CARACTERIZAÇÃO DAS REDES DE INFRATORES EXTRAÍDAS DE OCORRÊNCIAS POLICIAIS E IDENTIFICAÇÃO DE PESSOAS-CHAVE

WELLICE DOS SANTOS FRAGA

CARACTERIZAÇÃO DAS REDES DE INFRATORES EXTRAÍDAS DE OCORRÊNCIAS POLICIAIS E IDENTIFICAÇÃO DE PESSOAS-CHAVE

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dorgival Olavo Guedes Neto

Belo Horizonte

Setembro de 2009

dos Santos Fraga, Wellice
F811c Caracterização das redes de infratores extraídas de ocorrências policiais e identificação de pessoas-chave / Wellice dos Santos Fraga. — Belo Horizonte, 2009
xxiv, 55 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais
Orientador: Prof. Dorgival Olavo Guedes Neto

1. Redes sociais - Teses. 2. Criminalidade - Teses. 3. Estrutura de Redes Complexas - Teses. 4. Redes Small-World - Teses. 5. Grafos - Teses. I. Orientador. II. Título.

CDU 519.6*62(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Caracterização de infratores e identificação de pessoas-chave em ocorrências policiais

WELLICE DOS SANTOS FRAGA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. DORGIVAL OLAVO GUEDES NETO - Orientador
Departamento de Ciência da Computação - UFMG

PROF. JOÃO JOSÉ VASCO PEIXOTO FURTADO
Universidade de Fortaleza

PROF. WAGNER MEIRA JÚNIOR
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 28 de setembro de 2009.

Ao meu irmão, à minha família e aos meus amigos, pilares da minha vida.

Agradecimentos

Foram tantas pessoas que me ajudaram, direta ou indiretamente, que não posso deixar de agradecê-las. De formas distintas, essas pessoas fizeram diferença na minha vida pessoal, acadêmica e profissional.

Primeiramente, agradeço a Deus pela vida, por colocar pessoas maravilhosas no meu caminho, por tudo, mas principalmente por ser responsável pela constante renovação da minha esperança. A quem eu sempre recorro nos momentos mais difíceis e, inexplicavelmente, me sinto tranquila e convicta de que vai dar tudo certo.

Agradeço ao meu irmão, meu melhor amigo, pelo amor, companheirismo e pela confiança que ele sempre depositou em mim. Sua presença me dá segurança e certeza de que nunca estou sozinha. À minha família, agradeço pelo amparo, educação e constante incentivo na vida acadêmica. Com certeza, desempenharam papéis importantes na trajetória que tracei.

Ao longo de toda a minha vida estudantil, foram vários mestres e, cada um, de uma forma ou outra, contribuíram para as escolhas que fiz na vida acadêmica. Agradeço não só pelo repasse de conhecimento, mas também por me instigarem a querer saber mais, por me ensinarem a buscar o caminho das pedras e a caminhar com as próprias pernas. Especialmente, agradeço a meu orientador Dorgival, pessoa fantástica que de forma muito inteligente consegue transformar trabalhos difíceis da computação em tarefas prazerosas que nos mantêm fascinados por essa ciência. Obrigada por me guiar e por ser um exemplo de profissional e pessoa. À Polícia Militar de Minas Gerais, agradeço por terem disponibilizado a base de dados que foi utilizada neste trabalho. Aos amigos do SIDS, pelo aprendizado e incentivo constantes. Saibam que esse apoio foi decisivo.

Muitos amigos e colegas não sabem o quanto foram e são importantes na minha vida. Mesmo que não saibam que esses agradecimentos são destinados a eles, quero agradecê-los pela amizade, conversas, troca de conhecimento, brincadeiras etc. As vezes um sorriso ou uma frase eram suficientes para melhorar o humor e revigorar as forças. Em especial, agradeço a amizade fiel das “lavadeiras” que estão sempre

presentes. Aos amigos do CEFET, que por vezes nos distanciamos mas continuamos mantendo o mesmo sentimento e amizade por anos. Não posso deixar de mencionar os amigos da Grad022 que compartilharam comigo muitas lutas e vitórias. Contribuíram para que minha vida acadêmica fosse mais feliz. Obrigada a vocês e, especialmente, às meninas super-poderosas, pelo companheirismo nos trabalhos e também nas viagens, pelos momentos difíceis e felizes, enfim, pela amizade.

A todos vocês, meu agradecimento sincero, muito obrigada por tudo!

Resumo

Entender como as organizações criminosas estão interligadas é de fundamental importância no desenvolvimento de ações e táticas de combate ao crime. Uma das formas de aumentar a quantidade de informações a respeito dessas organizações é explorar os dados que a própria polícia já possui, como os dados de registros de ocorrências policiais, e extrair deles informações que até então estavam ocultas.

Neste trabalho, foram utilizadas diversas redes de criminosos extraídas a partir da base de registros de ocorrências da Polícia Militar de Minas Gerais. Foram calculadas várias métricas visando caracterizar a rede e extrair informações importantes sobre seu funcionamento e forma como os indivíduos se relacionam.

A primeira parte consistiu no tratamento da base, com o objetivo de extrair os indivíduos armazenados na mesma e apontar as réplicas de forma a identificar os indivíduos unicamente na base. Após realizada a deduplicação da base, foi possível extrair as redes de criminosos obtidas de forma que dois infratores são relacionados caso eles tenham aparecido em qualquer registro de ocorrência juntos.

Análises mostraram que as redes de criminosos podem ser classificadas como *small-world* o que indica que o fluxo de informações nas redes funciona bem e que a comunicação é feita de forma eficiente. Com isso, identificar indivíduos que desempenham papéis importantes na comunicação dessas redes e isolá-los pode ajudar a desestruturar essas organizações criminosas e, por isso, esses indivíduos devem ser os alvos das ações policiais.

Neste trabalho, também foram comparados diferentes algoritmos utilizados na identificação de nós importantes. Foi possível mostrar que a utilização da métrica de centralidade do nó (*betweenness*) é a mais apropriada para indicar sua importância, pois obtém um resultado eficaz, muito próximo do ótimo, e é computacionalmente mais eficiente do que a solução ótima.

Esperamos que os resultados obtidos e a ferramenta para identificação de pessoas-chave possa auxiliar no trabalho investigativo da polícia ajudando a solucionar os casos com maior rapidez e a desestruturar as organizações com maior eficiência.

Palavras-chave: Redes sociais, Criminalidade, Estrutura de Redes Complexas, Redes Small-World, Grafos.

Abstract

Understanding how criminal organizations are connected is extremely important in the development of actions and tactics to fight crime. One way to increase the amount of information over these organizations is to explore the data that the police already has, such as police reports, and extract from them information that were previously hidden.

In this work we used the criminals' networks extracted from the database of police reports from the Military Police of Minas Gerais. Several empirical measurements were made to characterize the network and extract important information concerning it's operation and how the individuals are related.

The first part consisted in the preparation of the base, aiming to extract the individuals stored in that base and point out the replicas in order to identify the individuals uniquely in that base. After carrying out the deduplication of the base, it was possible to extract the criminal networks which was obtained in such a way that two offenders are related if they have appeared in any police report together.

Analysis showed that the criminals' networks can be classified as small-world which indicates that the flow of information in these networks works well and that communication is done efficiently. Thus, identifying individuals who play important roles in these networks and isolate them can help to dismantle these criminal organizations and therefore these individuals should be the targets of police actions.

In this work, we also compared different algorithms used in the identification of important nodes. It was possible to show that the use of the metric of betweenness is the most appropriate because it leads to an effective result, very close to the optimum, but it is computationally more efficient than the optimal solution.

We hope that the results obtained here and the tool for identifying key people may assist in the police's investigative work by helping them solve cases faster and disrupt the organizations more efficiently.

Keywords: Social networks, Criminality, Structure of Complex Networks, Small-World Networks, Graphs.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

CBMMG	Corpo de Bombeiros Militar de Minas Gerais
PCMG	Polícia Civil de Minas Gerais
PMMG	Polícia Militar de Minas Gerais
REDS	Registro de Eventos de Defesa Social
SIDS	Sistema Integrado de Defesa Social

Lista de Algoritmos

Sumário

Capítulo 1

Introdução

O problema da criminalidade é uma realidade e, atualmente, é um dos maiores problemas nas médias e grandes cidades em todo o mundo. Cresce cada vez mais o número de indivíduos que acabam se tornando marginais e se ligando a organizações criminosas, como quadrilhas de assaltantes, traficantes de drogas e outros tipos de gangues. Para combater essas organizações, a polícia trabalha com recursos escassos e carece de ferramentas que a auxiliem em suas funções. Por exemplo, não existem nem policiais nem cadeias suficientes para tantos criminosos e, portanto, é necessária uma abordagem mais focada, estratégica e eficiente que consiga combater o crime organizado sem que seja necessário prender todos os criminosos, pois isso é totalmente inviável, senão impossível.

A criminalidade pode ser dividida em dois grupos básicos:

- O crime desorganizado que se caracteriza por crimes avulsos, ocasionais, como assaltos em sinais e sequestros relâmpagos que dependem mais das situações oportunas para o bandido do que de uma estratégia bem elaborada.

- O crime organizado, por outro lado, que é caracterizado por ações minuciosamente planejadas. É o caso do assalto a banco, que antes de ser executado exige dos bandidos um estudo detalhado de todo o esquema de funcionamento do banco, descobrindo, por exemplo, qual o horário de abertura do cofre. Outro exemplo claro do crime organizado é o do tráfico de drogas, que possui logísticas muito bem definidas para fazer com que a droga chegue até os consumidores.

Com o aumento da criminalidade e da violência, cresce a sensação de insegurança da população que é intensificada pela sensação de impunidade. É necessário que a polícia se organize melhor e utilize técnicas de inteligência no trabalho investigativo

para ser capaz de apontar com maior eficiência os culpados de cada crime para que eles possam receber as devidas punições.

Por esses motivos, entender como as organizações criminosas estão estruturadas é de fundamental importância no desenvolvimento de ações e táticas de combate ao crime. Para atingir esse objetivo, as organizações criminosas podem ser modeladas como redes para que seja possível analisar como os criminosos se relacionam e quais são seus papéis dentro das redes.

1.1 Objetivos

Este trabalho tem como objetivo caracterizar a rede de infratores contida na base de boletins de ocorrência da Polícia Militar de Minas Gerais e projetar e criar uma ferramenta capaz de apontar dentre esses indivíduos quais são aqueles que desempenham papéis-chave no funcionamento de suas respectivas redes de criminosos. Esses indivíduos poderão ser os alvos de ações policiais para desestruturar, com maior eficiência, o funcionamento das redes das quais eles fazem parte.

1.2 Motivação

Existe um grande volume de dados armazenados sobre os fatos policiais atendidos pela Polícia Militar de Minas Gerais, no entanto há poucas iniciativas de tratamento desses dados e recuperação de informações importantes para o desenvolvimento do trabalho policial.

A importância do combate à criminalidade fala por si mesma. No entanto, tão importante quanto o combate, é a prevenção. Foge do escopo deste trabalho discutir a causa do aumento da criminalidade percebido nos últimos anos. Entretanto, é fundamental salientar o benefício que pode trazer uma ferramenta capaz de identificar quais são as redes de criminosos "escondidas" na base e apontar, dentro dessas redes, os indivíduos-chave. Assim, a principal motivação para este trabalho é que a utilização de técnicas de computação e análise de redes sociais auxiliam na tarefa de identificação das redes de criminosos permitindo entender sua estrutura e características comuns. A identificação dessas características é fundamental no trabalho de desestruturação dessas redes.

