



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
NÍVEL MESTRADO**

MANOEL CAMILO DE SOUSA NETTO

**PREVENÇÃO CRIMINAL
POR MEIO DE GRAFOS E ANÁLISE DE REDES SOCIAIS**

Florianópolis

2019

MANOEL CAMILO DE SOUSA NETTO

**PREVENÇÃO CRIMINAL
POR MEIO DE GRAFOS E ANÁLISE DE REDES SOCIAIS**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Ciência da Informação.

Orientador: Prof. Adilson Luiz Pinto, Dr.

Coorientador: Prof. Thiago Magela Rodrigues Dias, Dr.

Florianópolis

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

de Sousa Netto, Manoel Camilo
PREVENÇÃO CRIMINAL POR MEIO DE GRAFOS E ANÁLISE DE
REDES SOCIAIS / Manoel Camilo de Sousa Netto ; orientador,
Adilson Luiz Pinto, coorientador, Thiago Magela Rodrigues
Dias, 2019.
81 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro de Ciências da Educação, Programa de Pós
Graduação em Ciência da Informação, Florianópolis, 2019.

Inclui referências.

1. Ciência da Informação. 2. Prevenção Criminal. 3.
Teoria dos grafos. 4. Análise de redes sociais. I. Pinto,
Adilson Luiz. II. Rodrigues Dias, Thiago Magela. III.
Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Ciência da Informação. IV. Título.

MANOEL CAMILO DE SOUSA NETTO

PREVENÇÃO CRIMINAL

POR MEIO DE GRAFOS E ANÁLISE DE REDES SOCIAIS

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Emerson Silva Barbosa, Dr.
Academia Nacional de Polícia Federal

Prof. Enrique Muriel Torrado, Dr.
Universidade Federal de Santa Catarina

Prof. Thiago Magela Rodrigues Dias, Dr.
Centro Federal de Educação Tecnológica de Minas Gerais

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Ciência da Informação.

Prof. Edgar Bisset Alvarez, Dr.
Subcoordenador do Programa de Pós-Graduação em Ciência da Informação

Prof. Adilson Luiz Pinto, Dr.
Orientador

Florianópolis, 15 de outubro de 2019

Dedico este trabalho à Ká, à Lourdinha e ao Pedro...as razões da minha existência.
Cada conquista, inclusive esta, é por vocês!

AGRADECIMENTOS

Um agradecimento especial à minha esposa Karla Karine e aos meus filhos Pedro Lucas e Lourdes Manoela pelo amor cotidiano e pela compreensão durante as eventuais ausências necessárias aos estudos. Agradeço aos meus pais Wilson e Socorro pelo seu amor incondicional que me acompanhou não apenas durante o mestrado, mas durante todos os anos de minha vida. Obrigado aos demais amigos e familiares que compreenderam as faltas aos eventos sociais e as mensagens respondidas com atraso.

Um agradecimento especial à minha segunda casa, a Polícia Federal do Brasil (PF), instituição da qual com muito orgulho faço parte. Por meio da Academia Nacional de Polícia (ANP), a PF assinou o acordo de cooperação técnica na área educacional com a UFSC, sem o qual essa conquista não seria possível. Ao agradecer à PF lembro que, ao adentrar pela primeira vez pelos portões da Academia Nacional de Polícia, recém-aprovado no concurso, me deparei com a frase posta em uma placa na entrada daquela instituição: “*Entrada para realização de um sonho*”. Eu diria “...*realização de vários sonhos*”, inclusive o que ora se concretiza. Meu muito obrigado ao Comitê Gestor de Capacitação e aos amigos que nele labutam, vocês fazem a diferença na vida dos servidores que buscam a capacitação como uma forma de serem melhores seres humanos e profissionais. Toda gratidão à Superintendência de Polícia Federal do Piauí, em especial aos servidores do Núcleo de Inteligência Policial, aos Superintendentes, Diretores e demais servidores que sempre me incentivaram desde a minha candidatura ao mestrado até o momento da defesa da dissertação.

Aos servidores e professores do Programa de Pós-Graduação em Ciência da Informação (PGCIN-UFSC), em especial Dr Adilson Luiz Pinto, Dr. Enrique Muriel, Dra. Ana Clara Cândido, Dr. Gustavo Medeiros, Dr. Márcio Matias, Dr. Angel Freddy e Dr. Douglas Dyllon, suas disciplinas foram ministradas com maestria, muito obrigado a todos pelos ensinamentos.

Aos colegas da PF da turma do mestrado, meu muito obrigado por aprendermos juntos, como também pelos bons e divertidos momentos de convívio.

Minha gratidão à Universidade Federal de Santa Catarina, instituição de excelência internacionalmente reconhecida como uma casa onde a ciência prospera. Agradeço ao Dr. Adilson Luiz Pinto, muito mais do que um professor e um orientador, um amigo que responde prontamente as minhas dúvidas, mesmo aquelas enviadas aos fins de semana e feriados. Um professor é um condutor de almas e sonhos e é isso que o senhor representa aos seus alunos e orientandos.

Um agradecimento especial ao bom Deus pelo dom da vida, sem o qual nenhuma realização humana seria possível.

“We all connect, like a net we cannot see.”

Mickenberg e Dugan (2016, p. 127)

RESUMO

A pesquisa ora apresentada, Prevenção Criminal por Meio de Grafos e Análise de Redes Sociais, tem como objetivo propor um modelo protocolar, interdisciplinar, genérico e adaptável de prevenção anticrime que identifique uma Rede Vulnerável (RV) a um Crime-Objeto (CO) por meio da análise de redes sociais e grafos. Como fundamentação teórica, o trabalho utiliza interdisciplinaridade, paradigma cognitivo da Ciência da Informação, aprendizado visual, teoria dos grafos, análise de redes sociais, análise exploratória, banco de dados baseado em grafos, estatística e notação BPMN. Como método, a pesquisa indica a fundamentação policial pela seleção de um Crime-Objeto e determinação de Variáveis Frisadas, Critérios de Vulnerabilidade e Classes Relevantes. Em seguida, o método usa preparação da rede, análises exploratórias e estatísticas, avaliação da Rede Vulnerável. Como resultado, a pesquisa apresenta um fluxograma de policiamento preventivo por meio de grafos e análise de redes sociais, apresentando a aplicação em um estudo de caso que segue todas as etapas propostas no fluxograma.

Palavras-Chave: Prevenção Criminal. Teoria dos grafos. Análise de redes sociais.

ABSTRACT

The research presented here, Criminal Prevention through Graphs and Social Network Analysis, aims to propose a protocol, interdisciplinary, generic and adaptive model of crime prevention that identifies a Vulnerable Network to an Object Crime by through the analysis of social networks and graphs. As theoretical basis, the work uses interdisciplinarity, cognitive paradigm of Information Science, visual learning, graph theory, social network analysis, exploratory analysis, graph database, statistical foundation and BPMN notation. As a method, the research indicates the police grounds for selecting an Object Crime and determining Focused Variables, Vulnerability Criteria, and Relevant Classes. The method then uses network preparation, exploratory analysis and statistics, Vulnerable Network assessment. As a result, the research presents a preventive policing flowchart through graphs and social network analysis, presenting the application in a case study that follows all the steps proposed in the flowchart.

Keywords: Criminal Prevention. Theory of graphs. Social network analysis.

LISTA DE FIGURAS

Figura 1 – Autógrafo social de uma estudante de 18 anos	22
Figura 2 – Diversos tipos de grau de um vértice	27
Figura 3 – Uso do leiaute <i>circle-pack</i>	29
Figura 4 – Uso do leiaute <i>circle-pack</i> em uma rede social.....	29
Figura 5 – Uso do leiaute <i>dual circle</i>	30
Figura 6 – Grafo tratado como um sistema de molas, até atingir equilíbrio	31
Figura 7 – leiaute direcionado por força aplicado a uma rede social	32
Figura 8 – Leiaute geográfico de malha aérea dos Estados Unidos	33
Figura 9 – Rede ego.....	35
Figura 10 – Desempenho superior dos bancos de dados de grafos	35
Figura 11 – Exemplo de Grafo de Modelagem de uma RS.....	44
Figura 12 – RV: uma filtragem da RS	45
Figura 13 – Aplicação do 1º CV [Número de passageiros da reserva = 1]	46
Figura 14 – Modelo de Policiamento Preventivo por meio de Grafos e ARS (em BPMN).....	51
Figura 15 – 23,14% dos reincidentes respondem por 50,89% dos indiciamentos	53
Figura 16 – Rede Vulnerável segundo os CVs das VFs	56
Figura 17 – Grafo de Modelagem de Rede – CO: Desvio de Recursos Públicos	58
Figura 18 – Rede Supergrafo (RS): uma <i>hairball</i>	58
Figura 19 – Leiaute <i>circle-pack</i> aplicado à rede Supergrafo	59
Figura 20 – Leiaute direcionado por força aplicado à Rede Supergrafo	61
Figura 21 – Fornecedores em Geral - percentual de <i>Outliers</i>	65
Figura 22 – Percentual de acusados e não acusados classificados como <i>Outliers</i> em relação a 2ª VF	66
Figura 23 – Percentual de acusados e não acusados classificados como <i>Outliers</i> em relação a 3ª VF	66
Figura 24 – Aplicação do 1º CV: O fornecedor possui sócios indiciados	67
Figura 25 – Aplicação do 2º CV: Fornecedor recebeu mais do que R\$ 496.199,38	68
Figura 26 – Aplicação do 3º CV: Fornecedor foi contratado por mais de 6 (seis) órgãos públicos.....	68
Figura 27 – Rede Vulnerável (RV): uma <i>hairball</i>	69
Figura 28 – Leiaute direcionado por força aplicado a Rede Vulnerável	70
Figura 29 – Leiaute <i>dual circle</i> e comportamento anormal dos fornecedores.....	71
Figura 30 – Composição de leiautes circular e geográfico do estado do Piauí	72

LISTA DE TABELAS

Tabela 1 – VFs e CVs para um tipo de CO	41
Tabela 2 – VFs e CVs para o CO do estudo de caso	53
Tabela 3 – Dados utilizados na pesquisa	57
Tabela 4 – Média e Mediana das VFs.....	63
Tabela 5– Variância, Desvio Padrão e Coeficiente de Variação das VFs	63
Tabela 6 – Quartis, Distância Interquartil e Limites.....	64
Tabela 7 – Números da Rede Supergrafo e da Rede Vulnerável	73

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	PROBLEMATIZAÇÃO E QUESTÃO DE PESQUISA.....	15
1.2	MOTIVAÇÕES E JUSTIFICATIVA	16
1.3	OBJETIVOS	16
1.3.1	<i>Objetivo Geral</i>	<i>17</i>
1.3.2	<i>Objetivos Específicos.....</i>	<i>18</i>
1.4	ORIGINALIDADE DA PESQUISA	18
2	DESENVOLVIMENTO	19
2.1	REFERENCIAL TEÓRICO	19
2.1.1	<i>Interdisciplinaridade.....</i>	<i>19</i>
2.1.2	<i>Paradigma cognitivo da Ciência da Informação</i>	<i>20</i>
2.1.3	<i>Aprendizado visual.....</i>	<i>21</i>
2.1.4	<i>Teoria dos grafos</i>	<i>23</i>
2.1.5	<i>Análise de redes sociais e centralidade de grau.....</i>	<i>24</i>
2.1.6	<i>Análise exploratória de dados</i>	<i>27</i>
2.1.7	<i>Análise exploratória de grafos usando leiautes</i>	<i>28</i>
2.1.8	<i>Bancos de dados baseados em grafos</i>	<i>33</i>
2.1.9	<i>Fundamentação estatística: medidas de centralidade e de dispersão</i>	<i>36</i>
2.1.10	<i>Notação BPMN.....</i>	<i>38</i>
2.2	MÉTODOS E MATERIAIS.....	38
2.2.1	<i>Fundamentação policial: seleção do CO, VFs, CVs e CRs.....</i>	<i>39</i>
2.2.2	<i>Preparação da Rede</i>	<i>42</i>
2.2.3	<i>Análise exploratória das redes</i>	<i>46</i>
2.2.4	<i>Análise estatística das redes</i>	<i>48</i>
2.2.5	<i>Avaliação da Rede Vulnerável</i>	<i>49</i>
3	RESULTADOS.....	50
3.1	FLUXOGRAMA DE POLICIAMENTO PREVENTIVO POR MEIO DE GRAFOS E ARS	50
3.2	ESTUDO DE CASO – APLICAÇÃO DO MODELO PROPOSTO.....	52
3.2.1	<i>Seleção do Crime-Objeto.....</i>	<i>52</i>
3.2.2	<i>Delimitação de VFs, CVs e CRs</i>	<i>52</i>

3.2.3	<i>Motivação do 1º CV: fornecedor possui sócios indiciados</i>	53
3.2.4	<i>Motivação do 2º CV: fornecedor recebeu grande quantia em recursos públicos</i>	54
3.2.5	<i>Motivação do 3º CV: ser fornecedor de uma grande quantidade de órgãos</i>	55
3.2.6	<i>Associação dos três CVs</i>	56
3.2.7	<i>Coleta e tratamento dos dados</i>	56
3.2.8	<i>Modelagem e criação da Rede Supergrafo</i>	57
3.2.9	<i>Análise Exploratória da Rede Supergrafo</i>	58
3.2.10	<i>Análise estatística da Rede Supergrafo</i>	62
3.2.11	<i>Extração da Rede Vulnerável</i>	67
3.2.12	<i>Análise exploratória da Rede Vulnerável</i>	69
3.2.13	<i>Análise estatística da Rede Vulnerável</i>	73
3.2.14	<i>Avaliação da Rede Vulnerável</i>	75
4	CONCLUSÕES	76

1 INTRODUÇÃO

A pós-modernidade tem compelido a humanidade à uma metamorfose descomedida. Esse novo ambiente foi descrito por Bauman (2007, p.11) como uma “vida líquido-moderna” ao equiparar as frenéticas mudanças sociais em curso com o estado físico da matéria líquida, posto que a sociedade contemporânea se defronta com pessoas, coisas e eventos essencialmente disformes, mutantes, efêmeros e voláteis. Esse mundo de obsolescências programadas e amanhã incertos converteu os Estados-Nações em entes incapazes de se adaptarem tempestivamente às frequentes mudanças impostas por uma agenda mundial globalizada e sem controle específico. Segundo Bauman (2007, p.8) houve um divórcio entre a política local e o – agora recém-emancipado – poder. Instala-se um ambiente de profundas incertezas porque as políticas públicas perdem grande parte do controle sobre os cursos sociais vindouros. Os revides estatais aos problemas sociais são tardios porque, no momento em que estão prontos para implementação, configuram-se natimortos inservíveis aos fins para os quais foram planejados: o cenário – decorrente do ambiente mutante – não é mais o mesmo do início da percepção estatal ao problema. O Estado, ao reagir de forma pouco proativa e muito reativa, fá-lo em ações atardadas e, portanto, de aplicabilidade obsoleta. Soluções antigas isoladas mostram-se inadequadas, pois os problemas pós-modernos são complexos e absolutamente diversos daqueles enfrentados algumas décadas atrás, época em que os ambientes eram mais estáveis (sólidos) e as soluções aplicadas serviam não apenas como aprendizados, mas como modelos aos enfrentamentos pósteros. Entretanto, na sociedade pós-moderna do hoje, as soluções do ontem, isoladas, são insuficientes aos óbices do amanhã. Urge a associação de técnicas já conhecidas às novas incorporadas ao rol de disponibilidade, reforçando o enfoque multidisciplinar.

Dentre todos os problemas sociais que exigem ação estatal, a hodierna atuação criminosa complexa e organizada destaca-se eminente. Dotada de transfiguração frenética, exige respostas céleres aos entes da persecução criminal. O Estado pode – e deve – munir-se de instrumentos que viabilizem seu amoldamento combativo ao adversário mutante; decerto a ciência pode ser uma poderosa parceira estatal, por ser capaz de fornecer um impreterível valor agregado às ações de combate ao crime: o poder do conhecimento.

Santos (2010, p. 65) caracteriza o conhecimento contemporâneo como sendo dotado de uma fragmentação pós-moderna não disciplinar, mas temática, cujos temas são galerias por onde os conhecimentos progridem ao encontro uns dos outros. Ainda Santos (2010, p. 66) reforça que a ciência pós-moderna não segue um estilo unidimensional, facilmente

identificável, mas uma configuração de estilos construída segundo o critério e a imaginação pessoal do cientista, com tolerância discursiva e pluralidade metodológica. Wersig (1993, p. 234) afirma que, desde o seu início, a ciência tentou resolver medos humanos como fome, solidão e doenças, mas que a própria ciência fez surgir novos temores mais complexos, tais como aqueles relacionados à poluição, tecnologia genética, inteligência artificial, tecnologia militar, etc.

O rol dos problemas inauditos e complexos elencado por Wersig por óbvio é não exaustivo, portanto é coerente considerar o crime pós-moderno (disforme, líquido e complexo) como um dos medos cuja mera aplicação das ciências clássicas isoladas revela-se insuficiente e ineficaz. Isso o torna um problema ideal à aplicação das ciências pós-modernas. Desta feita, abordagens sobre a problemática devem ir além da especificidade e singularidade de método, o que explica em parte o surgimento de ciências pós-modernas: problemas mais complexos não podem ser combatidos por ciência tradicional. Wersig (1993, p. 234) também destaca como indicativo do surgimento dessas novas ciências a ocorrência de uma grande quantidade de discussões acerca de interdisciplinaridade, multidisciplinaridade ou transdisciplinaridade. A Ciência da Informação, como uma ciência pós-moderna, reveste-se dessa característica e se mostra adequada ao enfrentamento de problemas deveras complexos para ciências clássicas isoladas.

1.1 PROBLEMATIZAÇÃO E QUESTÃO DE PESQUISA

Os avanços da tecnologia da informação aliados à queda dos custos de processamento, armazenamento e transmissão têm tornado disponíveis a boa parte da humanidade um grande número de oportunidades de negócios alicerçados em dados. Essa realidade disruptiva, propulsora das capacidades de realização humanas, também está disponível às organizações criminosas contemporâneas, eliminando suas fronteiras espaciais de obscura atuação. Diante desses adversários sociais, faz-se necessário a adoção de estratégias combativas. Uma delas é a prevenção criminal.

Baseada no que já se sabe historicamente, a prevenção utiliza técnicas que apontam cenários mais vulneráveis ao crime. O cenário deixa de ser reativo e posterior para tornar-se proativo e prévio. A prevenção também promove a economicidade ao evitar que o erário público arque com os prejuízos decorrentes da consumação do crime e da repressão estatal posterior. Considerando que crimes porventura cometidos podem alçar alta complexidade e difícil precaução, características típicas dos delitos pós-modernos, o problema da pesquisa é investigar uma estratégia interdisciplinar reutilizável de prevenção criminal.

1.2 MOTIVAÇÕES E JUSTIFICATIVA

Não bastassem as crises econômicas que vários países têm enfrentado, o crime, agora tecnológico e complexo, se junta ao cenário e maleficia as sociedades em duas vertentes danosas: os prejuízos próprios da consumação criminal (pelo desvio irrecuperável do erário, danos ao patrimônio público e privado, prejuízos a bens e direitos, etc.) e as despesas decorrentes da persecução penal compulsória. Esses dispêndios agravam os quadros de pobreza, pois ocorrem ao passo que minguem recursos estatais que garantam as necessidades mais básicas à população.

Segundo o Programa das Nações Unidas para o Desenvolvimento (PNUD), 3,8% da população brasileira - aproximadamente 7,7 milhões de pessoas - encontrava-se em condições de pobreza multidimensional em 2015 (REINO UNIDO, 1996). O termo multidimensional foi usado como uma quebra de paradigmas, pois o PNUD adotou outras variáveis além da baixa renda como influenciadoras da medição do grau de pobreza. Dentre as novas variáveis estão a saúde, a educação e a, agora protagonista, segurança pública.

Pressupondo o Estado como promotor primeiro do bem-estar social, mas com haveres limitados, a escrupulosa aplicação dos recursos públicos torna-se ainda mais imperiosa, especialmente em quadros como o da crise econômica ora vigente em nosso país. Entretanto, apesar desse cenário, a atuação da polícia judiciária na segurança pública continua se pautando majoritariamente na repressão criminal. Isso significa que primeiro os criminosos geram os prejuízos, posteriormente a ação estatal surge, muitas vezes tardia. Atuar por reatividade é prenúncio de danos permanentes, pois o crime consumado pode realizar perdas irreparáveis aos bens e aos direitos, sejam eles públicos ou privados. Nesse caso, quando muito, o Estado oferece a expectativa incerta da punição penal contra o criminoso. Essas são motivações relevantes que justificam a escolha da questão de pesquisa.

1.3 OBJETIVOS

A identificação das interconexões sociais vulneráveis às organizações criminosas pós-modernas é uma tarefa complexa. Diligências policiais clássicas, embora importantes ao conjunto supedâneo da obra investigativa, são relativamente ineficazes na identificação de entidades e interconexões vulneráveis aos crimes mais complexos em períodos prévios a consumação delitiva. Isso ocorre porque tais diligências não manipulam os dados como o que eles de fato são: uma rede.

Uma das estratégias para mitigação desse impasse é criar um protocolo formal, preventivo, adaptável e exequível de técnicas que tratem os dados como uma rede social de

atores permeada de padrões ocultos que aumentam a vulnerabilidade a eventuais delitos futuros. Isto posto, o presente trabalho justifica-se como relevante ao propor tal protocolo.

Embora a atuação da polícia judiciária seja, em regra, repressiva, a prevenção tem sido gradativamente valorizada pelo potencial de minimizar os custos danosos do crime e da sua repressão posterior. Entretanto, a prevenção não pode basear-se exclusivamente em percepções sensoriais. É indispensável a aplicação de uma sólida fundamentação conceitual e científica. Esse cenário, caracterizado pela necessidade de uma prevenção criminal respaldada, favoreceu a definição dos objetivos do trabalho.

1.3.1 Objetivo Geral

O conhecimento no âmbito atividade policial, seja de natureza preventiva ou repressiva, é obtido quando o policial realiza a travessia de uma rede cognitiva de aprendizado acerca de entidades envolvidas em crimes, estejam eles em fase dos atos preparatórios, executórios ou de consumação. Formada por atores sociais – pessoas, coisas, organizações ou eventos – essa rede aumenta e se torna mais complexa a medida em que o investigador reúne os dados que a compõem. Nas polícias brasileiras esse aumento de entidades e suas interconexões é uma composição majoritariamente mental e minoritariamente computacional. Entretanto, as características de uma rede a tornam um insumo propício a implementações computacionais aptas à análise, especialmente quando um grafo – instrumento dotado de definição matemática precisa – é o modelo escolhido para representá-la.

Contemporaneamente os atores sociais registram seus dados em muitos sistemas informacionais, seja na forma de coleta compulsória, seja na forma de depósito voluntário. Esses grandes repositórios informacionais detêm o registro existencial de redes sociais do mundo real, inclusive as de natureza excepcional classificadas como criminais. Sob certo aspecto, esses sistemas são o novo “local de crime”, pois neles estão indícios vários e concordantes que, quando unidos às diligências policiais clássicas (vigilância, dados de informantes, entrevistas, etc.) podem formar conjuntos informacionais úteis como prova de crime ou informação de inteligência voltada à prevenção.

Sem dificuldade, é possível citar diversos sistemas cujos dados formam redes sociais do cotidiano humano: voos, registros de imigração, vínculos societários empresariais, vídeos e imagens de câmeras de segurança, envio e recebimento de e-mails, redes sociais cibernéticas, passaportes, relações de trabalho, dados de benefícios sociais, registros de armas, dados sobre licitações, vínculos familiares, antecedentes criminais, etc. Considerando que várias dessas redes – inclusive as delitivas – estão registradas em sistemas sob posse estatal, é coerente que os seus rastros informacionais sejam analisados por órgãos de controle quando a motivação seja,

indubitavelmente, o benefício social¹. Entretanto, não basta essa análise ser instintiva. Ela deve seguir um protocolo mínimo, um processo composto por tarefas que possam ser reproduzidas em casos reais. Especialmente em casos de prevenção criminal.

Nesse contexto, o objetivo geral da presente pesquisa é cobrir esse vácuo: propor um modelo protocolar, interdisciplinar, genérico e adaptável de prevenção anticrime que identifique uma Rede Vulnerável (RV) a um Crime-Objeto (CO) por meio da análise de redes sociais e de grafos. Conhecidas as vulnerabilidades, o Estado poderá agir por antecipação proativa e mitigar a ação repressiva porvindoura.

1.3.2 Objetivos Específicos

A pesquisa explora a interdisciplinaridade entre as Ciências Policiais, a Ciência da Informação, a Estatística, a Sociologia e a Ciência da Computação para obter um modelo de prevenção anticrime concebido pelos seguintes objetivos específicos:

- Delinear conceitos de base que permitam criar um protocolo policial acerca do tema prevenção criminal, tais como Crime-Objeto, Variável Frisada, Critério de Vulnerabilidade, Classes Relevantes, Rede Supergrafo e Rede Vulnerável;
- Descrever um método de obtenção de uma Rede Vulnerável a delitos a partir de uma rede inicial denominada Supergrafo;
- Orientar o uso de análises exploratórias e estatísticas como métodos aplicáveis a prevenção criminal;
- Organizar as tarefas do método num processo passo-a-passo através de um fluxograma;
- Aplicar a metodologia em um estudo de caso real para comprovar sua viabilidade prática.

1.4 ORIGINALIDADE DA PESQUISA

Contemporaneamente, a informação é uma das mercadorias mais preciosas do mundo. Na esfera privada trata-se de um bem mais valioso que o dinheiro e o poder, pois a partir daquela podemos obter ambos. As organizações públicas, apesar de não possuírem como propósito auferir dinheiro ou poder, objetivam prestar ao cidadão um serviço público de qualidade, meta que também requer a posse de informações. Não poderia ser diferente no âmbito da Polícia Federal Brasileira (PF). Para cumprir seus objetivos, a PF necessita provar a autoria e materialidade de crimes por meio da transformação de dados em informações mediante uma

¹ Por óbvio, ainda que o Estado disponha desses dados, eventuais desvios de conduta acerca do acesso abusivo devem resultar em punição exemplar aos agentes públicos envolvidos.

atividade que se denomina investigação criminal. O enfoque dessa atividade é primordialmente repressivo. Contemporaneamente os órgãos de controle estatal tem agido em conjunto para reforçar a prevenção. Entretanto, os policiais que labutam na prevenção criminal ainda buscam por conhecimentos meramente textuais em sistemas corporativos. Esse comportamento contrasta com tendências de uso de ferramentas de visualização, tanto nas atividades de natureza repressivas quanto preventivas.

