza un análisis de la variación de los montos de las transacciones entrantes hacia el nodo centralizador, encontrando que la mayoría de estas transacciones se concentran en un rango de montos bajos. Sin embargo, se observan varios picos significativos que destacan por su magnitud, alcanzando valores cercanos a los 100 millones, en algunos casos, como se observa en la **Figura 7**. Estas transacciones atípicas, que representan un total de 719 eventos individuales, reflejan montos considerablemente superiores al promedio y merecen una atención especial debido a su posible

implicación en actividades inusuales o de alto riesgo. Este comportamiento podría indicar una alerta por otros factores de riesgo de uso interno de la organización.

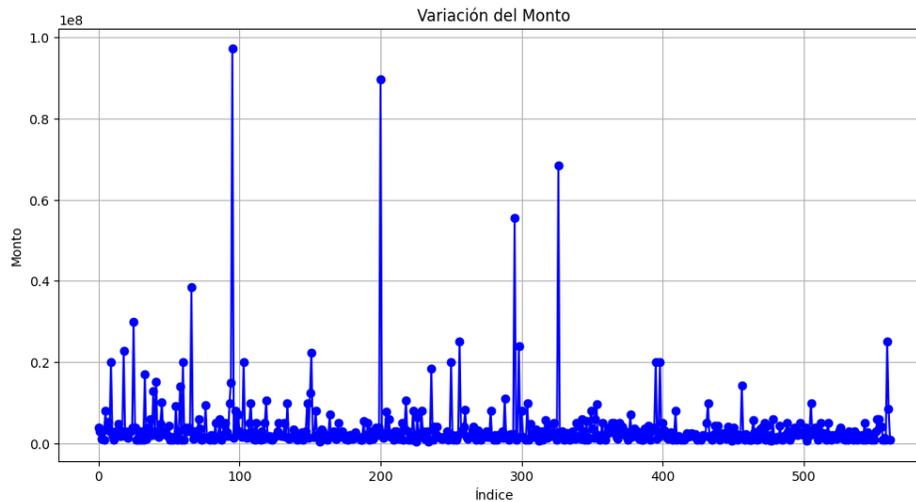


Figura 7. Variación monto transaccional hacia el nodo concentrador.

Fuente: *Elaboración Propia*

En el análisis de las salidas de los nodos que tienen una relación transaccional directa con el nodo concentrador, ver **Figura 8**, se observa que estas transacciones están dirigidas principalmente hacia personas con vínculos directos con la entidad financiera. Además, se identifica un patrón de triangulación dentro del mismo subgrafo, en el cual los fondos parecen circulares entre múltiples nodos interrelacionados. Este comportamiento sugiere la posibilidad de estructuras complejas de flujo de transacciones, que podrían indicar la presencia de actividades coordinadas o prácticas que requieren una investigación más detallada para descartar riesgos de fraude o conflicto de interés.

