UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

BIANCA MARIANA MIGLIORINI DE CAMPOS

USO DE GRAFOS NO AUXÍLIO DA IDENTIFICAÇÃO DE FRAUDES EM CARTÕES DE CRÉDITO

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

DOIS VIZINHOS
2022

BIANCA MARIANA MIGLIORINI DE CAMPOS

USO DE GRAFOS NO AUXÍLIO DA IDENTIFICAÇÃO DE FRAUDES EM CARTÕES DE CRÉDITO

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientadora: Prof. Dra. Rosângela de Fátima Pereira Marquesone

DOIS VIZINHOS
2022

BIANCA MARIANA MIGLIORINI DE CAMPOS

USO DE GRAFOS NO AUXÍLIO DA IDENTIFICAÇÃO DE FRAUDES EM CARTÕES DE CRÉDITO

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de aprovação: 09/setembro/2022

Rosangela de Fátima Pereira Marquesone
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

Francisco Pereira Junior
Mestrado
Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

Paulo Júnior Varela
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Francisco Beltrão

DOIS VIZINHOS
2022

AGRADECIMENTOS

Agradeço à minha orientadora sempre presente no decorrer desse trabalho e à minha família que entendeu toda a minha ausência durante o período do curso.

RESUMO

Este trabalho tem o intuito de estudar sobre as Fraudes em Cartão de Crédito nas Instituições Financeiras, de forma a tentar identificar de maneira mais ágil e assertiva indicando que uma dada operação pode ser uma possível fraude. Para este objetivo, inicialmente foi realizada revisão da literatura abrangendo conceitos de Fraudes, *Open Banking* e Teoria de Redes. A partir deste levantamento, foi proposta uma abordagem de redução de escopo das milhares de operações de cartão, divididas por setor de atuação das empresas que possuem maior probabilidade de ocorrer dentre o universo de operações. Como resultado foram obtidos alguns grafos que, dentro do universo de operações fraudulentas, evidencia no caso dessa amostra, além de fraudes eventuais causadas por fraudadores de fora dos estabelecimentos, o caso de empresas com alto potencial de serem comparsas na aplicação das fraudes dentro dos estabelecimentos, além de termos também clientes aplicando as chamadas auto fraudes.

Palavras-chave: grafo; fraude; *Open Banking*.

ABSTRACT

In this work we study credit card fraud in financial institution with the purpose of identifying nimbly and assertively when a given operation could be a fraud. To do so we started with a literature review covering the concepts of fraud, Open Banking and Network Theory. From this review we proposed an approach of reduction of the thousands credit card operations, sorting it by company sectors with higher likelihood of fraud within its operations. As a result we obtained some graphs that show us, within the sample studied, not only the eventual fraud caused by fraudsters from outside of the establishment, but also the fraud cases from fraudsters within the establishment, as well as the cases of users applying the so-called self frauds..

Palavras-chave: graph; fraud; Open Banking.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ABECS	Associação Brasileira das Empresas de Cartões de Crédito e Serviços
ABNT	Associação Brasileira de Normas Técnicas
BACEN	Banco Central do Brasil
DECOM	Departamento de Computação

SUMÁRIO

1 INTRODUÇÃO

Cartões de crédito e cartões de débito são, atualmente, uns dos meios de pagamento mais utilizados, tanto nos comércios físicos como no comércio eletrônico. O cartão de crédito tem como premissa permitir o pagamento por bens e serviços, com a promessa de que o titular do cartão honrará a esta operação ao banco emissor do cartão (podendo ou não ter outros tipos de encargos envolvidos, a depender da operação). Os cartões, atualmente, também podem ter diversos tipos de sistemas interligados. Antigamente era somente a tarjeta ou as informações numéricas contidas no plástico; hoje, os cartões podem contar com sistemas de chip, aproximação ou nem precisam necessariamente do plástico em si, podendo ser virtuais ou contidos em relógio de aproximação.

1.1 Problema de Pesquisa

Dada a quantidade de indivíduos que utilizam seus cartões diariamente como meio de pagamento para suas compras, sejam das mais básicas como um café, ou até parte de pagamento na aquisição de um carro, há uma elevada quantidade de operações diárias sendo realizadas, e isso pode acarretar o aumento do risco que o banco do credor pode incorrer em fraudes. A ClearSale, uma empresa autoridade em soluções anti-fraudes, analisou de janeiro a junho 2021 dados referentes a operações de cartões. De um universo de 165 milhões de operações, 2,8 milhões foram tentativas de fraudes, que somavam 2,9 bilhões de reais em valor (FINSIDERS, 2022).

Um ponto sobre a detecção das fraudes, de o porquê ser tão complicado detectar, é que, em um universo de vastas operações realizadas, a quantidade de operações fraudulentas é baixa, como é possível perceber pelos dados analisados pela ClearSale de janeiro a junho de 2021, que somente 1,7% das operações analisadas eram tentativas de fraudes.

Quando uma operação fraudada é bem-sucedida, todo o conjunto das empresas envolvidas assumem com os custos, mas, em consequência, acabam repassando essas perdas para os seus consumidores em geral. Dessa forma, o combate às fraudes de cartão pode trazer os mais diversos benefícios para a sociedade. Apesar da sua complexidade e dimensão, deve ser feita de forma contínua, com grande nível de eficiência e eficácia (OLIVEIRA, 2016).

A diminuição dos impactos dos prejuízos consequente de fraudes no sistema financeiro depende da tempestividade em detectar essas ocorrências, pois nessa janela de tempo o fraudador vai continuar as tentativas de efetuar a quantidade máxima de operações ilegítimas que ele puder fazer. Na detecção tardia da fraude, a instituição financeira fica limitada ao prejuízo do limite máximo disponibilizado ao cliente para efetuar compras no cartão. No entanto, a não detecção da fraude pode acarretar outros desdobramentos e não só o dano financeiro, como por exemplo, a conta da vítima pode ser utilizada de maneira ilícita (“laranja”) para

operações de outros tipos de créditos fraudulentos e a instituição financeira pode incorrer em risco de imagem a depender do dano alcançado, e, como consequência, a possibilidade de perda de clientes e de carteiras.

Desta forma, o envolvimento de todos estes cenários é uma das maiores motivações para o aperfeiçoamento contínuo das técnicas de detecção de fraudes, buscando tentar ao máximo combater as operações fraudulentas.

1.2 Objetivos

1.2.1 Objetivo Geral

Tem-se por objetivo geral o estudo de fraudes em operações de cartões de crédito, um problema persistente na realidade das instituições financeiras e que as técnicas utilizadas para isso são aprimoradas ao longo do tempo.

Para a obtenção do objetivo geral, foi utilizada uma base sintética, com finalidade de auxiliar o estudo de maneiras de se aprimorar a identificação de possíveis operações fraudulentas dentre um universo grande de operações, e como técnica foi utilizada a Teoria de Grafos, pois complexidade do problema permite essa abordagem, além de fácil visualização.

Em complemento, a realização de levantamento de estudos relacionados ao tema, de maneira que os resultados possam ser aplicados ou confrontados ao contexto deste trabalho.

1.2.2 Objetivo Específico

Tem-se por objetivo específico o de propor uma regra que possa auxiliar a identificação de operações fraudulentas dentro do universo de operações disponibilizados na amostra da base sintética. Além disso, se utilizar da Teoria de Grafos para reforçar de maneira visual as constatações alcançadas no estudo.

Para a inferência dos resultados finais, os seguintes objetivos específicos foram definidos:

- Realização de análise na literatura sobre: fraudes em cartão, *Open Banking*, Teoria de Grafos;
- Investigação de propostas de técnicas de detecção de fraudes em cartão de crédito por meio de trabalhos anteriores;
- Utilização de dados de Cartão de Crédito por meio de base sintética;
- Utilização da Teoria de Grafos;
- Propor uma regra de identificação de possíveis operações fraudadas;

1.3 Justificativa

O tema de detecção de fraudes de operações em cartão de crédito foi escolhido pois, justamente no período de escrita dessa monografia, o Banco Central está implementando o sistema de *Open Banking*. Segundo o Bacen (2022), órgão responsável no Brasil pela

regulamentação do *Open Banking*, esse sistema é o compartilhamento opcional de dados bancários pessoais de uma instituição financeira com outra, por iniciativa do cliente. Isso implicaria na possibilidade de verificar o cliente como um todo, como, por exemplo, operações realizadas em um ou outro cartão de um banco X ou Y. Por exemplo, o cliente tem o cartão X do banco X cadastrado em seu aplicativo da Amazon, uma facilidade atualmente. Ele acaba finalizando todas essas operações com esse cartão de determinado banco. No entanto, o cartão venceu, sem que ele percebesse e, em uma sexta-feira à noite ele acaba utilizando o cartão Y do banco Y, que habitualmente não utiliza para compras nesse aplicativo, para finalizar a compra de seu Iphone, com valor elevado. Pode ser que o banco Y nesse momento identifique e acuse uma operação fraudulenta erroneamente, pois saiu do padrão de compras que se era esperado para esse cartão, cujas operações eram baixas e somente em aplicativo de entrega de refeições. Se o banco tem uma visão do todo, pode ser mais fácil identificar uma possível fraude. Um estudo mais recente da ClearSale (MELO, 2022) com dados já de 2022, apontam que de janeiro a junho foram mais de 3 milhões de tentativas de fraudes, e que esse número foi maior que o observado em 2021. O público mais afetado nas tentativas de fraudes estão as pessoas de até 25 anos, seguidos pelo grupo de 26 a 35 anos, os públicos mais digitalizados atualmente.

1.4 Materiais e Métodos

Para o estudo desenvolvido será utilizada uma base sintética de transações de cartão de crédito pública, obtida por meio do site Kaggle[1]. Após a obtenção da base, a extração e preparação dos dados, além das análises estatísticas será executada utilizando a linguagem de programação Python combinado com o Excel. Para a construção dos grafos e análises dos resultados esperados com o estudo, será utilizado o *software* Gephi. Dessa forma é possível resumir o passo a passo como:

- Captura dos dados;
- Preparação dos dados;
- Análise dos dados;
- Construção dos grafos;
- Análise dos resultados.

1.5 Organização do Trabalho

No Capítulo 2, foi realizada uma revisão da literatura com o propósito de dar direcionamento ao trabalho, garantindo assim uma base em relação aos seguintes fundamentos utilizados no decorrer do trabalho: fraudes em cartão de crédito, *Open Banking* e Teoria dos Grafos. No Capítulo 3 foi realizado um estudo sobre trabalhos correlatos, possibilitando compreender melhor o estado da arte do contexto proposto. No Capítulo 4 é apresentada uma

[1] https://www.kaggle.com/datasets/ealaxi/banksim1

análise exploratória dos dados contidos na base de operações, além dos recursos que foram utilizados para a realização do estudo. No Capítulo 5 são apresentados os resultados e as análises obtidas da solução proposta. No Capítulo 6 são apresentadas as considerações finais com proposição de trabalhos futuros a medida que novos dados sejam obtidos.

2 REVISÃO DE LITERATURA

Para a elaboração deste trabalho foi necessário ter um panorama do problema de fraudes em operações executadas por cartões de créditos. Este levantamento teve por objetivo verificar como a fraude em cartões pode afetar o sistema financeiro, além de se verificar o que existia de trabalhos desenvolvidos, seus resultados e como isso pode ajudar a minimizar os impactos gerados pelas fraudes.