Na Polícia Federal brasileira, estudos sobre técnicas de investigação ainda carecem do uso de métodos visuais que utilizem grafos. Essas ferramentas são mais capazes de gerar e transmitir conhecimento do que massivos textos de difícil cognição. Associados a métodos da análise de redes sociais, os grafos ganham métricas de análise de centralidades, técnica amplamente utilizada nesse trabalho.

Com vistas a confirmar a originalidade da proposta e a percepção de vácuo temático, foram realizadas buscas nas bases de conhecimento da Academia Nacional de Polícia (ANP). Em 29/08/2019, o autor realizou uma pesquisa no Sistema de Biblioteca Digital da Academia Nacional de Polícia (ANP), pelos termos “Análise de Redes Sociais”, “*Social Network Analysis*”, “Grafo” e “*Graph*”. Os referidos sistemas informaram não haver, em seus repositórios, publicações vinculadas aos citados assuntos. Destarte, não há disponibilidade, em tese, de publicações decorrentes dos cursos de pós-graduação realizados pela ANP ou pela Escola Superior de Polícia acerca dos temas. Releve-se que a Polícia Federal é a maior referência nacional acerca de métodos investigativos inovadores na Ciência Policial, o que faz supor que nas polícias estaduais – tanto civil quanto militar – a temática possivelmente também seja pouco explorada. Assim, presume-se confirmada a originalidade da pesquisa.

2 DESENVOLVIMENTO

2.1 REFERENCIAL TEÓRICO

O referencial teórico da pesquisa fundamenta-se essencialmente em conceitos afetos à interdisciplinaridade característica das Ciências Policiais (CPs) e da Ciência da Informação (CI) e seu paradigma cognitivo, teoria dos grafos, aprendizado visual, análise de redes sociais, bancos de dados baseados em grafos, estatística e *Business Process Model and Notation* (BPMN).

2.1.1 Interdisciplinaridade

Tanto a Ciência da informação (CI) como as Ciências Policiais (CP) possuem forte base interdisciplinar e, em consequência, estabelecem fronteiras com diversas ciências circunvizinhas. Nos seus limites, as fronteiras científicas funcionam como membranas

semipermeáveis, uma espécie de agente equilibrante. O cenário resultante permite a troca de conhecimento (ferramentas, dados, métodos, etc.) entre as ciências envolvidas, tal qual uma osmose invertida² que busca igualar as “concentrações” de sapiência.

O conhecimento da ciência-origem, mais denso, flui pela membrana-fronteira em direção à – comunicável e periférica – coirmã vizinha, contribuindo ao aumento da densidade científica da receptora ciência-destino; posteriormente, o equilíbrio se reestabelece, mas não definitivamente, pois importantes agentes de mudança atuarão para alterar o quadro de efêmero equilíbrio temporário: os cientistas.

Sistematicamente, esses atores sociais atingem novos degraus de conhecimento nos diversos campos das ciências nas quais labutam. Assim, as novas contribuições científicas causam a mudança da densidade do conhecimento e, naturalmente, o ciclo de demanda pelo reequilíbrio reinicia. A habitual recorrência desse processo gera vantagens científicas óbvias: produtos de uma ciência são matéria-prima que reforçam outra. A Ciência Policial não pode desprezar esse potencial. Destarte, na sua missão de gerar sapiência anticrime, deve permeabilizar suas fronteiras a outras ciências e empreender suas osmose inversas de conhecimento. Nenhuma ciência isolada, atuando *per si*, atingiria a sinergia de um subconjunto comutável delas. O trabalho ora abordado fundamenta-se nesse princípio pois, em seu cerne, há o uso da interdisciplinaridade como método e peça chave na obtenção de um modelo básico de prevenção criminal.

2.1.2 Paradigma cognitivo da Ciência da Informação

A Polícia dispõe de muitos e diversificados dados, mas a mera posse desses elementos – estacionários em uma mídia de armazenamento computacional – não garante um estado de cognição minimamente proveitoso. Os padrões criminais camuflam-se em um amontoado de bits infrutíferos e os indicativos dos crimes futuros, de existência prévia à consumação delitiva, continuam velados. É possível realizar previsão criminal se conhecermos esses indicativos? Quiçá a Ciência Policial e as ciências fronteiriças, empregadas em parceria, transformem o estado anômalo de conhecimento em um patamar cognitivo que preveja alguns crimes porvir. Certamente não predirão a sua totalidade, mas talvez uma parcela significativa deles. Antes, entretanto, é necessário obter um estado de conhecimento não anômalo.

A fuga do estado anômalo do conhecimento é questão de estudo desde sempre, mas foi formalizada por Brookes (1980, p. 131) quando elaborou a expressão denominada *Equação Fundamental da Ciência da Informação*, cuja formulação é ilustrada pela seguinte expressão

² Na osmose real o fluxo de solvente atravessa a membrana semipermeável no sentido (meio menos denso) → (meio mais denso), contrário ao da metáfora ora empregada.

algébrica: $K(S) + \Delta I = K(S + \Delta S)$. A equação exibe um estado geral do conhecimento anômalo denominado $K(S)$, o qual é modificado para um novo estado do conhecimento $K(S+\Delta S)$ em decorrência do incremento da informação ΔI . No caso da prevenção de crimes, o estado final $K(S+\Delta S)$ é aquele em que, baseado em métodos científicos, revelam-se cenários onde a probabilidade de crimes é maior. Noutras palavras, é obtenção do conhecimento acerca da amostra que será monitorada preventivamente para evitar crimes futuros.

Tentar prever crimes é uma atividade semelhante ao modelo descrito por Capurro (2003, p. 8) quando afirma que, no paradigma cognitivo da CI, os processos informativos transformam ou não o usuário, entendido em primeiro lugar como “sujeito cognoscente” possuidor de “modelos mentais” do “mundo exterior”. É razoável pensar o policial exercendo o papel de sujeito cognoscente, a experiência acumulada como sendo os modelos mentais e o mundo exterior expresso pela manifestação da própria atividade policial. Senão, vejamos: contemporaneamente é comum, no meio policial (mundo exterior), a obtenção de demonstrações científicas daquilo que outrora já se conhecia nas práxis. Por meio da sapiência resultante da experiência, o policial (sujeito cognoscente) acumula mais referências empíricas (modelos mentais), o que o torna uma fonte importante de proposições inaugurais das Ciências Policiais. Assim, conclui-se que o policial se torna um ator importante no processo de prevenção criminal embasado nas ciências.

2.1.3 Aprendizado visual

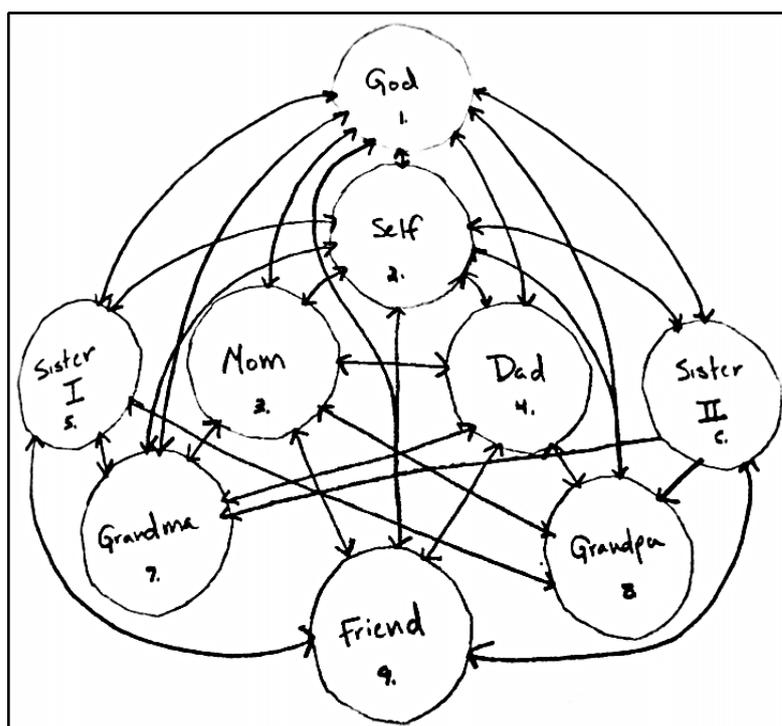
Os povos Sumérios, da Mesopotâmia, inventaram a escrita, àquela época talhada por cunhas em placas de barro. Segundo Cunningham (2013, p. 95) a natureza da escrita suméria mudou daquilo que é denominado sistema logográfico, isto é, signos representando palavras, para um sistema misto, parcialmente logográfico e parcialmente fonográfico (onde um número limitado de sinais é usado para representar sequências sonoras).

A mudança do sistema de escrita elevou a transmissão do conhecimento humano a outros patamares de reprodução, confiabilidade e preservação. Na contemporaneidade, a linguagem escrita continua como uma das principais formas de comunicação. Igualmente, os atores estatais da persecução criminal – Polícia, Justiça e Ministério Público – comunicam-se entre si utilizando documentos formalizados em linguagem predominantemente escrita em texto puro, com pouco uso de recursos visuais. Isso também ocorre na maioria das diligências realizadas nos sistemas informacionais: texto excessivo é o resultado corriqueiro das pesquisas nesses repositórios. Entretanto, muitos desses sistemas contêm redes sociais em sua essência e o uso de textos massivos vai de encontro ao fato de que as imagens são mais facilmente assimiladas, conforme alguns demonstram alguns casos citados a seguir.

Um experimento realizado por Standing, Conezio e Haber (1970) exibiu, a cada um dos componentes de um grupo de pessoas, 2.560 (duas mil, quinhentas e sessenta) imagens por dez segundos cada. A memória de reconhecimento das imagens foi testada em seguida e indicou a retenção de mais de 2.000 (dois mil) itens, mesmo após três dias decorridos entre o aprendizado e o teste. Em algumas variantes do mesmo experimento nas quais o tempo de apresentação foi reduzido para um segundo por imagem, não houve mudança significativa na performance dos indivíduos. Buzan e Buzan (1994) afirmam que as imagens são mais evocativas do que palavras, mais precisas e potentes em desencadear uma ampla gama de associações, ampliando o pensamento criativo e a memória. A visão, portanto, é um sentido extremamente relacionado à memória e, em consequência, ao aprendizado. Ao que parece, a sabedoria popular que afirma valer mais uma imagem a mil palavras está certo.

Um modo de ilustrar como um ser humano mentaliza seu mundo foi desenvolvida pela psicóloga Catalin Mamali, cujo experimento foi brevemente descrito por Christakis e Fowler (2010, p. 211). Os indivíduos eram instados a pensar nas pessoas que lhes eram “próximas” ou “muito significativas”, estabelecendo suas conexões e gerando uma espécie de desenho denominado de “autógrafo social”. Um dos desenhos, reproduzido por Christakis e Fowler (2010, p. 212), foi criado por uma estudante colegial de 18 anos, e ilustra como ela se sente conectada a pessoas importantes em sua vida, conforme a figura 1:

Figura 1 – Autógrafo social de uma estudante de 18 anos



Fonte: Extraído de Christakis e Fowler (2010, p. 212)

Portanto, ao lembrar de suas relações, a estudante colegial priorizou sua memória visual. Ela provavelmente não sabia o que é um grafo no momento em que criou o desenho, entretanto usou esse instrumento, ainda que intuitivamente, para representar seu autógrafo social.

O aprendizado visual é uma prática que exige dois principais papéis: a do produtor de conteúdo visual e ao observador cognoscente posterior. Ambas as habilidades devem ser desenvolvidas pelos policiais, pois tanto produzem quanto consomem conhecimento em suporte pictórico. A utilização do aprendizado visual nas investigações repressivas e preventivas deve ser uma prática contínua que pode ser potencializada pelo uso de grafos, ferramenta amplamente usada para representar redes.

2.1.4 Teoria dos grafos

O conceito de grafo remonta ao acontecimento descrito por Caldarelli (2007, p.7), quando relata que o matemático Leonard Euler queria responder a uma pergunta popular de seu tempo: estando no centro da cidade de Königsberg, é possível fazer um passeio cruzando cada uma das sete pontes apenas uma vez? Ao responder essa questão, Euler tornou-se o primeiro matemático a introduzir a noção dos grafos.

Matematicamente, segundo Chartrand, Lesniak e Zhang (2015, p.3), um grafo G é formado por um conjunto não-vazio e finito V de objetos chamados vértices junto a um conjunto possivelmente vazio E de subconjuntos de dois elementos de V chamados arestas. Concisamente, a notação apropriada é $G(V,E)$. Vértices por vezes são chamados de pontos ou nós, enquanto as arestas são designadas como linhas ou ligações.

Um grafo pode ser representado por uma matriz de adjacência $A(n,n)$. Erciyes (2015, p. 15) afirma que a matriz de adjacência de um grafo $G(V, E)$ com n vértices é uma matriz $n \times n$ que possui elementos a_{ij} definidos da seguinte forma:

$$a_{ij} = \begin{cases} 0 & \text{se } i = j \\ 1 & \text{se o nó } v_i \text{ estiver conectado ao nó } v_j \end{cases}$$

Segundo Erciyes (2015, p. 15) um grafo direcionado $G(V, E)$ consiste em um conjunto não vazio de vértices V e um conjunto de arestas direcionadas E , onde cada $e \in E$ é associado a um conjunto ordenado de vértices e uma aresta (u, v) que inicia do vértice u e finaliza no vértice v .

Erciyes (2015, p.15) também relata que as arestas são rotuladas com pesos em um grafo ponderado $G(V, E, w)$, onde $w: E \rightarrow \mathbb{R}$. Esses pesos podem estar representando a

importância de um *link*; essa valoração que pode ser, por exemplo, uma quantia monetária no caso em que a aresta representa um pagamento entre as entidades representadas pelos vértices.

Segundo Erciyas (2015, p. 11), dados dois grafos $G(V,E)$ e $G'(V',E')$, se $V' \subseteq V$ e $E' \subseteq E$ então G' é chamado subgrafo de G que, por sua vez, é chamado de supergrafo de G' . Se $\forall(x,y) \in E'$; $x \in V'$ e $y \in V'$, então G' é um subgrafo induzido de G .

A travessia de pontes inspirou o nascimento da teoria dos grafos. Travessia similar constrói o investigador criminal quando modela seu “grafo cognitivo” que cresce e se torna mais complexo à medida em que os dados conhecidos são reunidos e interconectados. Essa atividade seria mais eficazmente executada de modo computacional, por meio de softwares que implementem grafos, sua travessia e suas métricas características.

A procura pelo conhecimento nas investigações policiais assemelha-se ao que descreveu Capurro (2003, p. 8) quando afirma:

(...) a busca de informação tem sua origem na necessidade (“need”) que surge quando existe o mencionado estado cognitivo anômalo, no qual o conhecimento ao alcance do usuário, para resolver o problema, não é suficiente.

Ao investigar, o policial procura sair desse estado cognitivo anômalo de conhecimento insuficiente para elucidar ou prever crimes. Essencialmente, investigar crimes consumados significa estabelecer relações – outrora desconhecidas – entre entidades. Uma investigação pode vir a descobrir que x (uma pessoa) esteve em y (um local) portando w (um objeto) onde aconteceu o crime z (evento). As interligações ($x \leftrightarrow y \leftrightarrow w \leftrightarrow z$) transformam dados em informação e, posteriormente, caso as conexões recém-criadas sejam relevantes, alçarão o patamar de conhecimento. Quando os crimes ainda não vieram a ocorrer, determinados padrões de interconexões configuram-se como cenários que aumentam a probabilidade delituosa. As investigações criminais preventivas indicam a ocorrência desses cenários de provável ocorrência de crimes ulteriores.

Pessoas, objetos, locais e eventos são os principais tipos de entidades candidatas às interconexões, cuja operacionalização exige um instrumento hábil a efetivá-las. Os grafos são ferramentas adequadas para tal.

2.1.5 Análise de redes sociais e centralidade de grau

A origem das redes sociais remonta a própria existência humana, mas a Análise de Redes Sociais (comumente denominada como ARS) com enfoque científico é recente, tendo suas origens na Sociometria cujo fundador, segundo Marineau (1989, p. 7), foi Jacob Levy

Moreno. Nascido em Bucareste, capital da Romênia, Moreno cursou a faculdade de Medicina de Viena e emigrou para os Estados Unidos da América.

Freeman (2004, p. 48) relata que Jacob Levy inicialmente publicou dois livros. Ambas as obras de Moreno realizaram coleta sistemática de uma coleção de dados de indivíduos para análise e, em uma delas, ele usou o termo *network* (rede) pela primeira vez. Essa inovação fez surgir uma abordagem estrutural denominada *Sociometry* (Sociometria), precursora da ARS.

Segundo Latora, Nicosia e Russo (2017), a análise de redes sociais é a disciplina que fornece métodos e ferramentas quantitativos para entender os sistemas sociais, representando-os e estudando-os como redes. Robins (2015, p. 13) também afirma que para as redes sociais, as entidades estudadas são mais frequentemente pessoas, embora possam ser outros atores sociais, como organizações. Robins (2015, p. 19) afirma que uma rede social compreende (pelo menos) um conjunto de atores sociais e uma relação entre eles na forma dos laços relacionais diádicos. Uma rede social pode ser representada por um grafo, entretanto essa é uma das formas disponíveis para ilustrar uma rede. Portanto, o grafo não é a rede social em essência, mas sim uma forma de representá-la.

Wasserman e Faust (1994, p. 3) afirmam que do ponto de vista da análise de redes sociais, o ambiente social pode ser expresso como regularidades nas relações entre unidades de interação. Essas unidades podem possuir graus de importância variáveis dentro da rede. A centralidade é um conjunto de métricas que representam algumas formas de obter esses graus de importância, pois as medidas de centralidade indicam os atores mais importantes na rede.

Segundo Latora, Nicosia e Russo (2017, p. 31) o conceito de centralidade e as primeiras medidas relacionadas ao tema foram introduzidas no contexto da ARS. Sobre centralidade de atores, Wasserman e Faust (1994, p. 3) afirmam que atores proeminentes são aqueles que estão amplamente envolvidos no relacionamento com outros atores. Esse envolvimento os torna mais visíveis.

Erciyes (2015, p. 71) afirma que a centralidade é uma medida da importância de um vértice ou uma aresta em uma rede complexa. Assim, atribui-se um valor de importância para cada vértice ou aresta com base em sua posição topológica na rede. Segundo Ilhan, Gündüz-Ögüdücü e Etaner-uyar (2014, p.4), a centralidade mede a importância relativa de um vértice e fornece uma indicação sobre a sua influência na rede, sendo as seguintes suas variações: centralidade de grau, centralidade de intermediação, centralidade de proximidade e centralidade de autovetor. A pesquisa ora conduzida utiliza apenas a centralidade de grau como fundamento.

É necessário, portanto, o entendimento acerca do conceito da métrica nominada grau e das suas variações.

Muitas métricas da ARS estão relacionadas ao grau de um vértice, cujo conceito Caldarelli (2007, p.14) expõe como sendo o número de arestas de um vértice, ao tempo em que esclarece que a soma de todos os graus em um grafo é o dobro do número de suas arestas. Isso ocorre porque cada aresta contribui duas vezes na contagem do grau: uma unidade para cada um dos vértices aos quais se conecta. Dada uma matriz de adjacência $A(n,n)$, Caldarelli (2007, p. 14) relata que o grau pode ser expresso por:

$$k_i = \sum_{j=1,n} a_{i,j}$$

Latora, Nicosia e Russo (2017, p. 31) afirmam que, se o grafo for direcionado, o grau do vértice possui dois componentes: o número de links de saída k_i^{out} (grau de saída ou *out-degree*), e o número de links de entrada k_i^{in} (grau de entrada ou *in-degree*) do nó i . O grau total do nó é então definido como:

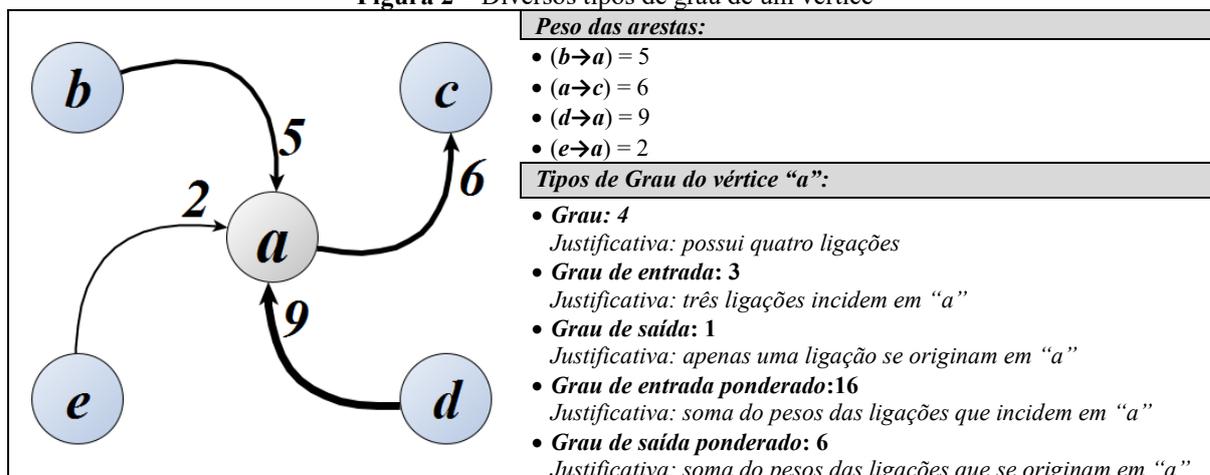
$$k_i = k_i^{out} + k_i^{in}$$

Nos grafos ponderados – cujas arestas possuem pesos – também podem ser computados o *weight in-degree* e o *weight out-degree*. Caldarelli (2007, p. 14) afirma que essas duas métricas são uma extensão do conceito de grau e contabilizam os pesos das arestas ao invés da simples contagem delas, sendo calculadas por:

$$k_i^w = \sum_{j=1,n} a_{ij}^w$$

A figura 2 ilustra todos os conceitos da métrica nominada como grau (que acima foram apresentados de forma matemática) de uma maneira mais intuitiva e concisa:

Figura 2 – Diversos tipos de grau de um vértice



Fonte: Elaborada pelo autor

2.1.6 Análise exploratória de dados

Segundo Pinheiro (2009, p. 12), a análise exploratória é um conjunto de técnicas de tratamento de dados que, sem implicar em uma fundamentação matemática mais rigorosa, contribui na sondagem do terreno, ou seja, na obtenção de um primeiro contato com a informação disponível. É livre o uso de recursos visuais, tais como grafos e gráficos. A aplicação concreta da pesquisa ditará quais ferramentas visuais fornecem melhor entendimento inicial do contexto.

Sobre a análise exploratória de dados, Bernabeu, Castro e Godino (1991, p. 2) afirmam:

Como indicamos, nos deparamos com uma nova filosofia na aplicação de métodos de análise de dados, embora, juntamente com ela, algumas técnicas específicas tenham sido desenvolvidas para sua aplicação. Essa filosofia consiste no estudo de dados de todas as perspectivas e com todas as ferramentas possíveis, inclusive as existentes. O objetivo é extrair o máximo de informação possível, gerar novas hipóteses, no sentido construir conjecturas sobre as observações que dispomos.

Bernabeu, Castro e Godino (1991, p. 1) afirmam que antes da análise exploratória a análise dos dados baseava-se principalmente no cálculo estatístico, levando diminuição da importância visual da representação dos dados, e atribuindo-a exclusivamente aos cálculos; em segundo lugar, a análise foi equiparada ao modelo confirmatório. Essa afirmação significa dizer que os estudos recaíam imediatamente sobre os cálculos. As percepções visuais iniciais, muito poderosas, eram deixadas ao segundo plano. Enxergar uma rede em determinadas formas e layouts pode ser uma técnica exploratória eficaz que pode guiar os cálculos confirmatórios posteriores.

O poder do aprendizado visual não deve ser subestimado. Assim, a pesquisa sugere uma análise visual de redes como uma importante estratégia na prevenção criminal. O embasamento é, principalmente, o uso de leiautes.

2.1.7 Análise exploratória de grafos usando leiautes

Uma das técnicas mais interessantes de análise exploratória é o uso de leiautes de grafos. Eles representam os formatos que um grafo pode assumir em um espaço de plotagem e que têm a aptidão de dotá-lo de algum conteúdo semântico que pode ser captado pela análise exploratória, a depender da estética.

Cherven(2015, p. 28) assim descreve a importância dos leiautes:

A seleção de um layout apropriado pode fazer a diferença entre criar um grafo impenetrável, que não comunica uma história, e uma visualização facilmente acessível que não apenas comunica, mas também tem apelo estético. Ao ler a literatura de grafos de rede, é provável que se depare com o termo *hairball*, uma descrição para uma rede muito densa com muitas conexões que são praticamente indecifráveis quando se usa os algoritmos comuns de grafos.

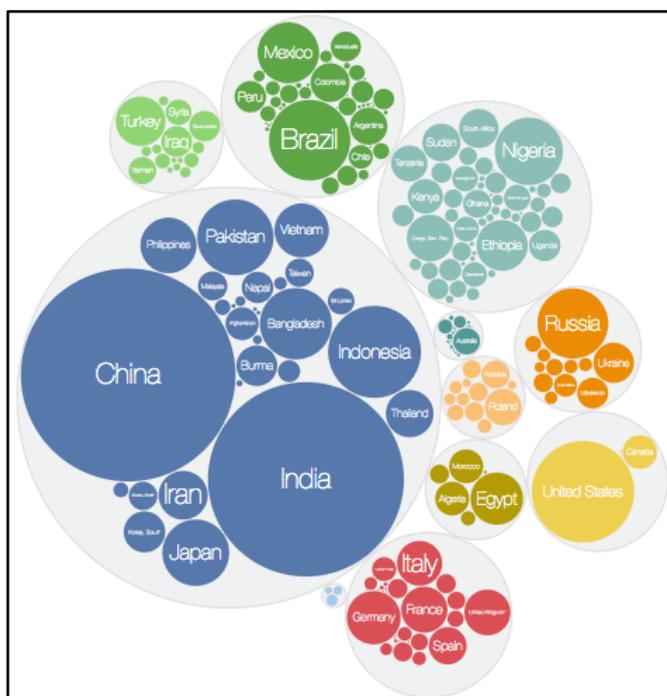
Sobre critérios para obter uma melhor estética dos grafos, Hu (2006, p. 37) afirma que:

A utilidade da representação depende da estética do desenho. Embora não existam critérios estritos para a estética, aceita-se que o mínimo cruzamentos entre arestas, vértices uniformemente distribuídos e a representação da simetria gráfica são desejáveis.