Ao se desestabilizar uma organização criminosa, ela é impedida de entrar em ação, mesmo que esse impedimento seja temporário. Assim, focar as ações policiais na desestabilização das redes de criminosos é também uma forma de prevenção e não só

de combate ao crime.

A utilização das técnicas que aqui serão discutidas possibilitará o desenvolvimento de estratégias mais eficientes que tenderão a maximizar o benefício dos recursos alocados para as entidades responsáveis pela segurança pública.

1.3 Contribuição

Para realizar a caracterização das redes de criminosos extraídas da base de ocorrências policiais, recorreu-se a fundamentos da Teoria dos Grafos e a modelagem dessas redes como redes complexas, o que será detalhado melhor nas seções seguintes. O estudo de redes complexas é bastante amplo e serve para modelar coisas de diferentes naturezas e áreas do cotidiano e da ciência. Com isso, acreditamos que este trabalho traz uma contribuição para a área de redes complexas uma vez que foram comparados diferentes algoritmos utilizados na identificação de nós importantes de uma rede. Foi possível mostrar que a utilização da métrica de centralidade do nó (*betweenness*) é a mais apropriada para indicar sua importância, pois obtém um resultado eficaz, muito próximo do ótimo, e é computacionalmente mais eficiente do que a solução ótima.

1.4 Estrutura do Texto

O conteúdo desta dissertação está dividido da seguinte forma: o capítulo 2 descreve a base de dados dos boletins de ocorrências policiais, introduz os conceitos básicos relacionados à análise de redes sociais e também discute alguns dos trabalhos relacionados. A caracterização da rede de infratores é mostrada no capítulo 3, onde são analisadas algumas estatísticas das redes encontradas na base e também analisa-se a estrutura topológica dessas redes. O capítulo 4 discute diferentes algoritmos para identificação de pessoas importantes e qual o impacto causado na rede após a remoção dessas pessoas da rede. Por fim, o capítulo 5 conclui o trabalho apresentado e mostra algumas opções de continuação para este trabalho.

Capítulo 2

Conceitos e Trabalhos Relacionados

2.1 Base de Dados dos Boletins de Ocorrências Policiais

Este trabalho parte do princípio de que é primordial entender como as organizações criminosas estão interligadas para que seja possível melhorar as formas de combate ao crime. Além disso, entende-se que uma das formas de aumentar a quantidade de informações a respeito dessas organizações consiste em explorar os dados que a própria polícia já possui, como, por exemplo, os dados de registros de ocorrências policiais, e extrair deles informações que até então estavam ocultas.

2.1.1 História da Base

No município de Belo Horizonte, os registros de ocorrências policiais que antes eram feitos manualmente em um formulário de papel próprio para esse fim, passaram a ser digitais, a partir de Agosto de 2005. Nesta data, foi implantado um sistema Web, o Registro de Eventos de Defesa Social (REDS), para realização do registro eletrônico das ocorrências policiais, popularmente conhecidas como B.O. - Boletim de Ocorrência. O sistema REDS é destinado ao registro de qualquer fato que tenha necessitado ou necessite da intervenção dos seguintes órgãos de defesa social: Polícia Militar de Minas Gerais (PMMG), Polícia Civil de Minas Gerais (PCMG) e Corpo de Bombeiros Militar de Minas Gerais (CBMMG). O REDS serve também para registrar fatos que podem acarretar problemas futuros para o indivíduo como é o caso da perda ou extravio de documentos e objetos pessoais. O registro no REDS desses fatos é fundamental para livrar a pessoa de responsabilidades civis e penais injustas.

As unidades policiais passaram a utilizar o REDS gradativamente e, hoje, todas as unidades policiais de Belo Horizonte e região metropolitana já realizam os registros eletronicamente através do REDS. O projeto de expansão da utilização do REDS nas demais cidades do estado de Minas Gerais já está em andamento sendo que os grandes centros urbanos, como Juiz de Fora e Montes Claros, também já começaram a registrar os fatos diretamente nos formulários eletrônicos do REDS e estão abolindo a utilização dos formulários de papel com preenchimento manual. É importante citar que a impressão do registro policial continua existindo, pois ele é considerado um documento e é necessário em muitos casos para que seja possível a adoção de providências legais. O relatório com os dados completos do registro é formatado pelo sistema para que o policial responsável possa realizar sua impressão.

A base de dados que armazena os dados do sistema REDS foi a base utilizada no desenvolvimento deste trabalho. Ela está armazenada no Sistema Integrado de Defesa Social (SIDS) que é um sistema modular, integrado, que, após a implantação de todos os seus módulos, permitirá a gestão das informações de defesa social do estado de Minas Gerais que digam respeito a ocorrências policiais e de bombeiros, a investigação policial, ao processo judicial e à execução penal, respeitadas as atribuições legais dos órgãos que o compõem.

2.1.2 Processo de Geração da Base da Dados

O processo de geração da base se inicia de duas maneiras diferentes: uma em que há a atuação direta dos órgãos de defesa social realizando o atendimento à ocorrência no próprio local onde o evento ocorreu e outra em que o(s) indivíduo(s) procura(m) diretamente uma unidade policial para registrar um fato e solicitar as providências cabíveis.

O atendimento direto à ocorrência também pode se iniciar de maneiras diferentes. O meio mais conhecido é através do Centro de Comunicações para o qual o cidadão liga em caso de emergência. É o caso do 190 para a Polícia Militar, 197 para a Polícia Civil e 193 para o Corpo de Bombeiros Militar. O atendimento também pode ter sido solicitado diretamente a um órgão policial ou a um policial que já estava na rua. O policial, ainda, pode ter se deparado com a ocorrência e por iniciativa, ter realizado o atendimento. Outras formas são em decorrência de operações policiais rotineiras ou ainda motivadas por denúncias anônimas.

O registro detalhado desses fatos é de fundamental importância para a instauração futura de inquéritos e possíveis investigações que possam se fazer necessárias. Assim, o sistema REDS provê uma forma de inserir os dados na base de maneira estruturada

para permitir um processamento posterior desses dados. Por exemplo, para a geração de estatísticas, é fundamental que os dados estejam parametrizados. Essas estatísticas são importantes não só para medir o nível da criminalidade, mas também para melhorar as operações policiais visando combater e prevenir esses crimes de forma mais eficiente.

Para tentar evitar que dados importantes não sejam cadastrados como ocorria facilmente com o boletim de ocorrência feito no papel, em que o policial ficava muito mais livre para lançar as informações da forma que lhe conviesse, o REDS está organizado em diferentes seções de forma a facilitar o preenchimento correto dos dados do evento ocorrido. A seção de Dados Gerais contém informações como data e hora do evento, endereço completo onde o mesmo ocorreu, natureza do evento (tipo do delito/ocorrência), dentre outros. A seção de envolvidos permite informar o tipo de envolvimento da pessoa com a ocorrência, seus dados pessoais e características físicas de forma a realizar a caracterização completa da pessoa. Através do tipo de envolvimento de um indivíduo na ocorrência, é possível separar os envolvidos em grupo de autores, vítimas, testemunhas, dentre outros. A seção de veículos possui campos específicos para o lançamento de dados como placa, chassi, renavam, marca, modelo, ano, condutor, enfim, todos os dados necessários para identificar e descrever o veículo envolvido no fato sendo registrado. Outras seções do sistema são destinadas ao cadastramento de documentos pessoais, cheques e/ou cartões, materiais utilizados, armas de fogo, equipe que realizou o atendimento e viaturas. Todas possuem campos estruturados de forma a facilitar a recuperação de informações através desses dados.

A partir dos dados dos envolvidos cadastrados nos registros serão identificadas pessoas que estarão relacionadas entre si sempre que aparecerem em um mesmo registro policial. Esses relacionamentos farão com que as redes sociais dos envolvidos surjam e são essas redes que foram objeto de estudo do trabalho descrito aqui. Através de análises topológicas das redes é possível apontar as pessoas mais importantes das redes por ocuparem posições estratégicas. Essas pessoas deverão ser o foco das ações policiais para desestruturar as organizações criminosas com maior eficiência.

Apesar do sistema REDS armazenar registros de ocorrências da Polícia Militar, Polícia Civil e Corpo de Bombeiros, no desenvolvimento deste trabalho foram utilizados somente os dados dos registros gerados pela Polícia Militar e pelo Corpo de Bombeiros. Além disso, foram utilizados somente os dados dos registros cadastrados até setembro de 2008 o que equivale a um total de 774.063 registros nos quais foram cadastradas 1.465.074 pessoas.

2.1.3 Classificação dos Tipos de Ocorrências

Todos os registros cadastrados no REDS devem ser classificados de acordo com o tipo de fato sendo relatado e esses tipos são denominados natureza do fato.

Cada natureza equivale a um tipo de fato específico como FURTO, ROUBO, CALÚNIA, ESTUPRO, INCÊNDIO, dentre outras. Existe um total de 1.190 naturezas catalogadas e que estão agrupadas em 21 grandes grupos, os quais estão descritos no apêndice A. Para exemplificar, citaremos algumas das categorias mais importantes para o escopo deste trabalho.

- *INFRAÇÕES CONTRA A PESSOA*: esse grupo contempla todas as naturezas relacionadas a pessoas e inclui, por exemplo, homicídio, lesão corporal, abandono de incapaz, recusa/dificulta/não assiste idoso, dentre outras;
- *INFRAÇÕES CONTRA O PATRIMÔNIO*: como o próprio nome já diz, contempla as infrações relacionadas ao patrimônio, por exemplo, roubo, furto, extorsão, estelionato, dano, dentre outros;
- *INFRAÇÕES CONTRA OS COSTUMES E FAMÍLIA*: contempla naturezas como estupro, corrupção de menores, entrega de filho menor a pessoa inidônea, jogo de azar, perturbação da tranquilidade etc;
- *INFRAÇÕES REFERENTES A SUBSTÂNCIAS ENTORPECENTES*: inclui tráfico de substância entorpecente, associação para o tráfico, adquirir/guardar/trazer droga para uso próprio, dentre outras;

2.2 Análise de Redes Sociais

Ultimamente tem-se notado vários trabalhos envolvendo o estudo de redes sociais e o trabalho aqui descrito também envolve a análise de uma rede social. Mas, exatamente, do que se trata o termo "Rede Social"?

A resposta para essa pergunta começa pelo entendimento sobre o que é uma rede. Uma rede é constituída por um grupo de elementos que estão conectados entre si de alguma forma. É possível enumerar vários tipos de redes, por exemplo, a rede de computadores que formam a Internet, a rede formada pelas rotas entre os diferentes aeroportos, ou ainda a rede formada pelos neurônios que compõem o cérebro humano, e também a rede de pessoas com as quais um indivíduo mantém contato. Enfim, as redes podem ser de diversas naturezas, mas são ditas sociais quando os elementos que as constituem representam seres humanos. As conexões entre os indivíduos também

podem representar diferentes tipos de ligações, podendo, por exemplo, se referir a relacionamentos profissionais, interesses em comum ou qualquer outra característica que se deseje representar.

Essas redes, em geral, só se tornam objeto de estudo quando os elementos que as constituem, sozinhos, não são suficientes para explicar a lógica de funcionamento da rede como um todo. Nesses casos, é necessário avaliar a maneira como os elementos da rede estão conectados e como eles interagem entre si para que se possa entender o comportamento completo da rede. Por exemplo, para entender o funcionamento do cérebro não basta analisar um único neurônio, é necessário analisar a estrutura completa. Ou seja, o foco não é sobre os itens da rede, mas sobre o todo que ela constitui.