2.1 Fraudes em cartão de crédito

Segundo a Associação Brasileira das Empresas de Cartões de Crédito e Serviços (ABECS, 2022), a fraude em operações de cartão de crédito é o causador de movimentação de bilhões de dólares ilegalmente, todos os anos, com implicações dispendiosas tanto para clientes, quanto para instituições financeiras, seguradoras e credenciadoras de cartão de crédito.

Com o aumento exponencial do crescimento das tecnologias observadas nos últimos anos, também se observa o desenvolvimento de novas tecnologias que facilitam a realização de operações fraudulentas em cartão. Em geral, os fraudadores são pessoas bem-organizadas, inteligentes e com grande conhecimento, e se aprimoram de maneira a acompanhar as novas tecnologias, pois sempre buscam mecanismos mais fáceis e baratos para se obter vantagens. A fraude é um negócio rentável, estável e muito bem-organizado e administrado (HAND, 2002).

No Brasil, a pandemia ajudou a aumentar o *boom do e-commerce* nos anos de 2020 e 2021 e, com isso, para fins de segurança digital, é solicitado que as pessoas sejam mais atentas, pois, com o ganho da facilidade na execução das compras online, existe o aumento do risco em incorrer em uma situação indesejada de fraude eletrônica, principalmente pelo meio eletrônico mais disseminado atualmente, o cartão de crédito.

> Em 2021, o comércio eletrônico brasileiro registrou um faturamento recorde, que totalizou mais de R$ 161 bilhões, alta de 26,9% comparado ao ano anterior. O número de pedidos aumentou em 16,9% com 353 milhões de entregas, segundo a Neotrust, empresa de segurança e monitoramento online. O *ticket* médio também registrou alta de 8,6% em 2021 quando comparado a 2020, atingindo média de R$ 455 por compra.
>
> Na mesma direção, as tentativas de fraudes em 2021 aumentaram 74%, aponta o estudo "Mapa da Fraude", que foi divulgado nesta quarta-feira (02) pela ClearSale, empresa de soluções antifraude.
>
> O estudo analisou mais de 375,5 milhões operações de alguns segmentos incluindo o *e-commerce*, mas também telecomunicações e mercado financeiro. Somados, o valor das tentativas de fraude chega a R$ 5,8 bilhões, 61% acima dos R$ 3,6 bilhões registrados em 2020.
>
> "O que estamos observando é que as fraudes estão mais dinâmicas e sofisticadas e são feitas por quadrilhas especializadas que buscam brechas em sistemas de segurança ou na inexperiência de ingressantes. O nível dos criminosos vem aumentando. Aquela história de uma pessoa inexperiente aplicando os golpes não existe mais, o processo está ficando cada vez mais profissional",

> diz Marcelo Queiroz, head de Estratégia de Mercado da ClearSale. (SUTTO, 2022).

Com relação aos tipos de fraudes mais comuns atualmente, são observadas 3 categorias, que são conhecidas por: fraude efetiva, fraude amigável e a auto fraude.

- Fraude efetiva - o fraudador faz a compra em uma loja virtual e, no momento de realizar o pagamento, usa dados roubados de cartões de crédito de outros consumidores. Como os dados são verdadeiros, essa fraude é conhecida também como "fraude limpa".
- Fraude amigável - acontece quando alguém próximo do titular, que convive no mesmo ambiente, por exemplo, usa os dados do cartão sem consentimento do dono. Sem saber que isso aconteceu, o titular não reconhece a compra e pede o estorno.
- Auto fraude - diz respeito a uma compra feita pelo próprio fraudador. O titular efetua a compra com o próprio cartão e, após receber o produto ou serviço, entra em contato com a administradora do cartão para contestar o lançamento na fatura como se não reconhecesse a dívida.

Por meio de monitoramento contínuo das operações em cartão de crédito, pode-se detectar a fraude ou tentativa de fraude no cartão. Isso porque uma radical alteração no padrão de consumo pode indicar a fraude, nos casos das categorias de fraude efetiva e fraude amigável, principalmente. A auto fraude pode ser mais difícil de detectar, pois em alguns casos, não costuma sair do padrão de compras do fraudador.

Muitas vezes, o desafio associado a tarefas como detecção de fraude e spam é a falta de todos os padrões prováveis necessários para treinar modelos de aprendizado supervisionado adequados. Esse problema se acentua quando os padrões fraudulentos não são apenas escassos, mas também mudam com o tempo. A mudança no padrão fraudulento ocorre porque os fraudadores continuam a inovar em novas maneiras de contornar as medidas adotadas para evitar fraudes. Dados limitados e padrões em constante mudança tornam o aprendizado significativamente difícil (PORWAL; MUKUND, 2018).

Considerando a ótica de relacionamentos, muitos sistemas sociais e econômicos podem ser representados como redes que codificam as relações entre entidades que são eles próprios descritos por diferentes atributos de nó. Encontrar anomalias nesses sistemas é crucial para detectar abusos como fraudes de cartão de crédito, *spams* na *web* ou invasões de rede. Intuitivamente, nós anômalos são definidos como nós cujos atributos diferem nitidamente dos atributos de um certo conjunto de nós de referência, chamados de contexto da anomalia (GUTIéRREZ-GóMEZ; BOVET; DELVENNE, 2020).

2.2 Open Banking

Segundo definições do Banco Central do Brasil (BACEN, 2022) no Brasil, o sentido ou a tradução literal de *Open Banking* é "banco aberto", mas também pode ser entendido como "Sistema Financeiro Aberto". Em linhas gerais, o *Open Banking* permite que o cliente autorize

e manifeste seu consentimento em compartilhar seus dados financeiros entre as instituições financeiras de sua preferência e que estão integradas ao sistema, e ainda, como consequência, ele poderá receber ofertas mais assertivas acerca de produtos e serviços, ou seja, mais adequadas ao seu perfil, com maior agilidade, conveniência e segurança.

Com o *Open Banking*, o consumidor poderá comparar preços de produtos e serviços, melhorando o controle de sua vida financeira e acessando produtos personalizados em condições mais vantajosas. Atualmente, uma instituição não "enxerga" o relacionamento de seu cliente com uma outra instituição, então pode enfrentar dificuldade em competir por ele com melhores prestações de serviços e condições de taxas.

Com a permissão de cada correntista, as instituições se conectam diretamente às plataformas de outras instituições participantes e acessam somente os dados autorizados pelos clientes. Todo esse processo é feito em um ambiente seguro e a permissão poderá ser cancelada a qualquer tempo pelo cliente, da mesma maneira e local onde aderiu a permissão (BACEN, 2022).

Nesse sistema, existe a padronização do processo de compartilhamento de dados e serviços financeiros pelas instituições autorizadas a funcionar pelo Banco Central do Brasil, por meio de abertura e integração de plataformas e infraestruturas de tecnologia.

O *Open Banking* parte do pressuposto que o consumidor é titular de seus dados cadastrais e financeiros, e que pode transferir essas informações que lhe pertencem para outra instituição, a qualquer momento, em busca de melhores produtos ou serviços a preços mais baixos. É importante ter em mente que a disponibilização de dados por parte dos consumidores gera um valor para as instituições financeiras, em termos de informação.

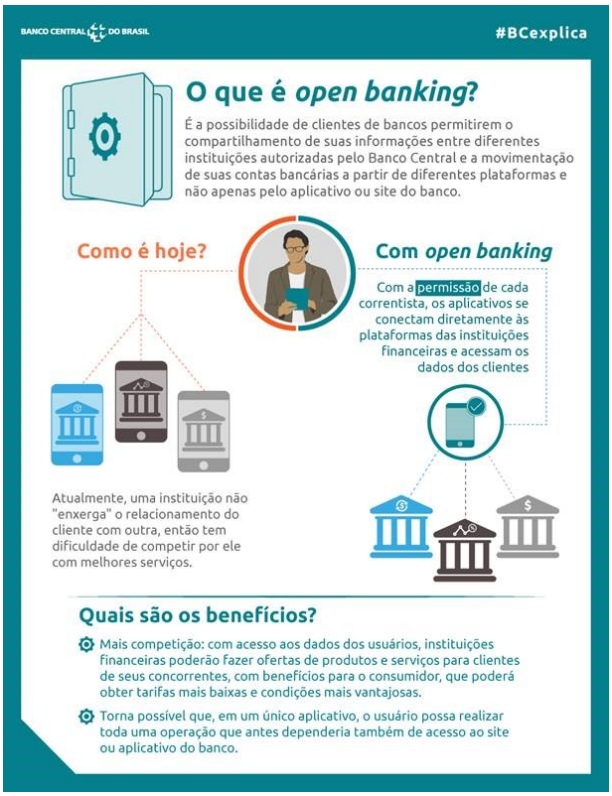
Desta forma, é possível pontuar os principais benefícios do *Open Banking* como sendo:

- Com a implementação e integração total do sistema do *Open Banking*, uma parte desse benefício será revertida para quem disponibiliza os dados, ou seja, para os próprios consumidores;
- A instituição financeira também tem a vantagem de ter a integração de serviços financeiros às diferentes jornadas digitais dos clientes, facilitando a contratação de produtos e serviços financeiros em ambientes mais convenientes para o consumidor, de forma ágil e segura;
- O aumento da transparência e a redução da assimetria de informações (ou a precificação da assimetria de informações de forma mais eficiente), diminuindo, assim, as barreiras à entrada no sistema financeiro e favorecendo um ambiente de negócios mais inclusivo;
- A entrega de serviços customizados aos diferentes perfis de clientes, levando em consideração os interesses, objetivos e necessidades de cada público;
- A assistência ao planejamento das famílias e das empresas;
- O surgimento de novos modelos de negócios e novas formas de relacionamento entre as instituições participantes, seus clientes e parceiros.

Na Figura 1 é apresentado o esquema de como se estrutura o *Open Banking* no Brasil, a partir das definições do Banco Central do Brasil, além das informações dos benefícios

envolvidos dada a sua implementação.

Figura 1 – Informações a respeito do novo sistema *Open Banking*



Fonte: Banco Central do Brasil

Segundo o estudo publicado pela Accenture (ACCENTURE, 2021), o grande pioneiro na adoção do sistema *Open Banking* foi o Reino Unido, tendo a implementação da sistemática em 2018, tornando-se, atualmente, exemplo e referência para a composição dos regulamentos acerca do tema para outros países, incluindo o Brasil. Características como incentivo à inovação por parte dos bancos, facilidades de uso, inclusão financeira da população e aumento da concorrência com efeito nas condições oferecidas, foram os principais pontos que favoreceram o sucesso mundial do sistema.

A publicação ainda traz como referência os Impactos do *Open Banking* e PIX no Brasil, que comprovam o sucesso de sua implementação. Atualmente o regulador observa mais de 1 milhão de usuários do *Open Banking* cadastrados em sistemas de compartilhamento, mais de 240 provedores de serviços regulados e cerca de 40 Instituições Financeiras e *Fintechs* que aderiram ao sistema.