O uso dos leiautes é uma boa estratégia de análise de redes representadas por grafos. A busca por simetria, menos cruzamentos de ligações (arestas) e distribuições uniformes podem conferir significado a uma rede que, representada de modo aleatório, pouco significa visualmente. A seguir explana-se alguns leiautes que podem potencializar a análise de redes.

O leiaute *circle-pack* pertence a um grupo de leiautes que ajudam a visualizar dados hierárquicos. A sua nomenclatura é oriunda do seu visual: cada nível de dados é reunido em um pacote (NGUYEN, 2018). Cada pacote é ilustrado como um círculo. Abaixo uma representação de uso do *circle-pack* (sem uso de arestas), no qual alguns países do mundo são agrupados por região, sendo que o tamanho do círculo representa a população de cada um deles (figura 3):

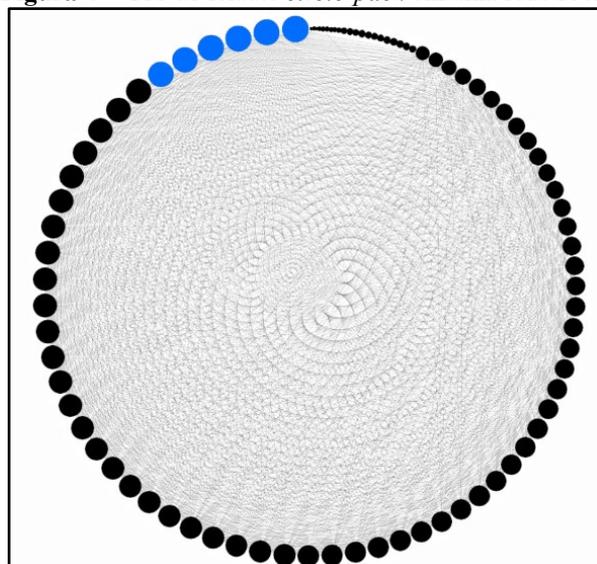
Figura 3 – Uso do leiaute *circle-pack*



Fonte: Cook (2019)

Outro leiaute, o circular, deve ser usado quando o número de vértices é pequeno e quando a ideia é mostrar classificações intuitivas. As métricas de grau relacionadas aos vértices podem ser usadas como critério de ordenação no arco do círculo, mas também é possível usar o valor de qualquer dos atributos em específico. O diâmetro dos círculos pode ser ajustado para criar visuais mais compactos e a ordenação dos vértices pode ser executada no sentido circular horário ou anti-horário. A figura 4, a seguir, apresenta um leiaute circular ordenado em sentido horário.

Figura 4 – Uso do leiaute *circle-pack* em uma rede social

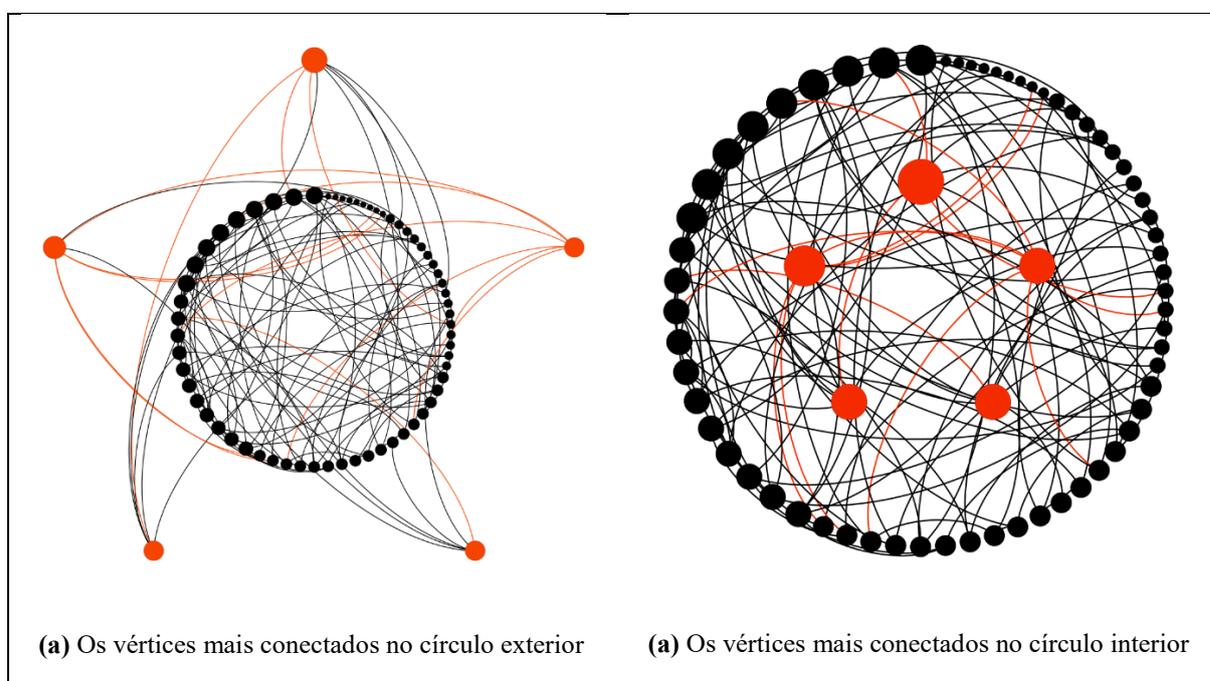


Fonte: Elaborado pelo autor

Na figura 4, cada vértice representa um amigo do autor da pesquisa no site Facebook³. A parte central do círculo é atravessada por diversas interconexões (arestas). Cada uma delas representa uma amizade entre dois amigos na referida rede social. A circunferência contém os vértices ordenados pela métrica nominada grau do vértice, o que significa que aqueles mais interconectados são representados pelos maiores círculos. Não por acaso, seis familiares do autor da pesquisa, representados na cor azul, ocuparam as seis primeiras colocações. Eles possuem uma grande rede social em comum com o autor e, em consequência, estão muito conectados aos demais vértices da rede, fato que eleva o grau dos seus vértices.

O leiaute *dual circle*, segundo Cherven (2015, p. 80), possibilita especificar um número de vértices de ordem superior para inseri-los dentro ou fora do círculo principal. Esse pode ser um leiaute eficaz quando houver um número específico de vértices a serem focados como parte da história, pois é uma forma de isolá-los visualmente dos demais. Abaixo, a figura 5 exibe duas redes que dão enfoque aos cinco vértices mais interconectados usando o *dual circle*. No grafo 5(a) optou-se por movê-los para o círculo exterior e, no grafo e(b) eles foram movidos para o círculo interior.

Figura 5 – Uso do leiaute *dual circle*

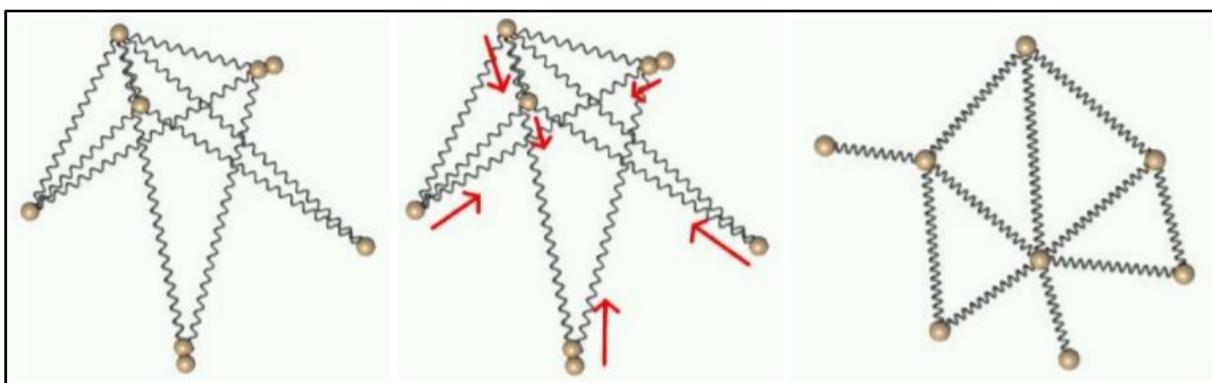


Fonte: Elaborado pelo autor

³ Rede social acessível em <http://www.facebook.com>

Outro leiaute que pode ser aplicado em redes segue princípios dos algoritmos direcionados por força. Eles se baseiam nas ideias de forças de atração, de repulsão e da gravidade. Cherven (2015, p. 70) define atração como aproximação dos vértices com base em seu relacionamento, repulsão como o processo que tende a manter os nós distantes uns dos outros e a gravidade como a propriedade que define como os vértices são desenhados em relação ao centro do grafo. Kobourov (2012, p.3) ilustra a atuação dos algoritmos direcionados por força através da ação das molas e da repulsão elétrica até que seja obtido o equilíbrio do sistema, conforme a figura 6:

Figura 6 – Grafo tratado como um sistema de molas, até atingir equilíbrio



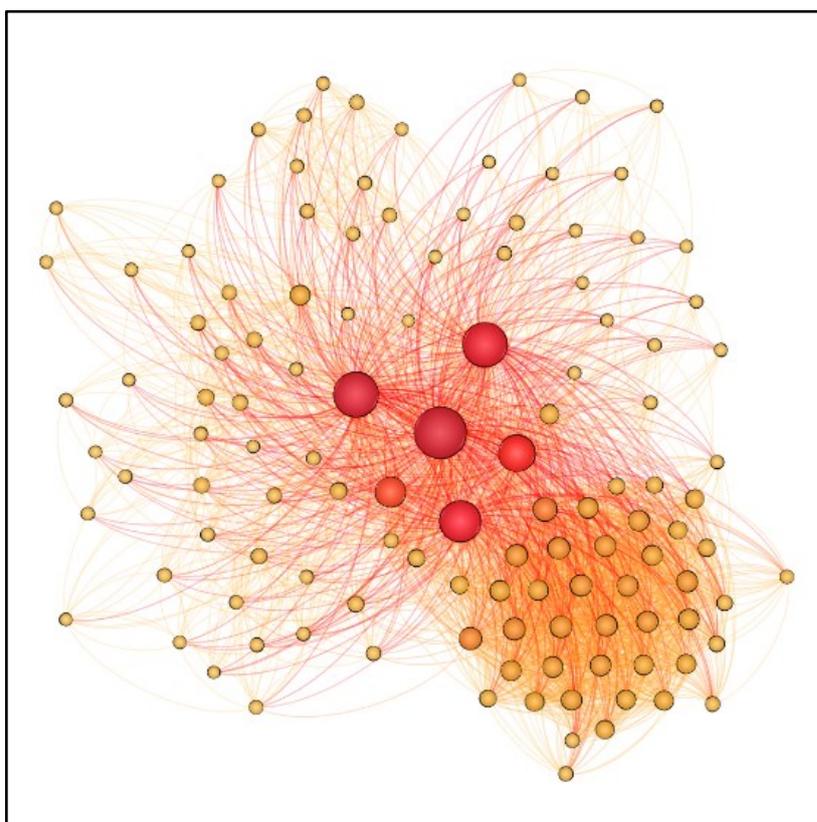
Fonte: Extraído de Kobourov (2012, p.3)

Um dos leiautes direcionados por força mais usados é o *Yifan Hu*, que recebeu o nome do seu criador. Cherven (2015, p. 76) assim define esse leiaute:

O *Yifan Hu* original produz resultados mais rápidos em comparação com outros métodos direcionados por força, concentrando-se na atração e repulsão no nível da vizinhança (e não em toda a rede), colocando assim uma carga computacional muito menor na máquina local. Ele também tem a vantagem de se interromper usando o resfriamento adaptativo, de modo que geralmente possa correr muito mais rapidamente do que métodos como o Force Atlas⁴. Entre as várias opções para *Yifan Hu*, está a capacidade de definir os níveis de distância ideal, com valores mais altos afastando os vértices, sem definir explicitamente um nível de repulsão. Essa é uma maneira mais genérica de definir os comprimentos das molas usadas para espaçar a rede.

Abaixo, a figura 7 exhibe o resultado da aplicação do leiaute *Yifan Hu* a uma parte da rede social do autor da pesquisa. Os seis familiares, que são pessoas ultraconectadas, migram para o centro da rede:

⁴ Force Atlas é um outro leiaute da mesma classe do Yifan-Hu. Ambos usam algoritmos direcionados por força.

Figura 7 – leiaute direcionado por força aplicado a uma rede social

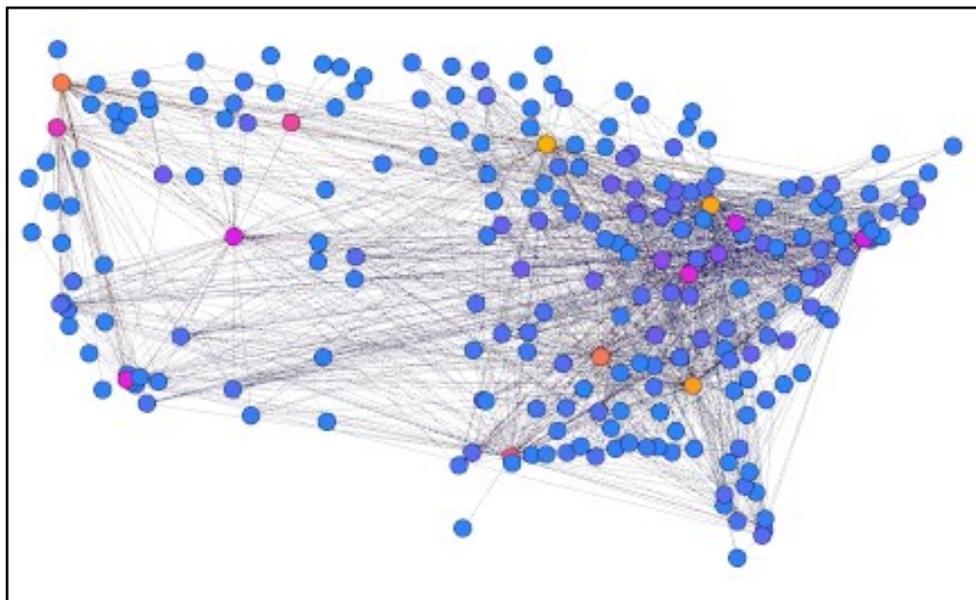
Fonte: Elaborado pelo autor

Percebe-se que, além de uma certa simetria, vértices muito conectados ocupam posições mais centrais do grafo (em vermelho). A migração dos vértices mais conectados – os nominados *hubs* – para regiões centrais é uma consequência dos algoritmos direcionados por força, que simulam atração (similares a da mola) exercidas pelas arestas (ligações) e, em contraposição, a repulsão entre vértices (similares à força elétrica de cargas de mesmo sinal). Em consequência, os *hubs* migram ao centro, enquanto os vértices menos conectados são afastados, por repulsão, à periferia. Estar bem conectado significa possuir alta centralidade de grau. Para melhor percepção da migração dos *hubs* ao centro, os vértices foram coloridos em uma espécie de *escala de calor* que varia do amarelo ao vermelho, de acordo com a quantidade de conexões (grau). Os vértices mais periféricos, cuja centralidade de grau é menor, assumem tons amarelos. Os *hubs*, muito conectados, assumem tons vermelhos.

Os leiautes geográficos aplicam uma abordagem que permitem usar coordenadas geográficas com redes, mistura que possui um grande potencial. Essas redes podem tirar proveito da capacidade inata dos usuários de visualizar e interpretar dados espaciais, sobrepondo grafos em uma base geográfica.

Para que os vértices sejam projetados geograficamente, os dados de latitude e longitude devem estar presentes entre os seus atributos. Abaixo, a figura 8 ilustra a malha aérea dos Estados Unidos da América usando um grafo onde os aeroportos são vértices e as arestas são as rotas aéreas que os interligam.

Figura 8 – Leiaute geográfico de malha aérea dos Estados Unidos



Fonte: Extraído de Heymann (2011)

Outra possibilidade de visualização de redes é o uso da composição de dois leiautes em um só grafo, de forma que as vantagens de dois deles podem ser concentradas em uma única rede. Por exemplo, um leiaute geográfico pode ser combinado a um leiaute circular.

A pesquisa não almeja esgotar todas os leiautes aplicáveis aos grafos, pois há uma grande variedade deles, cada um com suas peculiaridades. Pretende-se apenas destacar o seu uso na análise exploratória criminal e preventiva mediante análise exploratória e unicamente visual (inicial, unicamente para sondar a rede e sem utilização de métricas). Outros leiautes, diferentes dos aqui listados, podem ser esmiuçados em casos concretos, a depender do Crime Objeto a ser explorado.

2.1.8 Bancos de dados baseados em grafos

Durante as últimas décadas os modelos de bancos de dados chamados de “relacionais” (RDBMS⁵) têm sido o paradigma padrão dessa área da tecnologia da informação. Esse modelo

⁵ Acrônimo para Relational Database Management Systems

foi concebido, inicialmente, com a ideia de armazenar planilhas em formatos tabulares. Essa tarefa é executada extremamente bem pelos modelos relacionais. Entretanto, com o avanço da tecnologia da informação, novas necessidades surgiram. Houve demanda por mais dados, mais agilidade, mais velocidade e mais interconectividade. A *World Wide Web* e, em particular, as redes sociais, onde as conexões são forte supedâneo, são um exemplo mais familiar dessa nova era.

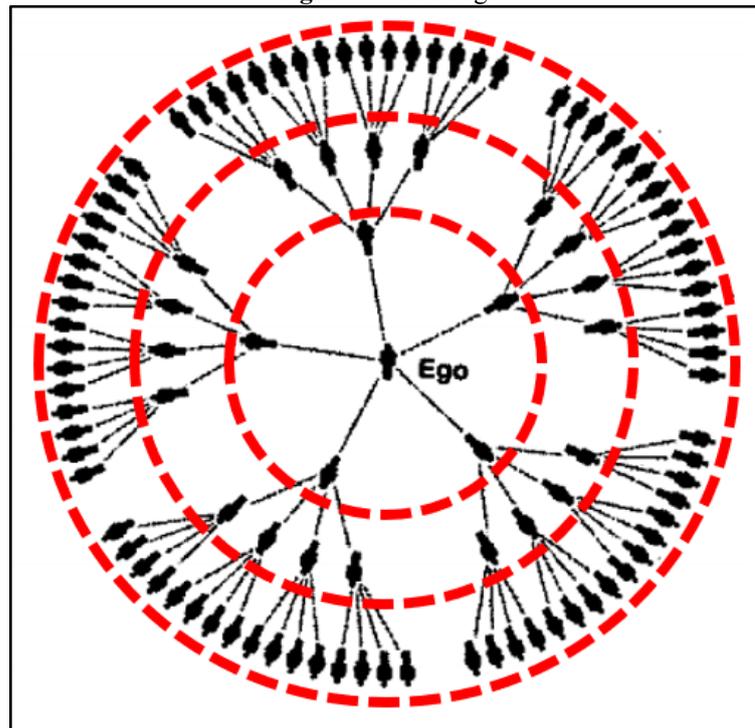
As demandas de interconectividade da informação não mais podem ser atendidas por bancos de dados relacionais por diversos motivos. Dentre eles, o fato de que há um conceito equivocadamente associado aos RDBMS. Ao contrário do que o próprio nome indica, os bancos de dados relacionais não lidam bem com relacionamentos concretos. O termo *relacional* remonta à álgebra de mesmo nome e, por isso, em nada se origina dos relacionamentos (ou interconexões) presentes nas entidades do mundo real que podem ser armazenadas em um banco de dados. Para suprir tal vácuo, surgiram os bancos de dados baseados em grafos, cujos primeiros adeptos reinventaram seus negócios em torno do valor das relações de dados. Estas empresas agora se tornaram líderes do setor: LinkedIn, Google, Facebook e PayPal.

Os bancos de dados relacionais são essencialmente formados por tabelas e, um dos problemas que eles são incapazes de solucionar são as junções aninhadas delas (tecnicamente conhecidas como “*joins*”), situação que regularmente causa performance degradada. Em alguns casos esse problema inviabiliza a obtenção de respostas. Noutra via, por ter o manejo de relacionamentos entre dados como foco principal, os bancos de dados baseados em grafos são eficientes no quesito velocidade de consulta. Nesse caso há excelente performance mesmo quando há dados interligados a serem recuperados mediante travessias recursivas e em profundidade, como ocorre rotineiramente nas redes sociais.

Numa rede social, denominamos os vizinhos de um vértice fixo como “*alters*” e, o nó central e focal, como *ego*. Hansen, Shneiderman e Smith (2011, p. 36), explicam que uma rede egocêntrica inclui apenas indivíduos (*alters*) que estão conectados a um ego específico. Embora as redes egocêntricas (ou rede ego, ou ego network) possam se estender qualquer número de graus de separação a partir do vértice focal, a rede básica dita *grau-1* consiste apenas no ego e seus alters.

A figura 9, a seguir, exemplifica uma rede *ego grau-3* (três graus de separação do vértice focal).

Figura 9 – Rede ego



Fonte: Extraído e modificado de Watts (2009, p. 20)

Vukotic et al. (2015, p.11) realizaram uma experiência para testar a performance um banco de dados relacional e em outro baseado em grafos. Basicamente o experimento consistiu em recuperar, em uma rede social básica, relações do tipo *amigos-de-amigos* (uma rede ego) a uma profundidade de cinco graus (ego grau-5). Seu conjunto de dados incluiu 1.000.000 (um milhão) de pessoas, cada delas com uma quantidade aproximada de 50 (cinquenta) amigos. Os resultados da experiência foram consolidados por Sasaki (2018), conforme a figura 10:

Figura 10 – Desempenho superior dos bancos de dados de grafos

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

Fonte: Extraída de Robinson, Webber e Eifrem (2015, p. 21).

Percebe-se, pelos resultados expostos na figura 10, a degradação de performance do uso de banco de dados relacionais em casos que exigem travessias recursivas em profundidades

(*Depth*) maiores que 4 (quatro) graus de separação. Assim, o tempo de execução inviabiliza as consultas. Entretanto, nos grafos, o tempo de execução permanece ótimo mesmo com o aumento da profundidade.

Os sistemas investigativos não são redes sociais, mas, no desenvolvimento da proposta surgem situações similares cujos desafios relacionados à travessia em profundidade são obstáculos a serem transpostos. Nesse caso os bancos de dados baseados no modelo relacional são incapazes de prover travessia em profundidade de vértices e, por isso, os grafos são a solução mais adequada.

2.1.9 Fundamentação estatística: medidas de centralidade e de dispersão

A média é uma medida de cálculo simples e a mais conhecida dentre as medidas de centralidade. Seu cálculo se resume a razão entre a soma das observações pela sua quantidade, conforme a seguir:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

A mediana, outra medida de centralidade, é a realização que ocupa a posição central da série de observações quando estão elas ordenadas em ascendência. É uma medida mais resistente (ou robusta), no sentido que ela não é muito afetada pelos valores discrepantes. O mesmo não ocorre com a média, que sofre distorções consideráveis causadas por eventuais *Outliers*. Quando a mediana está muito distante da média, os valores em questão não estão bem distribuídos ao redor desta.

Segundo Pinheiro et al. (2012, p. 247) a mediana pode ser descrita matematicamente como:

$$md(X) = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ ímpar,} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{se } n \text{ par} \end{cases}$$

Para um melhor entendimento sobre a dispersão e a tendência central de um conjunto de dados, é recomendável realizar o cálculo dos quartis. Devore (2006, p. 28) relata que os quartis dividem o conjunto de dados em quatro partes iguais e que aproximadamente 1/4 dos dados recaí sobre ou abaixo do primeiro quartil q_1 , metade dos dados sobre ou abaixo do

segundo quartil q_2 (a mediana) e aproximadamente 3/4 dos dados sobre ou abaixo do terceiro quartil q_3 .

Larson e Farber (2016, p. 81) afirmam que a distância (ou amplitude) interquartil (d_q) de um conjunto de dados é uma medida de variação que fornece a amplitude da porção central (aproximadamente metade) dos dados. A d_q , portanto, é a diferença entre o terceiro e o primeiro quartis, conforme abaixo:

$$d_q = q_3 - q_1$$

Segundo Bussab e Morettin (2010, p. 48), outliers são valores maiores do que o limite superior (Ls) e menores que o limite inferior (Li) os quais, por sua vez, são determinados respectivamente por:

$$\begin{cases} Ls = q_3 + 1,5d_q \\ Li = q_1 - 1,5d_q \end{cases}$$

Outra forma recomendável de compreensão de dados pode ser obtida por meio do estudo da sua dispersão ou variabilidade. A interpretação de um conjunto de observações estatísticas por uma única medida representativa de posição central pode ser distorcida porque não considera toda a informação sobre a variabilidade do conjunto. Alguns autores destacam os perigos do uso isolado das medidas de centralidade, tal qual demonstrado a seguir:

Informar apenas a medida de tendência central fornece apenas informações parciais sobre um conjunto de dados ou uma distribuição. Diferentes amostras ou populações podem ter medidas de tendência central idênticas e apresentar diferenças entre si em outros aspectos importantes (Devore, 2006, p. 31).

Essa dificuldade pode ser superada pelo uso das medidas de dispersão. Segundo Pinheiro (2012, p. 37) uma medida de dispersão para uma variável quantitativa é um indicador do grau de espalhamento dos valores da amostra em torno da medida de centralidade. Maiores dispersões indicam menos representatividade dos valores centrais.

Uma das medidas de dispersão, a variância populacional de um conjunto de dados com n elementos, segundo Larson e Farber (2016, p. 81) pode ser calculada da seguinte forma:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Uma das desvantagens da variância como medida de dispersão é que, por ser elevada ao quadrado, sua unidade é diversa daquela usada pelo conjunto de dados. A solução é o uso do desvio padrão, outra medida de dispersão. Segundo Haslwanter (2016, p.93) o desvio padrão é

calculado pela raiz quadrada da variância. De uma forma mais detalhada, o desvio padrão pode ser obtido por:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Um desvio-padrão perto de zero aponta que os pontos tendem a estar próximos da média e, quanto maior o desvio padrão, mais longe da média eles tendem a estar. O desvio padrão analisado por si, entretanto, não carrega consigo ideia da variação em relação à média dos dados. Segundo Pinheiro (2009, p. 32) a interpretação da magnitude do desvio padrão deve ser feita pelo coeficiente de variação, uma grandeza adimensional. Uma solução para esse óbice é a utilização do Coeficiente de Variação, definido por Larson e Farber (2016, p. 81) como sendo a razão entre o Desvio Padrão e a Média dos dados, conforme abaixo:

$$Cv = \frac{\sigma}{\bar{x}}$$

2.1.10 Notação BPMN

A notação BPMN (Business Process Model and Notation) versão 2.0 foi criada para especificar um padrão, uma linguagem comum para modelagem de processos negociais. Através dessa notação é possível usar símbolos universais inteligíveis por uma gama de profissionais. A BPMN 2.0 é amplamente difundida e tem uma excelente solução para a representação de processos, bem como uma baixa curva de aprendizado por ser predominantemente iconográfica.