Conforme mencionado anteriormente, vários estudos já foram feitos sobre diferentes tipos de redes. Matematicamente, o estudo das redes está intimamente ligado à Teoria dos Grafos, que estuda as relações entre os objetos de um determinado conjunto. Um grafo é definido como uma estrutura G(V,A) onde V é um conjunto não vazio de objetos denominados vértices e A é um conjunto de pares não ordenados de V, denominados arestas. As arestas podem ter propriedades associadas, sendo que a mais comum é o peso da aresta que, em geral, representa algum custo na ligação entre os elementos ou o número de vezes que a interação ocorreu. Além disso, dependendo do que o grafo represente, as arestas podem ou não ser direcionadas. Por exemplo, na utilização de um grafo para representar uma rede de estradas, o sentido das arestas seria utilizado para mostrar o sentido do fluxo dos carros permitindo representar se a estrada é de mão única ou mão dupla. Na utilização de um grafo para modelar uma rede social, os elementos da rede serão representados pelos vértices, também chamados de nós, e as arestas serão as ligações entre os elementos. E dentro dessa linha de pesquisa de Teoria de Redes ou Grafos, uma frente que tem ganhado muito espaço é o estudo sobre Redes Complexas. Uma rede complexa é aquela em que suas propriedades topológicas não são triviais, ou seja, não apresentam padrões regulares ou puramente aleatórios. De forma geral, uma rede complexa é definida como aquela em que a distribuição de graus dos nós não obedece a uma distribuição normal. Sendo que grau do nó corresponde ao número de ligações que cada nó possui. Ou seja, não se trata de uma rede com ligações aleatórias. Na literatura consta ainda que os estudos realizados mostraram que várias das redes reais são classificadas como redes complexas. As redes sociais, por sua vez, também representam um tipo de rede complexa que são formadas por pessoas que mantêm entre si algum tipo de relacionamento. Esses relacionamentos podem ser uma forma para modelar as idéias em comum dos indivíduos representados, ou seus vínculos afetivos, familiares, dentre outros. A ligação entre os nós depende da

característica que se quer estudar.

Existem várias métricas de redes que levam em consideração seus relacionamentos (arestas) e que são de extrema importância no entendimento de como a rede se comporta.

2.2.1 Propriedades Topológicas

As propriedades topológicas são utilizadas para estudar as características estruturais das redes complexas e permitem entender como seus nós estão interligados. Algumas propriedades são de um nó específico, outras refletem informações da rede (grafo) completa. A seguir são definidas as principais propriedades utilizadas nas análises de redes sociais e a forma como essas métricas são calculadas.

Grau do nó

O grau é a propriedade mais simples de um nó e corresponde ao número de nós aos quais um elemento se conecta diretamente. É basicamente o número de arestas incidentes nesse nó. Algumas vezes o grau é também denominado *conectividade* do nó. Em grafos com pesos, existe o grau ponderado que contabiliza não só o número de arestas do nó como também seus respectivos pesos.

Caminho Mínimo

O caminho mínimo entre dois nós quaisquer é a menor soma de arestas pertencentes a qualquer dos caminhos que existem interligando os dois nós em questão. O cálculo do caminho mínimo não é trivial, mas em Cormen et al. [2009] são explicados alguns algoritmos utilizados para encontrar o caminho mínimo entre dois nós.

Distância média

A distância média da rede é definida como a média dos caminhos mínimos de todos os possíveis pares de vértices da rede.

Diâmetro

O diâmetro da rede é o maior valor encontrado dentre os caminhos mínimos calculados para todos os pares de nós da rede. Na literatura sobre redes complexas, existe uma ambiguidade comum associada ao termo *diâmetro*. Alguns autores o utilizam para referenciar a métrica *distância média* apesar de corresponderem a características diferentes da rede. Então para evitar equívocos ambas as métricas já foram definidas e esses termos serão utilizados neste texto respeitando essas definições.

Closeness

Essa propriedade é definida para um determinado nó, como o inverso da média dos caminhos mínimos entre esse nó e todos os demais nós da rede. Ela corresponde à habilidade do nó em acessar outros nós (com menos saltos).

Betweenness

É uma medida padrão do quão central está a posição de um nó na rede (sua centralidade). Foi originalmente introduzida para quantificar a importância de um indivíduo em uma rede social. Essa métrica indica o controle que um indivíduo exerce sobre a rede, ou seja, se ele ocupa uma posição central que permite que ele interrompa o fluxo de informação ao longo da rede.

A métrica de *betweenness* de um nó pode ser definida de forma grosseira como sendo a contagem de quantos caminhos mínimos do grafo passam pelo referido nó. Dessa forma, essa métrica mede a importância do nó na comunicação da rede. Mas precisamente, a métrica é calculada para um determinado nó i da seguinte forma:

$$B_i = \sum_{j,k \in N, j \neq k} \frac{c_{jk}(i)}{c_{jk}}, \tag{2.1}$$

sendo que c_{jk} é o número de caminhos mínimos conectando j e k. $c_{jk}(i)$, por sua vez, representa o número de caminhos mínimos conectando j e k e que passam pelo vértice i.

Coeficiente de Aglomeração/Agrupamento

O coeficiente de aglomeração/agrupamento (*clustering*) quantifica o nível de agrupamento dos nós da rede. Usualmente, é dado pela razão entre o número de conexões entre vizinhos comuns a um nó de referência, dividido pelo número de possíveis conexões entre os vizinhos comuns ao nó. Segundo Watts & Strogatz [1998], o coeficiente de aglomeração de um determinado vértice i é dado por:

$$C_i = \frac{numero\ de\ arestas\ entre\ os\ vizinhos\ de\ i}{numero\ total\ de\ possiveis\ arestas\ entre\ os\ vizinhos\ de\ i} = \frac{2V_i}{d_i(d_i - 1)}, \tag{2.2}$$

sendo que d_i é o número de vizinhos de i e, portanto, $d_i(d_i - 1)/2$ é o máximo de arestas possíveis entre eles.

Por sua vez, o coeficiente de aglomeração global do grafo é dado por:

$$C_G = \frac{1}{N} \sum_{i \in N} C_i. \tag{2.3}$$

Ou seja, equivale à média entre todos os coeficientes (locais) calculados. Os grupos de nós em que o coeficiente de aglomeração é alto são conhecidos como *clusters*.

Eficiência

Latora & Marchiori [2001] definiram a eficiência global da rede como sendo a média dos inversos dos comprimentos dos caminhos mínimos entre todos os pares de nós do grafo:

$$E = \frac{1}{N(N-1)} \sum_{i \neq j \in N} \frac{1}{d_{ij}}, \tag{2.4}$$

sendo que d_{ij} é o comprimento do caminho mínimo conectando o vértice i ao j.

No estudo de redes complexas, a eficiência desempenha o papel de medir a habilidade da rede de transmitir informação e reflete a resposta da rede a perturbações em sua estrutura. Ao contrário da distância característica, a métrica da eficiência tem se mostrado adequada também em grafos não-conectados [Crucitti et al., 2004]. A distância característica é definida como a mediana das médias de todos os caminhos mínimos conectando cada um dos vértices $v \in$ V(G) a todos os demais vértices, sendo que V(G) é o conjunto de vértices do Grafo.

2.2.2 As organizações criminosas como um grafo

Para utilizar a teoria dos grafos no entendimento das organizações criminosas, o objeto de estudo é a rede formada pelas pessoas que integram essas organizações e como essas pessoas estão relacionadas. Assim, um grafo que represente essa rede terá seus vértices representando os indivíduos da rede e as arestas representando as ligações entres eles. É necessário identificar características nesse grafo que permitam definir métricas de análises dos relacionamentos entre os nós dessas redes (os bandidos). Será que o grafo que representa as organizações criminosas apresenta algum padrão de comportamento? Será que suas características e topologia obedecem a algum modelo de rede conhecido? Essas perguntas serão respondidas ao se realizar a análise das propriedades topológicas do grafo. Técnicas de análise de redes complexas permitirão entender a estrutura das redes de criminosos, o que é fundamental para a elaboração de táticas de combate bem fundamentadas e eficientes para desmanchar essas organizações. Essas técnicas permi-

tem, ainda, apontar possíveis quadrilhas de criminosos que ainda não estavam claras para a polícia. Isso é possível analisando a conectividade dos nós e os agrupamentos (clusters) formados, caso eles existam.

2.2.3 Identificação de Pessoas-Chave

Existem vários estudos na literatura como os realizados por Albert et al. [2000] e Crucitti et al. [2004], que mostram que a maioria das redes do mundo real apresentam grande vulnerabilidade a ataques. Isso quer dizer que existem elementos importantes na rede cuja remoção desestabilizaria profundamente essas redes. Os ataques são assim chamados porque a escolha do elemento a ser removido da rede não pode ser feita de forma aleatória. O elemento escolhido precisa estar estrategicamente posicionado na rede e, por isso, se torna o "alvo" do ataque.

Assim, se puder ser comprovado que as redes de criminosos se comportam da mesma forma que tantas outras redes reais já estudadas e que suas propriedades topológicas obedecem a determinados padrões comumente encontrados nos diferentes tipos de redes do mundo real, será possível também identificar pessoas que desempenham papéis importantes no funcionamento da organização. Essa informação permitirá abordagens mais elaboradas contra o crime organizado.

É fácil perceber, por exemplo, que em uma ação de combate ao tráfico de drogas, pouco adianta prender um usuário consumidor, pois a rede continuará existindo. No entanto, se forem presas pessoas ligadas ao traficante ou distribuidor, o impacto será obviamente maior. Esse exemplo de uma rede de tráfico de drogas pode ser usado para explicar melhor a importância na identificação das pessoas-chave da rede. Considere que existe uma quadrilha responsável pela produção da droga, uma outra quadrilha de traficantes responsável pela venda para o consumidor e uma terceira, responsável por realizar a ponte entre as duas anteriores, ou seja, levar a droga do produtor para o traficante. A abordagem mais eficiente para combater o tráfico seria atacando a quadrilha de distribuidores porque assim o produtor deixaria de se comunicar com o vendedor e a rede como um todo seria abalada. No entanto, um ataque somente aos produtores ou aos traficantes por ser mais pontual, daria mais tempo para a quadrilha se reorganizar antes que os impactos fossem sentidos na demais partes da rede.

2.3 Trabalhos Relacionados

Após terem sido introduzidos os principais conceitos relacionados à análise de redes sociais, este capítulo apresenta uma visão geral sobre os trabalhos relacionados ao que

é descrito neste texto. A seção 2.3.1 aborda as principais referências bibliográficas sobre conceitos e técnicas de análise de redes sociais. Essa literatura inclui tanto a parte de caracterização topológica das redes quanto seu comportamento na presença de falhas ocasionais e também nas causadas intencionalmente. A seção 2.3.2, por sua vez, apresenta os trabalhos que tratam especificamente de organizações criminosas retratando seu comportamento e particularidades.