No Brasil, o Banco Central programou a implantação do sistema de forma a acontecer de maneira gradual e foi dividida em 4 fases, sendo:

- 1ª fase – Implantação de compartilhamento de Informações relacionadas a produtos bancários. Iniciada em 01/02/2021.
- 2ª fase – Implantação do compartilhamento e troca de informações sobre conta e de operações dos clientes, mediante consentimento. Iniciada em 13/08/2021.

- 3ª fase – Implantação da iniciação de pagamentos, ou seja, permissão para que pagamentos sejam efetuados em páginas fora dos aplicativos dos bancos (o exemplo mais clássico é um site de compra virtual, no qual o cliente poderá fazer a compra ali mesmo, sem precisar migrar para o *app* do banco na hora de finalizar o pagamento). Iniciada em 29/10/2021.
- 4ª fase - Compartilhamento de informações adicionais sobre investimentos, câmbio e seguros. Iniciada em 25/03/2022.

Apesar da divisão em 4 fases para a implantação do sistema, o Banco Central afirma que ainda não é exaustivo o programa e que após o cumprimento de toda a programação, novas funcionalidades deverão existir, mas ainda sem previsão de quais e quando.

2.3 Teoria de Grafos

A teoria de grafos vem evoluindo de maneira ágil nas últimas décadas, se colocando como uma forma dinâmica de ver o mundo, sendo que o principal foco são as conexões existentes (PARKHE; WASSERMAN; RALSTON, 2006). A teoria foca no estudo das relações existente entre os objetos do estudo (indivíduos) dentro de uma rede, por meio de estruturas intituladas grafos. Um grafo é um conjunto de pontos (vértices, nós) que contenham relações (arestas, links) entre si.

O objetivo da teoria é o entendimento de quais indivíduos são mais influentes dentro do todo, ou seja, dentro da rede observada, por meio das medidas de centralidade. Essas medidas podem indicar quais vértices têm mais influência dentro da rede, de várias formas. Por exemplo, a determinação do indivíduo dominante observando-se a quantidade total de arestas (relações entre os vértices) relacionado com outros vértices do grafo. Essa medida é denominada centralidade de grau.

Tendo em vista quais análises pode-se executar corretamente dentro de uma rede (ANDERSON; VONGPANITLERD, 2006), a teoria de grafos pode contribuir em diversas áreas, tais como na área da economia, com estudos sobre crises econômicas (AMIN; THRIFT, 1992); na área de tecnologia, com estudos de redes e mídias sociais (KIM; HASTAK, 2018) e fluxo de informação (NEWMAN, 2018) e na área da saúde, com estudos de proliferação de doenças (BROWN et al., 2016).

2.3.1 Modelos de Grafos

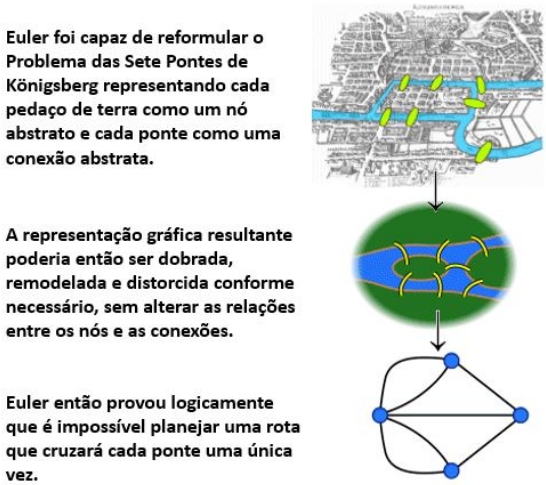
A teoria de grafos teve início em meados do século XVIII em Königsberg, pois nessa cidade havia o questionamento de seus cidadãos se seria possível atravessar as 7 pontes existentes na cidade, de maneira a não se passar pela mesma ponte mais de uma vez e retornar ao início do passeio. O questionamento foi respondido em 1741, quando o matemático Leonhard Euler publicou seu artigo em que concluiu que não seria possível nenhum caminho para cruzar as 7 pontes uma única vez.

Como resultado, Euler concluiu que só seria possível percorrer o caminho passando somente 1 vez por cada ponte se:

- Não existisse nenhum vértice com quantidade de arestas ímpares; ou
- Exatos 2 vértices com número de arestas ímpares.

A publicação do matemático Euler, respondendo ao questionamento dos habitantes da cidade de Königsberg, foi que embasou toda a teoria de Grafos conhecida atualmente. Na Figura 2 (SILVA, 2013), tem-se a representação do problema elucidado pelo matemático sobre as Sete Pontes de Königsberg e o respectivo grafo.

Figura 2 – Problema das Sete Pontes de Königsberg e o respectivo grafo



Fonte: Blog Tarcízio Silva

Passados 2 séculos, a teoria se expandiu e não se resume mais a problemas e estruturas fixas. Conforme Watts (2004) analisa, o estudo de redes atualmente associa desde redes sociais a redes econômicas, não aceitando o conceito de estruturas estáticas, mas sim o de redes dinâmicas, pois os elementos que compõe o conjunto de informações a serem estudadas evoluem e mudam suas características no tempo e espaço.

Nessa mesma linha de análise, dois matemáticos Erdos e Renyi (1960), expõem o Modelo de Redes Aleatórias, que define que no caso de existir pelo menos uma ligação de um integrante com outro, no final todos os integrantes da rede estarão conectados. Essas conexões acabam sendo aleatórias, o que define a característica randômica da rede. Adicionando mais integrantes a rede, aumentara a chance de criação de grupos ou *clusters*, podendo tornar a rede ainda mais aleatória.

Ainda na década de 60, Travers e Milgram (1969) propuseram a Teoria dos Seis Graus de Separação, se tornando a primeira experiência prática que propôs estudar, dentro de uma mesma rede, o grau de distância entre os indivíduos que a compõe. Para a realização desse estudo, foram enviadas cartas aleatoriamente para várias pessoas. Nessas cartas existia a indicação de uma determinada pessoa para quem a aquela carta deveria chegar especificamente.

Se o indivíduo que recebeu a carta não tivesse uma relação diretamente com o alvo, deveria enviar a carta para alguém que julgasse conhecer ou estar mais próxima de conhecer aquela pessoa. Como conclusão, as cartas acabaram chegando ao seu destino, e haviam passado, em média, por outros 5 indivíduos anteriormente. Dessa forma, verificou-se que existiam em média no máximo 6 elos na cadeia, desde a pessoa que enviou a carta, até o recebimento pela pessoa correta. Nesse experimento foi possível concluir que o mundo é relativamente pequeno, pois todos os indivíduos se contactam com poucos graus de separação entre eles (BARABASI, 2003).

Com isso, pode-se dizer que redes dizem respeito à conectividade, às inter-relações presentes entre os componentes de um sistema, sendo que os tradicionais modelos de risco empregam seus esforços em entender o comportamento individual da parte e assume que seu risco total é o agrupamento das partes dos riscos individuais (DIEBOLD; YILMAZ, 2014).

Logo, com o reconhecimento da existência de diferentes conexões entre os distintos componentes de um grupo, pode-se chegar à conclusão de que o fator total é causado como consequência direta da interdependência entre os fatores intermediários.

Com isso, a teoria de grafos ajuda a visualizar as premissas que não se ajustam ao mundo em que atualmente estamos inseridos, infinitamente conectado, descartando as teses de que as relações lineares ou determinísticas funcionam no contexto atual.

São características intrínsecas às redes a não-linearidade e complexidade. Dado isso, analisando um pequeno grupo de indivíduos, o resultado pode ser centenas de milhares de conexões diversas entre eles.

Dessa forma, é importante destacar que a utilização de redes, assim como diversas outras teorias, é apenas uma aproximação da realidade, que se mostra tão complexa quanto mais se aprofunda em seu entendimento.

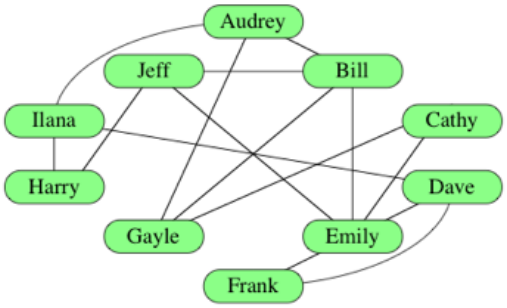
2.3.2 Definição Geral de um Grafo

Segundo informações da Khan Academy (ACADEMY, 2022), a Figura 3 é uma maneira de representar uma rede social. Uma linha que une os nomes entre 2 indivíduos implica que eles se relacionam ou se conhecem. Se não existir nenhuma relação entre dois indivíduos, representada por uma linha, eles não se conhecem. Também tem-se que a relação de conhecer um indivíduo é bidirecional, sendo que pode-se exemplificar por como Jeff conhece o Harry, isso implica que o Harry também conhece o Jeff. Esta rede social citada no exemplo da figura é denominada um grafo. Os indivíduos (representados pelos nomes) são os vértices desse grafo. Cada uma das linhas é uma aresta ou link, que estabelece uma ligação entre dois vértices ou nós.

Os grafos podem ser representados através de um diagrama onde os vértices são representados por pontos e cada aresta é representada por uma linha ligando os pares de vértices que a definem, como exemplo a Figura 4.

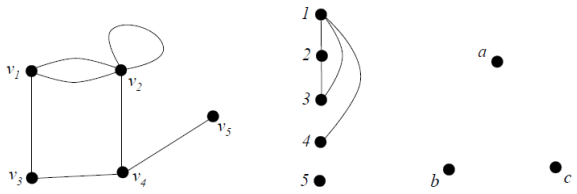
Em algumas aplicações, como indicado na Figura 5, as arestas são definidas como

Figura 3 – Rede social representada por meio de um grafo



Fonte: Khan Academy

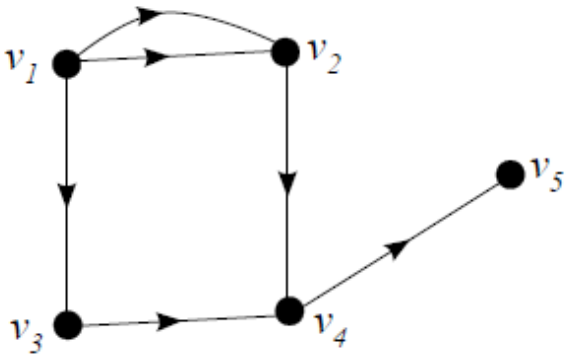
Figura 4 – Exemplos de grafos



Fonte: Própria

pares ordenados de vértices. Neste caso, diz-se que o grafo é orientado ou direcionado e se chama Dígrafo.

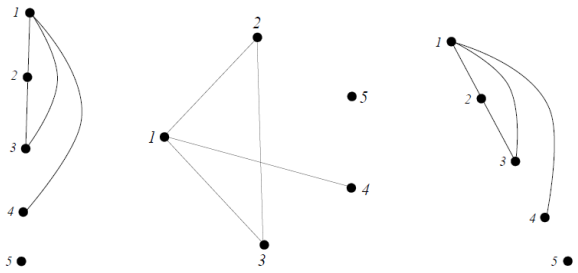
Figura 5 – Exemplo de um Dígrafo



Fonte: Própria

Assim, pode-se dizer que a teoria de grafos é um ramo da matemática que estuda as relações entre os elementos de um determinado conjunto, independentemente de como se dão essas relações, tal como pode ser verificado por meio da Figura 6.