O Object Management Group (2013,p .1) relata que o objetivo principal do BPMN é fornecer uma notação que seja prontamente compreensível por todos os usuários de negócios, desde os analistas de negócios que criam os rascunhos iniciais dos processos até os desenvolvedores técnicos responsáveis pela implementação da tecnologia que executará esses processos e, finalmente, às pessoas de negócios que irão gerenciar e monitorar esses processos. Assim, o BPMN cria uma ponte padronizada para a lacuna entre o design do processo de negócios e a implementação do processo. A notação BPMN será utilizada para gerar um fluxograma que represente o modelo que ora será proposto.

2.2 MÉTODOS E MATERIAIS

A pesquisa ora conduzida é de natureza quali-quantitativa, pois utiliza técnicas exploratórias mais sensoriais, mas também usa técnicas estatísticas. Em relação ao tipo, é classificada como exploratória, fato que resta evidenciado pela aplicação do modelo: um estudo de caso acerca de crimes relacionados ao desvio de recursos públicos. O modelo de prevenção

criminal proposto é um processo que pode ser adaptado a qualquer tipo de crime que envolva relações entre atores em uma rede social.

2.2.1 Fundamentação policial: seleção do CO, VFs, CVs e CRs

A etapa inicial do método exige reflexões acerca da fundamentação policial do crime que se quer prevenir. Esse processo engloba as seguintes tarefas: Seleção do Crime-Objeto (CO), Determinação das Variáveis Frisadas (VFs) e Critérios de Vulnerabilidade (CVs) e Classes Relevantes (CRs). Todos esses conceitos serão esmiuçados a seguir.

Crime-Objeto (CO) é a terminologia que o presente trabalho adotou para os delitos cuja ocorrência se quer prevenir, ou seja, os crimes candidatos à aplicação do modelo de prevenção proposto pela da pesquisa. A seleção do CO deve fundamentar-se em algum critério para que a escolha tenha importância e legitimidade. Como exemplo desses critérios, cita-se os delitos que causam maior dano ao bem-estar social, ou mesmo aqueles que incidem sobre determinados espaços (como é o caso do narcotráfico em aeroportos). Também é relevante que o CO incida sobre uma rede social cujos dados – ou ao menos parte deles – estejam acessíveis em sistemas informacionais, pois serão esses os insumos usados na concepção da prevenção.

O CO não é um delito em concreto, mas em abstrato. Isso porque ele ainda não ocorreu. Delitos consumados não são adequados à pesquisa porque o modelo ora proposto visa obter a prevenção daquilo que ainda não se materializou no mundo real. Assim, uma prisão em flagrante do indivíduo x por porte de drogas ilícitas no aeroporto y em uma data z não é um delito adequado para ser adotado como CO da aplicação da pesquisa. Entretanto, a *prevenção do uso de aeroportos como rota de narcotráfico* é um bom candidato a CO. Com vistas a prevenir o CO, é necessário entender quais são as variáveis que impactam na probabilidade de que esse delito ocorra, questão que se desvencilha a seguir.

A probabilidade da ocorrência delitiva pode aumentar ou diminuir por influência de alguns fatores. A pesquisa nominou esses fatores influentes como Variáveis Frisadas (VFs). Considerando que o enfoque da pesquisa é a prevenção criminal, os valores das VFs que contribuem para o aumento da probabilidade de crime são os mais relevantes. Ao contrário, os valores da VF que contribuem para que a probabilidade de crime diminua perdem importância no contexto preventivo da pesquisa. Para melhor entendimento, considere um caso concreto em que o CO é o de *narcotráfico internacional por meio de aeroportos*. Suponha que a estatística histórica demonstre que a compra de passagens aéreas pelos narcotraficantes, na maioria dos casos, ocorre em *até 48 horas antes do embarque*. Nesse caso, a VF pode ser descrita como sendo o *tempo de antecedência de compra da passagem*. O valor que aumenta as chances de

que o crime ocorra é *até 48 horas antes do embarque*, parâmetro que a pesquisa nominou como Critério de Vulnerabilidade (CV). Se o tempo de antecedência de compra é maior que 48 horas, a probabilidade de ocorrer o crime diminui. São ocorrências que não atendem ao CV que, por isso, passam a ser irrelevantes para a prevenção proposta pela pesquisa. As ocorrências que não atendem ao CV, portanto, não devem compor uma amostra vulnerável ao crime.

Obviamente, a análise isolada de um único CV é insuficiente para se afirmar que o cenário é vulnerável ao crime. Entretanto, vários CVs simultâneos podem aumentar significativamente as chances de que os delitos ocorram, o que justificaria a adoção de ações preventivas mais austeras. De forma compulsória, a escolha de cada um dos CVs deve ser motivada, caso contrário será fruto de aleatoriedade *nonsense*. Por isso, após listar cada um dos CVs, é necessário indicar quais foram as motivações que levaram a propô-los. Esse procedimento justifica a necessidade de prevenção.

Após a definição do CO, VFs e dos CVs, é necessário obter as Classes Relevantes (CRs). Elas são unidades de informação que possuem, em sua essência, conceitos acerca do domínio de conhecimento do Crime Objeto, ou seja, coisas importantes que se relacionam com o problema tratado. Normalmente as CRs são eventos, pessoas, objetos, locais ou coisas.

Para descobrir quais são as CRs, uma boa estratégia é gerar uma lista de sintagmas nominais ou de substantivos presentes nos textos que descrevem as VFs.

Segundo Pinilla, Rigoni e Indiani ([20--]):

A combinação das palavras para formarem as frases não é aleatória; precisamos obedecer a determinados princípios da língua. As palavras se combinam em conjuntos, em torno de um núcleo. E é esse conjunto (o sintagma) que vai desempenhar uma função no conjunto maior, que é a frase.

Eventualmente algum substantivo ou sintagma presente nas VFs não representará uma CR, mas essas poderão ser descartadas posteriormente se for demonstrada sua inutilidade.

Tomando ainda como exemplo o CO *narcotráfico internacional usando aeroportos*, suponha que as VFs e os CVs foram apontados por uma equipe de policiais. Com a finalidade de registrar esses apontamentos, a pesquisa propõe que seja criada uma lista que contenha o CO, as VFs e os CVs, conforme exemplificado pela tabela 1:

Tabela 1 – VFs e CVs para um tipo de CO

Crime-Objeto: Narcotráfico internacional em aeroportos ⁶		
Ordem	VFs (Variáveis Frisadas)	CVs (Critérios de Vulnerabilidade)
1º	[NÚMERO DE PASSAGEIROS] da [RESERVA]	Deve ser igual a 1 (um), o que significa que o passageiro está desacompanhado
2º	[PERMANÊNCIA] no [AEROPORTO DE DESTINO]	Voo de retorno em até 48 horas após o voo da chegada
3º	[ROTA] entre o [AEROPORTO DE ORIGEM] e [AEROPORTO DE DESTINO]	Devem ser uma das rotas já mapeadas como de uso comum pelo narcotráfico internacional
4º	[ANTECEDÊNCIA DA RESERVA] dos [VOOS DO TRECHO]	Passagem comprada em menos de 48h antes do embarque

Fonte: Elaborado pelo autor

É recomendável que, tal como na tabela 1, as Classes Relevantes sejam identificadas no texto das Variáveis Frisadas (por maiúsculo, colchetes, negrito ou outro recurso de ênfase textual). Alguns Sintagmas Nominais podem ser quebrados em partes menores – os substantivos – para aumentar a especificidade das CRs. Outros Sintagmas podem se fundir e criar uma única CR.

No caso acima, um bom conjunto de Classes Relevantes seria: Número de Passageiros, Passageiro, Reserva, Permanência no Aeroporto de Destino, Aeroporto de Destino, Aeroporto de Origem, Aeroporto, Trecho, Voo, Rota. Determinadas as CRs, é necessário informar todas as motivações que levaram a escolha dos Critérios de Vulnerabilidade. Essas motivações legitimam as escolhas. A seguir, listam-se alguns exemplos simplificados dessas motivações.

Uma motivação plausível para a 1º CV (passageiro deve estar desacompanhado) seria: *Embora possa ocorrer exceções, os narcotraficantes normalmente não viajam em grupos formados por membros da família, amigos ou colegas de trabalho. Há dados estatísticos que demonstram essa tendência de viajarem sozinhos.*

Uma motivação plausível para a 2º CV seria: *Os passageiros contratados para transportar drogas, após a entrega do entorpecente no destino, procuram retornar o mais rápido possível aos aeroportos de origem. Há dados estatísticos que demonstram que, na esmagadora maioria dos casos, o voo de retorno à origem estava programado para um período não superior a 48 horas após o voo de ida ao destino.*

Obviamente as motivações acima são apenas exemplos imaginários e simplificados. Nos casos reais os motivos devem ser mais longos e detalhados. Podem incluir estatísticas e

⁶ As Variáveis Frisadas, os Critérios de Vulnerabilidade e o Crime Objeto da Tabela de Vulnerabilidades são meramente exemplificativos e figurativos. Em nada se relacionam com casos reais.

outros conceitos advindos do empirismo policial. Cada um dos CV deve ter uma motivação até que todos sejam exauridos.

A escolha das VFs e dos CVs é uma tarefa baseada em paradigmas construídos pelo empirismo investigativo e, portanto, fundamentada na prática desde muito vivenciada pelo policial experiente, um ser-repositório de modelos mentais obtidos junto ao ambiente laboral que o cerca. Essa é, precisamente, a manifestação real do modelo cognitivo da Ciência da Informação proposto por Capurro (2003, p. 8).

2.2.2 Preparação da Rede

A etapa de Preparação das Redes diz respeito a coleta e tratamento dos dados, a modelagem e criação da rede Supergrafo usando os dados coletados e a extração da Rede Vulnerável.

Inicialmente, para realizar a coleta e tratamento dos dados, é necessário entender as Classes Relevantes como representações das coisas importantes que se relacionam com o CO. Essas Classes em geral possuem também relevância registral na sociedade e, por isso, são assinaladas em sistemas informacionais que contém os dados a serem coletados. Como exemplo, cita-se um passageiro de um voo: para fazer usufruto do sistema de transporte aéreo, seu registro deve, compulsoriamente, ser depositado nos bancos de dados das companhias permissionárias que operam as linhas comerciais aeroportuárias.

Para que a CR seja útil a prevenção criminal, seu registro nos bancos de dados deve estar acessível aos investigadores. Sem esse acesso, os dados não poderão ser coletados, o que inviabiliza qualquer estratégia de prevenção baseada na CR. Por exemplo, se os dados da CR *Passageiro* estiverem inacessíveis aos policiais, resta inviabilizada qualquer estratégia de prevenção que a utilize.

Garantido o acesso aos bancos de dados, a tecnologia usada na coleta pode ser qualquer uma dentre as disponíveis. A única exigência é que seja capaz de extrair registros dos Bancos de Dados e salvá-los em formatos universais, tais como planilhas, arquivos-texto tipo *comma-separated values* (valores separados por vírgulas – os arquivos *.csv*), etc. Não é objetivo da pesquisa esmiuçar esses aspectos tecnológicos, pois os processos de extração, transformação e carga de dados são comuns, de forma que o detalhamento deles é desnecessário ao escopo do presente trabalho. Isso porque a pesquisa não tem essência nos métodos mecânicos e tecnológicos, mas naqueles que são predominantemente conceituais e cognitivos.

Coletados os dados de todas as Classes Relevantes, a próxima etapa é transformá-los em uma rede. A parte mais importante de uma rede são suas interconexões, não o detalhamento esmiuçado das informações isoladas - função essa que já é exercida pelos sistemas originais de onde os dados advêm. Assim, a coleta de dados para criação da rede não se destina a gerar cópias completas dos bancos de dados dos sistemas originários. Ao contrário, adota uma estratégia minimalista no destino: apenas as informações cruciais serão alvo de reprodução. Em consequência o volume de dados armazenado na rede (o repositório destino) será consideravelmente reduzido, pois, em geral, a maior parte das informações dos sistemas originais é desnecessária. A prioridade de coleta são os campos que identificam as entidades de forma única (campos-chave), os que estabelecem as interconexões e os que são necessários para aplicação dos Critérios de Vulnerabilidade das Variáveis Frisadas.

Após a coleta, o tratamento dos dados deve remover eventuais inconsistências e valores nulos. Em seguida, os dados devem ser modelados como uma rede denominada Supergrafo, etapa que será desvincilhada a seguir.

Finda a fase de coleta e tratamento dos dados das Classes Relevantes, faz-se necessário transformá-los em uma rede, a qual foi nominada como Rede Supergrafo (RS). Ela é expressão dos dados coletados em um formato que pode ser denominado de *network*. Criar a RS faz parte da estratégia que permitirá que uma análise futura possa extrair dela uma Rede Vulnerável ao CO, a qual será alvo de monitoramento preventivo.

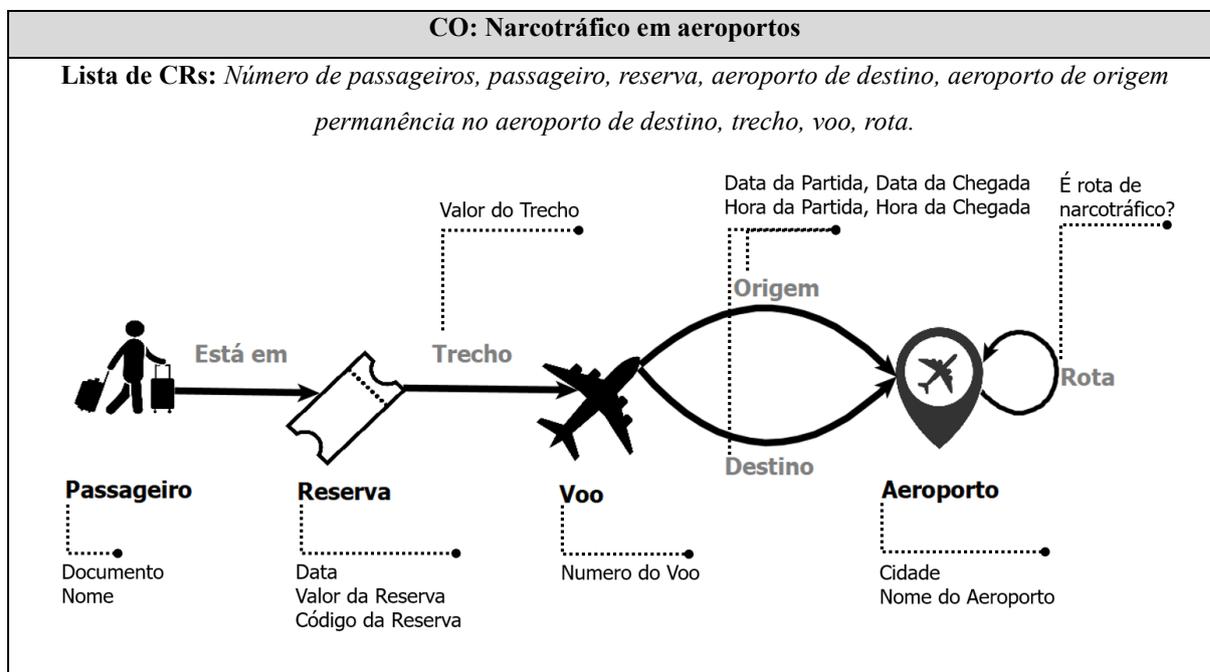
A RS é inteiriça, contém todos os dados advindos dos sistemas originários, o que significa dizer que nela não foram aplicados os filtros nominados como Critérios de Vulnerabilidade. Antes de preencher a RS com os dados coletados, é necessário projetá-la em um modelo denominado *Grafo de Modelagem*. A primeira ação para a criação do Grafo de Modelagem é a analisar a lista de CRs. Cada uma delas será representada, no Grafo, por uma dentre as escolhas abaixo:

- Um vértice de um grafo;
- Uma aresta entre vértices;
- Um atributo de um vértice ou de uma aresta;
- Uma métrica de um vértice ou de uma aresta.

A escolha será baseada na semântica de cada CR. Por exemplo, uma *rota aérea* pode ser considerada como uma aresta (ou ligação) entre dois vértices do mesmo tipo (o aeroporto de origem e o aeroporto de destino). Nesse caso, a rota será representada por uma aresta

direcionada que interliga dois vértices de aeroportos. A figura 11 ilustra um Grafo de Modelagem completo:

Figura 11– Exemplo de Grafo de Modelagem⁷ de uma RS



Fonte: Elaborado pelo autor

No grafo de modelagem da figura 11, percebe-se que um (ou mais de um) passageiro está numa reserva que, por sua vez, pode ter vários trechos. Cada trecho está ligado a um voo. O voo possui um aeroporto de origem e um aeroporto de destino. Esses aeroportos podem ser (ou não) uma rota de narcotráfico.

Agora que o Grafo de Modelagem está pronto, a próxima etapa é, seguindo rigorosamente seu modelo, criar via software um banco de dados (BD). Dentre os disponíveis, não à toa foi escolhido um software baseado em grafos. Isso porque a natureza interconectada dos dados usados nas investigações policiais torna a capacidade de travessia um requisito importante à recuperação da informação. Conforme a fundamentação teórica expôs, a capacidade de travessia é uma característica dos grafos.

A simples representação textual da situação do exemplo já demonstra a capacidade de travessia como requisito *sine qua non*: **passageiro** → **reserva** → **trecho** → **voo** → **aeroporto** → **aeroporto**. Ademais, conforme a fundamentação teórica da pesquisa apontou, a travessia é

⁷ Embora as informações referentes a passageiros, reservas, voos e aeroportos possam ser mais complexas do que o exposto no Grafo de Modelagem do exemplo, o propósito da figura 11 é meramente ilustrativo e, por isso, a simplificação das informações foi proposital.

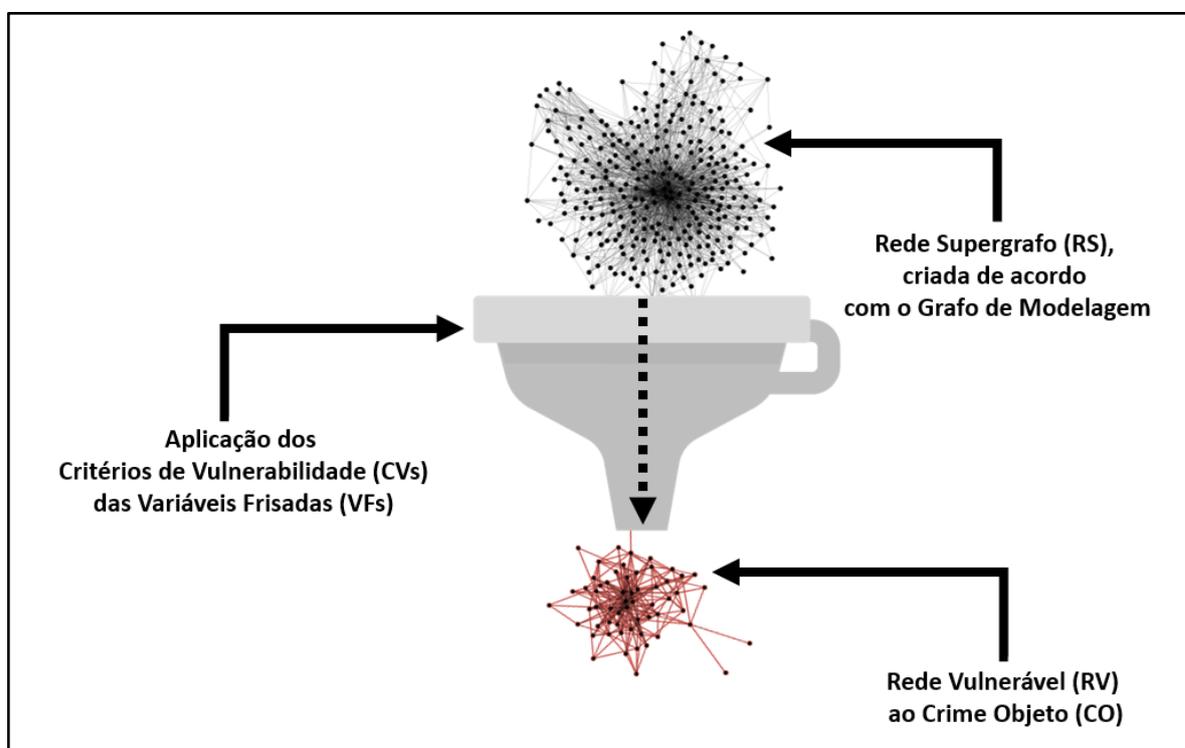
uma característica intrínseca aos grafos desde a sua concepção baseada nas pontes de Königsberg.

Após a criação do BD, a carga dos dados originais pode ser executada por *scripts*. Os detalhes tecnológicos acerca da criação desses *scripts* dependem do tipo de Banco de Dados e essa minúcia tecnológica e mecânica foge ao escopo essencial da pesquisa.

Ao término da criação da Rede Supergrafo, se o objetivo é prevenir crimes, pergunta-se: quais partes dessa rede estão mais vulneráveis ao Crime-Objeto? A resposta a essa pergunta tem como resultado outra rede, filtrada, menor e mais específica, resultante da aplicação dos CVs: a Rede Vulnerável.

Agora que a RS foi criada, a próxima tarefa é a extração da Rede Vulnerável (RV) ao CO. Ela é um subgrafo da RS, mas formada apenas pelos vértices e arestas que atendem ao CVs, os quais funcionam como filtros que capturam apenas as situações que aumentam a probabilidade que o delito ocorra. Para melhor entendimento da RV, é necessário percebê-la como o resultado de uma filtragem da RS, conforme ilustra a figura 12:

Figura 12 – RV: uma filtragem da RS

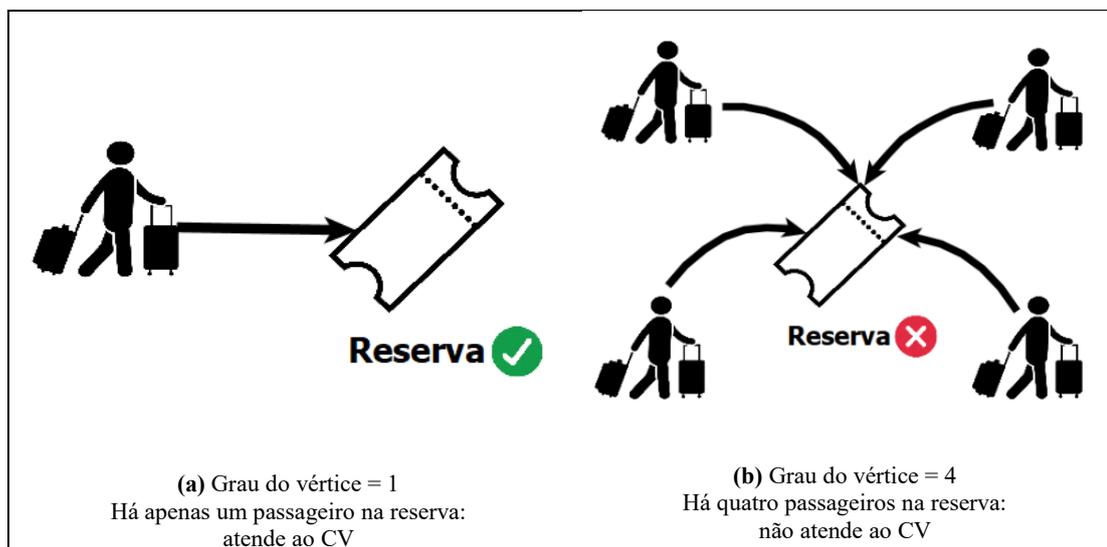


Fonte: Elaborado pelo autor

Assim, a aplicação dos CV à RS gera a RV. Considerando que a RS é formada apenas por vértices e arestas, para nela aplicar os CVs é necessário mapear sobre quais métricas ou atributos de vértices e arestas esses filtros irão incidir. A melhor forma de fazer esse

mapeamento é o uso de imagens iconográficas que representam a aplicação do CV. Como exemplo, a figura 13 ilustra um cenário em que a Variável Frisada Número de passageiros da reserva é obtida pela métrica Grau de Entrada de um vértice. O CV estabelece que essa VF seja igual a 1(um), o que significa que o passageiro está desacompanhado em sua reserva.

Figura 13 – Aplicação do 1º CV [Número de passageiros da reserva = 1]



Fonte: Elaborado pelo autor

Após a criação das imagens que representam a aplicação de todos os CVs, a filtragem da Rede Supergrafo já pode ser executada, e a forma de executar esse filtro depende do software adotado para a implementação da rede, método que não é objeto da pesquisa. A aplicação de todos os CVs à Rede Supergrafo terá como resultado a Rede Vulnerável. Agora o policial dispõe de duas redes que podem ser alvo de análise exploratória e estatística: RS e RV.

2.2.3 Análise exploratória das redes

Terminada a preparação das duas redes (Supergrafo e Vulnerável), é necessário fazer análise de ambas. A análise exploratória é uma técnica apropriada para obter *insights* (intuições) sobre as redes, especialmente quando se usa leiautes de grafos, cuja aplicação não altera os dados e muito menos as estatísticas a eles associadas. Leiautes são meras formas diferentes de visualizar os mesmos dados, sem alterá-los.

Os *insights* não são frutos apenas dos olhares aleatórios. Eles devem ser pensados como um meio de demonstrar que a rede analisada possui características, ainda que inicialmente intuitivas, que consigam apontar que a sua concepção tem um fim. No caso da pesquisa, o olhar sobre as redes deve procurar *insights* que revelem características quer permitam usá-las para prever crimes. Com olhar atento às VFs e aos CVs, o policial deve tentar obter intuições úteis à prevenção do Crime-Objeto.

As intuições que cada um dos leiautes pode fornecer podem ser drasticamente diferentes, a depender do resultado visual obtido. É uma tarefa que requer uma certa prática reiterada para ser aprimorada. A análise exploratória visual, como parte do método, permitiu concluir que apenas a repetição aumenta a capacidade intuitiva do policial em perceber se um leiaute visual será útil ou não. E essas repetições são livres, considerando que a aplicação dos leiautes não implica em modificação dos dados. Importante ressaltar que as métricas de rede independem do leiaute, entretanto a abordagem da pesquisa visa primeiramente obter insights visuais, e só depois confirmá-los através das métricas e das estatísticas associadas.