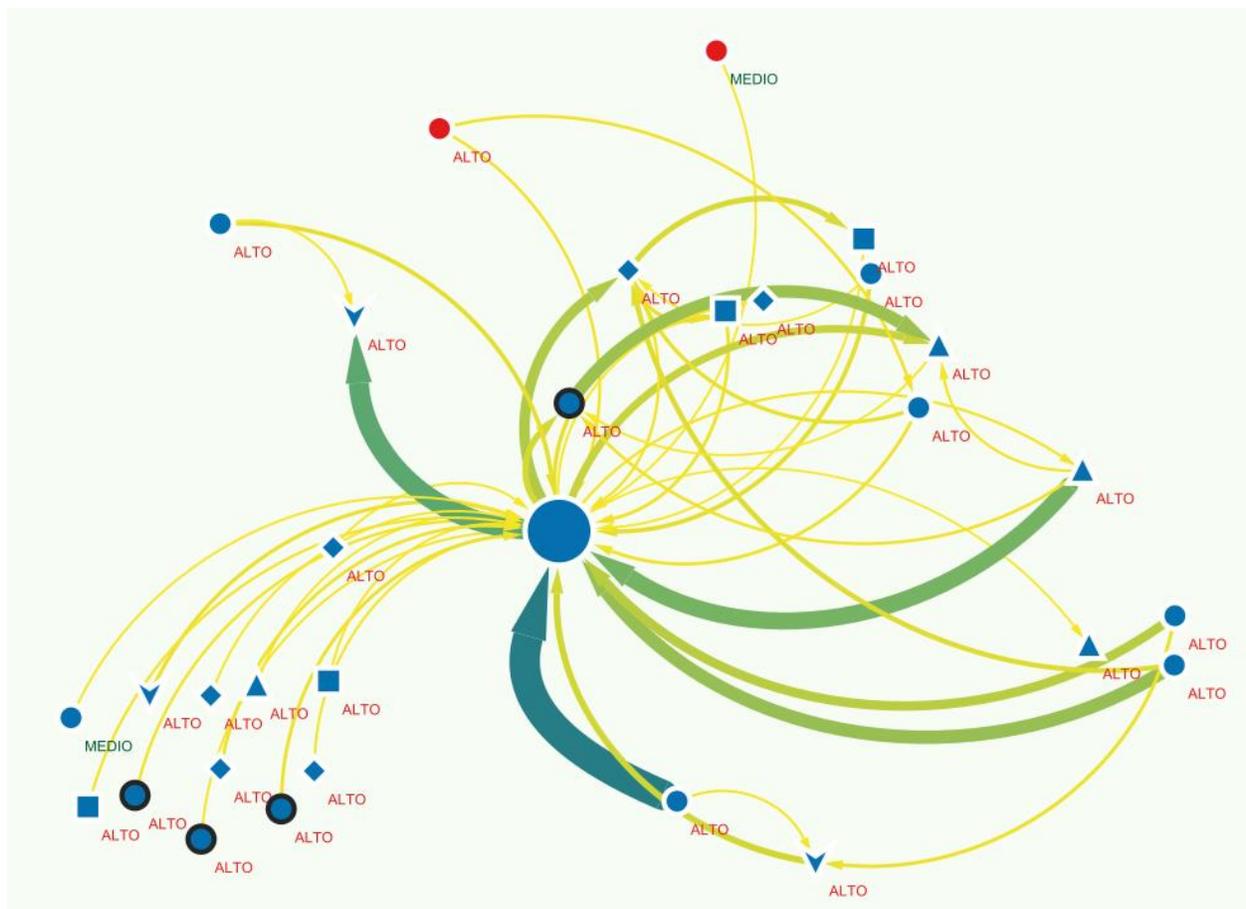


Figura 8 Salidas Nodo Concentrador & Entradas Nodos De Interés Interno.

Fuente: *Elaboración Propia*

Estos resultados sugieren que el uso de técnicas como la Ley de Benford y PageRank permite identificar patrones anómalos en las colocaciones financieras, destacando nodos con alta centralidad y transacciones atípicas. En particular, el análisis de PageRank identificó al Nodo N1 como el más relevante (PageRank: 0.05%), con un 98.93% de conexiones indirectas (grado 2), lo que lo posiciona como un concentrador clave en la red. Otros nodos, como N2 y N3, combinaron relaciones y transacciones, actuando como intermediarios mixtos.

Como resultado del uso de esta metodología, se generaron alertas con una cobertura del 70% sobre la red total, enviándose al equipo de Investigaciones de Fraude y Lavado de Activos para su revisión y validación. Un análisis más explícito de las métricas y patrones de la evolución del PageRank pueden ser consultadas en el **Anexo C**.