Figura 6 – Exemplos de um grafo com diferentes representações gráficas



Fonte: Própria

2.3.3 Banco de Dados de Grafos

Banco de dados de grafos permitem armazenar entidades e relacionamentos entre essas entidades. Entidades também são conhecidas como nodos, os quais possuem propriedades (SADALAGE; FOWLER, 2013).

O Gephi é um software livre e gratuito, ou seja, de código aberto, desenvolvido em linguagem Java e na plataforma NetBeans, que é utilizado para estruturação, manipulação, análise e visualização de redes complexas. Seu funcionamento pode ser em sistemas operacionais Windows, MacOS e Linux.

O Gephi tem em sua base um grande número de *plug-ins* e métricas para serem utilizados nas análises de redes, sendo que sua aplicabilidade pode se dar para análise de dados tanto na área da biologia, por exemplo, quanto na análise de informações de redes sociais, entre outras.

Um efeito colateral interessante, proveniente do uso de banco de dados de grafos para recomendações, é que, à medida que o tamanho dos dados aumenta, o número de nodos e relacionamentos disponíveis para fazer recomendações também aumenta, rapidamente. Os mesmos dados também podem ser utilizados para encontrar informações – por exemplo, quais produtos são sempre comprados juntos ou quais itens são sempre faturados juntos. Alertas podem ser disparados quando essas condições não forem satisfeitas. Assim como outros mecanismos de recomendação, os bancos de dados de grafos podem ser utilizados para pesquisar padrões em relacionamentos a fim de detectar fraudes em operações.

3 TRABALHOS CORRELATOS

Neste capítulo foi realizada uma revisão de trabalhos correlatos com a temática de fraudes em cartões de créditos. Foi observado como diferentes autores escolheram técnicas estatísticas ou de *machine learning*, avaliando os resultados obtidos por meio de seus trabalhos.

3.1 Artigo 1 - FControl®: Sistema inteligente inovador para detecção de fraudes em operações de comércio eletrônico

Neste artigo (COELHO; RAITTZ; TREZUB, 2006), foi verificado que o problema de combate à fraude é antigo e persiste ao longo do tempo. A abordagem aqui estudada é a de utilização de inteligência computacional a partir das técnicas de redes neurais, lógica *fuzzy* e computação evolutiva para detecção e classificação da possível operação fraudada. Nesta publicação foram utilizados dados reais, combinando as técnicas citadas e o resultado originou o método denominado FControl®.

Em 2006, o problema das fraudes em negócios realizados recaia principalmente sobre o estabelecimento comercial, e nunca sobre o comprador, sendo que em sua maioria, as operações eram realizadas por meios eletrônicos, como sites de internet. O cliente que não reconhecia a compra efetuada, entrava em contato com o banco que cancelava o crédito com o estabelecimento ou ainda efetuava um débito em sua conta. Independentemente de ser uma auto fraude ou uma fraude legitima, o processo era executado dessa forma. Os custos para o estabelecimento comercial, além do valor do produto, incluíam em perda da mercadoria, perdas com embalagens e frete, além de perda com taxas bancárias.

O problema de detecção desde a época do artigo, envolve a identificação e classificação dessas fraudes de maneira mais assertiva e com a maior agilidade possível. Com isso a principal contribuição do projeto foi o sistema FControl® desenvolvido para dar suporte ao lojista em operações efetuadas por meio de comercio eletrônico.

Os sistemas convencionais da época se utilizavam de análises de informações dos pedidos, de maneira a se identificar desvios de padrões nas compras efetuadas para poder se classificar o risco da venda; checagens de algumas informações por meio de banco de informações externos; verificações dos dados nas chamadas listras negras, que poderiam ser desenvolvidas pelos próprios estabelecimentos, e por último, quando a informação se afastasse muito de alguns padrões, uma checagem diretamente com o cliente para verificação e confirmação da compra, além de solicitação de cópia de documentos.

Por outro lado, as administradoras de cartão e os bancos também se utilizavam de regras para tentar dirimir o índice de fraudes. Contudo, os métodos eram um pouco frágeis e possuíam diversos problemas, acarretando para o cliente alguns transtornos. Por exemplo, os sistemas eram ineficientes na detecção e dessa forma rejeitavam pedidos validos; também

poderiam utilizar de artifícios que invadiam a privacidade do consumidor e; o processo era moroso, então gastava-se muito tempo na análise, interferindo no fechamento do pedido, acarretando atraso na entrega dos pedidos.

O sistema desenvolvido e foco de objeto de estudo do artigo conta com um conjunto de regras e utilização de informações de modo tempestivo de maneira a agilizar a tomada de decisão por parte do estabelecimento comercial. Conta com módulos tecnológicos que conciliam informações de listas brancas / negras, se utilizam de regras de negócios, travas de seguranças, combinado com um modelo de rede neuro-nebulosa e pontuação de risco, agregados ainda com um sistema especialista que toma uma decisão sem necessitar de uma pessoa física exercendo essa decisão e por último, faz um gerenciamento das solicitações de análises do risco.

No artigo, foram desenvolvidos diversos estudos e uma das abordagens utilizadas foi a análise de *cluters*, fazendo a categorização das operações em normal, suspeita ou fraudulenta. O Banco de dados utilizado continham 2916 dados reais de operações sendo que o modelo teve uma taxa de acerto de 88,9% para operações fraudulentas e 89,5% de acerto para operações legais.

As informações das regras utilizadas no modelo eram renovadas de 14 em 14 dias de maneira a manter o sistema sempre atualizado e alerta para o surgimento de possíveis mudanças nos padrões observados.

À época (maio de 2005) o sistema FControl® contava com um rol de 250 estabelecimentos comerciais como clientes que utilizavam o sistema para ajudar na identificação das fraudes e um banco de dados com 45670 operações realizadas.

Atualmente verifica-se que o problema de identificação de fraudes em cartão ainda persiste, mas que a forma de responsabilização se alterou e muito ao longo do tempo. O que antes era um problema do comerciante, atualmente é um problema das Instituições Financeiras em conjunto com as redes credenciadas de cartão de crédito.

3.2 Artigo 2 – Aplicação de Técnicas de Inteligência Computacional para Detecção de Fraude em Comércio Eletrônico

Neste artigo (LIMA; PEREIRA, 2012), já existia a verificação e o apontamento do crescimento exponencial do comercio eletrônico e juntamente com ele, os casos de fraudes tendo por consequência prejuízos de bilhões em todo o mundo. Da mesma forma o artigo tem por objetivo fornecer mecanismos de identificação de maneira ágil para que esse tipo de situação seja prontamente identificado e não concluído, de modo a poupar valores de perdas operacionais.

Para a elaboração do trabalho, foram utilizados dados reais disponibilizados pelo serviço de pagamento eletrônico da UOL, no período de 6 meses, sendo 10/2010 a 03/2011. A base obtida, assim como esperado, se observa um desbalanceamento nos dados, comparando as operações de fraudes e não fraudes.

Para a execução desse trabalho, foram desenvolvidos modelos de redes *bayesianas* e

de regressão logística, que comparados entre si demonstrou que o modelo de redes *bayesianas* teve uma acurácia melhor, de modo a identificar melhor as operações fraudulentas. Também se utilizou do conceito de Eficiência Econômica - EE (além da acurácia, é observado em termos monetários qual dos modelos teria melhor desempenho quando colocado para performar) para medir a eficácia entre os 2 modelos de maneira a deixar a comparação mais fidedigna.

A validação do trabalho contou com a aplicação do modelo em cenário real, tendo bons resultados, com ganhos de até 35,6% comparados ao cenário real da empresa sem os mecanismos de atuação embutidos.

3.3 Artigo 3 – Crimes e fraudes eletrônicos: Perspectivas de Ações Empresariais Adotadas por Instituições Financeiras

Neste artigo (CARVALHO et al., 2015), foi analisado que o problema de fraudes eletrônicas e consequentemente fraudes em cartão de crédito ainda é uma realidade bastante frequente no dia a dia da população brasileira e instituições financeiras.

O trabalho teve por objetivo fazer uma análise estatística de dados de fraudes ocorridos no ano de 2009 na cidade de São Paulo. O estudo foi desenvolvido a partir de uma pesquisa de análise qualitativa e exploratória, além de pesquisa documental e teve por finalidade de identificar a influência de inúmeras variáveis, tais como, localidade, sazonalidade, horário, que gerou um modelo estatístico de previsão para esses crimes de fraudes eletrônicas. Os dados foram levantados pelos autores diretamente nas instituições financeiras, privadas e públicas, pelas pessoas responsáveis nos setores de fraude de cada instituição. Os dados foram descaracterizados, pois são dados sensíveis, e no trabalho levou-se em conta o fator do sigilo bancário.

Após a coleta, foram feitos diversos estudos com as informações, tais como: análise descritiva e teste de interdependência para verificar quais das informações coletadas estariam mais correlacionadas com a variável *target*. Na sequência foi desenvolvido um modelo utilizando a técnica de Regressão Logística, para previsão das fraudes.

Além do modelo desenvolvido, o artigo ainda conta com uma extensa revisão de literatura, que traz informações detalhadas de como podem ocorrer as Fraudes na Internet, e os meios que os fraudadores se utilizam para isso. Ações denominadas *Phishing*, Boatos e falsos *E-mails* são umas das técnicas utilizadas pelos fraudadores. Mas não são as únicas formas. Tentativas de golpes em estabelecimentos físicos também ocorrer quando existe a figura da empresa ou de algum funcionário fraudando *in loco*. Essa modalidade visa na troca do cartão magnético, cópia dos dados do cartão físico, além da obtenção da senha do cliente, ou ainda quando o cartão fica "retido" indevidamente no caixa eletrônico.

Como resultado, o artigo indicou para a época analisada, que poderia existir sazonalidade na efetuação das fraudes, quando verificou que existiam períodos que se observava mais fraudes em relação a outros períodos, como dia do mês, até o dia 10 era o período que mais se concentrava as execuções das fraudes; dia da semana em que se ocorriam mais fraudes,

no caso do estudo o final de semana foi onde se observaram mais fraudes sendo ocorridas; horários mais visados para aplicação das fraudes, sendo as fraudes se concentravam no horário de 6h da manhã até 12h; discrepância entre os trimestres, sendo que o $2^{\underline{o}}$. trimestre foi o que se observou mais fraudes sendo realizadas; e, localidade na Região de São Paulo, sendo que a região da Zona Leste + Centro + ABCD, foi a região que na época apresentava o maio número de fraudes.

Como conclusão, o artigo mostrou que a nossa legislação ainda é muito frágil e imprecisa quando se trata de problemas relacionados a fraudes eletrônicas, sendo necessário modificação na legislação de maneira a auxiliar as instituições financeiras a se prevenir contra esses tipos de crimes. Além disso, o modelo de previsão desenvolvido mostrou-se adequado no auxílio de prevenção contra as fraudes, dado que verificou que as variáveis que mais influenciavam nas ocorrências das fraudes eram a região, mês e faixa de horário.