Na aplicação do modelo proposto, portanto, é normal que ocorra um período significativo de testes e descobertas obtidas pelo aprendizado visual dos leiautes. Ocorrem pequenos – e até desejáveis – fracassos cognitivos decorrentes da aplicação de leiautes inadequados ou desajustados. O processo de erro e acerto tem o potencial de, aos poucos, refinar as escolhas. Ao término, o *feeling*⁸ policial aprimorado pela repetição resulta na descoberta de oportunidades visuais cognitivas. As tentativas infrutíferas iniciais, portanto, não devem ser encaradas com frustração, mas, ao contrário, como janelas de oportunidades rumo à consolidação do aprendizado visual. Assim, ao fim e ao cabo, o êxito será decorrência desse caminho a ser percorrido. É necessário considerar, na aplicação do modelo, que a avaliação visual de uma rede possui um componente subjetivo próprio e decorrente do julgamento visual e interpretativo. Entretanto, algumas recomendações sistemáticas do método ora proposto podem fornecer uma orientação inicial.

Uma recomendação essencial é que o policial deve conhecer muito bem o Crime Objeto, as Classes Relevantes, as Variáveis Frisadas e os Critérios de Vulnerabilidade. O Grafo de Modelagem, peça já explanada e que dá uma visão iconográfica das CRs, também deve ser familiar ao policial. Conhecer o funcionamento conceitual dos leiautes que se pretende usar também é fundamental.

Recomenda-se que, sempre que o objetivo for entender a dinâmica de atributos numéricos expressos pelas Variáveis Frisadas nas redes, considere vinculá-los a três aspectos visuais cujas intensidades podem ser alvo de quantificação: o tamanho do vértice e a espessura da aresta, bem como as cores de ambos. Quanto à cor, é possível utilizar uma *escala de calor de cores* para expressar valores progressivos (por exemplo: valores menores iniciando com

⁸ Intuição que nos leva por um determinado caminho

branco, passando por amarelo, laranja, vermelho claro, vermelho escuro e, finalmente, os maiores valores em preto).

Ainda em relação aos atributos numéricos, é aconselhável verificar se alguma métrica da análise de redes sociais não já os traz consigo, de forma intrínseca e no próprio grafo. Isso evitará esforço desnecessário de buscar de informações que já estão na própria rede. Por exemplo, o número de amigos de uma pessoa x em uma rede social é expressa pela métrica relativa ao grau do seu vértice. Assim, o grafo já faz essa contagem de forma intrínseca no próprio vértice que representa x , sendo desnecessário buscá-la por outros meios.

Em relação ao uso de cores das arestas nos grafos direcionados, importante salientar que, se o objetivo é destacar o vértice destino, a aresta deve acompanhar a cor desse vértice. Caso contrário, se o objetivo é destacar a origem, a cor da aresta será igual à do vértice onde a aresta se origina.

Se a percepção que se busca é de ordenação de vértices no espaço de plotagem, o leiaute *circular* ou o *dual circle* são boas alternativas. Entretanto, eles são mais úteis quando não há uma grande quantidade de vértices, caso contrário o leiaute acaba resultando em um visual poluído, que empobrece a semântica.

Tanto o *circular* como *dual circle* são leiautes que trabalham bem quando essa quantidade chega a, no máximo, algumas centenas de vértices e arestas. Assim, é interessante usá-los apenas na Rede Vulnerável, já que seu tamanho é menor que a Rede Supergrafo.

O leiaute *circle-pack* deve ser utilizado para evidenciar destacar hierarquias e grupos. Ao contrário dos leiautes circular e dual circle, o circle-pack consegue representar bem uma grande quantidade de vértices. Então, é possível utilizá-lo com desenvoltura tanto na rede Supergrafo quanto na Rede Vulnerável.

Por sua vez, os leiautes geográficos têm como pré-requisito a disponibilidade das coordenadas geográficas na rede, caso contrário seu uso estará inviabilizado. Recomenda-se usá-lo quando o objetivo for destacar ocorrências criminais georreferenciadas, sejam elas preventivas ou repressivas. Nesse caso a escolha dos tamanhos dos vértices deve ser cuidadosa para que respeite a dimensão do mapa. Isso evitará localizações distorcidas.

2.2.4 Análise estatística das redes

Após a análise exploratória dos grafos, é necessário realizar a análise estatística. O objetivo principal é confirmar os *insights* e percepções oriundas da análise exploratória, pois essa, apesar de direcionar o investigador, carece de confirmação estatística mais precisa.

Para realizar uma análise estatística dos dados é aconselhável que o investigador se utilize métodos da estatística descritiva. Segundo Devore (2006, p. 3):

Outros métodos descritivos envolvem o cálculo de medidas numéricas, como médias, desvios padrão e coeficientes de correlação. A ampla disponibilidade de pacotes de softwares estatísticos facilitou bastante essas tarefas. Os computadores são muito mais eficientes que os seres humanos em cálculo e na criação de imagens (depois de terem recebido as instruções corretas do usuário). Isso significa que o investigador não tem de perder muito tempo com “trabalho braçal” e terá mais tempo para estudar os dados e extrair mensagens importantes.

De forma mais objetiva, segundo Larson e Farber (2016, p. 5), a estatística descritiva é o ramo da estatística que envolve a organização, o resumo e a representação dos dados. Larson e Farber (2016, p. 5) também descrevem a estatística inferencial, segundo os quais é o ramo da estatística que envolve o uso de uma amostra para chegar a conclusões sobre uma população. Assim, o uso das estatísticas descritivas e inferenciais podem trazer conhecimento mais precisos acerca das características importantes sobre as redes, algumas delas já evidenciadas instintivamente pela análise exploratória.

Os conceitos básicos de estatística descritiva foram apresentados na fundamentação teórica. Outros conceitos oriundos da estatística inferencial também podem ser incorporados. Entretanto não é objetivo da pesquisa descrever todas as inúmeras possibilidades estatísticas disponíveis ao uso. A escolha de quais técnicas usar dependerá do campo de domínio de conhecimento acerca do Crime-Objeto.

A análise estatística pode ser realizada na Rede Supergrafo ou na Rede Vulnerável. Em qualquer uma delas, um dos objetivos dessa análise é confirmar estatisticamente os *insights* obtidos nas análises exploratórias dessas redes. Em particular, no caso da Rede Vulnerável, a análise estatística também serve, compulsoriamente, para demonstrar que o seu monitoramento preventivo é necessário. Isso porque essa rede é mais suscetível ao Crime-Objeto.

2.2.5 Avaliação da Rede Vulnerável

Ao término das análises exploratória e estatística, o policial deve avaliar se o conhecimento adquirido com ambas as redes comprovou que a RV apresenta características que a diferenciam e que a tornam mais suscetível ao crime. Caso o policial não encontre as

características que diferenciam a RV, significa que ela não é dissimilar aos padrões normais e, desta forma, não há motivos para monitorá-la preventivamente. Nesse caso, certamente, houve erros na determinação das Variáveis Frisadas, dos Critérios de Vulnerabilidade ou das Classes Relevantes. As tarefas de delimitação desses itens deverão ser novamente executadas de forma mais criteriosa. O fluxo deve retornar a essa etapa inicial, já explanada anteriormente.

Encontrando as características que a diferenciam, o policial terá justificado o monitoramento preventivo da RV pela sua conformidade aos propósitos pelos quais ela foi concebida: a prevenção contra o Crime-Objeto. Nesse caso, as VFs e os CVs não precisam ser reavaliados e o fluxo do processo pode seguir à próxima tarefa: a entrega da RV às equipes de policiamento preventivo. O resultado final da aplicação do modelo é a prevenção cotidiana incidente sobre a própria Rede Vulnerável.

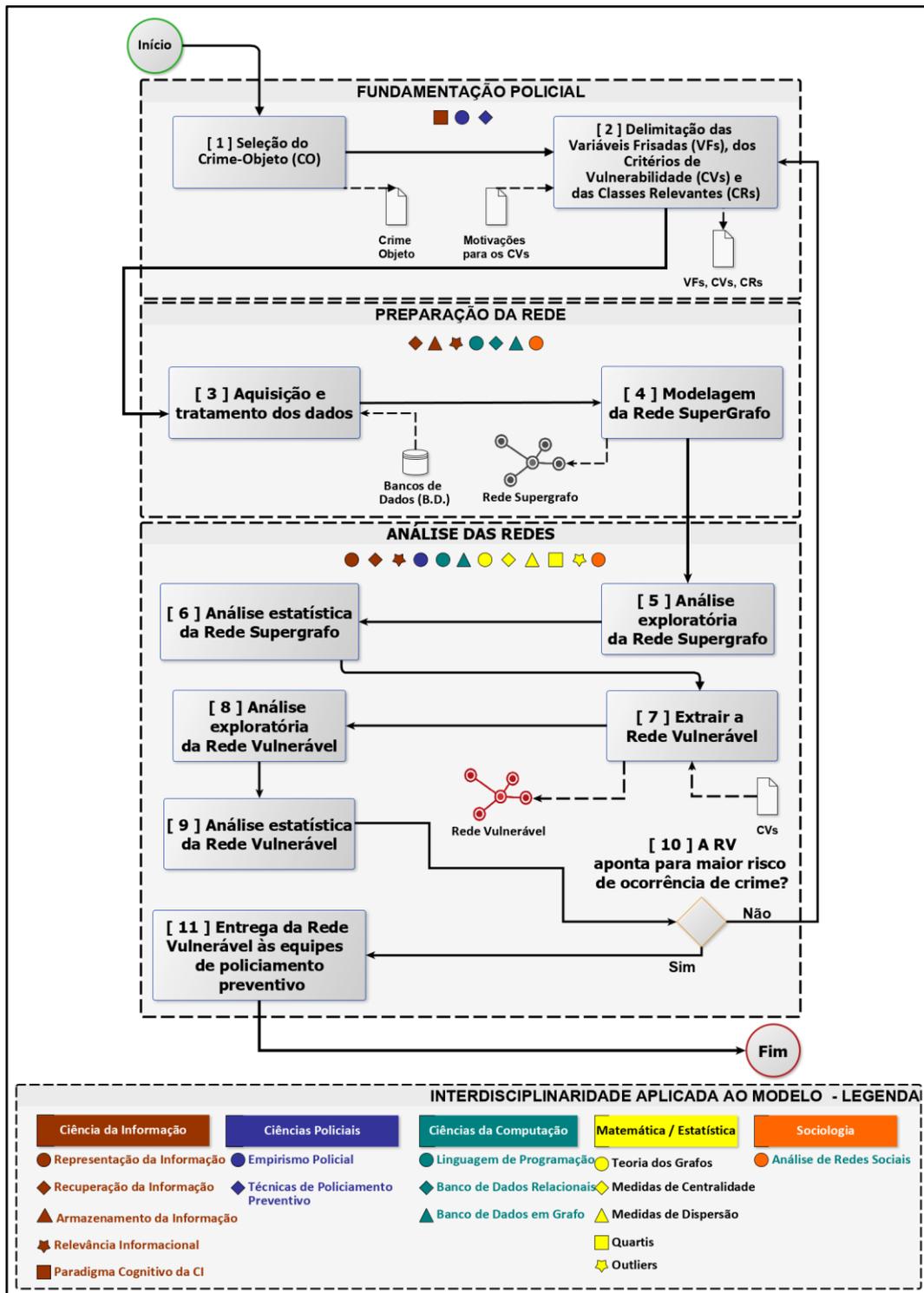
As equipes de policiamento preventivo, ao receber a Rede Vulnerável, iniciam as atividades de prevenção. As formas como o policiamento preventivo será realizado (diligências, compartilhamento com órgãos de auditoria ou setores de inteligência, etc.) não são objeto do presente trabalho. Em alguns casos, essas equipes podem ser compostas por servidores de órgãos de controle administrativo e auditoria, pois não há exclusividade policial a fatos candidatos a crimes, pois esses são apenas informações de inteligência. O termo policiamento preventivo, nesse caso, tem sentido *lato senso*, não se referindo especificamente a equipes exclusivamente formadas por policiais.

3 RESULTADOS

3.1 FLUXOGRAMA DE POLICIAMENTO PREVENTIVO POR MEIO DE GRAFOS E ARS

Após a definição do método, todas as tarefas foram formalizadas em um fluxograma modelo. Trata-se de um processo composto por uma sequência de tarefas bem definidas e que podem ser adaptadas a qualquer tipo de CO cujos dados estejam manifestos em sistemas informacionais. A notação adotada no fluxograma é a BPMN. Há amplo uso da interdisciplinaridade característica das pesquisas em CI. O modelo de prevenção baseado em grafos e ARS segue o fluxograma da figura 14, a seguir:

Figura 14 – Modelo de Policiamento Preventivo por meio de Grafos e ARS (em BPMN)



Fonte: Elaborado pelo autor

Percebe-se que o modelo inicialmente executa um conjunto de tarefas nominadas como *Fundamentação Policial*. Em seguida, a *Preparação da Rede*. As tarefas posteriores se relacionam à *Análise das Redes* e, em seguida, o fluxo termina (ou retorna à *Fundamentação Policial*, a depender da decisão expressa no passo 10).

O modelo utiliza ampla interdisciplinaridade em cada grupo de tarefas, conforme revela a legenda, a qual contém cada ciência ou disciplina representada por um ícone de determinada forma e cor.

Para melhor entendimento da pesquisa, foi executado um estudo de caso real para o Crime-Objeto desvio de recursos públicos, o qual será esmiuçado a seguir.

3.2 ESTUDO DE CASO – APLICAÇÃO DO MODELO PROPOSTO

O estudo de caso aplica todas as etapas apresentadas no fluxograma da figura 14. A discussão desses resultados ocorre de forma simultânea à aplicação do modelo.

3.2.1 Seleção do Crime-Objeto

O Brasil tem sido alvo de verdadeira pilhagem aos cofres públicos em todas as esferas de poder: o Ministério Público Federal estima em 200 bilhões em prejuízos causados por esse tipo de crime aos cofres estatais brasileiros anualmente⁹. Parte considerável desses delitos são operacionalizados através das contratações estatais de fornecedores sob manto de legalidade aparente, todavia acompanhada da obscura utilização de artifícios criminosos.

Pelas razões expostas, o delito de desvio de recursos públicos foi escolhido como Crime-Objeto em um estudo de caso para testar o modelo. A principal fonte dos dados são as contratações onde os poderes executivos de cada município de uma unidade da federação (habitualmente conhecidos como prefeituras municipais) figuram como órgãos públicos contratantes. Por sua vez, as empresas privadas figuram como seus fornecedores bens ou prestação de serviços, todos pagos com recursos públicos.

Doravante, no decorrer do estudo de caso, essas prefeituras municipais são nominadas simplesmente como *órgãos públicos*. As empresas, por sua vez, são nominadas como *fornecedores*.

3.2.2 Delimitação de VFs, CVs e CRs

Determinado o Crime-Objeto, a próxima tarefa, seguindo o fluxograma do modelo (ilustrado pela figura 14), é a delimitação de VFs, CVs e CRs.

O Crime Objeto do estudo de caso incide sobre uma rede social de contratações públicas, formada por fornecedores, pagamentos, órgãos públicos, etc. Três variáveis frisadas foram fixadas. Com base nessas VFs, foram definidos os CVs e as CRs. A tabela 2, abaixo, exhibe a lista de VFs e CVs.

⁹ Segundo <https://istoe.com.br/brasil-perde-cerca-de-r-200-bilhoes-por-ano-com-corrupcao-diz-mpf/>

Tabela 2 – VFs e CVs para o CO do estudo de caso

Crime-Objeto: Desvio de recursos públicos		
Ordem	VFs (Variáveis Frisadas)	CVs (Critérios de Vulnerabilidade)
1º	[FORNECEDOR] possui (ou não) ao menos um [SÓCIO] [PESSOA FÍSICA] com [INDICIAMENTO] em [INQUÉRITOS POLICIAIS]	Sim, fornecedor possui sócios indiciados
2º	[VALOR EM RECURSOS PÚBLICOS] pagos a cada [FORNECEDOR]	Fornecedor recebeu uma grande quantia em recursos públicos
3º	[QUANTIDADE DE ÓRGÃOS PÚBLICOS CONTRATANTES] de cada [FORNECEDOR]	Fornecedor foi contratado por uma grande quantidade de órgãos públicos

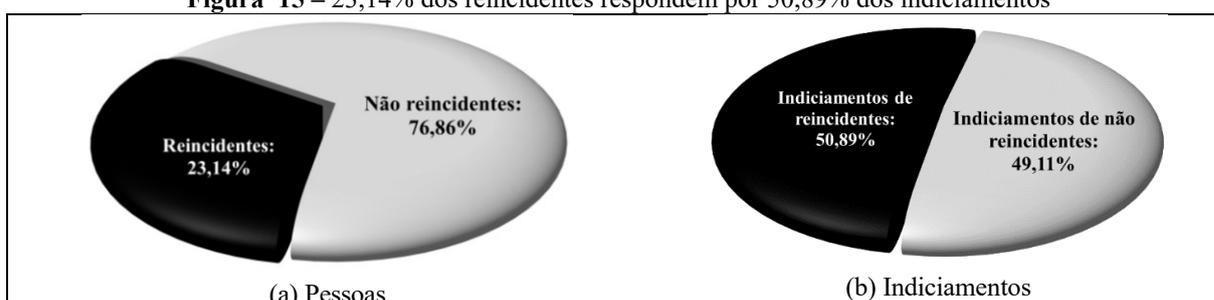
Fonte: Elaborado pelo autor

As CRs identificadas são: fornecedor, inquérito policial, órgãos públicos contratantes, indiciamento, recursos públicos, sócios, pessoas físicas. Seguindo as orientações do método, a próxima etapa prevê a motivação que levou a escolha de cada um dos Critérios de Vulnerabilidade, conforme se esmiúça a seguir.

3.2.3 Motivação do 1º CV: fornecedor possui sócios indiciados

Durante o estudo de caso, a reincidência criminal se evidenciou, estatisticamente, como um fator relevante na detecção de crimes futuros. Isso porque há um grande percentual de novos crimes praticados por pessoas outrora já acusadas (reincidentes). Para chegar a tal conclusão, foram analisados os dados dos indiciamentos criminais de uma Superintendência de Polícia Federal de uma Unidade da Federação (UF) brasileira. Nesse repositório, das 1.327 (mil trezentas e vinte e sete) pessoas indiciadas em pelo menos um inquérito policial, 307 delas (trezentas e sete), ou seja, 23,14% do total, possuíam mais de um indiciamento, o que significa que são reincidentes. E, embora representem uma minoria do total de pessoas indiciadas, os reincidentes respondem a um total de 1.056 (mil e cinquenta e seis) indiciamentos, ou seja, 50,89% de um total de 2.075 (dois mil e setenta e cinco). Assim, na população considerada, os reincidentes, os quais representam 23,14% das pessoas respondem por mais de 50% de todos os indiciamentos. Esse cenário é condensado na figura 15, abaixo:

Figura 15 – 23,14% dos reincidentes respondem por 50,89% dos indiciamentos



Fonte: Elaborado pelos autores

Um questionamento natural pode vir à tona: qual a relação entre a reincidência criminal e o crime objeto escolhido (desvio de recursos públicos)? Para respondê-la, é preciso considerar a possibilidade de as pessoas indiciadas serem - ou virem a se tornar - sócias de empresas que, por sua vez, são contratadas por órgãos públicos, ou seja, fornecedores estatais. Assim, como o indiciamento é um fator que aumenta as chances de reincidência, podemos concluir que as contratações de pessoas previamente acusadas aumentam as chances da prática de desvios de recursos públicos pelos envolvidos nesses contratos.

Outro provável questionamento é aquele segundo o qual o fornecedor – uma pessoa jurídica contratada pelo Estado – não se confunde com as pessoas físicas que compõem seu quadro societário. Em resposta, entretanto, considere que uma pessoa jurídica é uma ficção social que não possui consciência e vontades *per si*. Nesse sentido, o fornecedor, como pessoa jurídica que é, pratica atos pela orientação dos sócios, sendo um produto das suas decisões conscientes deles voltadas a um fim específico. Por esse motivo, e também por simplificação textual, doravante o texto nominará as empresas que possuem em seu quadro societário pessoas indiciadas por crimes simplesmente como *fornecedores acusados*.

Considerar os indiciamentos como um fator de reincidência é uma questão que implica em outro dilema: a presunção da inocência resguarda, inclusive, as pessoas indiciadas, o que em tese impediria tratamento diferenciado aos fornecedores previamente acusados. Isso não invalida o estudo de caso, pois não há previsão de consequências negativas – sejam penais ou administrativas. Também não há intenção de supressão ou mitigação de direitos que prejudiquem os fornecedores acusados. O que se pretende é tão somente o maior controle, pelos órgãos de fiscalização e policiamento, sobre as contratações mais vulneráveis. Tal medida é o zelo mínimo que se espera no tocante ao erário. A primeira VF, portanto, é o indicativo booleano (binário ou lógico) que aponta se o fornecedor possui sócios acusados de crimes. Seu CV corresponde ao valor “SIM”, ou seja, apenas os fornecedores com acusações atendem ao referido critério.

3.2.4 Motivação do 2º CV: fornecedor recebeu grande quantia em recursos públicos

A segunda VF escolhida é o total em recursos públicos recebidos por cada fornecedor, independentemente da quantidade de pagamentos ou contratações. O Critério de Vulnerabilidade exige que o fornecedor tenha recebido uma grande quantia. Receber valores vultosos é um objetivo dos criminosos como forma de compensação aos perigos penais e sociais decorrentes do delito (reprovação social, prisão e condenação). Portanto, essa variável numérica é impactada pelo comportamento do criminoso. Entretanto, resta definir precisamente

o que é uma grande quantia. Não se pode responder tal questão sem analisar os dados. O valor, portanto, permanece em aberto até que os dados sejam coletados e analisados estatisticamente. A partir de então será possível afirmar qual montante é um piso inferior a partir do qual todos os valores excedentes serão considerados grandes quantias, ou seja, atendem ao 2º Critério de Vulnerabilidade.

Salienta-se que o mero recebimento de grandes quantias em recursos públicos por si é um indicador insuficiente para aumentar a probabilidade criminal, pois os indivíduos sem intenções delitivas também buscam auferir ganhos elevados. Entretanto, esse não é o único CV que será aplicado para filtrar a RS. Para obter a RV, aplicam-se como filtros todos os CVs, não apenas o ora analisado.

3.2.5 Motivação do 3º CV: ser fornecedor de uma grande quantidade de órgãos

A terceira VF é a quantidade de órgãos públicos contratantes de cada fornecedor. O CV, por sua vez, exige que o fornecedor tenha sido *contratado por uma grande quantidade de órgãos públicos*. Tal critério, para que tenha uma definição precisa, precisa estabelecer o que é *uma grande quantidade de órgãos*, mas não é possível responder tal questão sem analisar os dados estatisticamente. O valor, portanto, permanece em aberto até que os dados sejam coletados e uma análise estatística possa fornecer essa informação. A partir dela será possível definir qual quantidade é considerada um piso inferior a partir do qual todos os demais valores atendem ao Critério de Vulnerabilidade, ou seja, possuem grandes quantidades de contratantes.

A motivação da escolha do 3º CV se relaciona com a alta seletividade das contratações públicas. Contemporaneamente, ser contratado por inúmeros órgãos públicos é uma tarefa hercúlea, pois as seleções dos fornecedores são realizadas por certames públicos, em geral muito concorridos: as licitações.

Segundo Justen Filho (2014, p. 495):

A licitação é um procedimento administrativo disciplinado por lei e por um ato administrativo prévio, que determina critérios objetivos visando a seleção da proposta de contratação mais vantajosa e a promoção do desenvolvimento nacional sustentável, com observância do princípio da isonomia, conduzido por um órgão dotado de competência específica.

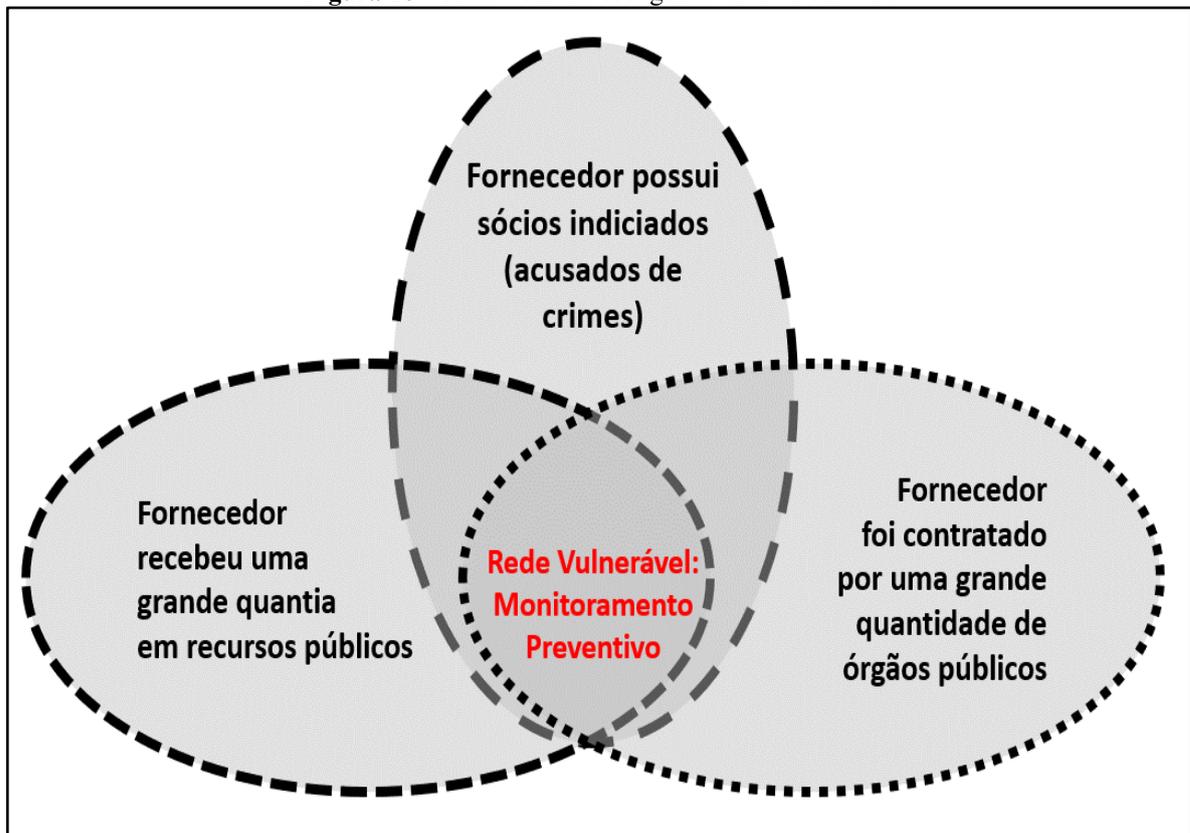
Assim, fornecedores que conquistam muitos órgãos públicos como clientes são excepcionalíssimos. Entretanto, conluíus criminosos entre agentes públicos e privados podem impulsionar um fornecedor a colecionar múltiplos contratantes estatais e, com isso, aumentar

significativamente o valor da VF a um patamar que atenda ao CV. Salienta-se que um fornecedor ter sido contratado por uma quantidade expressiva de órgãos públicos é, por si, é um indicador insuficiente para aumentar a probabilidade criminal, pois fornecedores sem intenções delitivas também buscam obter uma grande quantidade de contratantes oriundos do poder público. Entretanto, esse não é o único CV que será aplicado. A associação de todos eles é que determinará a vulnerabilidade.