2.3.1 Análise de Redes Sociais

Existe uma imensidão de sistemas encontrados na natureza e na sociedade que podem ser descritos através da modelagem por redes complexas. Em função disso, esse ramo de pesquisa tem crescido muito e é possível encontrar vários trabalhos sobre o assunto. Um trabalho completo foi apresentado por Albert & Barabási [2002], no qual mostrou-se que essas redes complexas têm princípios organizacionais robustos, apesar de terem sido modeladas durante muito tempo como grafos aleatórios. Esse trabalho discute a topologia de várias redes reais através de dados empíricos e também explica os princípios de cada categoria (grafo aleatório, *small-world* e *scale-free*) utilizados na modelagem das redes complexas. Mostra, ainda, a tolerância de alguns tipos de redes à remoção de nós. Trabalho similar foi apresentado em Newman [2003] e Dorogovtsev & Mendes [2002]. Boccaletti et al. [2006] além de explicar os conceitos envolvidos nas análises de redes sociais mostram como essas idéias são aplicadas na prática em diferentes tipos de sistemas e como eles se comportam.

Ao invés de utilizar o grau do nó na classificação de uma rede *scale-free* como é feito tradicionalmente, Goh et al. [2002] propõe uma nova forma utilizando a métrica *betweenness* do nó que se apresenta mais robusta com uma distribuição obedecendo a uma lei de potência bem definida.

Exemplificando os conceitos de *small-world*, Adamic [1999] mostra como encontrar as propriedades de *small-world* na web e realiza a caracterização dessa rede. Já Marteleto [2001] mostra a aplicação da análise de redes sociais em estudos de transferência de informação.

Albert et al. [2000] contrasta o comportamento de uma rede exponencial (aleatória) ao de uma rede *scale-free*. Mais especificamente, mostra-se como esses tipos de redes se comportam diante da remoção aleatória de nós em oposição a remoção de nós cuidadosamente selecionados e que desempenham papel importante na conectividade da rede. O impacto das remoções é medido avaliando-se o efeito causado em três métricas da rede: distância média, fração do componente gigante remanescente e média dos tamanhos dos componentes menores. Crucitti et al. [2004] realiza comparação se-

melhante mas, ao invés de utilizar as três métricas citadas anteriormente para medir o impacto da remoção, utiliza-se a eficiência global. Crucitti et al. [2003] estende este estudo incluindo o impacto causado pelas remoções no coeficiente de aglomeração. Todos mostraram que as redes aleatórias se comportam de forma similar independente da forma como os nós são removidos e que as redes *scale-free* toleram muito bem a remoção aleatória enquanto que são drasticamente desestruturadas pela remoção de nós específicos.

2.3.2 Organizações Criminosas

Os trabalhos direcionados ao estudo específico de organizações criminosas não são numerosos e a maioria é focada em redes de terrorismo principalmente após o atentado de 11 de setembro. Mas como as organizações terroristas correspondem a um tipo de organização criminosa, esses trabalhos também foram levados em consideração no desenvolvimento desse projeto.

Klerks & Smeets [2001] mostra que as organizações criminosas não devem ser analisadas como hierarquias mas sim levando em consideração conceitos de redes sociais, pois essas organizações são bastante distribuídas e não obedecem a uma estrutura estática bem definida como a maioria das organizações regulares. Assim, é necessário identificar posições estratégicas na rede e quais são os indivíduos ocupando esses lugares para que seja possível lidar de maneira eficiente contra o crime organizado. Dentro desta mesma linha de raciocínio, Carley et al. [2003] mostra como desestabilizar organizações criminosas. Para tal, é necessário levar em consideração a dinamicidade da rede e eliminar as pessoas mais importantes de acordo com algumas métricas de redes complexas. De maneira semelhante, Krebs [2002] mostra como mapear redes clandestinas através de dados disponíveis em fontes de notícias. Em seguida, é feita a análise de algumas redes mostrando quais são os indivíduos principais e consequentemente os pontos sensíveis da rede.

Por sua vez, Fellman & Wright [2004] mostram como modelar redes terroristas levando em consideração algumas características comuns a essas redes e levantam três limitadores básicos nas análises de organizações criminosas:

1. Incompletude: A ausência de alguns nós e relacionamentos é quase sempre inevitável, pois nem sempre os investigadores conseguem identificar todos os envolvidos;

2. Limites confusos: Dependendo do nível de detalhe com o qual se deseja modelar a rede, pode ser tarefa difícil decidir se um elemento deve ou não ser incluído na

rede;

3. Dinamismo: as redes evoluem constantemente, mas a maioria dos modelos de análise de redes complexas existentes considera um *snapshot* da rede, ou seja, a estrutura da rede em determinado instante de tempo.

Chen et al. [2003] descreve um sistema que integra as bases de dados de diversas aplicações policiais e provê uma interface que facilita no acesso a esses dados o que auxilia no processo de investigação. Ele também analisa, de forma rápida, grandes quantidades de dados que estão inicialmente desconexos e mostra as relações presentes entre eles. Este trabalho é mais focado no processo de recuperar informações das bases policiais do que de análise das redes formadas pelos indivíduos.

O trabalho desenvolvido apresentado em Ozgul et al. [2007] é semelhante ao que está sendo descrito neste texto uma vez que ele utilizou os dados extraídos de uma base policial que armazena os dados das prisões realizadas pela polícia local. Nesse trabalho, os autores mostram como identificar os grupos de criminosos, ou seja, como extrair as quadrilhas escondidas na enorme massa de dados.

Em Xu & Chen [2008] é feita a caracterização de 4(quatro) diferentes tipos de redes ocultas as quais eles denominam "dark networks". Eles mostram através de dados experimentais que essas redes podem ser classificadas como *small-world* e *scale-free*. Em seguida, mostra-se como elas se comportam diante da remoção de nós importantes da rede contrastando a remoção baseada nos graus e na métrica *betweenness* dos nós e qual foi o impacto na eficiência global da rede. As redes *scale-free* se mostraram mais suscetíveis à remoção baseada na métrica *betweenness*. Já o trabalho apresentado por Latora & Marchiori [2004] mostra o comportamento da rede frente à remoção de nós baseando a importância dos indivíduos em função da queda na eficiência da rede causada pela remoção individual de cada nó. O nó mais importante será o que causar maior diferença entre a eficiência da rede completa e a eficiência da rede sem o referido nó. E com base nesses resultados, os trabalhos mostram como desenvolver estratégias para combater as organizações criminosas.

Capítulo 3

Caracterização da Rede

Para identificar os relacionamentos entre as pessoas armazenadas na base é preciso saber qual pessoa de um determinado registro equivale a mesma pessoa cadastrada em um outro registro para conseguir identificar como que as pessoas de uma ocorrência estão relacionadas a pessoas cadastradas em outras ocorrências. Para isso, é necessário identificar as réplicas da base.

3.1 Identificação dos Indivíduos

O primeiro desafio é conseguir identificar os indivíduos unicamente na base. Apesar de existirem campos como RG(Carteira de Identidade) e CPF (Cadastro de Pessoa Física) que potencialmente serviriam para identificar precisamente o envolvido, esses dados nem sempre estão presentes. Por se tratar de um registro de ocorrência, os dados são muitas vezes incompletos. Portanto, é necessário lidar com problemas como falta de nomes, nomes digitados de formas diferentes, apelidos diferentes etc.

Para resolver este problema, utilizou-se o PAREIA, algoritmo de deduplicação proposto por Santos et al. [2007]. Basicamente, o algoritmo compara as entidades duas a duas e atribui notas ao par de acordo com sua similaridade. Para computar o grau de similaridade entre as entidades, é utilizada a comparação probabilística, na qual é necessário definir, além de quais atributos devem ser comparados, qual a contribuição (peso) desse atributo para o resultado final, ou seja, dadas duas entidades com o mesmo valor para o atributo X, qual a probabilidade de serem a mesma entidade? Esse método é eficiente na tarefa de deduplicação, pois tende a contornar situações cotidianas em que duas pessoas distintas acabam tendo o mesmo atributo. Por exemplo, o simples fato das pessoas terem o mesmo nome, não quer dizer que sejam a mesma pessoa, como é o caso dos homônimos. Outra situação que mostra a importância em se pontuar a

relevância de um atributo é o caso dos documentos. Apesar de se ter a tendência natural de afirmar que pessoas com mesmo documento se referem ao mesmo indivíduo, isso nem sempre é verdade. Vale lembrar que muitas vezes os documentos dos pais são utilizados nos cadastros dos filhos, quando estes ainda não possuem documentos próprios. No entanto, é necessário respeitar as proporções e levar em conta que existem muito mais homônimos do que pessoas que utilizam os documentos alheios. Isso deve ser refletido no peso de cada atributo utilizado na computação da nota de similaridade das entidades. Ao realizar a comparação textual dos atributos, também são aplicados alguns algoritmos para casamento parcial que levam em consideração muitos erros de digitação comuns.

A abordagem de se avaliar o maior número de atributos é importante quando não se dispõe de todos os atributos críticos para identificar um envolvido unicamente. É comumente aceitável afirmar que dois registros são referentes à mesma pessoa quando eles possuem o mesmo nome, nome da mãe e data de nascimento. No entanto, quando todos esses dados não estão completos, outros dados podem ser utilizados de forma a agregar e permitir a deduplicação dos indivíduos. Por exemplo, quando não se tem o nome da mãe, mas se tem o nome do pai e o endereço completo e eles são iguais, aumenta consideravelmente a probabilidade de se tratar da mesma pessoa.

O resultado do PAREIA é uma lista de pares com sua respectiva pontuação de similaridade. É necessário avaliar os pares encontrados para extrair as réplicas dos registros. Fazendo-se um histograma da distribuição dos pontos, é possível separar 3 (três) conjuntos de pares. Para o intervalo com as maiores pontuações, pode-se afirmar com certeza se tratar de réplicas. Para o intervalo com os menores valores, pode-se afirmar que os pares não são réplicas. No entanto, o intervalo intermediário é incerto e quanto menor for, melhor será a precisão dos resultados encontrados. A análise do histograma ajuda a encontrar o ponto de corte, mas não é conclusiva, é necessário voltar aos dados para comparar os pares encontrados e definir se naquela pontuação os pares encontrados realmente são réplicas.

Após identificado o ponto de corte, é necessário agrupar os pares, as réplicas. Para tal, foi desenvolvida uma rotina que recebe como entrada os pares identificados pelo PAREIA e atribui novos identificadores aos indivíduos de forma que as réplicas recebam o mesmo identificador. O resultado obtido foi que os 1.465.074 indivíduos iniciais foram agrupados em 1.015.925 indivíduos distintos, ou seja, mais de 30% da base inicial era composta por réplicas. Neste ponto, após a deduplicação da base, ela está pronta para ser processada com técnicas voltadas para a análise de redes complexas.

3.2 Construção das Redes

Feita a extração dos indivíduos envolvidos nas ocorrências é necessário apontar em que situação uma pessoa está ligada a outra para que seja possível identificar as redes sociais de criminosos que estão "escondidas" na base de boletins policiais. Para o contexto da base de ocorrências policiais, uma pessoa estará ligada a outra sempre que elas tiverem tido participação na mesma ocorrência. Assim, para realizar a extração das redes dos infratores, cada uma das pessoas que aparecem na base será um nó da rede. E haverá uma aresta entre dois nós quaisquer da rede sempre que as respectivas pessoas tiverem aparecido no mesmo registro de ocorrência. Se duas pessoas aparecerem juntas em mais de um registro isso será expresso através do peso da aresta, mas é importante citar que esse peso não deverá ser considerado para o cálculo do caminho mínimo entre os nós da rede.