Após a revisão dos três trabalhos selecionados, concluiu-se que o problema de fraudes em cartão é existente de longa data e a cada tecnologia desenvolvida para que exista uma maior proteção para esse meio de pagamento, com o tempo, é desenvolvido também um mecanismo para se bular os mesmos.

Ao final, este trabalho também tem por objetivo de desenvolver uma regra de identificação de possíveis fraudes em operações de cartões de crédito, contudo, a diferença está na abordagem escolhida para isso. Neste trabalho tem-se por objetivo a utilização de uma regra inicialmente visual e por meio dela, desenvolver uma regra de identificação, ou gerar um alerta, de uma possível operação fraudulenta.

4 MATERIAIS E MÉTODOS

Neste capítulo foram abordados os materiais e os métodos utilizados no trabalho. Inicialmente, na Seção 4.1 são apresentados detalhes referentes ao banco de dados selecionado e as ferramentas utilizadas para isso. Já na Seção 4.2, foram tratados os resultados da análise exploratória dos dados.

Segundo Wazlawick (2009), o método de pesquisa consiste na execução de uma cadeia de procedimentos necessários para demonstrar que os objetivos do estudo serão atingidos, e como acontecerão. Na continuação, serão trazidos os passos que descrevem as características da pesquisa e a forma de obtenção dos resultados.

4.1 Materiais

4.1.1 Objeto de estudo

O objeto de estudo é uma base de dados sintética contendo operações de cartão de crédito de clientes de um banco fictício X.

4.1.2 Procedimento de obtenção dos dados

Dados de operações de cartões dentro do sistema *Open Banking* ainda não estão consolidados, pois o sistema ainda é recente. Logo, para os estudos neste trabalho, será utilizada uma base sintética que foi obtida por meio do site Kaggle.

4.1.3 Descrição dos dados disponíveis

BankSim é um simulador de operações com cartões bancários criado com base em uma amostra de dados transacionais agregados, fornecidos por um banco na Espanha. Seu principal objetivo é a geração de dados sintéticos que podem ser usados para pesquisas de detecção de fraudes.

Para o desenvolvimento e calibragem do simulador, foram utilizadas estatística e análise de redes sociais das relações entre comerciantes e clientes. O objetivo final é que o BankSim seja utilizado para modelar cenários relevantes que combinem pagamentos normais e assinaturas de fraudes conhecidas injetadas. Os dados gerados pelo simulador não contêm informações pessoais ou divulgação de operações legais e privadas de clientes. Portanto, pode ser compartilhado pela academia, e outros, para desenvolver e raciocinar sobre métodos de detecção de fraudes.

Os dados sintéticos têm o benefício adicional de serem mais fáceis de serem adquiridos, com mais agilidade e menor custo para experimentação, mesmo para quem tem acesso aos seus próprios dados.

O sistema foi executado por 180 etapas (ou seja, simulando 180 dias e aproximadamente seis meses), e os parâmetros foram calibrados para obter uma distribuição próxima o suficiente para ser confiável para teste. Dentre vários testes, a base final é composta por um conjunto de dados com valores que exprimem consistência e apresentam dados mais precisos.

Na simulação dos dados foi acrescentado um cenário contendo ladrões que visam roubar em média três cartões por etapa e realizar cerca de duas operações fraudulentas por dia. Dessa forma, foram produzidas 594.643 observações no total, sendo 587.443 referentes a pagamentos normais e 7.200 a operações fraudulentas. Uma vez que esta é uma simulação aleatória, os valores obviamente não são idênticos aos dados originais.

Os dados da amostra foram disponibilizados em uma tabela no formato CSV, contendo 10 colunas de informações.

4.1.4 Ferramentas de análise

A análise e tratamento dos dados foi realizada por meio do banco de dados PostgreSQL e linguagem SQL, software Jupyter utilizando-se da linguagem Python, além do Microsoft Excel. Após tratamento, a modelagem das redes foi realizada por meio do software Gephi.

Gephi é um software de código aberto, distribuído gratuitamente e serve para visualização, além de análises e manuseamento de redes e grafos. Originado em 2006 com um protótipo denominado Graphiltre, que foi construído pois seu criado estava insatisfeito com resultados de softwares livres da época e as ferramentas pagas eram inacessíveis. Na internet encontram-se ainda diversos tutoriais que auxiliam na execução da criação e interpretação das redes.

4.2 Métodos

4.2.1 Critérios de inclusão e exclusão

Com relação aos dados, observou-se na base a não presença de informações faltantes, ou seja, a base utilizada não continha *missings*. No entanto, foi verificado que as informações de *ZipCode*, tanto dos clientes, quanto das empresas, tinham o mesmo código. Logo, essas informações foram descartadas, pois não acrescentariam nenhuma informação relevante e nenhuma correlação. Ou seja, não discriminariam os dados.

4.2.2 Análise Exploratória e Descritiva dos Dados

O Quadro 1 apresenta as variáveis da tabela de dados e sua respectiva descrição.

Com relação às informações dos clientes, foi verificado que a base é composta por 4.112 clientes distintos, e que nos 180 dias foram efetuadas 594.643 operações de pagamentos em cartão de crédito.

Em uma análise geral das operações que compõe a base, verifica-se que nos 180 dias foram observadas 594.643 operações que totalizaram 7.200 fraudes e um montante gasto total

Quadro 1 – Informações das variáveis constante na Base de Dados estudada.

Variável	**Descrição**	**Domínio da variável**
step	Passo ou a simulação do dia que foi efetuada a transação com cartão	Varia de 0 a 179
customer	Cliente que provavelmente efetuou a transação	São 4.112 clientes representados por códigos distintos
age	"Idade" do cartão do cliente	A idade do cartão varia de 0 a 6 e U
gender	Gênero do cliente	M=masculino – F=feminino – O=outros
zipcodeOri	Identificação postal do cliente	Representado pelo código único 28007
merchant	Empresa	50 empresas representadas por códigos distintos
zipMerchant	Identificação postal da empresa	Representado pelo código único 28007
category	Categoria de atuação da empresa onde foi efetuada a transação	15 categorias distintas
amount	Valor da transação	Mínimo 0 – Máximo $8.329,96
fraud	valor binário de identificação da fraude	0 ou 1

Fonte: Própria

de $22.531.103,73, sendo que desse valor, $3.822.671,71 são referentes a operações fraudadas, cerca de 20% do valor total.

O dia, ou *step*, que contém o máximo de observações na série, foi o *step* 175, com 3.774 operações. O dia que se observou a menor quantidade de operações foi no *step* 1, com 2.424 operações. Monetariamente, o *step* que se observou o maior gasto conjunto foi o de número 152, com valor de $151.773,23 gastos; já no *step* 0, o valor observado foi de $92.563,27.

A operação que apresentou o valor máximo de fraude foi executada no *step* 158, no valor de $8.329,96 e a operação que apresentou o menor valor de fraude foi no *step* 159, no valor de somente $0,03.

Com relação às fraudes, estas estão distribuídas uniformemente, sendo que das 7.200 observações em 180 dias acabam por perfazer um total de 40 fraudes diárias.

Com relação aos clientes, tem-se que dos 4.112 clientes distintos, 1.483 (36,03%) tiveram pelo menos 1 operação marcada como fraude na base. O cliente com maior número de fraudes observadas foi o C1350963410 (*gender* = F e *age* = 5) com 144 operações fraudadas, das 191 operações que efetuou ao longo do período, sendo que dos $56.217,36 gastos, $54.274,78 foram referentes a operações fraudadas. Ou seja, 75% das operações efetuadas foram fraudadas e 96,54% do valor gasto por esse cliente também se refere a operações fraudadas.

O cliente com maior valor observado gasto em operações fraudadas foi o C806399525 (*gender* = F e *age* = 2), totalizando $80.324,04 (do total de $83.755,49, ou seja, 95,9%) num total de 125 operações fraudadas, dentro do universo de 237 operações efetuadas, ou seja,

52,74

Analisando as informações no geral, com relação ao gênero dos clientes, observa-se que aproximadamente 55% são do gênero feminino e o restante se dividem em gênero masculino e outros, como indica a Tabela 1, que traz as informações da quantidade de clientes distribuídos por gênero.

Tabela 1 – Quantidade de clientes distribuídos por gênero.

Gender	Quantidade	Percentual
F	2.256	54,86%
M	1.844	44,84%
Outros	12	0,29%
Total	4.112	100%

Fonte: Própria

Ainda sobre os clientes, associando às informações da totalidade das compras no período observado, observa-se que a proporcionalidade no número das operações se mantém, com relação ao gênero, sendo os clientes do gênero feminino os que mais executaram operações ao longo dos 180 dias, cujos números podem ser observados na Tabela 2.

Tabela 2 – Quantidade de operações efetuadas ao longo do tempo pelos clientes e distribuídos por gênero.

Gender	Quantidade Operações	Percentual
F	324.565	54,58%
M	268.385	45,13%
Outros	1.693	0,28%
Total	594.643	100%

Fonte: Própria

Com relação às fraudes, a proporcionalidade se distingue, sendo que as fraudes se concentram mais nos clientes do sexo feminino, conforme os dados da Tabela 3.

O mesmo ocorre quando são analisados os dados com relação aos valores de operações fraudadas, segundo a Tabela 4 a proporção no valor agregado de operações fraudadas se mantém em relação a quantidade de operações fraudadas.

Analisando o tempo de existência do cartão, verifica-se que mais de 50% dos cartões dos clientes possuem entre 2 e 3 anos e 33,73% apresentam mais de 4 anos. Apenas pouco mais de 10% são considerados "novos", com idade até 1 ano. Os números correspondentes ao tempo de existência dos cartões podem ser observados na Tabela 5.

Tabela 3 – Quantidade de fraudes observadas ao longo do tempo e distribuídas por gênero.

Gender	Quantidade Fraudes	Percentual
F	4.758	66,08%
M	2.435	33,82%
Outros	7	0,10%
Total	7.200	100%

Fonte: Própria

Tabela 4 – Valor das transações observadas divididas pelo gênero e fraude observados.

Gender	Fraude			
	0	%	1	%
F	10.223.626	54,65%	2.503.556	65,49%
M	8.428.747	45,05%	1.315.801	34,42%
Outros	56.060	0,30%	3.314	0,09%
Total	18.708.433	100%	3.822.671	100%

Fonte: Própria

Tabela 5 – Quantidade de cartões distintos distribuídos por sua idade.

Age	Quantidade	Percentual
0	22	0,54%
1	403	9,80%
2	1.291	31,40%
3	1.024	24,90%
4	757	18,41%
5	424	10,31%
6	182	4,43%
U	9	0,22%
Total	4.112	100%

Fonte: Própria

Com relação as compras efetuadas pelos clientes com os cartões, a Tabela 6 traz as informações referentes a proporcionalidade de compras que se mantém em relação a idade do cartão.

Por outro lado, pode-se observar que com relação às fraudes, estas ocorrem em cartões mais novos, sendo que quase 67% ocorrem em cartões com até 3 anos de existência, conforme dados observados na Tabela 7.

Analisando as informações com relação ao montante gasto, pode-se observar por meio da Tabela 8, que os cartões entre 2 e 5 anos concentram 85% do montante gasto, e 86% das

Tabela 6 – Quantidade de operações efetuadas ao longo do tempo pelos clientes e distribuídos pela idade do cartão.