3.2.6 Associação dos três CVs

Para o CO desvio de recursos públicos, segundo os CVs das VFs, a Rede Vulnerável que se busca é a intersecção entre os conjuntos ilustrados pela figura 16:

Figura 16 – Rede Vulnerável segundo os CVs das VFs



Fonte: Elaborado pelo autor

3.2.7 Coleta e tratamento dos dados

Os dados que foram coletados para viabilizar o estudo de caso estão quantificados e qualificados quanto à fonte da coleta, conforme a tabela 3:

Tabela 3 – Dados utilizados na pesquisa

UNIDADE DE INFORMAÇÃO	QUANTIDADE	FONTE DA COLETA (BANCO DE DADOS)	BREVE EXPLANAÇÃO
Fornecedores	11.421	Secretaria da Fazenda da UF ¹⁰	Empresas que fornecem bens e/ou serviços aos órgãos públicos
Inquéritos policiais	380	Sistemas Cartorários da Polícia Federal ¹¹	Procedimentos administrativos que apuram autoria e materialidade de crimes
Indiciamentos	4.139	Sistemas Cartorários da Polícia Federal	Relação entre uma pessoa física e um inquérito policial que qualifica esse indivíduo como acusado de crime
Sócios de empresas	309.653	Secretaria de Fazenda da UF	Relação entre uma pessoa física e uma empresa
Pagamentos de recursos públicos	583.072	Tribunal de Contas da UF	Pagamentos realizados por órgãos públicos às empresas como contraprestação pelo fornecimento de bens e/ou serviços
Órgãos públicos	224	Tribunal de Contas da UF	Órgãos contratantes de bens e serviços. No estudo de caso esses órgãos são prefeituras (poderes executivos dos municípios)
Coordenadas geográficas	224	IBGE	Localização geográfica dos órgãos públicos
População de municípios	224	IBGE	População dos municípios geridos pelos órgãos públicos

Fonte: Elaborado pelo(a) autor(a).

O tratamento dos dados removeu algumas poucas inconsistências e valores nulos, ação que resultou numa redução de menos de 1% do total inicial. As inconsistências são: pessoas indiciadas cujo CPF não constava na base de dados (valor nulo no campo CPF), pagamentos realizados a empresas sem CNPJ (valor nulo no campo CNPJ). O abatimento desse percentual de 1% já está incluso na tabela 3 e não afetou significativamente os resultados. Das bases de dados da tabela 3, apenas as originárias da Polícia Federal e da Secretaria de Fazenda da UF não são públicas.

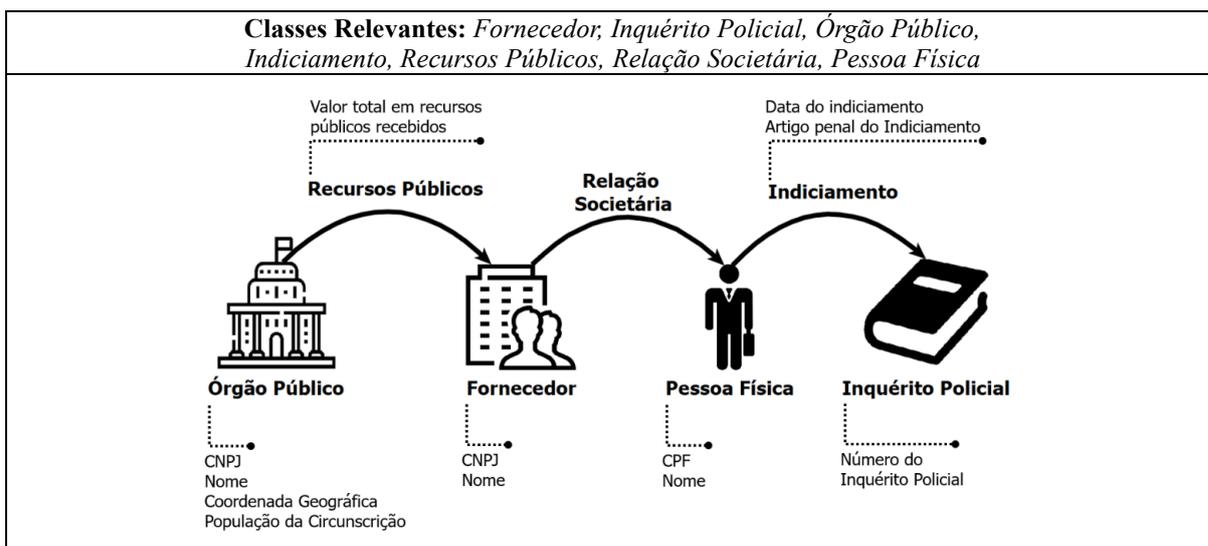
3.2.8 Modelagem e criação da Rede Supergrafo

Com base nas Classes Relevantes, já definidas em etapa anterior, foi criado o Grafo de Modelagem da Rede, conforme ilustra a figura 17:

¹⁰ UF: Unidade Federativa.

¹¹ Todos os dados foram tornados anônimos.

Figura 17 – Grafo de Modelagem de Rede – CO: Desvio de Recursos Públicos



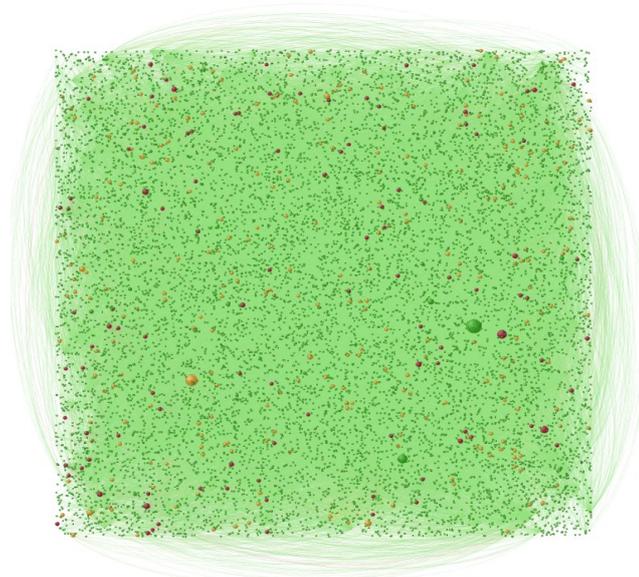
Fonte: Elaborado pelo autor.

A figura 17 fornece uma visão iconográfica e auto descritiva do modelo da Rede Supergrafo: um órgão paga recursos públicos a fornecedores, que por sua vez possuem uma relação societária com pessoas físicas. Eventualmente esses indivíduos estão na condição de indiciados em inquéritos policiais. Toda a semântica é oriunda do processo cognitivo que obteve as CRs. O modelo apresentado releva a necessidade de travessia, aptidão intrínseca aos grafos. Com base no Grafo de modelagem, o banco de dados de grafos foi criado e nele foram carregados todos os dados coletados na fase de coleta e tratamento.

3.2.9 Análise Exploratória da Rede Supergrafo

O Grafo a seguir representa a Rede Supergrafo criada a partir dos dados coletados. A plotagem inicial, ainda sem nenhum tratamento visual, possui o aspecto da figura 18:

Figura 18 – Rede Supergrafo (RS): uma *hairball*

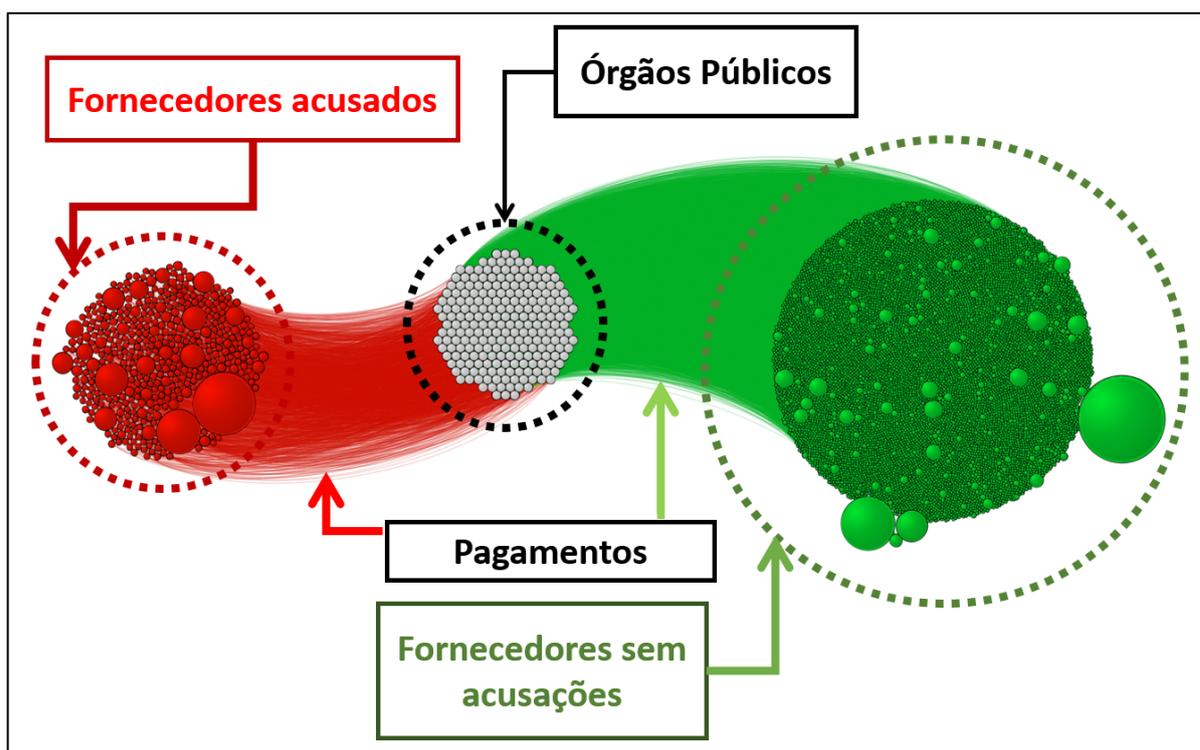


Fonte: Elaborado pelo autor

Por assumir um formato aleatório, sem uso de técnicas de apresentação da informação, a rede ainda carece de significado visual minimamente útil à análise exploratória. Trata-se do que Cherven (2015, p. 28) chamou de *hairball*, uma descrição para uma rede visualmente muito densa e com muitas conexões. Esse tipo de rede possui uma aparência indecifrável quando se usa os algoritmos padronizados de grafos. Embora as métricas de rede já possam fornecer algumas conclusões, o método adotado utilizou análise exploratória como foco no visual da rede, então a análise das métricas será alvo de estudo estatístico posterior. O objetivo inicial é a sondagem do terreno. E isso pode ser feito pela aplicação de leiautes.

Conforme visto, a rede em formato *hairball* é inútil à análise puramente visual. A primeira estratégia para avançar na tarefa de analisar a rede foi a aplicação do leiaute *circle-pack*, que divide os vértices em grupos hierárquicos de acordo com o valor de um ou mais atributos dos vértices. O grafo assumiu a forma reproduzida na figura 19:

Figura 19 – Leiaute circle-pack aplicado à rede Supergrafo



Fonte: Elaborado pelo autor

Para obter uma representação menos poluída, foram ocultados do grafo os inquéritos policiais e os sócios de empresas. Eles foram substituídos pela coloração dos vértices de fornecedores. Assim, os fornecedores que possuem sócios acusados em inquéritos policiais receberam a cor vermelha e, os demais, a cor verde. A rede, a partir dessa estratégia, passou a exibir apenas dois tipos de vértices: fornecedores e órgãos públicos. Essa abordagem permite

uma melhor inteligibilidade quando da aplicação dos leiautes e, por isso, será usada doravante como um padrão.

Percebe-se, pela figura 19, que três grupos foram apartados no espaço de plotagem, de acordo com o atributo que designa o tipo de vértice (órgão público, fornecedor acusado e fornecedor não acusado). As ligações entre os vértices (ou arestas) representam pagamentos de recursos públicos. Eles fluem no sentido dos órgãos públicos (representados pelos vértices em branco) para os fornecedores localizados em um dos dois possíveis grupos. As ligações (arestas) foram matizadas com a cor do vértice destino dos pagamentos.

Arestas em vermelho são pagamentos a fornecedores acusados. Por sua vez, aquelas em verde são pagamentos aos não acusados. Quanto maior o tamanho do vértice¹², mais recursos públicos o fornecedor recebeu, e esse montante é uma das Variáveis Frisadas. Assim, devido à grande quantidade de vértices diminutos, é possível afirmar que a grande maioria deles recebeu valores de pequena monta.

Há uma minoria de fornecedores cujos vértices possui tamanho consideravelmente maior que os demais, o que significa que esses receberam valores vultosos e discrepantes da ampla maioria.

Também é possível perceber que os fornecedores sem acusações (em verde) são o maior grupo e, apesar disso, os seus vértices são, em geral, menores do que aqueles que compõem o grupo dos fornecedores acusados (em vermelho). Assim, é possível afirmar que os valores recebidos pelos não acusados tem maior granularidade e, por isso, os valores individuais recebidos pelos acusados em geral são maiores.

Comparando os dois grupos de fornecedores é perceptível que, apesar do grupo dos não acusados ser muito mais numeroso, quando se analisa a presença de fornecedores que receberam as maiores quantias – representados pelos maiores vértices – percebe-se um maior equilíbrio entre os dois grupos. Assim, a partilha dos maiores fornecedores – os que receberam mais dinheiro – é mais equitativa entre os dois grupos.

Em relação ao tamanho dos vértices, é perceptível que nos grupos de fornecedores há uma amplitude de tamanho considerável: os maiores são completamente discrepantes em relação as menores deles. Isso pode representar alta dispersão (espalhamento) dos valores da

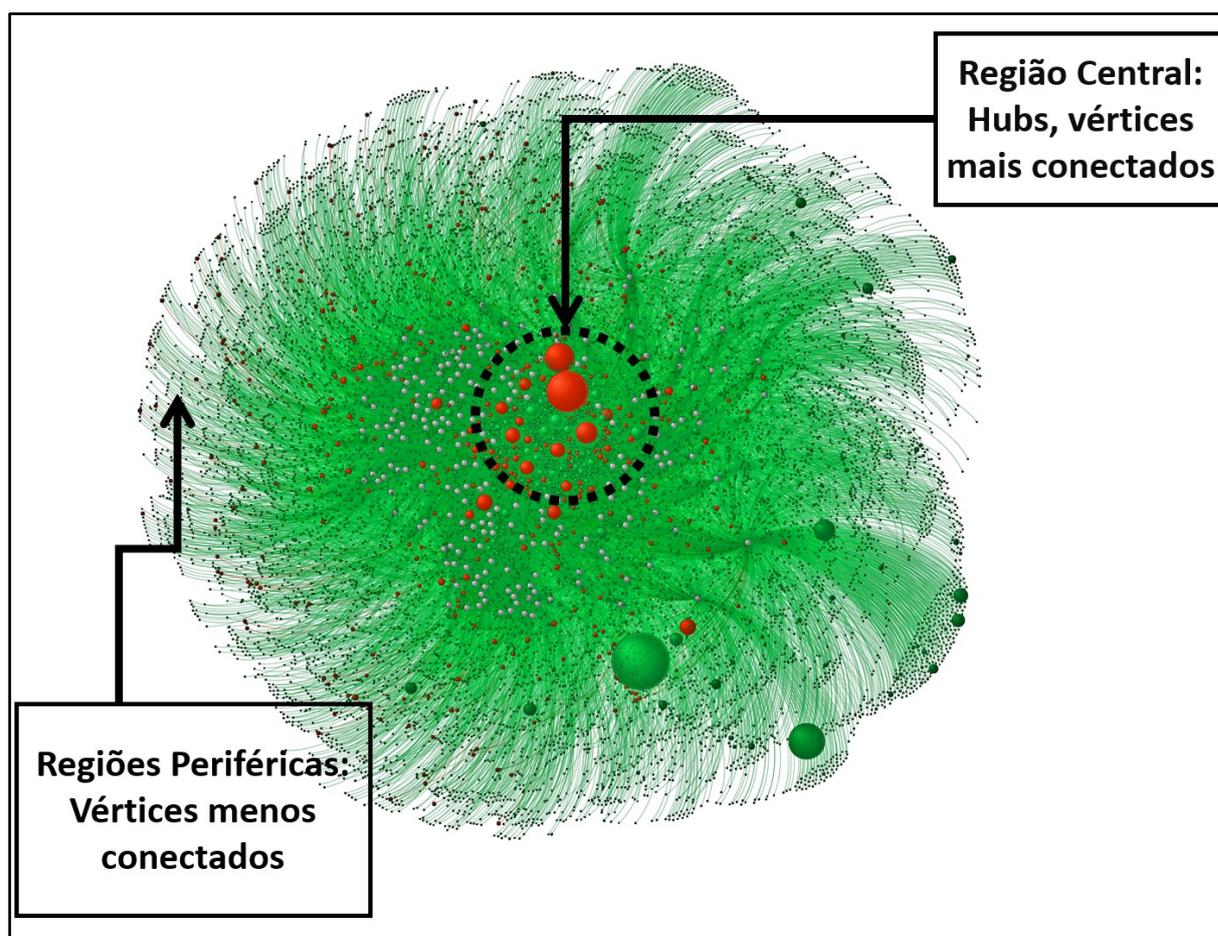
¹² O tamanho do vértice é proporcional à métrica de ARS nominada *Grau de Entrada Ponderado*.

Variável Frisada que se refere aos valores recebidos, o que aponta para uma provável presença de *Outliers* (valores atípicos).

As conclusões supracitadas serão, em etapa posterior, confirmadas por métodos estatísticos.

O próximo leiaute aplicado, denominado de direcionado por força, foi aplicado à rede Supergrafo, conforme ilustra a figura 20:

Figura 20– Leiaute direcionado por força aplicado à Rede Supergrafo



Fonte: Elaborado pelo autor

Segundo Hu (2006, p. 37), os elementos de estética fundamentais na representação da rede são simetria, menos cruzamentos de ligações (arestas) e distribuições uniformes, exigências que foram obtidas pela aplicação do leiaute ilustrado na figura 20. Além disso, é perceptível que os vértices muito conectados ocupam posições mais centrais do grafo. A migração dos vértices mais conectados – os nominados *hubs* – para regiões centrais é uma consequência do leiaute direcionado por força, o qual simula atração de pseudo-molas (as arestas ou ligações) e, em contraposição, a repulsão entre vértices (similares à força elétrica de cargas de mesmo sinal). Em consequência, os *hubs* (vértices muito conectados) migram ao

centro, enquanto os vértices menos conectados são afastados, por repulsão, à periferia. Estar bem conectado significa possuir alta centralidade de grau. No caso dos fornecedores significa ser contratado por uma grande quantidade de órgãos públicos.

Na figura 20, os vértices em branco representam órgãos públicos. Cada aresta que deles sai representa a soma de todos os pagamentos do órgão a um único fornecedor. Logo, a contagem das arestas ligadas a um fornecedor – representada pela métrica nominada grau – indica a o número de contratantes do fornecedor.

Postas essas considerações, percebe-se haver, na rede da figura 20, uma alta dispersão do número de contratantes dos fornecedores. Isso é perceptível porque na periferia do grafo há forte presença dos vértices, fraca presença de arestas e um grande espalhamento da rede. Enquanto isso, na região central o cenário muda. Há menos vértices, muito mais concentrados. Há uma quantidade maior de arestas no centro. Elas transformam os vértices centrais em *hubs*. Há alta densidade de ligações (representantes da grande quantidade de contratantes) cruzando a região central, a ponto de sobreporem, por completo, o plano de fundo preto da imagem.

Assim, na região periférica a grande maioria dos fornecedores foi contratada por uma pequena quantidade de órgãos, enquanto que, em contraste, uma minoria, mais central e compacta, foi contratada por uma grande quantidade deles. Provavelmente esses fornecedores mais centrais são discrepantes (ou *Outliers*) em relação a grande maioria que orbita a periferia da rede.

Outro *insight* relevante: os maiores vértices centrais são de fornecedores acusados (na cor vermelha). Por serem maiores, significa que receberam mais recursos públicos¹³ e, por estarem no centro¹⁴, foram contratados por muitos órgãos.

Os maiores fornecedores não acusados (na cor verde) são mais periféricos. Isso significa que foram contratados por poucos órgãos públicos, mas, como tem um tamanho relevante, receberam valores significativos desses poucos contratantes.

3.2.10 Análise estatística da Rede Supergrafo

O objetivo dessa tarefa é demonstrar os *insights* visuais que foram obtidos na análise exploratória da Rede Supergrafo. Assim, usar-se-á estatística descritiva para tanto.

¹³ Alto grau ponderado de entrada aumenta o tamanho do vértice.

¹⁴ Possuem alta centralidade de grau por estarem conectados a vários órgãos públicos. Assim, tornaram-se hubs e migraram ao centro da rede.

A análise exploratória da Rede Supergrafo indicou a possível existência de *Outliers*. Quando há ocorrência relevante desses discrepantes, há uma distorção no valor da média da população que, em consequência, deixa de ser uma medida representativa para o conjunto de dados. Inicialmente, para testar essa hipótese, uma alternativa é comparar a média e a mediana. Se forem muito distantes, os valores das observações estatísticas não estão bem distribuídos ao redor da média, fato que, além de descartá-la como medida representativa, é um indício da alta dispersão dos dados. A tabela 4, abaixo, faz essa comparação das medidas de centralidade (média e mediana) para as duas Variáveis Frisadas numéricas:

Tabela 4 – Média e Mediana das VFs

Variável Frisada↓	Medida estatística	
	Média (\bar{x})	Mediana (Md)
1ª VF (Valores recebidos em rec. públicos)	R\$ 919.215,80	R\$ 30.537,50
2ª VF (Número de contratantes)	3,24	1

Fonte: Elaborado pelo autor

A tabela indica disparidade entre os valores de médias e medianas. Assim, conclui-se que a média é inútil como medida representativa dos dados muito dispersos pela ação dos *Outliers*. Entretanto, com vistas a determinar a dimensão dessa dispersão, é necessário conhecer o Coeficiente de Variação de ambas as VFs, cujo valor, bem como as outras medidas utilizadas no seu cálculo, constam abaixo, na tabela 5:

Tabela 5– Variância, Desvio Padrão e Coeficiente de Variação das VFs

Métricas de Rede Associada→	Grau Ponderado de Entrada	Grau de Entrada do Vértice
Medida	2ª VF (Valores Recebidos)	3ª VF (Número de Contratantes)
Variância (σ^2)	37.160.004.557.207,50	55,80
Desvio Padrão (σ)	R\$ 6.095.900,64	7,47
Coeficiente de variação (C_v)	663,16%	230,91%

Fonte: Elaborado pelo autor

O σ foi calculado com base em σ^2 . O C_v é calculado com base no σ e na Média \bar{x} . Os valores recebidos pelos fornecedores possuem um C_v igual a 663,16%. Por sua vez, o número de contratantes dos fornecedores possui um C_v de 230,91%. Ambos confirmam: há alta dispersão dos dados em ambas as VFs. A confirmação dessas hipóteses, no contexto das VFs do estudo de caso, significaria que há fornecedores que superam em muito os demais sob dois aspectos quantitativos: nos valores recebidos e no número de órgãos públicos contratantes.

Considerando que alguns *Outliers* receberam grandes quantias em recursos públicos, serão eles, portanto, os fornecedores que atenderão ao CV da 2ª VF do presente estudo de caso, o qual exige justamente o recebimento de *altas quantias* (cujo piso mínimo ainda permanece desconhecido).

Considerando que outros *Outliers* possuem um grande número de órgãos públicos contratantes, serão eles, portanto, os fornecedores que atenderão ao CV da 3ª VF do presente estudo de caso. Esse critério exige justamente ter sido o fornecedor contratado por uma grande quantidade de órgãos (cujo piso mínimo ainda permanece desconhecido).

Ainda é necessário apontar precisamente quais fornecedores são *Outliers*, cuja definição do referencial teórico indica serem as observações estatísticas que superam o Limite Superior (L_s) da respectiva variável. Assim, a tabela 6 exibe os valores que delimitam os Quartis, o Limite Superior e os *Outliers*. O Limite Inferior foi ignorado em ambas as VFs porque não há observações estatísticas menores que ele, logo não há *Outliers* inferiores.