Neste ponto, de extração das redes contidas na base, surge outra questão importante que é decidir qual o nível de detalhamento que deve ser incluído na extração dessas redes sociais. Deve-se decidir se todos os envolvidos serão considerados, e por exemplo, se os policiais que atenderam à ocorrência serão retirados. Ou ainda se vítimas e testemunhas devem entrar no conjunto de dados a serem avaliados ou somente os autores e co-autores serão considerados. Essa decisões dependem de qual será o objetivo do estudo e influenciarão diretamente em quais características da rede se deve focar.

Como a base estudada é muito rica e abre um leque de linhas de pesquisa muito vasto, a estratégia adotada neste trabalho foi analisar somente a rede formada pelos infratores, sem considerar os policiais que fizeram o atendimento à ocorrência. Ou seja, foram incluídas somente as pessoas envolvidas diretamente na ocorrência, e que foram caracterizadas como tendo sido os autores, co-autores ou suspeitos da infração/crime. Foram incluídas também as pessoas enquadradas como condutoras do veículo para que não fossem excluídos os infratores que ficam responsáveis pela condução do veículo nas ocorrências em que os meliantes fazem uso de veículo automotor. Essas decisões foram tomadas visando focar a análise na rede dos infratores.

Uma vez que os policiais já são identificados de forma única na base do REDS, eles dispensam o processamento de deduplicação e poderão, de forma fácil, ser incorporados à rede posteriormente para a realização de trabalhos futuros.

Assim, levando-se em consideração somente os infratores, das 1.015.925 pessoas identificadas através da deduplicação restaram somente 424.591. Deste total também devem ser desconsiderados os nós isolados, pois eles não agregam informações à rede uma vez que estão desconectados dela. Restam, então, 265.964 pessoas que estão

relacionadas através de 228.905 ligações diretas e que formam a rede final que será analisada.

Para realizar a análise da rede é necessário avaliar os componentes conectados do grafo. Um componente conectado corresponde a uma porção de vértices do grafo na qual existe um caminho entre todos os possíveis pares de vértices dessa porção. Realizando a decomposição do grafo, extraído da base de ocorrências policiais, nos seus componentes conectados, pôde-se contabilizar que a rede é composta por 97.367 componentes. A figura 3.1 mostra os tamanhos dos componentes encontrados sendo que um deles é consideravelmente maior que os demais e atrapalha a visualização dos tamanhos dos demais componentes.

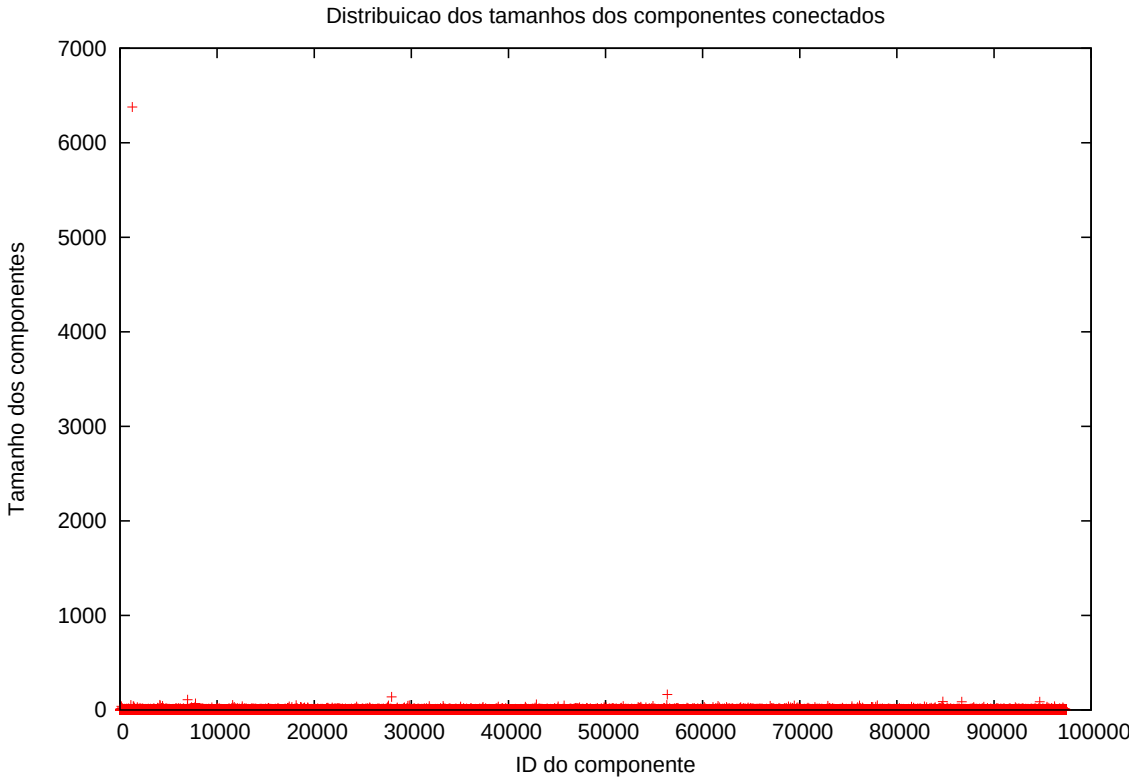

Figura 3.1. Tamanho dos Componentes Conectados

A figura 3.2 exibe novamente os tamanhos dos componentes encontrados ocultando o maior para que seja possível avaliar melhor os dados exibidos. Dessa forma, é possível perceber que a maioria dos componentes encontrados tem um tamanho pequeno.

Como o número de componentes é alto percebe-se o baixo nível de acoplamento da rede. Isso nos sugere que a rede não pode ser analisada como uma mas que é necessário

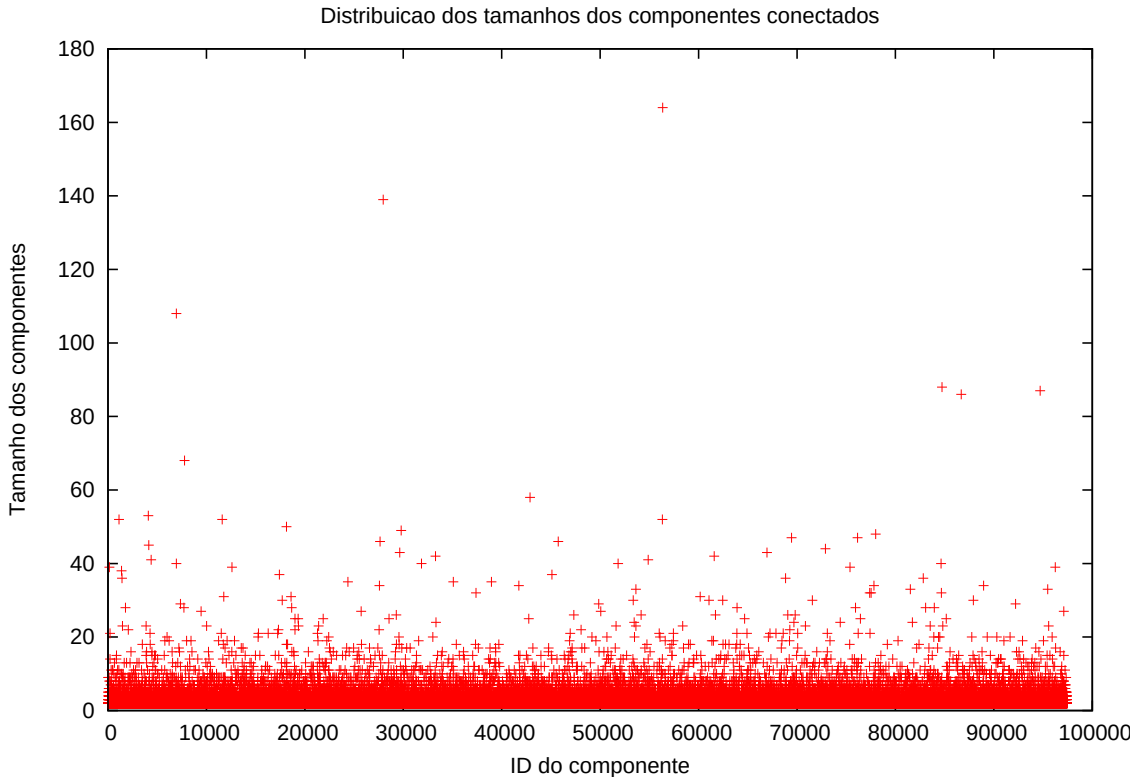

Figura 3.2. Tamanho dos Componentes Conectados Excluindo o Maior

avaliar os vários componentes encontrados para tentar extrair deles um padrão de comportamento.

Para confirmar a suspeita, é necessário verificar se a rede apresenta um componente gigante. Um componente gigante é aquele cujo tamanho é da mesma ordem de N [Boccaletti et al., 2006], onde N é o número total de nós no grafo completo. Quando uma rede apresenta um componente gigante, todas as análises podem ser feitas focadas no componente gigante, desconsiderando os demais componentes menores, uma vez que o componente gigante representa um montante significativo da rede e pode ser usado para representar a rede como um todo. A tabela 3.1 exibe uma lista identificando os maiores componentes conectados encontrados no grafo. São exibidos também o número de vértices de cada componente, assim como o número de arestas. A coluna *Fatia da rede* se refere à porcentagem em relação ao número total de nós que aquela fatia de nós da rede que formam o componente representa no grafo como um todo. Por fim, a última coluna apresenta rótulos que serão utilizados posteriormente para fazer referência aos componentes isolados. Por exemplo, ao falar do maior componente cujo ID é 1273, será citado o componente ou rede c1. O segundo maior componente será citado como

c2 etc. Essa forma facilita o entendimento sobre qual componente está sendo falado no momento.

Tabela 3.1. Maiores componentes conectados

ID.	No. Vértices	Fatia da rede	No. Arestas	Rótulo
1273	6378	2,4%	18458	c1
56359	164	0,062%	297	c2
27968	139	0,052%	264	c3
6954	108	0,041%	271	c4
84731	88	0,033%	3828	c5
94710	87	0,033%	194	c6
86678	86	0,032%	240	c7

Analisando a tabela 3.1 pode-se notar que a rede encontrada não possui um componente gigante, pois o maior componente encontrado representa uma parcela pequena da rede, que corresponde somente a 2,4% de seu tamanho total. Isso comprova a suspeita inicial de que a rede não pode ser analisada levando-se em consideração somente seu maior componente, mas que é necessário analisar separadamente cada um dos componentes encontrados e verificar se eles apresentam algum padrão ou propriedades em comum. Isso já era esperado, uma vez que é sabido que a rede de criminosos é composta por diversas quadrilhas e gangues que atuam em áreas diferentes da criminalidade e que algumas podem ter relação umas com as outras mas que existem várias que são independentes entre si.

Ainda com base na tabela 3.1, apesar do maior componente não ser significativo em relação à rede com um todo, é possível notar que seu tamanho é ordens de grandeza maior que o segundo maior componente encontrado. Além disso, o tamanho dos demais componentes é bem próximo do segundo maior e vai decaindo lentamente. A primeira dúvida que surge então é o que explicaria a discrepância no tamanho do maior componente em relação aos demais? Algumas das análises feitas nas seções seguintes tentarão esclarecer um pouco essa dúvida.

3.3 Análise Estatística das Redes de Criminosos

Antes de analisar as propriedades topológicas da rede, será feita aqui uma análise estatística dos dados sem levar em consideração a estrutura da rede.