Age	Quantidade	Percentual
0	2.452	0,41%
1	58.131	9,78%
2	187.310	31,50%
3	147.131	24,74%
4	109.125	18,33%
5	62.642	10,53%
6	26.774	4,50%
U	1.178	0,20%
Total	594.643	100%

Fonte: Própria

Tabela 7 – Quantidade de fraudes observadas ao longo do tempo e distribuídas pela idade do cartão.

Age	Quantidade	Percentual
0	48	0,67%
1	689	9,57%
2	2.344	32,56%
3	1.755	24,38%
4	1.410	19,58%
5	686	9,53%
6	261	3,63%
U	7	0,10%
Total	7.200	100%

Fonte: Própria

operações fraudadas também estão concentradas nessas faixas.

Analisando as informações sob a ótica do fornecedor, verifica-se que das 50 empresas distintas, estas se dividem nas seguintes categorias, conforme a Tabela 9.

Com relação às empresas, observa-se que das 50 empresas distintas, 30 (60%) tiveram pelo menos 1 operação marcada como fraude na base. A empresa com maior número de fraudes observadas foi a M480139044 (*category = "health"*), com 1.634 operações fraudadas, das 3.508 operações que efetuou ao longo do período, sendo que dos $858.388,22 efetuados em compras, $664.804,39 foram referentes a operações fraudadas. Ou seja, 45% das operações efetuadas, foram fraudadas e 77,44% do valor gasto nessa empresa também se refere a operações fraudadas. A empresa com maior valor observado recebido em operações fraudadas foi a M732195782 (*category = "travel"*), totalizando $1.350.979,31 (do total de $1.413.661,65, ou seja 95,5%) num total de 518 operações fraudadas, dentro do universo de 608 operações efetuadas, ou seja,

Tabela 8 – Valor das transações observadas divididas pela idade do cartão e fraude observados.

Age	Fraude			
	0	%	1	%
0	82.721	0,44%	31.549	0,83%
1	1.837.376	9,82%	344.333	9,01%
2	5.890.537	31,49%	1.294.386	33,86%
3	4.638.868	24,80%	934.412	24,44%
4	3.428.336	18,33%	736.585	19,27%
5	1.946.071	10,40%	335.741	8,78%
6	844.690	4,52%	142.350	3,72%
U	39.833	0,21%	3.314	0,09%
Total	18.708.433	100%	3.822.671	100%

Fonte: Própria

Tabela 9 – Quantidade de empresas distintas distribuídos por sua categoria de atuação.

Category	Quantidade	Percentual
es_barandrestaurants	1	2,00%
es_contents	2	4,00%
es_fashion	3	6,00%
es_food	1	2,00%
es_health	5	10,00%
es_home	5	10,00%
es_hotelservices	7	14,00%
es_hyper	1	2,00%
es_leisure	2	4,00%
es_otherservices	1	2,00%
es_sportsandtoys	6	12,00%
es_tech	3	6,00%
es_transportation	2	4,00%
es_travel	4	8,00%
es_wellnessandbeauty	7	14,00%
Total	50	100%

Fonte: Própria

85,2%.

A categoria de empresas que concentra o maior número de operações observadas é a *"transportation"*, com aproximadamente 85% das operações, cuja informação pode ser visualizada por meio da Tabela 10.

Por outro lado, observa-se na Tabela 11, que com relação às fraudes observas por categoria, tem-se que mais de 50% das operações foram efetuadas em empresas caracterizadas como de *"health"* e *"sportsandtoys"*.

Tabela 10 – Quantidade de transações distintas distribuídos por categoria de atuação da empresa.

Category	Quantidade Operações	Percentual
es_barandrestaurants	6.253	1,06%
es_contents	885	0,15%
es_fashion	6.338	1,08%
es_food	26.254	4,47%
es_health	14.437	2,46%
es_home	1.684	0,29%
es_hotelservices	1.196	0,20%
es_hyper	5.818	0,99%
es_leisure	25	0,00%
es_otherservices	684	0,12%
es_sportsandtoys	2.020	0,34%
es_tech	2.212	0,38%
es_transportation	505.119	85,99%
es_travel	150	0,03%
es_wellnessandbeauty	14.368	2,45%
Total	587.443	100%

Fonte: Própria

Curiosamente, a categoria *"transportation"*, em que se observa o maior número de operações efetuadas e com o maior montante observado ao longo do tempo, é a categoria que não apresenta nenhuma operação de fraude, conforme os dados na Tabela 12. A categoria que apresenta a maior concentração do montante fraudado é a *"travel"* com 40,23% do valor das fraudes.

As informações de *"zipcodeOri"* e *"zipMerchant"* foram descartadas de quaisquer análises, pois todo o conjunto de dados continha um valor único para essas variáveis, o que não acrescentaria e nem auxiliaria a discriminar nenhuma das conclusões.

Outra informação que necessariamente foi descartada foi o estudo de sazonalidade, que serviria de verificação se existia algum período em que se ocorresse mais fraudes em função de outros. A base sintética gerou a quantidade de 40 fraudes para cada um dos *steps* observados.

A questão de desbalanceamento dos dados também é um fator relevante, pois num universo grande de operações, as de fraude representam somente 1,21%.

Tabela 11 – Quantidade de transações fraudadas distintas distribuídos por categoria de atuação da empresa.

Category	Quantidade Fraudes	Percentual
es_barandrestaurants	120	1,67%
es_contents	-	0,00%
es_fashion	116	1,61%
es_food	-	0,00%
es_health	1.196	23,56%
es_home	302	4,19%
es_hotelservices	548	7,61%
es_hyper	280	3,89%
es_leisure	474	6,58%
es_otherservices	228	3,17%
es_sportsandtoys	1.982	27,53%
es_tech	158	2,19%
es_transportation	-	0,00%
es_travel	578	8,03%
es_wellnessandbeauty	718	9,97%
Total	7.200	100%

Fonte: Própria

Tabela 12 – Valores das transações fraudadas distintas distribuídas por categoria de atuação da empresa.

Category	Valor Transações x Valor Fraude			
	0	%	1	%
es_barandrestaurants	257.286	1,38%	19.691	0,52%
es_contents	39.425	0,21%		0,00%
es_fashion	395.160	2,11%	28.653	0,75%
es_food	973.246	5,20%		0,00%
es_health	1.497.654	8,01%	690.325	18,06%
es_home	190.862	1,02%	138.160	3,61%
es_hotelservices	127.432	0,68%	231.159	6,05%
es_hyper	232.936	1,25%	47.392	1,24%
es_leisure	1.831	0,01%	142.336	3,72%
es_otherservices	51.769	0,28%	72.155	1,89%
es_sportsandtoys	178.776	0,96%	684.517	17,91%
es_tech	221.033	1,18%	65.613	1,72%
es_transportation	13.617.092	72,79%		0,00%
es_travel	100.354	0,54%	1.537.944	40,23%
es_wellnessandbeauty	823.577	4,40%	164.725	4,31%
Total	18.708.433	100%	3.822.671	100%

Fonte: Própria

5 RESULTADOS E ANÁLISES

Neste capítulo serão descritos os resultados observados dado a aplicação da Teoria das Redes nos dados de operações de cartão de crédito.

5.1 Características da Rede

Para a modelagem da rede, foram executadas algumas alterações nos dados disponíveis tais como, agrupamento das informações independente do dia da operação. Pontuação das características da operação de maneira a se obter um peso de relevância, ou seja, o peso maior torna maior a probabilidade de que uma operação com as mesmas características seja uma fraude.

Dado isso, a rede foi constituída com as seguintes características:

- Tipo: Direcionada
- Nós: clientes e estabelecimentos comerciais
- Arestas: execução de uma compra
- Peso das arestas: características que podem influenciar na fraude a partir dos estudos de análise descritiva da base – regra própria.

5.2 Layout da Rede

O software Gephi é composto por uma seleção de algoritmos de distribuição, utilizados para a visualização dos dados na forma de redes.

Para os dados que estão sendo estudados, dentre os algoritmos disponibilizados, tem-se que o ForceAtlas2 seria um dos mais adequados para este contexto, pois tem a particularidade de considerar o peso das arestas. Também uma boa visualização pode ser obtida pelo algoritmo Fruchterman-Reingold, que evita a sobreposição dos nós, fazendo com que se tenha uma melhor visualização geral da rede.

Os grafos direcionados utilizam *layouts* desenvolvidos de forma a ilustrar esses resultados e de uma forma esteticamente agradável, dispondo os nós em duas ou até três dimensões. Essas disposições tendem a evitar o cruzamento das ligações ao máximo possível de maneira a facilitar a interpretação dos resultados.

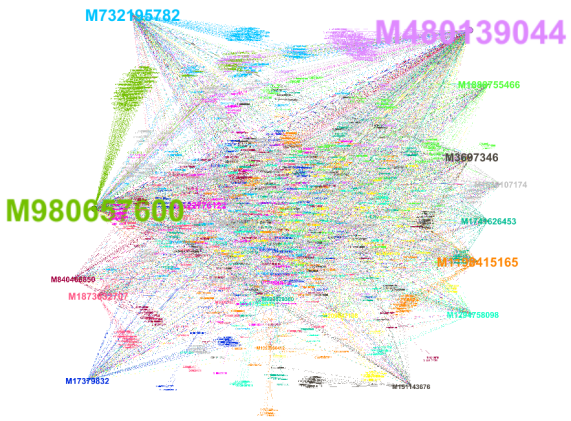
Um dos *plugins* desenvolvidos para o Gephi e que foram aqui utilizados é o ForceAtlas2 e tem por objetivo de desenhar um grafo mais compreensível para o usuário. Esse modelo segue um modelo orientado pela força em que as ligações se atraem e os nós se repulsam mutuamente. Para refinar o resultado foi esse *plugin* utilizado combinando seu resultado com o *plugin* Fruchterman-Reingold que consiste na distribuição dos vértices de maneira igual no espaço que tem disponível, minimiza os cruzamentos de arestas e uniformiza os seus tamanhos, e por fim também proporciona uma simetria ao grafo analisado.

5.3 Tratamento da Rede

Para a modelagem da rede, foi necessário utilizar uma amostra dos dados obtidos, em virtude da incapacidade de processamento de todos os dados disponibilizados pela base sintética do BankSim.

Inicialmente uma rede foi gerada e traz a informações das operações fraudadas e sem peso das arestas. Para a execução, foi calculado o valor da modularidade, ou seja, a indicação de grupos com caraterísticas semelhantes, nesse caso, indicando uma sensibilidade maior para ocorrências de fraudes. O cálculo se dá com base na proximidade maior entre os nós, e apresentam uma dependência maior entre eles quando se compara com os nós vizinhos. Para a identificação de possíveis *clusters* na rede, a coloração também foi aplicada por modularidade. Como resultado um grafo foi obtido com indicação de presença de três *clusters* principais.

Figura 7 – Grafo com coloração por modularidade, com indicação de três principais *clusters*.



Fonte: Própria

Para o cálculo do peso das arestas, foi levado em consideração uma regra própria que combinou as informações mais relevantes para a caracterização da fraude.