Tabela 6 – Quartis, Distância Interquartil e Limites

2ª VF: Valores recebidos em recursos públicos pelos fornecedores	
Medida	Valor da VF (em R\$)
Limite Inferior (L_i)	----
Q_1 (1º Quartil, 25% dos fornecedores receberam no máximo...)	7.000,00
Q_2 (2º Quartil ou mediana, 50% dos fornecedores receberam no máximo...)	30.537,50
Q_3 (3º Quartil, 75% dos fornecedores receberam no máximo...)	202.679,80
Dq (Distância Interquartil, diferença entre Q_3 e Q_1)	195.679,75
Limite Superior (L_s , 83,26% dos fornecedores receberam no máximo...)	496.199,38
<i>Outliers</i> (16,74% dos fornecedores receberam acima de...)	496.199,38
3ª Variável Frisada: Quantidade de órgão públicos contratantes do fornecedor	
Medida	Valor da VF
Limite Inferior (L_i)	----
Q_1 (1º Quartil, 25% dos fornecedores contratados no máximo por...)	01 órgão
Q_2 (2º Quartil, 50% dos fornecedores contratados no máximo por...)	01 órgão
Q_3 (3º Quartil, 75% dos fornecedores contratados no máximo por...)	03 órgãos
Dq (Distância Interquartil, diferença entre Q_3 e Q_1)	02 órgãos
Limite Superior (L_s , 90,47% dos fornecedores contratados no máximo por...)	06 órgãos
<i>Outliers</i> (9,53% dos fornecedores contratados por mais de...)	06 órgãos

Fonte: Elaborado pelo autor

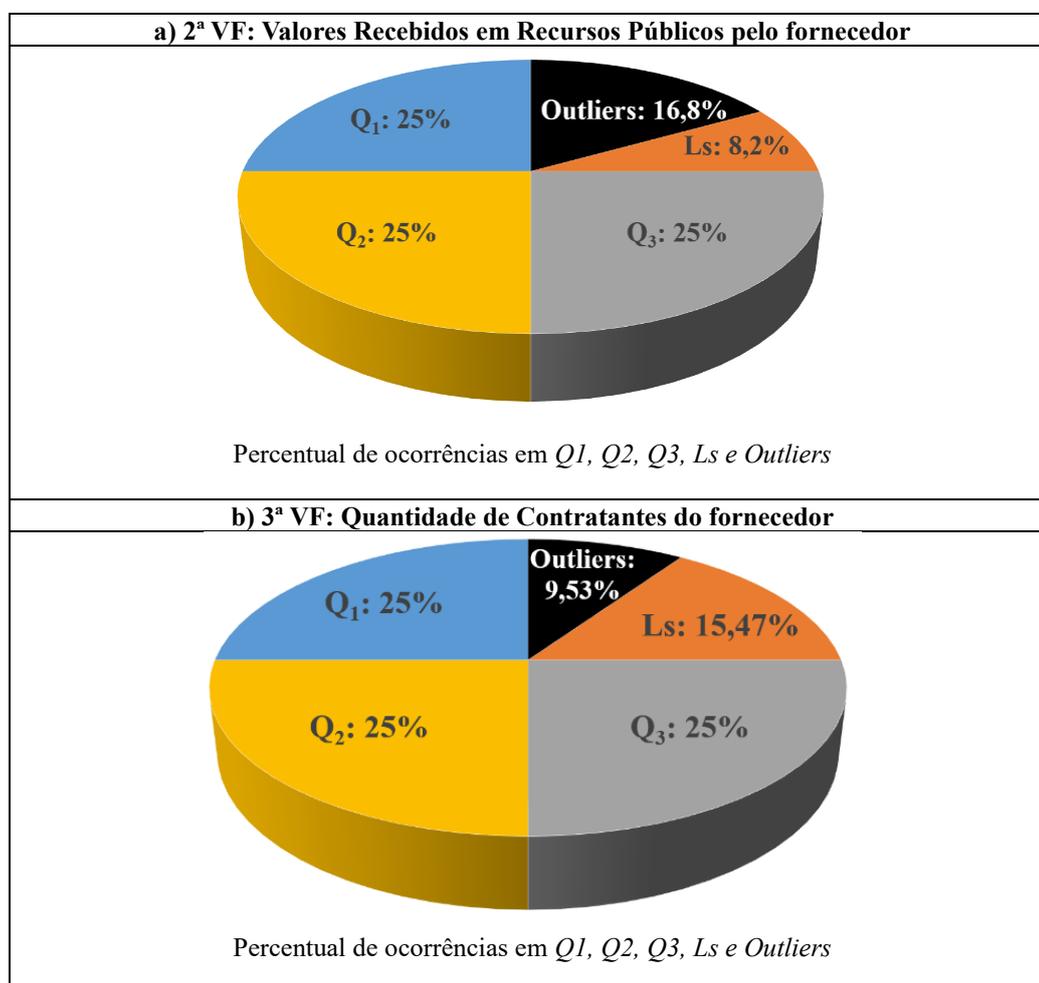
Assim, existem dois conjuntos de *Outliers*. O primeiro conjunto considera os valores recebidos em recursos públicos. Nele estão quaisquer fornecedores contemplados com valores acima de R\$ 496.199,30 (quatrocentos e noventa e seis mil, cento e noventa e nove reais e trinta centavos). O segundo conjunto considera o número de órgãos públicos contratantes. Nele estarão quaisquer fornecedores contratados por mais de 6 (seis) seis órgãos públicos.

Isto posto, podemos afirmar que a rede vulnerável que se busca é formada por aqueles fornecedores que, além de ter acusações criminais (1º Critério de Vulnerabilidade), estão

presentes nos dois conjuntos de *Outliers*, ou seja, que receberam mais de R\$ 496.199,30 (2º Critério de Vulnerabilidade, ter recebido uma grande quantia em recursos públicos) e foram contratados por mais de 6 órgãos públicos (3º Critério de Vulnerabilidade: ter sido contratado por uma grande quantidade de órgãos públicos).

Para prosseguir com a análise estatística, é relevante entender qual o percentual da população é *Outlier* e, posteriormente, se os fornecedores que possuem acusações possuem participações percentuais diferenciadas nesses *Outliers* em relação aqueles não acusados. A figura 21, abaixo, aponta qual o percentual de *Outliers* da população inteira para a 2ª e a 3ª Variáveis Frisadas em cada um dos quartis (*Q1*, *Q2* e *Q3*) e acima deles (*Limite Superior* e *Outliers*), desconsiderando a existência de acusações criminais.

Figura 21 – Fornecedores em Geral - percentual de *Outliers*

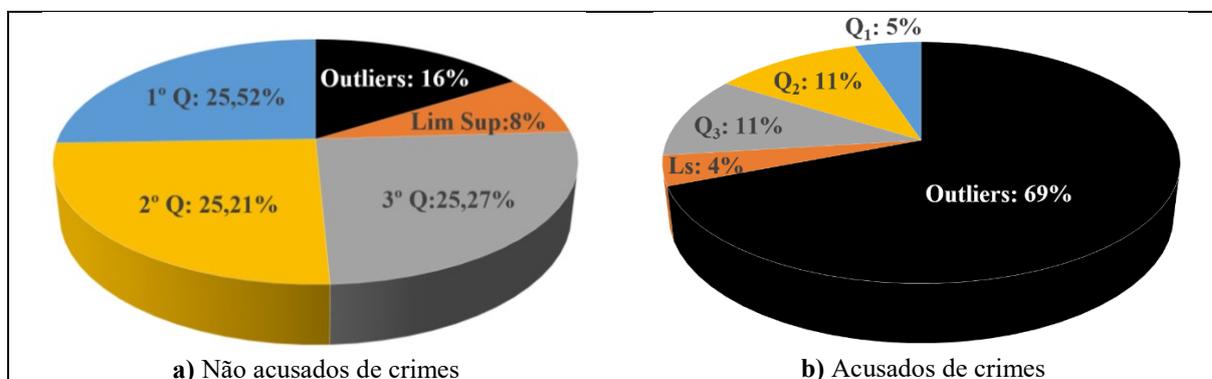


Fonte: Elaborado pelo autor

Pelo gráfico 21(a), conclui-se que 16,8% da população inteira de fornecedores são *Outliers* em relação aos valores recebidos em recursos públicos. Em relação ao número de contratantes, o gráfico 21(b) demonstra que apenas 9,53% dos fornecedores são *Outliers*.

Quando a análise considera o critério acusação, a distribuição dos acusados e dos não acusados em cada um dos quartis, no limite superior e *Outliers* para a 2ª VF (valores recebidos em recursos públicos) se apresenta conforme a figura 22:

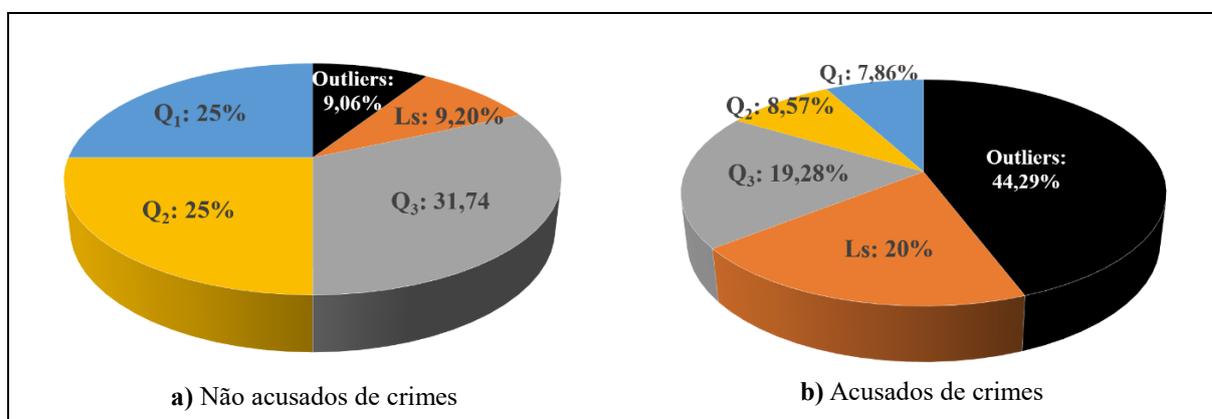
Figura 22 – Percentual de acusados e não acusados classificados como *Outliers* em relação a 2ª VF



Fonte: Elaborado pelo autor

Ainda considerando o critério acusação, a distribuição dos acusados e dos não acusados em cada um dos quartis, no limite superior e *Outliers* para a 3ª VF (quantidade de órgão públicos contratantes) se apresenta conforme a figura 23:

Figura 23 – Percentual de acusados e não acusados classificados como *Outliers* em relação a 3ª VF



Fonte: Elaborado pelo autor

Assim, os gráficos apontam que há distribuições nos intervalos interquartis populacionais muito similares quando se compara os fornecedores em geral (gráficos 21.a e 21.b) e aqueles não acusados de crimes (22.a e 23.a). Entretanto, quando se analisa os fornecedores acusados de crimes (gráficos 22.b e 23.b), a distribuição é anômala e tendente à discrepância: 69% deles são *Outliers* quanto aos valores recebidos e 44,29% deles são *Outliers* quanto à quantidade de contratantes. Melhor dizendo: *os fornecedores acusados de crimes, em geral, recebem muitos recursos públicos e são muito contratados (participação significativa de Outliers em ambas as VFs), situação completamente divergente dos não acusados que, por sua vez, se comportam de forma muito similar à população em geral.*

Assim, foram ratificadas, pela análise estatística, os *insights* da análise exploratória, conforme relata-se a seguir.

De fato, a grande maioria dos fornecedores recebeu valores de pequena monta e minoria recebeu valores vultosos. Assim também, poucos fornecedores foram contratados por muitos órgãos. A maioria deles foi contratada por poucos órgãos.

Há alta dispersão das observações estatísticas, tanto nos valores recebidos quanto números de contratantes. Essas dispersões são causadas pela presença confirmada de *Outliers*.

Também foi confirmado que os recursos públicos de fornecedores não acusados são mais pulverizados do que aqueles recebidos por fornecedores acusados. Isso porque a maioria percentual deles estão distribuídos nos três primeiros quartis, possuem um percentual menor de *Outliers* e, em consequência, em geral dividem recursos de menor monta nos quartis de valores inferiores, o que causa a pulverização em pagamentos menores.

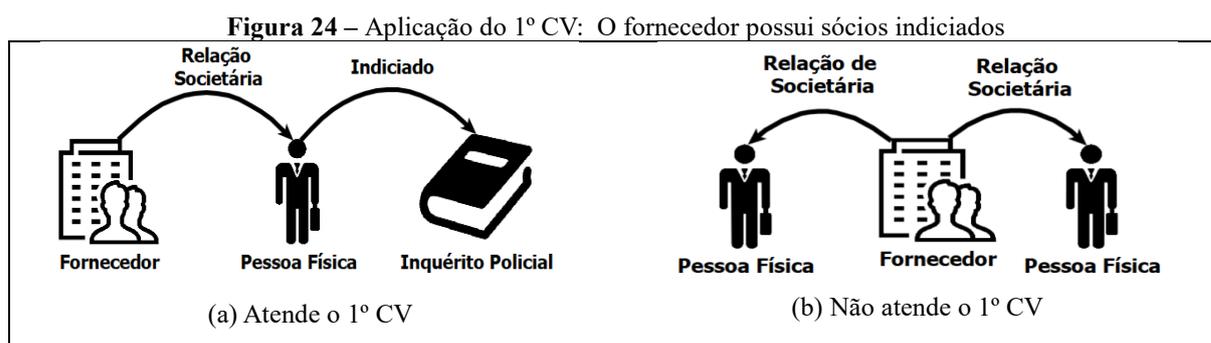
O grande percentual de acusados dentre os *Outliers* da 2ª VF (69%) também demonstra o equilíbrio observado pela distribuição mais equitativa dos maiores vértices nos dois grupos (acusados e não acusados) gerados pelo leiaute *circle-pack* durante a análise exploratória.

Os maiores vértices de alta centralidade, que representam os mais contratados e evidentes no leiaute direcionado por força da análise exploratória, são de fato de fornecedores acusados, pois o percentual deles que está no grupo de *Outliers* é de 44,29% em relação a 2ª VF (quantidade de órgão públicos contratantes).

Assim, finda-se a análise estatística da rede Supergrafo, confirmando todas os insights da análise exploratória.

3.2.11 Extração da Rede Vulnerável

Antes de extrairmos a Rede Vulnerável da rede supergrafo, é necessário criar a representação iconográfica da aplicação dos Critérios de Vulnerabilidade, conforme mostra a figura 24, a seguir:

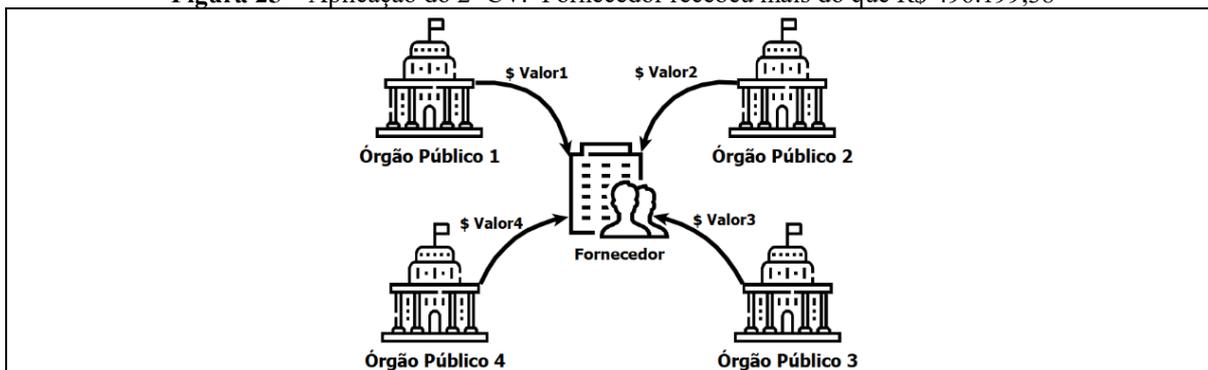


Fonte: Elaborado pelo autor

Na figura 24(a) existe ao menos uma aresta de indiciamento entre um dos sócios e um inquérito policial, logo o fornecedor possui sócios indiciados e, por isso, atende o 1º CV. Na figura 24(b), por sua vez, o fornecedor não atende ao 1º CV, pois não possui sócios indiciados (acusados).

Na figura 25, para determinar se o fornecedor atende ao 2º CV, é necessário determinar o Grau Ponderado do vértice.

Figura 25 – Aplicação do 2º CV: Fornecedor recebeu mais do que R\$ 496.199,38

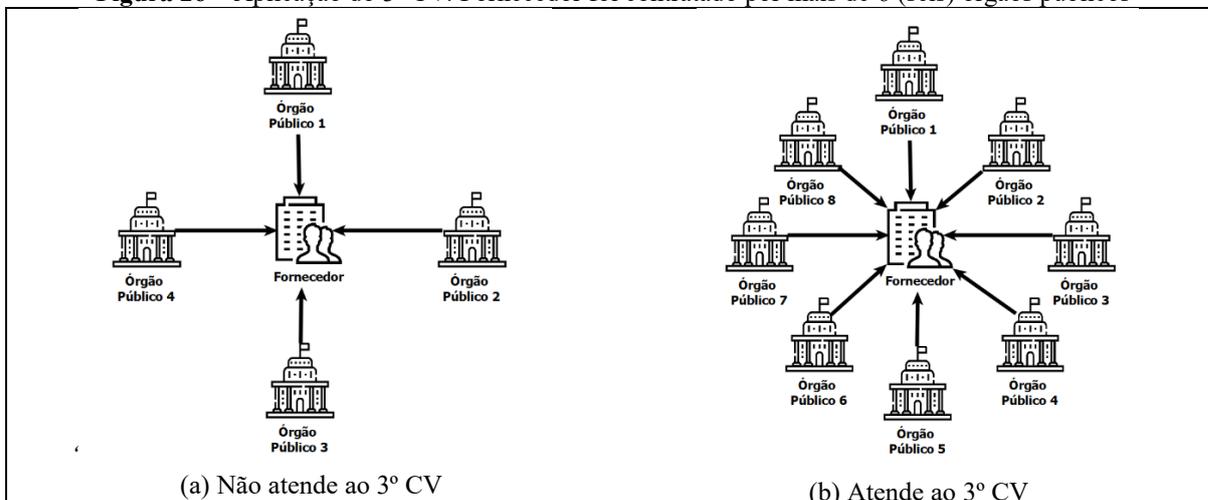


Fonte: Elaborado pelo autor

Assim, na figura 25, o valor total em recursos públicos recebidos será obtido pela soma ($Valor1+Valor2+Valor3+Valor4$). Se esse valor for maior do que R\$ 496.199,38, estará atendido o 2º CV.

Na figura 26, para determinar se o fornecedor atenderá o 3º CV, é necessário determinar o valor do grau do vértice. Se o valor for maior do que 6, estará atendido o 3º CV, pois o número de contratantes transforma o fornecedor em um *Outlier*, conforme apontou a análise estatística. Nesse caso, o fornecedor da figura 26(a) não atende ao 3º CV (possui 4 contratantes) e, por sua vez, o fornecedor da figura 26(b) atende ao 3º CV (possui 8 contratantes).

Figura 26 – Aplicação do 3º CV: Fornecedor foi contratado por mais de 6 (seis) órgãos públicos



Fonte: Elaborado pelo autor

Os fornecedores que atenderem aos três critérios de vulnerabilidade irão compor a RV. Todos os órgãos a eles conectados também comporão essa rede. Assim, a Rede Supergrafo sofre a filtragem dos três CVs (conforme detalhado nas figuras 24, 25 e 26). O resultado dessa filtragem é extração da Rede Vulnerável. Essa, por sua vez, está pronta para a análise exploratória, que será realizada a seguir.

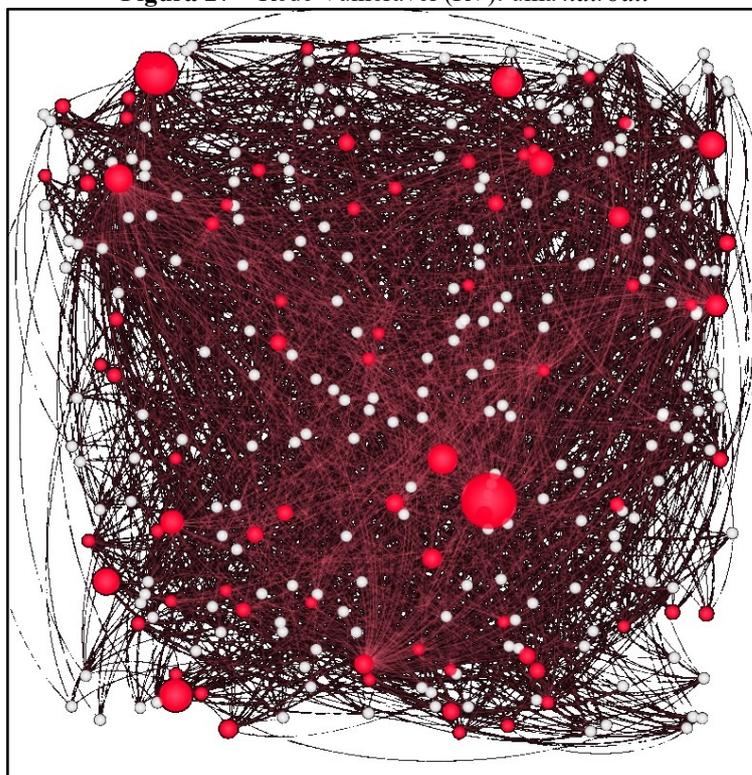
3.2.12 Análise exploratória da Rede Vulnerável

Conforme já explanado, a RV é extraída da Rede Supergrafo usando como filtros os CVs. Na RV estão apenas os fornecedores que atenderam a todos os CVs. Os órgãos públicos, por estarem a ele conectados, também comporão essa rede.

Para fins de despoluição visual, a plotagem da rede não exibiu os inquéritos policiais, os indiciamentos, as pessoas físicas e as relações societárias. Dessa forma o número de vértices ficou reduzido pela ocultação e, em consequência, a análise exploratória foi facilitada. Isso, entretanto, não significa perda de conteúdo, já que todas as informações que serão usadas na análise estão visíveis no grafo.

A plotagem inicial da RV, ainda sem nenhum tratamento visual, possui o aspecto *hairball* ilustrado pela figura 27:

Figura 27 – Rede Vulnerável (RV): uma *hairball*

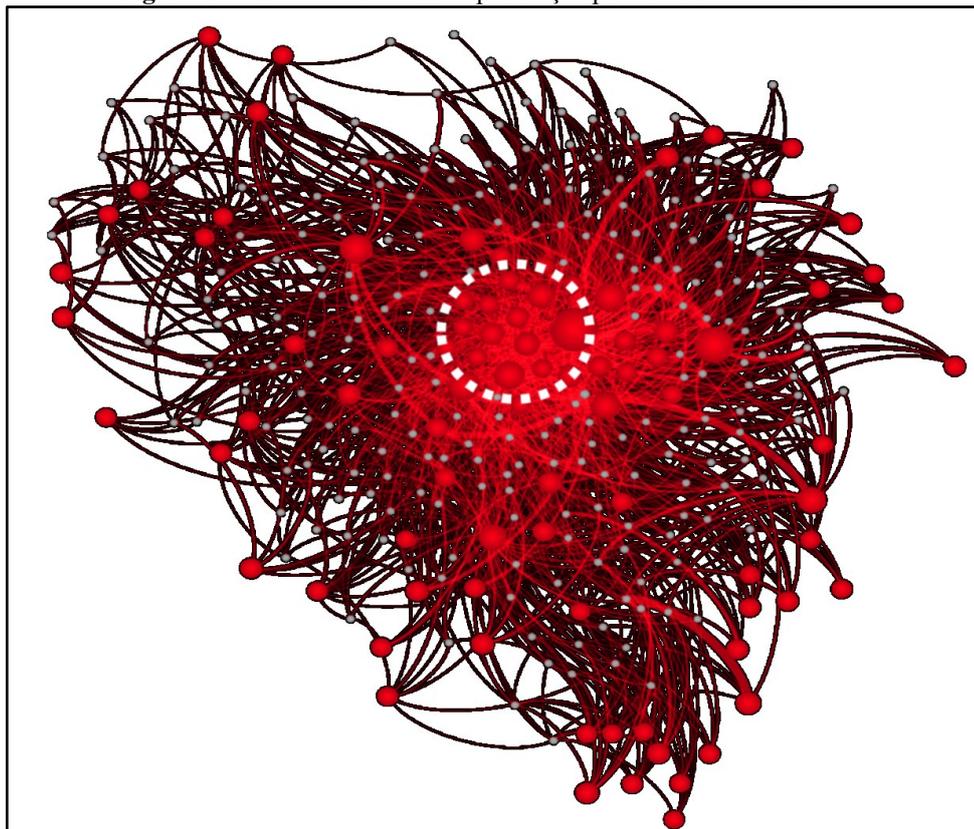


Fonte: Elaborado pelo autor

Por assumir um formato randômico, sem técnicas de apresentação da informação, a RV ainda carece de significado visual minimamente útil à análise exploratória. Trata-se do que Cherven (2015, p. 28) chamou de *hairball* (rede muito densa e indecifrável quando apenas o aspecto visual é usado como método).

Para tentar obter *insights*, foi realizada a aplicação de um leiaute direcionado por força à rede Vulnerável, conforme ilustrado na figura 28:

Figura 28 – Leiaute direcionado por força aplicado a Rede Vulnerável



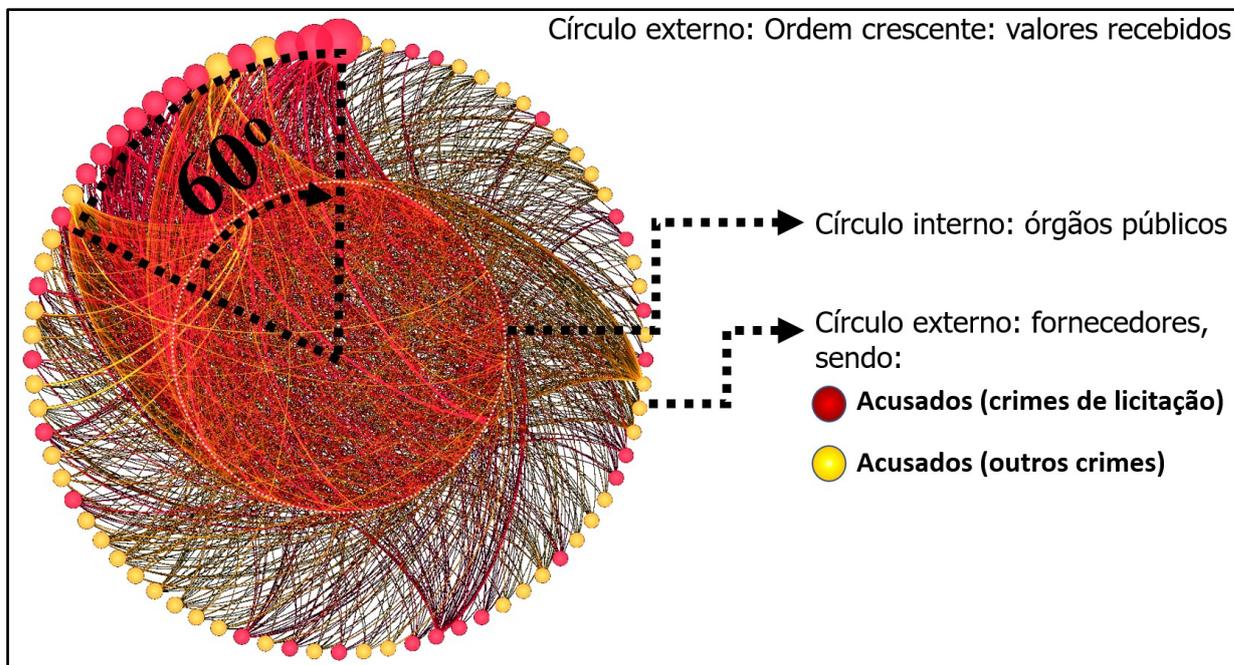
Fonte: Elaborado pelo autor

O algoritmo direcionado por força usado na RV da figura 29 foi o *Yifan-Hu*. Os vértices em vermelho representam os fornecedores; os vértices em cinza representam os órgãos públicos contratantes. Cada uma das conexões representa o total de pagamentos de um órgão para um fornecedor. A linha branca, circular e pontilhada no centro da rede, destaca que há grupo, representado por um núcleo central de alta densidade de interconexões. Esse grupo, que se diferencia do resto da rede, foi contratado por uma quantidade extrema de órgão públicos. São, dentre fornecedores da RV, os mais contratados.