3.3.1 Estatísticas baseadas nas pessoas (nós da rede)

Primeiramente foi calculado o número de vezes que cada pessoa aparece na base para verificarmos o percentual de pessoas reincidentes. A tabela 3.2 mostra a distribuição das pessoas de acordo com o número k de vezes que elas cometeram uma infração, ou seja, são reincidentes na base.

Tabela 3.2. Reincidência das pessoas na base

k	No. de Pessoas que cometeram k delitos	Porcentagem das pessoas
1	335985	79,13%
2	62891	14,81%
3	17016	4,01%
4	5480	1,29%
5	1858	0,44%
6	723	0,17%
7	310	0,07%
8	162	0,04%
9	78	0,02%
10	39	0,01%
+10	49	0,01%

Os dados da tabela 3.2 mostram que a parcela de infratores reincidentes é significativa e corresponde a mais de 20% dos infratores que aparecem na base. Se forem consideradas somente as pessoas que não agem sozinhas, ou seja, em alguma ocorrência essas pessoas aparecem como infratores juntamente com outros envolvidos também infratores que podem ser considerados cúmplices, o percentual de reincidência dos infratores sobe ainda mais e representa 33,32%. Ou seja, 1/3 das pessoas que aparecem nas redes de criminosos são reincidentes e, por isso, mesmo que seu delito seja leve, elas merecem mais atenção. Além disso, se forem levados em consideração os delitos cometidos, dentre os reincidentes, 40,75% das pessoas reincide no mesmo delito pelo menos uma vez. Ou seja, dentre os reincidentes, independente do número de vezes que eles cometeram algum delito, existe pelo menos um tipo de delito que foi repetido.

A distribuição das naturezas dos delitos cometidos pelas pessoas indica que mais de 53,17% são referentes a infrações e acidentes de trânsito. Isso é facilmente compreendido, pois também foram inseridos na rede os envolvidos qualificados como condutores. Após os delitos relacionados ao trânsito, as naturezas mais frequentes são:

- VIAS DE FATO / AGRESSÃO com pouco mais de 5% ;

- AMEAÇA com mais de 4% ;
- LESÃO CORPORAL também com 4% ;
- FURTO com quase 3% ;
- DANO AO PATRIMÔNIO com quase 2% ;
- ROUBO também com aproximados 2% ;

Uma curiosidade observada foi que dentre os envolvidos reincidentes 10% cometeu o delito de AMEAÇA e, depois, cometeu outro delito mais grave como VIAS DE FATO / AGRESSÃO, LESÃO CORPORAL, FURTO, ROUBO, DANO, dentre outros. Isso mostra que os delitos mais frequentes de forma geral (independente se o envolvido só aparece uma vez na base ou várias) são também os mais praticados dentre os envolvidos que cometem mais de um delito. No entanto, apesar dessa coincidência, não houve a predominância de nenhum tipo de natureza na base.

3.3.2 Estatísticas baseadas nos relacionamentos (arestas da rede)

De maneira semelhante ao que foi mostrado na seção anterior, foram feitos também alguns cálculos estatísticos levando em consideração os elos entre os indivíduos e o componente ao qual pertenciam. Lembrando que dois indivíduos foram ligados por arestas se esses indivíduos aparecem juntos no mesmo registro e, portanto, estão relacionados. As análises feitas sobre os componentes separadamente são importantes para tentar explicar a grande diferença no tamanho do componente c1, maior componente da rede, em relação aos demais.

3.3.2.1 Delitos Cometidos em Conjunto

Serão chamados de parceiros ou cúmplices, quaisquer pares de indivíduos que aparecem juntos na mesma ocorrência policial. Ou seja, na modelagem utilizando grafos, cada aresta corresponde a um par de cúmplices. Ao analisar o tipo de delito cometido pelos cúmplices, novamente grande parte é referente a delitos de trânsito. Em seguida, estão os delitos relacionados ao tráfico de drogas, lesão corporal, dentre outros. Os tipos de delito são muito específicos e estão organizados em uma tabela contendo 1.190 tipos de delito diferentes, dos quais, 288 estão presentes no grafo sendo analisado. A lista contendo a distribuição dos tipos de delito é extensa e, por isso, a tabela 3.3 mostra somente os delitos mais frequentes nas arestas do grafo.

Tabela 3.3. Principais Delitos Cometidos em Conjunto

Natureza do Delito	Frequência	Porcentagem
ACIDENTE DE TRANSITO SEM VITIMA	87031	38,02%
ACIDENTE DE TRANSITO COM VITIMA	32559	14,22%
TRAFICO ILÍCITO DE DROGAS	10231	4,47%
LESÃO CORPORAL	7772	3,40%
VIAS DE FATO / AGRESSÃO	7576	3,31%
FURTO	7534	3,29%
ROUBO	7521	3,29%
USO OU CONSUMO DE DROGAS	5228	2,28%
JOGO DE AZAR	4960	2,17%
POSSE IRREGULAR DE ARMA DE FOGO	4808	2,10%
AMEAÇA	3986	1,74%
DANO	3902	1,70%
REFERENTE A DROGA P/ USO PRÓPRIO	3653	1,60%
OUTRAS INFRAÇÕES CONTRA A PESSOA	3501	1,53%
TRAFICO DE SUBSTANCIA ENTORPECENTE	3425	1,50%
OUTRAS INFRAÇÕES CONTRA O PATRIMÔNIO	2830	1,24%
RIXA	2474	1,08%
OUTRAS OCORRÊNCIA DE TRANSITO	2135	0,93%
HOMICÍDIO	1944	0,85%
ATRITO VERBAL	1616	0,71%
OUTRA REF. SUBSTANCIAS ENTORPECENTES	1435	0,63%
PORTE ILEGAL DE ARMA DE FOGO	1424	0,62%
OUTRAS INFRAÇÕES RELATIVAS A FLORA	1007	0,44%
TODAS AS DEMAIS NATUREZAS	20353	8,89%

É possível notar que naturezas referentes ao uso/tráfico de drogas é frequente. Para facilitar a análise, os delitos foram agrupados em subcategorias o que resultou nos seguinte números:

- 12,15% são referentes às infrações contra a pessoa que constam no CÓDIGO PENAL dentre as quais estão HOMICÍDIO, LESÃO CORPORAL, RIXA, AMEAÇA, VIAS DE FATO / AGRESSÃO, dentre outras;
- 11,43% das arestas são relacionadas a TRAFICO/USO DE SUBSTANCIAS ENTORPECENTES;
- 10,60% são referentes às infrações contra o patrimônio constantes no CÓDIGO PENAL, dentre as estão FURTO, ROUBO, EXTORSÃO, DANO, ESTELIO-

NATO, dentre outras;

- as demais subcategorias não apresentaram um percentual significativo e serão desconsideradas.

Realizando cálculos similares para os maiores componentes, foi possível verificar que no maior componente (c1):

- 26,44% das arestas são relacionadas a TRAFICO/USO DE SUBSTANCIAS ENTORPECENTES;
- 20,49% são referentes às infrações contra o patrimônio que constam no CÓDIGO PENAL sendo que as naturezas mais frequentes foram FURTO, ROUBO e DANO;
- 20,31% são referentes a LEI DAS CONTRAVENÇÕES PENAIS sendo que quase sua totalidade (19,63%) foram referentes a JOGO DE AZAR.
- 12,41% são referentes às infrações contra a pessoa que constam no CÓDIGO PENAL sendo que as naturezas mais frequentes foram LESÃO CORPORAL, AMEAÇA, HOMICÍDIO e VIAS DE FATO / AGRESSÃO.

Enquanto que no segundo maior componente (c2):

- 65,90% das arestas são relacionadas a TRAFICO/USO DE SUBSTANCIAS ENTORPECENTES;
- 25,58% são referentes às infrações contra o patrimônio que constam no CÓDIGO PENAL sendo que as naturezas mais frequentes foram FURTO e ROUBO;
- 2,95% são referentes às infrações contra a pessoa que constam no CÓDIGO PENAL sendo que a natureza mais frequente foi HOMICÍDIO.

Análises semelhantes foram realizadas nos demais componentes e os resultados obtidos foram similares no que diz respeito às categorias com maiores porcentagens. Nos demais componentes, os delitos mais frequentes, em geral, são referentes ao TRAFICO/USO DE SUBSTANCIAS ENTORPECENTES, às infrações contra o patrimônio e às infrações contra a pessoa.

Num primeiro momento, ao comparar os resultados obtidos para o componente c1 com os resultados do componente c2, a grande diferença é que o maior componente é composto por uma parcela significativa de ligações provenientes de ocorrências relacionadas a Jogo de Azar. A primeira intuição é supor que essas ligações são responsáveis

por conectar pessoas que na verdade não têm ligação alguma e, consequentemente, conectar componentes que são originalmente desconexos. Essa suposição é baseada no fato de que ocorrências de Jogos de Azar estão comumente relacionadas a fechamentos de bingos e as pessoas ali participando não necessariamente têm relação umas com as outras. No entanto, por aparecerem na mesma ocorrência, terão arestas ligando-as umas às outras. Para verificar essa hipótese, foram removidas as arestas provenientes de ocorrências relacionadas a Jogo de Azar mas ainda sim o maior componente continuou muito maior que o segundo maior componente, pois continuou com mais de 5 mil nós sendo que antes possuía 6.378. Assim, a hipótese de que as ocorrências de jogos de azar poderiam ser a causa da diferença no tamanho entre as maiores redes não pode ser confirmada, pois mesmo retirando as ligações em decorrência desse tipo de delito, a maior rede continuou sendo composta por milhares de pessoas enquanto que a segunda maior possui somente 164, ou seja, a diferença nos tamanhos continuou discrepante.

A outra diferença significativa que se pode notar é que a porcentagem de arestas referentes ao código penal de infrações contra a pessoa é bem menor no segundo que no maior componente. Então, uma hipótese seria de que esse tipo de delito estaria fazendo a ponte entre diferentes componentes. Novamente, o maior componente continuou muito maior que o segundo.

Uma curiosidade observada foi que em grande parte dos componentes, não só nos maiores que foram apresentados aqui, uma parcela significativa das arestas estão relacionadas ao tráfico/uso de substâncias entorpecentes. A nova hipótese então é de que as arestas relacionadas a substâncias entorpecentes são responsáveis por conectar pessoas que, na prática, não têm nenhuma relação. Essa hipótese é fácil de compreender: usuários de drogas, sem nenhum relacionamento entre eles, podem ser pegos comprando droga de uma terceira pessoa com a qual eles também podem não ter relação alguma mas estarão conectados na rede por terem aparecido juntos na mesma ocorrência. Infelizmente, com a extração automática das redes não é possível separar os vínculos fortes e que são claramente reais dos links eventuais que culminaram por aparecer somente devido às circunstâncias. Foi feita então a remoção de todas as arestas relacionadas a substâncias entorpecentes, mas o maior componente continuou apresentando uma diferença de tamanho enorme em relação ao segundo componente.