Para a regras de pontuação considerou-se o percentual observado de cada informação, sendo:

- Tempo de Cartão
  - 0 a 2 anos – 42,8 pontos
  - 3 anos – 24,4 pontos
  - 4 anos – 19,2 pontos
  - 5 anos – 8,8 pontos
  - 6 anos – 3,7 pontos
  - Outros – 0 pontos
- Sexo do cliente
  - Feminino – 65 pontos
  - Masculino – 35 pontos

  - Outros – 0 pontos
- Categoria da Empresa
  - Sportsandtoys – 28 pontos
  - Health – 24 pontos
  - Wellnessandbeauty – 10 pontos
  - Travel - 8 pontos
  - Hotelservices - 7,6 pontos
  - Leisure - 6,58 pontos
  - Home – 4 pontos
  - Hyper - 3,8 pontos
  - Otherservices - 3,1 pontos
  - Tech – 2 pontos
  - Barsandrestaurants - 1,6 pontos
  - Fashion - 1,6 pontos
  - Transp – 0 pontos
  - Contents – 0 pontos
  - Food – 0 pontos

Após a marcação da pontuação, o valor final do peso foi composto pela multiplicação pelo valor da operação, para que o peso pudesse também refletir essa informação. Transações de valores mais baixos, pode ficar dentro de uma margem de segurança para perdas operacionais e não ser um problema para a Instituição Financeira, por outro lado operações com valores elevados, devem ter mais celeridade de identificação pois podem acarretar perdas operacionais que impactam diretamente dos resultados das instituições.

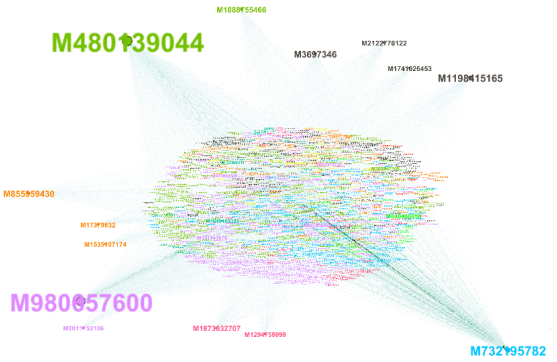
Para a composição do grafo da Figura 8, observa-se os comércios mais impactados. Também foi utilizado o cálculo da modularidade para a coloração, no entanto o tamanho dos títulos se deu por meio do grau de entrada entre os nós, ou seja, quais as empresas que têm mais operações fraudulentas dentre todas.

Para a composição do grafo da Figura 9, observa-se os clientes mais impactados. Também foi utilizado o cálculo da modularidade para a coloração, no entanto o tamanho dos títulos se deu por meio do grau de saída entre os nós, ou seja, quais os clientes que apresentaram mais operações fraudulentas dentre todas.

5.4 Análise das características da rede

Para analisar os nós mais relevantes, foram calculados os valores de PageRank, sendo que esse algoritmo analisa a importância do nó tanto pela qualidade quanto pela quantidade de links que possui dentro da rede. Como exemplo, uma pessoa X se conecta com 300 pessoas e uma pessoa Y se conecta com 50 pessoas. Numa análise imediata pode-se dizer que a pessoa X tem mais importância por apresentar mais relações, no entanto, a pessoa Y se conecta com alguém muito influente como o Governador de seu estado, dessa forma, levando essa

Figura 8 – Grafo com tamanho dos rótulos por grau de entrada.



Fonte: Própria

Figura 9 – Grafo com tamanho dos rótulos por grau de saída.

Fonte: Própria

informação em consideração, seu PageRank será maior.

Contudo, foi analisado o PageRank dos 3 nós que mais se destacaram nas figuras Figura 7 e Figura 8, coincidentemente são as mesmas empresas. A Tabela 13 a seguir, mostra as principais informações dessas empresas.

Tabela 13 – Informações dos 3 nós das empresas mais relevantes na rede.

Empresa	PageRank	Transações Totais	Transações Fraudulentas	% transações fraudulentas	Valor Total	Valor Fraudes	% valor fraudulento
M480139044	0,08677	3.508	1.634	46,6%	858.388,22	664.804,39	77,45%
M980657600	0,08289	1.769	1.472	83,2%	530.635,69	505.311,62	95,23%
M732195782	0,034829	608	518	85,2%	1.408.673,34	1.350.979,31	95,90%
Total		5.885	3.624		2.797.697,25	2.521.095,32	

Fonte: Própria

Dentre o universo de 7.200 operações fraudulentas na base sintética estudada, observa-se que as 3 empresas mais representativas acumulam mais de 50% das operações fraudadas. Observa-se também que mais de 68% referente ao valor das fraudes estão concentrados nessas 3 empresas, além de todas as operações estarem diluídas nos 180 *steps* da base.

Por outro lado, analisando o grafo obtido com as informações dos clientes, foi delimitado por meio do grau de saída, os 5 primeiros clientes mais significativos para análise.

Tabela 14 – Informações dos 5 nós dos clientes mais relevantes na rede.

Cliente	Grau de Saída	Transações Totais	Transações Fraudulentas	% transações fraudulentas	Valor Total	Valor Fraudes	% valor fraudulento
C806399525	9.741.071	237	125	52,7%	83.755,49	80.324,04	95,90%
C2004941826	7.932.876	126	119	94,4%	64.751,08	64.165,54	99,10%
C1849046345	7.211.853	171	127	74,3%	60.693,04	59.066,72	97,32%
C1572610482	7.071.376	101	89	88,1%	58.687,13	58.203,13	99,18%
C1350963410	6.578.861	191	144	75,4%	56.217,36	52.947,31	94,18%
Total		826	604		324.104,10	314.706,74	

Fonte: Própria

Dentre o universo de 7.200 operações fraudulentas na base sintética estudada, observa-se que os 5 clientes mais representativos acumulam pouco mais de 8% das operações fraudadas. Observa-se também que pouco mais de 8% referente ao valor das fraudes estão concentrados nesses 5 clientes, além de todas as operações estarem diluídas nos quase 180 *steps* da base.

5.5 Definição da regra para identificar uma possível fraude

Após os estudos realizados, e ainda com o apoio dos grafos obtidos, é possível definir uma regra para apontamento de possíveis fraudes sendo realizadas, de modo a emitir um alerta.

Considerando a retirada dos *outliers* como sendo as 3 empresas que se destacavam no universo das operações fraudadas, ainda restaram 3.576 operações para análise.

De fato, as características referentes a idade do cartão e sexo não se alteraram, sendo que cartões com até 2 anos e pertencentes a indivíduos do sexo feminino já acendem um alerta, conforme Tabelas Tabela 15 e Tabela 16, a seguir.

Tabela 15 – Quantidade de fraudes observadas ao longo do tempo e distribuídas por idade do cartão, após retirada de outiliers.

Age	Quantidade Fraudes	% Fraudes
0	23	0,64%
1	355	9,93%
2	1.120	31,32%
3	870	24,33%
4	709	19,83%
5	360	10,07%
6	133	3,72%
Outros	6	0,17%
Total	3.576	100%

Fonte: Própria

Tabela 16 – Quantidade de fraudes observadas ao longo do tempo e distribuídas por sexo, após retirada de outiliers.

Gender	Quantidade Fraudes	% Fraudes
F	2.387	66,75%
M	1.183	33,08%
Outros	7	0,17%
Total	3.576	100%

Fonte: Própria

Outro ponto a ser considerado são operações realizadas em estabelecimentos pertencentes a categoria "es_wellnessandbeauty", pois com a retirada dos *outliers* essa categoria se sobressaiu no restante das informações.

Tabela 17 – Quantidade de fraudes observadas ao longo do tempo e distribuídas por categoria da empresa, após retirada de *outliers*.

Category	Quantidade Fraudes	% Fraudes
es_wellnessandbeauty	718	20,08%
es_hotelservices	548	15,32%
es_sportsandtoys	510	14,26%
es_leisure	474	13,26%
es_home	302	8,45%
es_hyper	280	7,83%
es_otherservices	228	6,38%
es_tech	158	4,42%
es_barandrestaurants	120	3,36%
es_fashion	116	3,24%
es_health	62	1,73%
es_travel	60	1,68%
Total	3.576	100%

Fonte: Própria

Considerando esses pontos, nesta base de dados com as informações disponíveis, seria de se emitir um alerta, de forma a tentar comunicação com o cliente para confirmação, nas operações realizadas por clientes do sexo feminino, que possuem cartões de até 2 anos, efetuando compras em estabelecimentos da categoria "es_wellnessandbeauty", "es_hotelservices", "es_sportsandtoys" e "es_leisure". Essa regra atingiria 4.849 operações do total, sendo que abrangeria 664 operações fraudulentas, o que corresponde a 18,5% do total.

Em termos monetários, haveria uma economia de $21.106,37, o correspondente a 1,6% do valor observado referente as operações fraudadas.

Por outro lado, a construção de uma regra que leve em consideração o valor das operações fraudulentas, sendo que para isso considera-se um valor material acima de $100,00 por operação, restando um universo de 2.951 operações para análise.

Por essa regra a questão do alerta se mantém para clientes do sexo feminino, que possuem cartões de até 2 anos, mas abrangendo um rol maior de categorias das empresas sendo, além das já elencadas "es_wellnessandbeauty", "es_hotelservices", "es_sportsandtoys" e "es_leisure", acrescentando ainda as categorias "es_home" e "es_travel".

A regra proposta tem um impacto de emissão de alerta para 6.040 operações do total de 588.758 operações, a menos dos *outliers* retirados, representando pouco mais de 1% da base. Em termos financeiros, o alerta pode reduzir o valor das perdas em $280.954,19, o que representa 21,6% de redução nas perdas.

6 CONCLUSÃO

Após a execução de todo o estudo e levando em consideração a vasta presença de estudos que abordam sobre o tema de fraudes em cartão de crédito, tem-se que esse é um problema que merece atenção, dado que a questão de fraude em cartão de crédito já vem de longa data e é um problema persistente ao longo do tempo, mudando apenas as suas características, e que ainda deve permanecer sob os holofotes das instituições financeiras. Esse tipo de problema é persistente, pois vai se moldando ao longo do tempo em conjunto com as mudanças de tecnologia de proteção contra fraude das operadoras de cartão e instituições financeiras.

Por não ser possível o acesso a bases de *Open Banking*, que contam com a informação do cliente em diversas instituições financeiras, mostrando de forma completa todo o seu consumo no sistema financeiro nacional, o estudo foi limitado, pois a base sintética utilizada não continha muitos insumos, que seriam importantes para uma análise mais refinada e assertiva a respeito dos dados. Informações como localidade, data e horário da operação, bandeira do cartão, número do cartão para verificação se o cartão é o mesmo ao longo do tempo, tipo de cartão (físico / virtual), meio de pagamento (internet / físico), instituição financeira credora, etc. Com a agregação de todas essas informações, é possível refinar ainda mais as regras de indicação da fraude, diminuindo o custo da emissão de alerta de fraude.