5 Conclusiones

- La metodología propuesta constituye un enfoque novedoso y eficaz para la detección temprana de alertas que permitan la mitigación de riesgo de fraude asociada al desembolso de créditos al interior de la organización, puesto que la Ley de Benford permite identificar irregularidades en los montos de crédito asignados, destacando patrones numéricos que se desvían de las expectativas estadísticas. Entre tanto que, el análisis de grafos a través de PageRank, facilita la detección de concentradores transaccionales y relaciones sospechosas.
- La potencialidad de esta metodología radica en su capacidad para estructurar grandes volúmenes de datos, modelar relaciones complejas y generar resultados accionables en entornos financieros. Su enfoque reactivo basado en datos históricos es adecuado para identificar patrones que pueden pasar desapercibidos en análisis tradicionales, fortaleciendo las estrategias de monitoreo al interior de las entidades financieras.
- Esta metodología depende de datos históricos que implican un alto volumen de procesamiento y capacidad disponible con la cual no se cuenta al interior de la organización, por tanto, una extensión en su aplicación para la detección en tiempo cercano a real no está disponible. Sin embargo, esta forma de aplicación de la metodología resulta consistente con la estructura del proceso de investigación de las alertas obtenidas a través de esta.
- La metodología propuesta ofrece una base para detectar patrones de posible fraude interno y generar alertas en entornos complejos. Su aplicación contribuye a mitigar riesgos operativos y reputacionales, su efectividad puede incrementarse mediante la integración de técnicas proactivas y validaciones adicionales en el contexto de la organización.

6 Referencias

- Association of Certified Fraud Examiners. (n.d.). *ACFE*. Retrieved from <https://acfe-spain.com/recursos-contrafraude/que-es-el-fraude/arbol-fraude>
- Association of Certified Fraud Examiners. (n.d.). *ACFE*. Retrieved from <https://acfe-spain.com/recursos-contrafraude/que-es-el-fraude/triangulo-del-fraude>
- BBVA. (n.d.). *BBVA*. Retrieved from <https://www.bbva.com.co/personas/productos/prestamos/consumo/libranza.html>
- Castañeda, R. (2021, mayo 28). *CPA Informatics*. Retrieved from <https://cpainformatics.com/2021/05/lbycontabilidad/>
- Cloudera. (n.d.). Retrieved from Cloudera: <https://docs.cloudera.com/cdp-private-cloud-base/7.1.8/impala-reference/topics/impala-parquet.html>
- criado, E. (2019, 07 05). *Navarro Abogados*. Retrieved from <https://navarroabogados.net/garantias-reales-garantias-personales-caso-especial-pignoracion-acciones/#:~:text=La%20principal%20diferencia%20entre%20las,sujeto%20que%20responde%20con%20su>
- El Universo*. (2023, agosto 07). Retrieved from <https://www.eluniverso.com/noticias/economia/que-significa-tener-una-cartera-castigada-nota/#:~:text=La%20cartera%20castigada%20se%20refiere,con%20sus%20obligaciones%20de%20pago>.
- Financial Crime Academy. (2024, septiembre 27). *Financial Crime Academy*. Retrieved from <https://financialcrimeacademy.org/es/que-es-una-transaccion/>
- Graph everywhere. (n.d.). *Grapheverywhere*. Retrieved from <https://www.grapheverywhere.com/algoritmos-deteccion-rutas-cortas/#:~:text=El%20algoritmo%20de%20rutas%20m%C3%A1s,v%C3%A9rtices%20que%20conforman%20un%20grafo>.
- LexisNexis® Risk Solution. (2024, 06 22). Retrieved from <https://risk.lexisnexis.com/global/es/about-us/press-room/press-release/20240620-true-cost-of-fraud-colombia#:~:text=Comunidades%20de%20Prensa-,Cada%20peso%20perdido%20por%20fraude%20en%20Colombia%20cuesta%20a%20las,del%20fraude%20en%20Am%C3%A9rica%20Latina>

M, A. C. (2023, febrero 03). *Auditool*. Retrieved from <https://www.auditool.org/blog/fraude/ley-de-benford>

Neo4j. (n.d.). Retrieved from <https://neo4j.com/es/producto/graph-data-science>

Page, L. B. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Stanford InfoLab.

Scotiabank Colpatría. (n.d.). *Scotiabank Colpatría*. Retrieved from <https://www.scotiabankcolpatria.com/educacion-financiera/productos-y-servicios/credito-rotativo#:~:text=Un%20cr%C3%A9dito%20Rotativo%20es%20un,del%20cupo%20parcial%20o%20totalmente>.

UNICEF. (n.d.). Retrieved from <https://www.sbcguidance.org/es/comprender/ciencias-del-comportamiento-aplicadas#:~:text=Las%20ciencias%20del%20comportamiento%20aplicadas%20combinan%20la%20investigaci%C3%B3n%20y%20evidencia,motivar%20o%20impedir%20comportamientos%20espec%C3%ADficos>.