Como última alternativa, adotou-se uma abordagem drástica: foram removidas todas as arestas relacionadas aos delitos que apresentaram os maiores percentuais. As arestas removidas estavam relacionadas a jogos de azar, drogas, roubo, dano, furto e lesão corporal. O maior componente diminuiu consideravelmente e passou a conter somente 800 nós, mas, ainda assim, continuou muito maior que o segundo. A única conclusão que se pode chegar é que não existe um tratamento genérico que possa

ser feito na base para resolver essa diferença no tamanho dos maiores componentes, nem foi possível notar um padrão ou tipo de delito que explicasse essa diferença. É possível inclusive que isso não seja uma anomalia mas que reflita a realidade. Por exemplo, como a porcentagem de delitos relacionados ao patrimônio também foi alta, é perfeitamente plausível que moradores de rua sejam pegos furtando/roubando e em outro momento sejam pegos fazendo uso de substâncias entorpecentes. Esse quadro é bastante comum e é noticiado constantemente pelos jornais e revistas brasileiros. Como em cada momento a pessoa pode ser pega na companhia de uma pessoa diferente, isso faz com que a rede cresça. Diante do cenário retratado, é possível supor que o maior componente não constitui uma organização criminosa mas é sim formado por infratores que cometem delitos avulsos aproveitando-se de situações oportunas. Ou seja, não existe um fluxo de informação bem definido na rede, pois as pessoas que a compõem não têm papéis bem definidos buscando um objetivo comum para a rede como todo. Assim, essa hipótese de que o maior componente é formado por pessoas que cometem delitos avulsos é totalmente plausível uma vez que já se esperava que as redes identificadas fossem compostas tanto por organizações com formas de operacionalização bem definidas quanto das redes sociais puras formadas por delitos avulsos.

3.3.2.2 Cúmplices Reincidentes

Para os cálculos realizados nesta seção, foram atribuídos pesos às arestas como forma de representar o número de vezes que os infratores envolvidos apareceram juntos na base como cúmplices. Em seguida, foi feita a contabilização da frequência de cada peso no grafo e o percentual de cada peso na rede de infratores independente do delito cometido. A tabela 3.4 mostra a distribuição das arestas de acordo com seu peso.

Tabela 3.4. Reincidência das Arestas na Base

Peso	No. de Arestas	Porcentagem
1	224420	99,035%
2	2094	0,924%
3	83	0,037%
+3	10	0,004%

Ao contrário do que foi observado na reincidência das pessoas, o número de cúmplices reincidentes é bem menor e somente 1,0% dos cúmplices voltam a aparecer em ocorrências juntos. Isso não quer dizer que na prática os mesmos cúmplices não ten-

dem a cometer delitos juntos, mostra somente que na base eles não reaparecem com frequência.

No entanto, ao analisar os componentes de forma isolada, esse percentual de cúmplices reincidentes começa a aumentar. A tabela 3.5 mostra, para o componente c1, a distribuição dos pesos das arestas.

Tabela 3.5. Reincidência das Arestas no Componente c1

Peso	No. de Arestas	Porcentagem
1	17925	97,12%
2	507	2,75%
3	22	0,12%
+3	2	0,01%

No componente c1, o número de cúmplices reincidentes corresponde a 2,88% do total de arestas do componente e o percentual é quase 3(três) vezes maior que o percentual apresentado no grafo completo. Cálculos semelhantes foram feitos para os demais 6 maiores componentes do grafo sendo que o componente:

- c2 apresentou 3,39% de cúmplices reincidentes;
- c3 apresentou 1,91% de cúmplices reincidentes;
- c4 apresentou 0,74% de cúmplices reincidentes;
- c5 apresentou 0,00% de cúmplices reincidentes;
- c6 apresentou 1,04% de cúmplices reincidentes;
- c7 apresentou 1,68% de cúmplices reincidentes.

Com exceção do componente c5 todos os demais componentes apresentaram um percentual de arestas reincidentes significativo apesar de baixo. O componente c5 é composto por dados de somente uma ocorrência e, portanto, corresponde a um grafo completo que não acrescentará em nada nas análises realizadas. Ele foi mantido para respeitar a ordem no tamanho dos componentes encontrados. Mesmo sendo formado por pessoas de somente uma ocorrência, esse componente é grande porque o delito está relacionado a jogos de azar e em ocorrências desse tipo é comum um número elevado de envolvidos, pois são relacionadas na ocorrência todas as pessoas que estavam no estabelecimento ilegal no momento da "batida policial".

Levando-se em consideração a natureza do delito que os cúmplices cometeram é possível constatar que, de maneira geral, dentre os reincidentes, 53% deles voltam a atuar juntos no mesmo tipo de delito pelo menos 1 vez. Essa análise foi feita levando-se em consideração o próprio delito cometido e não as categorias de tipos de delitos. Para o componente c1, a porcentagem é ainda maior e 63% das arestas reincidentes contém delitos repetidos. O componente c2 também apresentou valor semelhante: 60%. Já para o componente c3 a porcentagem foi de 40%. 50% dos cúmplices reincidentes dos componentes c4 e c6 voltaram a cometer o mesmo tipo de delito pelo menos uma vez.

Os altos valores de reincidência no mesmo tipo de delito são um indício de que os cúmplices em questão cometem o referido delito com frequência. É de extremo interesse da comunidade policial poder identificar esses infratores reincidentes para que as autoridades possam providenciar sua penalização imediatamente mesmo que o delito seja pequeno. Afinal, são esses casos de reincidência e consequente liberação que geram a sensação de impunidade na sociedade. Por exemplo, mais de 3% das arestas reincidentes são de pessoas que foram detidas por USO DE DROGAS. Enquanto que 4% dos reincidentes foram detidos por TRÁFICO DE DROGAS mas esses delitos, em geral, já são tratados com mais rigor. Ainda, 5% das arestas reincidentes são de ocorrências ligadas a JOGOS DE AZAR. Outra natureza de delito que tem participação considerável é a de FURTO que corresponde a quase 4% das arestas reincidentes. Outras participações significativas são dos delitos de LESÃO CORPORAL e ROUBO que correspondem cada uma a 3% das arestas reincidentes. Dentre uma lista de 1.190 possibilidades para diferentes tipos de delito, esses 6(seis) tipos citados anteriormente aparecem com porcentagens significativas, e se sobressaem em relação aos demais. Isso leva a crer que os infratores desses tipos de delitos têm alta probabilidade de reincidirem no delito.

3.4 Análise Topológica das Redes

Neste ponto, é importante lembrar que a rede sendo estudada é composta por 97.367 componentes e não possui um componente gigante o que faz com que a análise de suas propriedades topológicas tenham que ser feitas para todos os componentes e não só para o maior componente encontrado. No entanto, apesar do alto número de componentes encontrados, nem todos são úteis para a análise. Os componentes que representam um grafo completo não acrescentam nada para o estudo realizado, pois todos os seus nós terão a mesma importância na rede. Então esses componentes podem ser desconsiderados e, assim, restam 4807 componentes úteis e são somente eles que devem ser

analisados.

Baseando-se nas propriedades topológicas que foram definidas anteriormente, as redes complexas podem ser caracterizadas em 3 (três) tipos: aleatórias, *small-world* e *scale-free* sendo que redes aleatórias não são comuns no mundo real [Newman, 2003]. Redes aleatórias são caracterizadas por distâncias médias pequenas, pequenos valores para o coeficiente de aglomeração global e a distribuição dos graus dos nós obedece a lei de Poisson [Albert & Barabási, 2002]. As redes *small-world* mantêm uma distância média pequena mas são caracterizadas por um coeficiente de aglomeração significativamente maior que uma rede aleatória gerada sinteticamente com o mesmo número de nós e arestas utilizando o modelo proposto por Erdös & Rényi [1959]. Já as redes *scale-free* são caracterizadas por apresentarem uma distribuição de graus que obedece a uma lei de potência, mesmo que assintoticamente. Isso quer dizer que a rede é caracterizada por apresentar um pequeno número dos nós com os valores mais altos de graus enquanto que uma grande porcentagem da rede possui graus muito baixos.

3.4.1 Propriedades Small-World

Entendidas as categorias das redes e as características que cada uma apresenta, é hora de analisar os resultados. Primeiramente, para verificar se as redes podem ser caracterizadas como *small-world*, foram calculadas as distâncias médias e os coeficientes de aglomeração para cada uma das redes. E também, para cada uma, foram geradas 10(dez) redes aleatórias com o mesmo número de nós e o mesmo número de arestas que a correspondente rede real, para que essas redes pudessem ser comparadas permitindo avaliar se as redes reais podem ser classificadas como *small-world* ou se apresentam padrões aleatórios. Constatou-se que todas apresentam um coeficiente de aglomeração significativamente maior que suas respectivas redes aleatórias. Os valores das redes reais são ordens de grandeza maiores se comparados aos aleatórios e, portanto, não há dúvidas na diferença encontrada. Já as distâncias médias das redes apresentam valores relativamente pequenos se comparados ao tamanho da rede, mas são ligeiramente maiores que os valores calculados para suas respectivas redes aleatórias. Com exceção dos componentes c4 e c6, que apresentaram valores muito próximos dos aleatórios. Esses dados estão listados na tabela 3.6 sendo que para cada grupo de componentes aleatórios foram calculadas as médias dos valores encontrados para que eles pudessem ser exibidos na tabela e são mostrados na linha cujo rótulo é *Aleatório*. Lembrando que, como o número de componentes é muito alto, é impossível realizar a análise dos resultados de todos eles e, por isso, a análise é feita focada nos 7 (sete) maiores componentes. A tabela mostra os componentes conectados ordenados de forma decrescente de acordo

com o número de membros que a rede possui.

Tabela 3.6. Propriedades relacionadas a redes *small-world*

Componente	*c1*	*c2*	*c3*	*c4*	*c6*	*c7*
Diâmetro	51	20	24	12	12	15
Distância média	18,89	8,23	8,02	4,90	4,71	6,49
(Aleatório)	8,18	6,83	6,45	4,93	4,99	4,04
Coef. Aglomeração	0,58	0,56	0,56	0,68	0,54	0,59
(Aleatório)	0,0005	0,0095	0,0121	0,0160	0,0235	0,0282

Como se pode ver, os dados da tabela 3.6 mostram que as redes sendo estudadas são classificadas como *small-world*, pois apresentam os padrões topológicos desse modelo. Na prática, isso quer dizer que o fluxo de informações nas redes funciona bem o que permite uma comunicação eficiente. Isso pode ser afirmado porque os valores altos para o coeficiente de aglomeração permitem que a comunicação local, ou seja, entre os nós mais próximos seja feita de forma eficiente. Além disso, se for necessário se comunicar fora de seu agrupamento, os baixos valores para distância média indicam que essa comunicação poderá ser feita com alguns poucos mediadores.

3.4.2 Propriedades Scale-free

Para verificar se as redes apresentam um padrão *scale-free*, é necessário calcular métricas relacionadas ao número de ligações dos nós. Então foram calculadas algumas estatísticas relacionadas aos graus dos nós e o resultado é exibido na tabela 3.7.

Tabela 3.7. Propriedades relacionadas a redes *scale-free*

Componente	*c1*	*c2*	*c3*	*c4*	*c6*	*c7*
Tamanho	6378	164	139	108	87	86
Grau médio	5,79	3,62	3,80	5,03	4,47	5,60
Grau máximo	53	12	15	18	17	22
	0,83%	7,32%	10,79%	16,67%	19,54%	25,58%

Além do grau máximo encontrado na rede, é exibida logo abaixo desse valor, a porcentagem do tamanho da rede que aquele número de nós representa. Os valores apresentados para essas porcentagens tendem a mostrar que a rede é formada por um número pequeno de nós que possuem muitas ligações e por um número alto de nós que possuem poucas ligações. Essa hipótese é baseada no fato de que as porcentagens

apresentadas são relativamente altas e se referem somente às conexões de um nó da rede. Como os valores médios dos graus são muito distantes dos valores máximos, a rede não pode apresentar muitos outros nós com graus tão altos.