Entretanto, após as análises a partir do grafo obtido, conjuntamente com os estudos das informações estatística dos dados disponibilizados na base sintética, pode-se concluir alguns pontos:

- As empresas M480139044, M980657600 e M73219578 apresentam indícios de serem coparticipantes na aplicação das fraudes, pois a quantidade de operações fraudadas nessas empresas totaliza mais de 50% das operações fraudulentas da base, além de mais de 68% do valor total observado referente a essas operações fraudulentas. Na verdade, essas 3 empresas podem ser consideradas um *outlier* e serem retiradas do estudo, pois elas descaracterizam as regras para as outras empresas pontualmente. Um forte indício que as empresas podem ser comparsas na aplicação, é que as operações fraudulentas estão diluídas ao longo do período objeto do estudo.
- O mesmo ocorre sob a perspectiva dos clientes, sendo que tem-se 5 clientes que se destacam, acumulando muitas fraudes num longo período de tempo. Com isso, é possível concluir que ou a instituição financeira não está sendo assertiva e diligente na identificação das fraudes desses clientes, ou que esses clientes estão de alguma forma aplicando as chamadas auto fraudes.

Outro ponto que não pode ser deixado de ser observado é a avaliação sob a ótica do consumo do cliente. Esse ponto deve continuar sendo analisado e um alerta também deve ser emitido quando o padrão da compra for muito fora do que se observa rotineiramente.

Observa-se ao longo de toda a pesquisa que o desbalanceamento entre operações normais é um problema a ser considerado. O universo de operações fraudulentas dentro do universo total é muito pequeno na maioria esmagadora dos casos. Isso dificulta a verificação dentre um universo tão grande de operações. Um ponto a ser estudado futuramente poderia ser a quebra da base em partes que façam sentido de ser analisadas, de modo a verificar regras de identificação por cada parte. Por exemplo, se for identificado que certa região tem características de fraudes distintas de outras regiões, isso pode indicar que a aplicação do golpe poderia estar sendo feita por núcleos distintos também. Dessa forma, a regra não seria a mesma para cada partição e com isso seria possível ter uma previsão melhor da indicação das fraudes. Mas para isso, a base teria que conter mais informações de forma a ser possível realizar quebras relevantes para o estudo.

Além disso, como essa situação é dinâmica, e a tempestividade necessária, há de ser necessário uma revisitação das regras em curtos períodos, para verificação se a acurácia do modelo se mantém, ou se necessário for uma recalibragem de todo o sistema. Nesse contexto, o sistema deve ser simples de modo a não ter problemas de demora em sua implementação.

6.1 Limitações

Infelizmente no momento da execução desse trabalho não foi possível obter dados de *Open Banking* para os estudos, contudo, o objeto do estudo continuará a ser estudado, dado que o tema que é vasto, é um problema para as instituições financeiras de longa data e é uma dificuldade ainda sem resolução, de forma que os artifícios de aplicação de fraude se modificam em conjunto com as novas tecnologias empregadas para proteção dos cartões de crédito.

Como uma das maiores limitações se refere à base de dados, pois foi utilizado no trabalho uma base de dados com informações sintéticas, a base era constituída por poucas variáveis, que nesse caso seria essencial para se traçar uma regra para o apontamento de uma possível operação fraudulenta. Dentre as informações existentes, duas variáveis precisaram ser descartadas pois continham a mesma informação para todo universo de operações, sendo que nesse caso a informação não seria discriminante.

Outro ponto a ser considerado como uma limitação foi o processamento de um grande volume de dados, necessário para a execução do trabalho. Inicialmente a proposta era de se trabalhar com o *software* Neo4j, no entanto sua utilização foi inviável diante dos recursos disponíveis para o estudo. Para o processamento dos dados por meio do *software* Gephi, foi necessário restringir as operações somente no universo de operações fraudulentas, para que fosse possível gerar as informações.

6.2 Trabalhos Futuros

Como trabalho futuro, tem-se como expectativa a possibilidade de trabalhar com uma base de dados de informações reais, de forma a ser uma base mais completa no sentido de

informações e possível traçar um perfil, ou uma regra, mais assertiva. Informações diversas como localidade da compra, meio de pagamento, horário da operação, são informações preciosas para se traçar o perfil de uma operação fraudulenta.

Também, para um trabalho futuro, sugere-se a atuação com demais softwares de geração de grafos, de modo a se comparar os resultados obtidos e métricas de desempenho.

6.3 Considerações Finais

Ao longo do trabalho, incluindo a revisão da bibliografia, foi possível observar que o problema de fraudes em cartões de crédito foram, são e continuarão a ser um problema que necessita de tempestividade e evoluções constantes.

Os mecanismos de fraudes acompanham as tecnologias de proteções desenvolvidas para esse meio de pagamento. Desde sempre, para cada tecnologia de prevenção a fraude, existe a criação de um mecanismo de burlar esses sistemas.

Diversos trabalhos foram desenvolvidos ao longo do tempo de maneira a tentar precaver essas operações, utilizando-se de modelos estatísticos e ao longo do tempo de *machine learning*. No entanto, na revisão da bibliografia, não foi localizado nenhum trabalho que abordasse o problema por meio de grafos.

Referências

ABECS. **Segurança e Prevenção a Fraudes**. [S.l.], 2022. Disponível em: <https://files.abecs.org.br/revista/023/Revista_Abecs_023/index.html#page=18>. Acesso em: 23 de abril de 2022. Citado na página 16.

ACADEMY, K. **Definição de Grafo**. [S.l.], 2022. Disponível em: <https://pt-pt.khanacademy.org/math/macs-11-ano/xab679065dfe43c0e:modelos-matematicos/xab679065dfe43c0e:modelos-de-grafos/a/describing-graphs>. Acesso em: 30 de abril de 2022. Citado na página 22.

ACCENTURE. **Impactos do Open Banking e PIX no Brasil**. [S.l.], 2021. Disponível em: <https://www.accenture.com/br-pt/insights/banking/muito-alem-dos-bancos>. Acesso em: 23 de abril de 2022. Citado na página 19.

AMIN, A.; THRIFT, N. Neo-marshallian nodes in global networks. **International Journal of Urban and Regional Research**, n. 16, p. 571–587, 1992. Citado na página 20.

ANDERSON, B. D. O.; VONGPANITLERD, S. **Network analysis and synthesis: a modern systems theory approach**. New York: Dover Publications Inc., 2006. Citado na página 20.

BACEN. **Open Banking**. [S.l.], 2022. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/openbanking>. Acesso em: 12 de março de 2022. Citado 3 vezes nas páginas 13, 17 e 18.

BARABASI, A.-L. The new science of networks. **J. Artificial Societies and Social Simulation**, v. 6, 2003. Citado na página 22.

BROWN, D. G. et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. **Cancer & Metabolism**, v. 11, n. 4, p. 1–12, 2016. Citado na página 20.

CARVALHO, A. et al. Crimes e fraudes eletrônicos: Perspectiva de ações empresariais adotadas por instituições financeiras. **Cadernos de Estudos Sociais**, v. 30, n. 1, 2015. Citado na página 27.

COELHO, L. S.; RAITTZ, R. T.; TREZUB, M. Fcontrol®: Sistema inteligente inovador para detecção de fraudes em operações de comércio eletrônico. **Gestão & Produção**, v. 13, n. 1, p. 129–139, 2006. Citado na página 25.

DIEBOLD, F. X.; YILMAZ, K. On the network topology of variance decompositions: Measuring the connectedness of financial firms. **J. Econom.**, v. 182, n. 1, p. 119–134, 2014. Citado na página 22.

ERDOS, P.; RENYI, A. On the evolution of random graphs. **PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES**, p. 17–60, 1960. Citado na página 21.

FINSIDERS, R. ao. **Transações financeiras têm 527 mil tentativas de fraudes até junho**. [S.l.], 2022. Disponível em: <https://finsiders.com.br/2022/07/25/transacoes-financeiras-tem-527-mil-tentativas-de-fraudes-ate-junho/>. Acesso em: 23 de abril de 2022. Citado na página 12.

GUTIéRREZ-GóMEZ, L.; BOVET, A.; DELVENNE, J.-C. Multi-scale anomaly detection on attributed networks. **Proceedings of the AAAI Conference on Artificial Intelligence**, v. 34, n. 01, p. 678–685, 2020. Citado na página 17.

HAND, D. J. Pattern detection and discovery. In: HAND, D. J.; ADAMS, N. M.; BOLTON, R. J. (Ed.). **Pattern Detection and Discovery**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 1–12. ISBN 978-3-540-45728-2. Citado na página 16.

KIM, J.; HASTAK, M. Social network analysis: Characteristics of online social networks after a disaster. **International Journal of Information Management**, v. 1, n. 38, p. 86–96, 2018. Citado na página 20.

LIMA, R. A. F.; PEREIRA, A. C. M. Aplicação de técnicas de inteligência computacional para detecção de fraude em comércio eletrônico. **Revista Eletrônica de Iniciação Científica**, v. 12, n. 3, 2012. Citado na página 26.

MELO, C. **Brasil teve mais de 3 milhões de tentativas de fraude no comércio só em 2022**. [S.l.], 2022. Disponível em: <https://mundoconectado.com.br/noticias/v/27095/brasil-teve-mais-de-3-milhoes-de-tentativas-de-fraude-no-comercio-so-em-2022>. Acesso em: 23 de abril de 2022. Citado na página 14.

NEWMAN, M. E. J. **Networks**. Oxford: Oxford University Press, 2018. Citado na página 20.

OLIVEIRA, P. H. M. A. **Detecção de fraudes em cartões: um classificador baseado em regras de associação e regressão logística**. Janeiro de 2016. 103 f. Tese (Mestre em Ciências) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016. Citado na página 12.

PARKHE, A.; WASSERMAN, S.; RALSTON, D. A. New frontiers in network theory development. **Acad. Manag. Rev.**, v. 31, n. 3, p. 560–568, 2006. Citado na página 20.

PORWAL, U.; MUKUND, S. Credit card fraud detection in e-commerce: An outlier detection approach. **ArXiv**, v. 1811.02196, 2018. Disponível em: <https://arxiv.org/abs/1811.02196>. Citado na página 17.

SADALAGE, P.; FOWLER, M. **NoSQL Essencial – Um guia conciso para o mundo emergente de persistência poliglota**. São Paulo: Novatec, 2013. Citado na página 24.

SILVA, T. **Pontos, Linhas e Métricas #05: O Grafo de Konigsberg**. [S.l.], 2013. Disponível em: <https://tarciziosilva.com.br/blog/pontos-linhas-e-metricas-05-o-grafo-de-konigsberg/euler_konigsberg/>. Acesso em: 30 de abril de 2022. Citado na página 21.

SUTTO, G. **Tentativas de fraudes sobem 74% no país e são comuns em compras nas madrugadas de dias úteis**. [S.l.], 2022. Disponível em: <https://www.infomoney.com.br/minhas-financas/tentativas-de-fraudes-sobem-74-no-pais-e-sao-comuns-em-compras-nas-madrugadas-de-dias-uteis-veja-o-que-nao-fazer/>. Acesso em: 15 de março de 2022. Citado na página 17.

TRAVERS, J.; MILGRAM, S. An experimental study of the small world problem. **Sociometry**, v. 32, n. 4, p. 425–443, 1969. Citado na página 21.

WATTS, D. J. The new science of networks. **Annual Review of Sociology**, p. 243–270, 2004. Citado na página 21.

WAZLAWICK, R. S. **Metodologia de Pesquisa para Ciência da Computação**. Rio de Janeiro: Elsevier, 2009. Citado na página 29.