Distribución del Estado de la Obligación

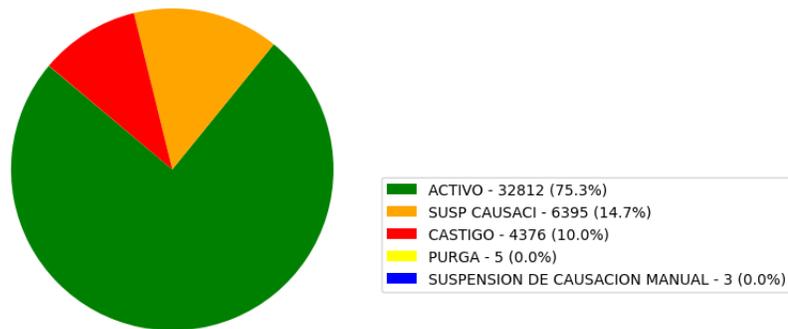


Figura 10. *Estado de los créditos asignados*

Fuente: *Elaboración propia*

En la **Figura 10**, se evidencia que la mayor parte de las obligaciones está en estado ACTIVO, lo cual es esperable en un portafolio saludable. Sin embargo, es notable que las categorías de Suspensión de Causación y Castigo, que representan créditos en mora significativa y aquellos formalmente reconocidos como incobrables respectivamente, tienen una cantidad significativa de casos. Esta cifra se encuentra "cerca" de los créditos activos según la proporción observada, lo que podría ser una señal inusual. Este comportamiento merece mayor investigación, ya que, en condiciones normales, se esperaría que las obligaciones en estos estados fueran considerablemente menores, reflejando un mejor manejo del riesgo crediticio.

mnt_tot_contratado	
count	115,473
mean	161,862,401
std	4,103,161,535
min	2,850
25%	8,800,000
50%	20,800,000
75%	50,000,000
max	487,629,071,404

Tabla 1. *Análisis Estadístico Monto de los Créditos*

Fuente: *Elaboración propia*

De acuerdo con la **Tabla 1** anteriores se puede evidenciar una alta concentración en montos de créditos relativamente bajos. El diagrama de cajas y bigotes de la derecha de la Figura 3, muestra una asimetría pronunciada hacia la derecha y algunos contratos de montos excepcionalmente altos que actúan como *outliers*, la media de los datos es mayor a la mediana y al percentil 75 corroborando lo anteriormente mencionado, la desviación estándar es muy alta más de 400 mil millones, lo que muestra una gran dispersión en los datos como se puede ver en el valor mínimo y en valor máximo del set de datos, este comportamiento en el monto de los créditos se puede considerar esperable ya que se tienen las colocaciones todos los clientes pertenecientes a los diversos segmentos comerciales la cual tienen características financieras diferentes.

cnt_dias_mora	
count	115,473
mean	47
std	100
min	0.0
25%	0.0
50%	0.0
75%	11
max	794

Tabla 2. *Análisis Estadístico Días de Mora*

Fuente: *Elaboración Propia*

La **Tabla 2** revela que la mayoría de los clientes tienen una mora baja, concentrándose principalmente en el rango de 0 a 50 días. Este comportamiento sugiere que una gran parte de la

cartera se encuentra sana, con clientes que cumplen con sus obligaciones de pago en tiempo o con atrasos leves. No obstante, se identifican algunos casos con moras extendidas, con días de atraso que superan los 100 días, e incluso algunos con más de 300 días. El 75% de los clientes presenta una mora menor a 11 días.