Outro leiaute que pode ser aplicado é o *dual circle*. Ele é especialmente útil quando queremos separar duas categorias de vértices numa plotagem de um grafo. A primeira delas vai compor um círculo interno e a segunda irá compor um círculo externo; assim, cada categoria de vértice ocupará um dos círculos. Como a Rede Vulnerável só possui dois tipos de vértice

(fornecedores e órgãos públicos), a aplicação de *dual circle* é ideal para separar essas duas categorias, conforme a figura 29, abaixo:

Figura 29 – Leiaute *dual circle* e comportamento anormal dos fornecedores



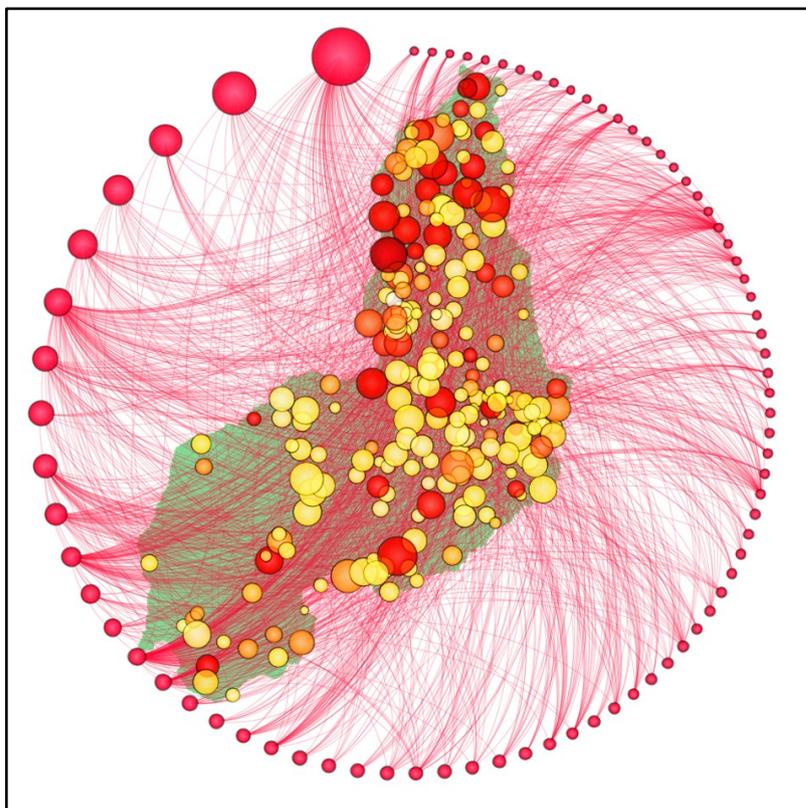
Fonte: Elaborado pelo autor

Um comportamento anormal e relevante foi evidenciado após a aplicação do leiaute *dual circle*. No grafo, os fornecedores foram separados em cores por categoria de crimes. Os fornecedores acusados de crimes relacionados a licitação, mais perigosos para um cenário do CO de desvio de recursos, foram matizados na cor vermelha. Qualquer outro fornecedor acusado de qualquer outra categoria de crime foi matizado em amarelo. Feito isso, e sabendo que as ligações (ou arestas) assumem a cor dos vértices de destino, é perceptível que surge um arco, de aproximadamente 60° (sessenta graus) onde há uma alta densidade de arestas na cor vermelha. Na circunferência desse arco estão localizados justamente os fornecedores que mais receberam recursos, devido à ordenação em sentido horário por valor recebido em recursos públicos, fato que é perceptível também pelo tamanho dos vértices que aumenta no mesmo sentido.

Assim, dos 15 (quinze) fornecedores que estão dentro do arco, 12 são acusados por crimes de licitação. O *insight*, portanto, indica que há preponderância no pagamento a fornecedores acusados por esse tipo de crime.

Outra possibilidade interessante de análise exploratória é a aplicação da composição de dois leiautes em um único grafo. Para ilustrar essa alternativa, um leiaute geográfico e outro circular foram aplicados a Rede Vulnerável do estudo de caso, conforme mostra a figura 30.

Figura 30– Composição de leiautes circular e geográfico do estado do Piauí



Fonte: Elaborado pelo autor

São necessárias algumas considerações sobre as convenções adotadas na rede da figura 30. Há, no grafo, dois leiautes independentes conforme será detalhado a seguir.

O primeiro leiaute, o circular, contém apenas os fornecedores de bens e serviços, todos da Rede Vulnerável ao Crime-Objeto (desvio de recursos públicos). Eles são representados pelos vértices do círculo externo (na cor vermelha, em um único tom). Nesse leiaute, o tamanho dos vértices é proporcional ao total de recursos públicos recebidos pelo fornecedor. É a expressão, portanto, do grau de entrada ponderado de cada vértice.

O segundo deles, o leiaute geográfico, contém apenas órgãos públicos, em cores variadas. Eles estão representados por vértices dentro do mapa do estado do Piauí, de acordo com suas coordenadas geográficas. O tamanho desses vértices é proporcional a quantidade de contratados por esses órgãos. É a expressão, portanto, do grau do nó. Assim, quanto maior é o vértice no interior do mapa, mais fornecedores da Rede Vulnerável ao Crime-Objeto ele

contratou. A cor desses vértices diz respeito a população da circunscrição. Essa cor obedece a uma *escala de calor*, que varia de amarelo claro, passando por laranja, vermelho claro e chegando até o vermelho escuro. Assim, órgãos públicos amarelos atendem a uma pequena população. No outro extremo, órgãos públicos em vermelho escuro atendem às maiores populações. A análise exploratória da rede da figura 31, indica uma anomalia relevante em relação aos órgãos públicos, conforme será detalhado a seguir.

Um órgão *X* atende a uma pequena população (e, por isso, está tingido em amarelo claro). O vértice de *X* está entre os maiores da rede. Isso indica que ele contrata muitos fornecedores da RV.

Um outro órgão, *Y*, atende a uma das maiores populações do estado (e, por isso, está tingido em vermelho escuro). O vértice de *Y*, assim como *X*, está entre os maiores da rede, o que implica que ele também contrata muitos fornecedores da RV.

Considerando que *Y* contrata muito mais bens e serviços do que *X* (decorrência do fato de que atende à uma população muito maior), então o que justifica o fato de ambos possuírem quantidades equivalentes de contratados da RV? A resposta a essa pergunta não é objeto do estudo de caso. Entretanto é perceptível que há muitos outros vértices no mapa semelhantes a *X*. Uma conclusão prévia da análise exploratória, portanto, seria: uma boa parte dos órgãos que atendem a uma pequena população, mesmo contratando relativamente poucos fornecedores, têm quantidade considerável deles advindos da RV ao CO.

A análises exploratória da Rede Vulnerável usando composição de dois leiautes pode render inúmeras descobertas. Certamente há outras anomalias dignas de registro no mesmo estudo de caso, entretanto a intenção era tão somente demonstrar como conduzir a análise com composição de dois leiautes em um caso concreto.

3.2.13 Análise estatística da Rede Vulnerável

Inicialmente, antes de avaliar os *insights* da análise exploratória, alguns números das duas redes (RS e RV) foram comparados, conforme a tabela 7 (abaixo):

Tabela 7 – Números da Rede Supergrafo e da Rede Vulnerável

Medição	Quantidade na Rede Supergrafo	Quantidade na Rede Vulnerável	Percentual da Rede Vulnerável em Relação a Original
Quantidade de Órgãos Públicos	224	224	100%
Quantidade de Fornecedores	11.421	78	0,68%
Valores movimentados	R\$ 10.486.405.860,00	R\$ 1.093.192.603,00	10,42%

Fonte: Elaborado pelo autor

O primeiro fator relevante é que todos os 224 (duzentos e vinte e quatro) órgãos presentes na Rede Supergrafo original continuaram presentes na Rede Vulnerável. Entretanto, o objetivo não é monitorar todos os órgãos (o que certamente inviabilizaria o monitoramento preventivo), mas apenas aqueles contratos onde os 78 (setenta e oito) fornecedores da Rede Vulnerável figuram na condição de contratados.

Outra informação importante diz respeito aos valores movimentados: apesar de representarem apenas 0,68% dos fornecedores, os que compõem a Rede Vulnerável movimentaram mais de 10% de todos os recursos públicos da população estatística considerada, o que equivale a 14 (catorze) vezes o seu próprio quantitativo.

Na análise exploratória, por meio do leiaute *Yifan-Hu*, ficou evidenciada a existência de um núcleo central muito interconectado e que por isso se diferencia do restante da Rede Vulnerável. A análise estatística mostrou que, dos dez componentes mais interconectados desse grupo, sete deles (setenta por cento) foram alvos de cinco Operações da Polícia Federal: *Geleira*, *Topique*, *Gangrena*, *Boca Livre* e *Rapina*. Ainda sobre esses dez fornecedores mais interconectados, apesar do universo total de fornecedores ser de 11.421 (onze mil, quatrocentos e vinte e um), todos os 224 (duzentos e vinte e quatro) órgãos públicos contrataram pelo menos um desses dez hubs centrais. O fornecedor mais interconectado dentre todos (o maior hub central), foi contratado sozinho por 154 (cento e cinquenta e quatro) dos 224 órgãos públicos existentes na população considerada, sendo que bastava que fosse contratado por apenas seis órgãos para ser considerado um discrepante. Assim, a percepção da análise exploratória de que esse grupo se diferencia, até mesmo dos demais fornecedores da própria Rede Vulnerável, está correta.

A análise exploratória, através do uso do leiaute *dual circle* na Rede Vulnerável, apontou que há preponderância no pagamento a fornecedores acusados por crimes de licitação. A análise estatística ratificou essa hipótese, pois mostrou que enquanto os 45 (quarenta e cinco) fornecedores acusados por outras categorias de crime receberam, na Rede Vulnerável, 34,60% dos valores, os 33 (trinta e três) fornecedores acusados de crimes de licitação receberam 65,40%, ou seja, quase o dobro. Conclui-se, portanto, que a análise exploratória estava correta, há preponderância de pagamentos a fornecedores acusados por crimes de licitação. Receberam quase o dobro, mesmo sendo uma minoria deles.

A análise exploratória, pelo uso de uma composição de leiautes (circular e geográfico), indicou que uma boa parte dos órgãos que atendem a uma pequena população, mesmo

contratando relativamente poucos fornecedores de bens e serviços, têm quantidade considerável deles advindos da RV ao CO.

Para analisar tal fato, tome-se como referência o órgão *A*. Ele atende à circunscrição mais populosa (861.442 pessoas). Naturalmente, *A* foi o primeiro lugar quanto à quantidade de fornecedores contratados: foram 23 (vinte e três) deles oriundos da Rede Vulnerável. É normal que isso ocorra, tendo em vista que uma maior população requer mais recursos públicos e investimentos. Em consequência, uma maior quantidade de bens e serviços são contratados. Assim, o universo de fornecedores aumenta e um número maior deles pode fazer parte da Rede Vulnerável.

A anormalidade surge ao analisar o órgão público *B*, que ficou em 2º lugar na quantidade de contratados advindos da Rede Vulnerável. Apesar administrar uma circunscrição com uma população equivalente a 0,62% de *A*, *B* contratou o equivalente a 86,95% da quantidade de fornecedores que *A* contratou.

Importante destacar que os fornecedores da RV representam apenas 78 no universo dos 11.421 da Rede Supergrafo (ou apenas 0,68% do total deles), mas alguns órgãos, tal como o anômalo *B*, são clientes de mais de ¼ de toda a Rede Vulnerável. É uma informação intrigante que pode levar a conjecturas acerca dos reais motivos dessa anomalia, embora esse não seja o objetivo do trabalho.

Também se constata não ser esse um caso isolado: 34 dos 224 órgãos (pouco mais de 15% do total) contrataram uma quantidade de fornecedores acima de 50% do quantitativo que *A* contratou. Entretanto, dentre esses, o de circunscrição mais populosa atende a uma população equivalente a apenas 7% daquela atendida por *A*.

Conclui-se que a análise exploratória por composição de leiautes estava certa. Uma boa parte dos órgãos que atendem a uma pequena população, mesmo contratando relativamente poucos fornecedores, têm quantidade considerável deles advindos da RV ao CO.

3.2.14 Avaliação da Rede Vulnerável

Nessa fase devemos avaliar se a Rede Vulnerável aponta para maior risco de ocorrência de crime. Pelas considerações e análises expostas, é oportuno afirmar que a Rede Vulnerável obtida é suficientemente incomum. Alguns exemplos de sua anormalidade são: movimentar valores percentuais discrepantes; 70% dos fornecedores mais contratados já foram alvo de operações policiais; há preponderância de pagamentos dos maiores valores a indiciados por

crimes de licitação; alguns órgãos, quando comparados a outros, contratam grandes quantidades de fornecedores advindos da RV, dentre outros.

Assim conclui-se pela sua conformidade da RV aos propósitos pelos quais foi concebida: a prevenção contra crimes relacionados ao desvio de recursos públicos. Nesse caso, as VFs e os CVs não precisam ser reavaliados e o fluxo do processo pode seguir com a próxima tarefa: a entrega da Rede Vulnerável às equipes de policiamento preventivo. O resultado do estudo de caso é a própria Rede Vulnerável gerada em conformidade com o esperado. Composta por 302 vértices, sendo 224 órgãos públicos e 78 fornecedores que atendem a todos os critérios de vulnerabilidades.

4 CONCLUSÕES

O fluxograma apresentado nos resultados é a produto mais importante da pesquisa, pois é a materialização do modelo proposto no objetivo do trabalho. Ele representa um processo que apresenta três grandes grupos de tarefas.

O primeiro grupo diz respeito a fundamentação policial. Esse grupo se inicia pela escolha do Crime-Objeto (aquele que se quer prevenir). Posteriormente, delimita as Variáveis Frisadas (que influenciam a probabilidade de o delito ocorrer), define os Critérios de Vulnerabilidade (cujos valores aumentam as probabilidades criminais) e as Classes Relevantes (conceitos acerca do domínio de conhecimento do Crime-Objeto). O segundo grupo são tarefas afetas à preparação da rede, aquisição dos dados, modelagem e criação da Rede Supergrafo (formada pelos dados coletados). O terceiro grupo, a análise das redes, realiza essas tarefas nas modalidades exploratória e estatística, como também extrai a RV da RS usando os CVs como filtro. Posteriormente é necessário, com base nas análises exploratória e estatística, avaliar se a RV aponta para um maior risco de crime. Se a resposta for negativa, as VFs e os CVs devem ser reavaliados. Caso contrário, ocorre a tarefa de entrega da Rede Vulnerável às equipes de policiamento preventivo e finda-se o fluxo.

As contribuições decorrentes da pesquisa são principalmente destinadas à inteligência policial. Isso porque há similaridades entre a execução do modelo proposto e a atividade inteligência policial na segurança pública. Senão, vejamos: A Doutrina Nacional de Segurança Pública – DNISP, assim define a Inteligência de Segurança Pública (DNISP, 2015, p. 15):

A atividade de Inteligência de Segurança Pública (ISP) é o exercício permanente e sistemático de ações especializadas para identificar, avaliar e acompanhar ameaças reais ou potenciais na esfera de Segurança Pública, basicamente orientadas para produção e salvaguarda de conhecimentos necessários para subsidiar os tomadores de

decisão, para o planejamento e execução de uma política de Segurança Pública e das ações para prever, prevenir, neutralizar e reprimir atos criminosos de qualquer natureza que atentem à ordem pública, à incolumidade das pessoas e do patrimônio.

A aplicação da pesquisa é uma ação para prevenir atos criminosos. Assim, parece adequado que a ISP a acolha com uma das suas atividades.

As contribuições da pesquisa objetivam fornecer à sociedade um protocolo de prevenção criminal que pode ser, mediante pequenas adaptações, aplicado a qualquer tipo de crime que se manifeste usando redes sociais expressas por dados. Assim, é possível citar algumas áreas de aplicabilidade: crimes previdenciários, crimes ambientais, crimes de narcotráfico, desvio de recursos públicos, crimes fazendários, etc.

Há, no estudo, uma limitação decorrente da inviabilidade de avaliação da sua eficácia. Isso porque a presente pesquisa concretizou a proposição de um modelo de prevenção criminal, entretanto não objetivou avaliar se a sua aplicação (a própria prevenção) será eficaz. O Crime Objeto (desvio de recursos públicos) cujos dados estavam disponíveis para o estudo de caso, se perpetra em longos períodos temporais. Assim, é inviável aplicar o modelo e medir os impactos durante o período da pós-graduação. A eficácia da aplicação da pesquisa seria alvo de um novo trabalho até mais complexo do que a proposição do modelo ora esmiuçado. Isso porque apenas após vários meses seria possível avaliar se houve melhora nas estatísticas criminais decorrentes da prevenção. Entretanto, há algumas poucas áreas cujo imediatismo criminal poderia viabilizar a avaliação rápida da aplicação do modelo. Cita-se como exemplo os crimes de tráfico internacional de drogas.

Como possibilidade de pesquisas futuras, sugere-se um estudo de caso que colete métricas capazes de avaliar a eficácia da aplicação do modelo. Essas métricas podem ser registradas e avaliadas. Como exemplos de métricas viáveis, é possível citar: o número de apreensões de drogas em aeroportos, o número de fraudes previdenciárias desbaratadas, etc.

Ao longo da pesquisa, foram construídos diversos conceitos que deram supedâneo a estruturação de um processo-modelo que, ao fim e ao cabo, busca dar uma singela contribuição ao bem-estar social por via da prevenção criminal.

Diversas mazelas sociais marcam forte presença em nosso país. Elas açoitam diuturnamente a população, em especial a mais carente. O crime é uma das mais maléficas. A escolha pelo desvio de recursos públicos como estudo de caso também tem o objetivo de capitanear um olhar diferente às injustiças do dia-a-dia. Que se enxergue esse delito como

corresponsável pela criança analfabeta, pela doença oriunda da falta de saneamento, pela morte decorrente da falta de atendimento em hospitais, pelos necessitados que gritam por socorro sem que ninguém os ouça. Por todos esses injustiçados, recorre-se a ciência para que as Redes Vulneráveis ao crime sejam meramente...vulneráveis, mas dificilmente se tornem crimes consumados.

REFERÊNCIAS

- BAUMAN, Z. **Tempos Líquidos**. Rio de Janeiro: Zahar, 2007.
- BERNABEU, María del Carmen Batanero; CASTRO, Antonio Estepa; GODINO, J. Díaz. Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria. **Suma**: Suma, Barcelona, n. 9, p.25-31, 1991. Disponível em: <<http://revistasuma.es/revistas/9-otono-1991/analisis-exploratorio-de-datos-sus.html>>. Acesso em: 10 set. 2019.
- BRASIL. Presidência da República. Ministério da Justiça. Secretaria Nacional de Segurança Pública. **Doutrina Nacional de Inteligência da Segurança Pública – DNISP**. – 4.ed. rev. e atual. – Brasília: Secretaria Nacional de Segurança Pública, 2015.
- BROOKES, B. C. **The foundations of information science: philosophical aspects**. Journal of Information Science, v.2, p.125-133, 1980.
- BUSSAB, Wilton de O.; MORETTIN, Pedro A.. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010.
- BUZAN, Tony; BUZAN, Barry. **The Mind Map Book**: How to use radiant thinking to maximize your brain's untapped potential. New York: Penguin Books, 1994.
- CALDARELLI, Guido. **Scale Free Networks**: Complex Webs in Nature and Technology. Oxford: Oxford University Press, 2007.
- CAPURRO, R. **Epistemologia e Ciência da Informação**. Belo Horizonte: V Encontro Nacional de Pesquisa em Ciência da Informação, 2003.
- CHARTRAND, Gary; LESNIAK, Linda; ZHANG, Ping. **Graphs & Digraphs**. 6. ed. Kalamazoo: Crc Press, 2015.
- CHERVEN, Ken. **Mastering Gephi Network Visualization**. Birmingham, UK: Packt Publishing Ltd, 2015.
- CHRISTAKIS, Nicholas A.; FOWLER, James H. **O Poder das Conexões**: A Importância do Networking e como ele molda nossas vidas. Rio de Janeiro: Campus - Elsevier, 2010.
- COOK, Peter. **Layouts**. 2019. Disponível em: <<https://www.d3indepth.com/layouts/>>. Acesso em: 15 set. 2019.
- CUNNINGHAM, Graham. The Sumerian Language. In: CRAWFORD, Harriet. **The Sumerian World**. Abingdon: Routledge, 2013.
- DEVORE, Jay L.. **Probabilidade e Estatística para Engenharia e Ciências**. 6. ed. São Paulo: Cengage Learning, 2006.
- ERCIYES, Kayhan. **Complex Networks**: An Algorithmic Perspective. Boca Raton: Crc Press, 2015.
- HANSEN, Derek; SHNEIDERMAN, Ben; SMITH, Marc. **Analyzing social media networks with NodeXL**: Insights from a connected world. Burlington: Elsevier, 2011.
- HASLWANTER, Thomas. **An Introduction to Statistics with Python**: With Applications in the Life Sciences. Berlin: Springer, 2016.
- HEYMANN, Sébastien. **Layouts in Gephi**. 2011. Disponível em: <<https://gephi.wordpress.com/2011/06/13/new-tutorial-layouts-in-gephi/>>. Acesso em: 15 set. 2019.

- HU, Yifan. Efficient, high-quality force-directed graph drawing. **The Mathematica Journal**. Champaign, p. 37-71. jan. 2006.
- ILHAN, Nagehan; GÜNDÜZ-ÖGÜDÜCÜ, Sule; ETANER-UYAR, A. Sima. Introduction to Social Networks: Analysis and Case Studies. In: GÜNDÜZ-ÖğÜDÜCÜ, Şule; ETANER-UYAR, A. Şima. **Social Networks: Analysis and Case Studies**. Londres: Springer, 2014. Cap. 1. p. 1-18.
- JUSTEN FILHO, Marçal. **Curso de Direito Administrativo**. 10. ed. São Paulo: Revista dos Tribunais, 2014.
- FREEMAN, Linton Clark. **The Development of Social Network Analysis: A Study in the Sociology of Science**. Vancouver: Empirical Press, 2004.
- KOBOUROV, Stephen G.. Spring Embedders and Force Directed Graph Drawing Algorithms. **Arxiv**, Ithaca, v. 1201, n. 3011, p.1-23, 14 jan. 2012.
- LARSON, Ron; FARBER, Betsy. **Estatística Aplicada**. 6. ed. São Paulo: Pearson, 2016.
- LATORA, Vito; NICOSIA, Vincenzo; RUSSO, Giovanni. **Complex Networks: Principles, Methods and Applications**. Cambridge: Cambridge University Press, 2017.
- MARINEAU, Rene. **Jacob Levy Moreno (1889-1974): Pai do Psicodrama, da Sociometria e da psicoterapia de grupo**. São Paulo: Editora Agora, 1992.
- MICKENBERG, Risa; DUGAN, Joanne. **Taxi Driver Wisdom**. San Francisco: Chronicle Books, 2016.
- NGUYEN, Trung. **Packs of circles to present hierarchical data**. 2018. Disponível em: <<https://travisnguyen.net/dataviz/2018/02/01/circle-pack/>>. Acesso em: 15 set. 2019.
- OBJECT MANAGEMENT GROUP (Estados Unidos). **Business Process Model and Notation (BPMN)**. 2. ed. Needham: OMG, 2013. 532 p. Disponível em: <<https://www.omg.org/spec/BPMN/2.0/About-BPMN/>>. Acesso em: 30 ago. 2019.
- PINHEIRO, João Ismael D. et al. **Estatística Básica: A Arte de Trabalhar com Dados**. Rio de Janeiro: Elsevier, 2009.
- PINHEIRO, João Ismael D. et al. **Probabilidade e Estatística: Quantificando a Incerteza**. Rio de Janeiro: Elsevier, 2012.
- PINILLA, Aparecida; RIGONI, Cristina; INDIANI, Maria Thereza. **Português - Ensino a Distância - UFRJ: A Classe dos Sintagmas**. [20--]. Disponível em: <<http://www.pead.letas.ufrj.br/>>. Acesso em: 19 set. 2019.
- REINO UNIDO. Phil Williams. Nações Unidas - Conselho Econômico e Social (Ed.). **The United Nations and Transnational Organized Crime**. Londres: Frank Cass, 1996.
- ROBINS, Garry. **Doing Social Network Research: Network-based Research Design for Social Scientists**. Londres: Sage, 2015.
- ROBINSON, I.; WEBBER, J.; EIFREM, E. **Graph Databases – New Opportunities for connected Data**. O'Reilly, Sebastopol, Califórnia, 2a edição, 2015.
- SANTOS, B. S. **Um discurso sobre as ciências**. 16. ed. Porto: B. Sousa Santos e Edições Afrontamento, 2010. 59p.

SASAKI, Bryce Merkl. **Graph Databases for Beginners: Why Connected Data Matters**. 2018. Disponível em: <<https://neo4j.com/blog/why-graph-data-relationships-matter/>>. Acesso em: 30 ago. 2018.

STANDING, Lionel; CONEZIO, Jerry; HABER, Ralph Norman. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. **Psychonomic Science**, [s.l.], v. 19, n. 2, p.73-74, ago. 1970. Springer Nature. <http://dx.doi.org/10.3758/bf03337426>.

VUKOTIC, Aleksa et al. **Neo4j In Action**. Shelter Island: Manning Publications Co., 2015.

WASSERMAN, Stanley; FAUST, Katherine. **Social Network Analysis: Theory and Applications**. Cambridge: Cambridge University Press, 1994.

WATTS, Duncan James. **Seis Graus de Separação: A Evolução da Ciência de Redes em uma era conectada**. São Paulo: Leopardo, 2009.

WERSIG, G. **Information Science: The Study of Postmodern Knowledge Usage**. Information Processing & Management, Vol. 29, pp 22-239. Berlin: Pergamon Press Ltd, 1993.