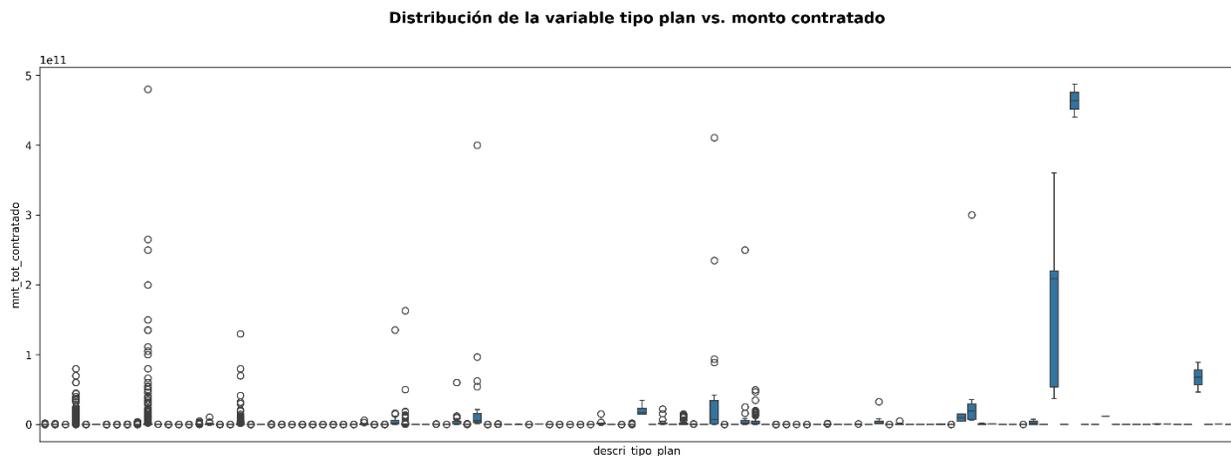


Figura 11. Tipo de plan crédito vs Monto Contratado

Fuente: Elaboración propia

La **Figura 11** muestra una alta variabilidad en los montos, con algunos planes presentando valores considerablemente altos y otros concentrándose en rangos de menor valor. Esta dispersión indica que la institución maneja una cartera de productos diversificada, diseñada para satisfacer distintas necesidades de financiamiento, desde pequeños créditos hasta grandes colocaciones de capital. Sin embargo, la presencia de valores atípicos en ciertos planes y la amplia dispersión en los montos contratados sugieren áreas de mayor riesgo. Los planes con montos elevados y variabilidad en sus valores contratados pueden ser más susceptibles a manipulaciones o asignaciones irregulares, lo cual justifica la implementación de políticas de control más estrictas y auditorías periódicas.

Por otro lado, los planes con montos bajos y baja variabilidad presentan un perfil de menor riesgo, ya que su homogeneidad facilita la supervisión y reduce las probabilidades de anomalías significativas.

Anexo B. Aplicación Ley de Benford

Como parte del análisis de los resultados de aplicar la ley de Benford, se describen los resultados numéricos de las métricas de las iteraciones que consideres más relevantes, junto con las configuraciones.

Primer Dígito	Frecuencia Observada	Frecuencia Esperada	Desviación (%)
1	29%	30%	-127%
2	18%	18%	-3%
3	14%	12%	111%
4	11%	10%	88%
5	13%	8%	478%
6	6%	7%	-90%
7	4%	6%	-173%
8	4%	5%	-113%
9	3%	5%	-170%

Tabla 3. Resultados del análisis de la Ley de Benford aplicados a los montos de colocaciones financieras

Fuente: Elaboración Propia

Nota. La columna "Frecuencia observada (%)" indica el porcentaje de montos cuyo primer dígito corresponde al indicado. "Frecuencia esperada (%)" refleja los valores teóricos según la Ley de Benford. "Desviación (%)" representa la diferencia porcentual entre las frecuencias observada y esperada.

Los resultados indican que:

- El dígito '3' tiene una desviación positiva del +111%, mientras que el dígito '5' presenta una desviación aún mayor del +478%. Estas desviaciones reflejan una frecuencia inusualmente alta, lo que podría estar relacionado con ajustes deliberados en los montos asignados.
- Los dígitos ('6', '7', '8' y '9') presentan desviaciones negativas significativas, lo que sugiere que los montos altos que comienzan con estos valores están siendo sistemáticamente evitados o subrepresentados."

Este análisis resalta la utilidad de la Ley de Benford como herramienta para detectar posibles irregularidades en las colocaciones financieras. Los dígitos '3' y '5', que presentan desviaciones positivas destacadas, justifican una investigación más profunda para determinar si estas anomalías

están relacionadas con fraudes internos o prácticas atípicas. Además, la subrepresentación de los dígitos '6' a '9' podría indicar estrategias para evitar valores llamativos en los montos de créditos asignados.

Continuando con la metodología propuesta se aplica el algoritmo PageRank (Page, 1999), para los créditos que no se acoplan a la distribución y por tanto son considerados anómalo, en cuyo caso se obtienen los nodos con los puntajes más altos o con un nivel de importancia más alta en las conexiones fueron aquellos que actuaban como concentradores de transacciones de primer y segundo grado, recibiendo dinero de varios clientes, comerciales y transfiriéndolo a personas dentro de la entidad financiera Figura 9.

Anexo C. Análisis score PageRank de los grafos

Nodo	PageRank	Grado Total Entrante	Porcentaje Grado 1 Entrante (%)	Porcentaje Grado 2 Entrante (%)	Grado Total Saliente	Porcentaje Grado 1 Saliente (%)	Porcentaje Grado 2 Saliente (%)	Tipo(s) de Transacción
Nodo 1	5E-04	562	1.07	98.93	8	0	100	transaccion
Nodo 2	4E-04	553	0.36	99.64	68	100	0	relacion, transacción
Nodo 3	4E-04	424	0.94	99.06	139	100	0	relacion, transacción
Nodo 4	3E-04	388	0.26	99.74	135	99.26	0.74	relacion, transaccion
Nodo 5	3E-04	340	0.29	99.71	2	0	100	transaccion
Nodo 6	3E-04	354	0.56	99.44	119	99.16	0.84	relacion, transaccion
Nodo 7	2E-04	236	0.42	99.58	108	100	0	relacion, transaccion
Nodo 8	2E-04	219	0.46	99.54	31	100	0	transaccion
Nodo 9	2E-04	225	0.44	99.56	1	0	100	transaccion
Nodo 10	2E-04	207	0.48	99.52	0	0	0	transaccion

Tabla 4. Métricas clave del grafo transaccional Algoritmo PageRank.

Fuente: *Elaboración Propia*

Nota.

- *PageRank (%)*: Indica la relevancia de cada nodo en la red transaccional.
- *Grado Total Entrante/Saliente*: Suma total de conexiones que recibe (entrante) o genera (saliente) cada nodo.
- *Grado 1 y 2 (%)*: Representa el porcentaje de conexiones de grado directo (1) e indirecto (2).
- *Tipo(s) de Conexión*: Clasificación de las conexiones asociadas a cada nodo (transacciones o relaciones).

PageRank y relevancia de los nodos

El nodo N1 presenta el mayor valor de PageRank (0.05%), lo que indica su alta relevancia dentro de la red transaccional. Sin embargo, este nodo tiene un grado total saliente reducido (8 conexiones), todas de grado 2 (indirectas). Este patrón sugiere que N1 podría actuar como un concentrador en transacciones indirectas, posiblemente vinculadas a intermediarios.

Por su parte, los nodos N2 y N3, con valores de PageRank ligeramente menores, combinan conexiones de tipo 'relación' y 'transacción.' Esto podría indicar que ambos nodos desempeñan un papel mixto en la red, posiblemente actuando como intermediarios entre comerciales y clientes.

Grados Entrantes y Salientes

En general, los grados entrantes predominan en todos los nodos analizados, con proporciones de grado 2 (indirecto) superiores al 98%. Esto sugiere que la mayoría de las conexiones en la red se originan a través de relaciones indirectas, lo que refuerza el papel de intermediarios en la red.

Por otro lado, los grados salientes son significativamente menores. Este comportamiento indica que los nodos clave, como N1, centralizan información o transacciones, pero redistribuyen un volumen limitado de estas conexiones, manteniendo un control sobre las relaciones.

Un patrón notable en la red es la alta proporción de conexiones de grado 2 (indirectas) en los nodos clave, como N1, N2 y N3. Este comportamiento destaca el papel de estos nodos como intermediarios, lo que podría estar relacionado con estrategias específicas de redistribución de transacciones para evitar detección directa o para consolidar su influencia dentro de la red.