A seguir são mostrados os gráficos com as distribuições de graus complementares para os maiores componentes. Ou seja, mostra para um dado valor k, qual a probabilidade do valor do grau de um nó ser igual ou maior que esse valor k.

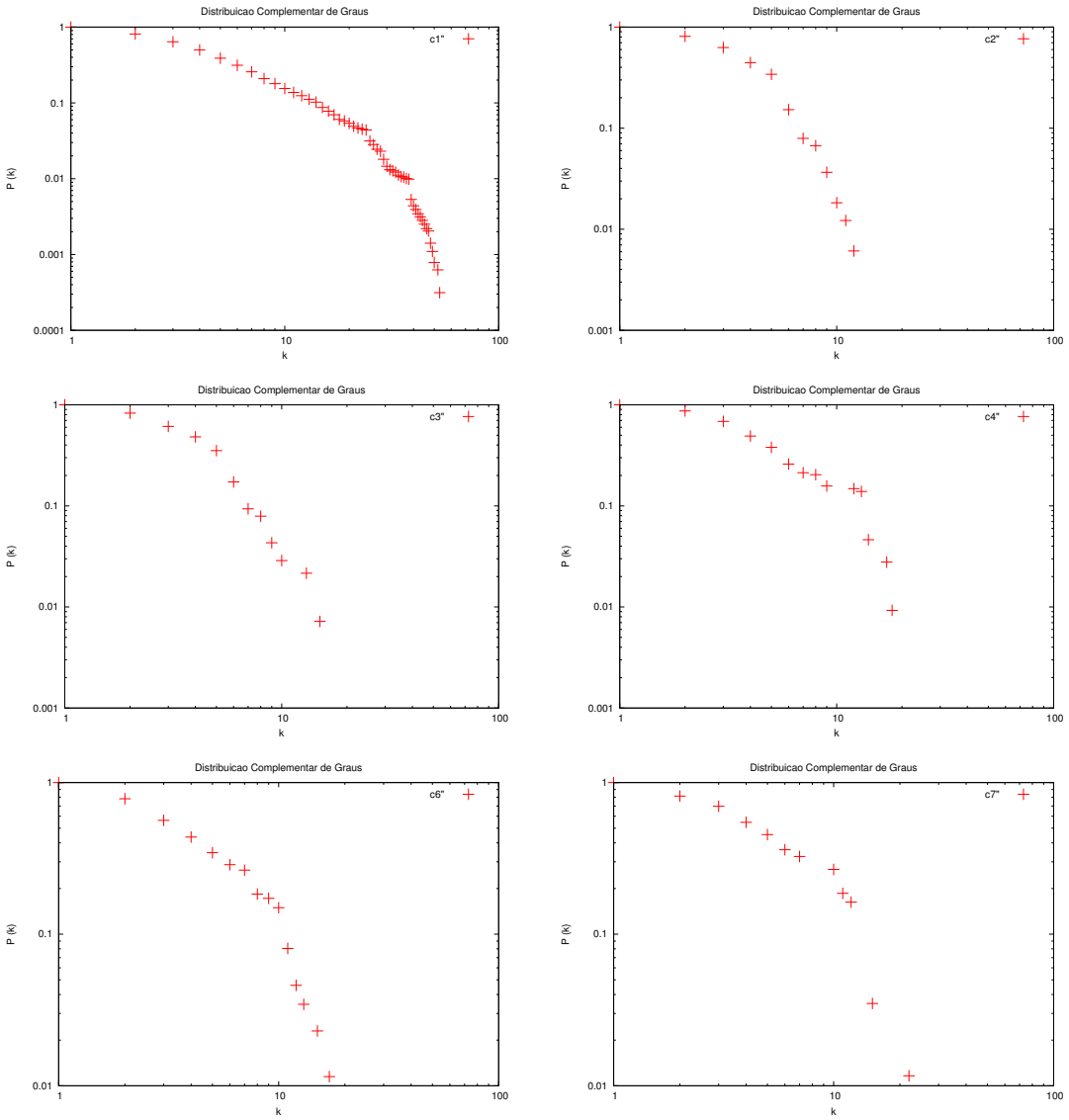

Figura 3.3. Distribuição Complementar de Graus dos Componentes

Ao avaliar a distribuição de graus da rede, mostrada na figura 3.3, não foi possível mostrar que a rede seguia uma distribuição que obedecia a uma lei de potência na forma $P(k) = \alpha k^{-\gamma}$. A lei de potência pode ser interpretada como uma linha reta em um

gráfico de escala log x log cuja equação anterior passa a ser descrita como

$$log(P(k)) = -\gamma log(k) + log(\alpha), \tag{3.1}$$

Foram feitas várias tentativas de regressão linear utilizando os dados dos componentes, para se encontrar o valor de γ que atenderia ao expoente da lei de potência. No entanto, não foi possível encontrar um valor satisfatório uma vez que para que a regressão fosse bem sucedida era necessário desprezar vários nós das extremidades dos gráficos o que seria incoerente com os estudos aqui apresentados uma vez que são os nós das extremidades dos gráficos que receberão atenção especial nos estudos mostrados na seções seguintes. Constatou-se então, que as redes não obedecem a uma distribuição de lei de potência, pois a equação 3.1 não define o comportamento dos nós das redes. Assim, é possível concluir que os componentes não podem ser categorizados como *scale-free*.

No entanto, através do gráfico 3.3, é possível notar uma característica importante da distribuição de graus dos componentes. Apesar de não seguir uma lei de potência, os gráficos mostram que a grande maioria dos valores são baixos, e uma fração muito pequena é composta de valores altos. Essa característica mostra que a conectividade das redes está concentrada em poucos indivíduos e permite afirmar que essas redes possuem integrantes que desempenham papéis-chave no funcionamento de suas organizações. Esta é a motivação para analisar o impacto causado na eficiência da rede pela remoção desses nós estratégicos.

Capítulo 4

Remoção de Nós Importantes

Trabalhos anteriores mostraram que as organizações criminosas evoluíram e estão bem melhor adaptadas aos modos modernos de colaboração, negociação e comunicação do que as tradicionais estruturas hierárquicas [Klerks & Smeets, 2001]. Por isso, tentar desestruturar essas organizações com estratégias baseadas em abordagens hierárquicas tendem a falhar. Essas abordagens hierárquicas consistem basicamente em atacar o chefe da organização, por exemplo. No entanto, estudos como o de [Klerks & Smeets, 2001] mostram que um novo chefe assumirá rapidamente e a organização continuará operando normalmente. Por outro lado, estratégias de desestruturação baseadas nas análises topológicas das redes tendem a ser mais eficientes. Sabe-se que diferentes tipos de redes observadas no mundo real apresentam um grau alto de tolerância a falhas. As falhas, nesse caso, são definidas como sendo a remoção de um nó aleatório ou ainda a falha ou queda de um nó da rede por mau funcionamento. Essas falhas raramente geram algum impacto no funcionamento do restante da rede (em redes do mundo real). Ou seja, elas apresentam robustez diante de falhas. Em contrapartida, as redes são altamente vulneráveis a ataques, que são caracterizados pela remoção proposital de nós escolhidos cuidadosamente segundo algum critério de importância. Albert et al. [2000] e Crucitti et al. [2004] mostraram que as redes *scale-free* apresentam essas características de tolerância a falhas e vulnerabilidade a ataques. Como as redes sendo estudadas apresentaram características semelhantes às *scale-free*, será testada a vulnerabilidade dessas redes diante de ataques.

4.1 Como medir o impacto da remoção do nó

Primeiramente, é necessário definir uma métrica para saber o impacto da remoção de um nó da rede. Ou seja, precisa-se de uma medida que possa quantificar o quão bem

organizada ou danificada a rede se encontra. Dorogovtsev & Mendes [2002] propõem que sejam calculadas as seguintes métricas, antes e depois do ataque, para se entender o impacto estrutural causado por esse ataque:

- Distância média da rede
- Tamanho relativo do maior componente conectado
- Tamanho médio dos componentes conectados, excluindo o gigante

Por outro lado, Latora & Marchiori [2004] propõem que seja utilizada a eficiência da rede, que quantifica a capacidade da rede de se comunicar e transmitir informações entre os nós. A eficiência já foi matematicamente definida na equação 2.4.

É importante explicar que a distância média não pode ser utilizada sozinha como forma de mensurar o impacto do ataque a rede, pois essa métrica analisada separadamente em um ataque pode levar a resultados equivocados. Isso pode ser compreendido ao se analisar a conectividade dos nós das redes sendo estudadas. Como esses componentes são formados por diversos clusters onde todos os nós estão ligados a todos os demais e com apenas alguns nós fazendo a interligação entre esses clusters, ao se isolar um nó com valor de *betweenness* alto e que estava realizando o papel de ponte entre os clusters, acaba-se quebrando o componente em componentes menores. Isso faz com que a distância média diminua e poderia dar a falsa sensação de que a remoção não foi eficiente. Por isso, a importância de se avaliar também o tamanho relativo do maior componente conectado remanescente após a remoção e do tamanho médio dos componentes conectados sem considerar o componente gigante.

A métrica de eficiência, por outro lado, consegue sintetizar as características das 3(três) métricas citadas anteriormente sendo capaz indicar que os nós de um sub-componente deixaram de se comunicar com os nós de outro sub-componente, pois apresentará valores mais baixos sempre que isso acontecer. Então, para facilitar as análises que têm que ser feitas e simplificar o processo, fazendo com que somente uma métrica tenha que ser analisada, a métrica de eficiência será utilizada para medir o impacto dos ataques.

4.2 Métricas da importância de um nó

A remoção de nós importantes da rede causará uma queda maior em sua eficiência do que a remoção de nós escolhidos aleatoriamente, pois já foi mostrado na literatura que redes *scale-free* são altamente robustas a falhas. Por isso, é fundamental saber

identificar os nós mais importantes para remoção. É necessário, então, definir qual métrica será utilizada para eleger qual é o nó mais importante para que ele possa ser o alvo do ataque. Existe quase que um consenso na literatura relacionada a redes de que a centralidade *betweenness* quantifica a importância do nó na rede por refletir a participação dele no processo global de comunicação da mesma. No entanto, existe uma nova métrica de centralidade proposta por Latora & Marchiori [2004] para indicar a importância de um nó da rede. Essa métrica, a importância I_i para um nó i qualquer pertencente ao grafo, foi definida como sendo a queda na eficiência da rede causada pela remoção do nó i:

$$I_i \equiv \Delta E = E(G) - E(G - i), \tag{4.1}$$

sendo que G - i representa a rede obtida pela remoção do nó i na rede inicial G. Assim, os nós mais importantes são aqueles que apresentarem os maiores ΔE. Apesar dos autores dessa nova métrica afirmarem que ela é mais eficaz na identificação de indivíduos-chave, eles só contrastam seu comportamento com a utilização dos graus do nó mas não com a centralidade *betweenness*. Então, por isso, foram simulados dois diferentes tipos de ataques: um baseado na centralidade *betweenness* dos nós e outro baseado na métrica de importância. Isso permitiu realizar a comparação entre as diferentes abordagens. Ainda para cada tipo de ataque, foram testadas duas estratégias de remoção diferentes. Uma utiliza os valores iniciais calculados para as métricas sendo utilizadas como critério de remoção sem que esses valores fossem atualizados a cada iteração do ataque simulando uma remoção simultânea dos nós. A outra estratégia atualiza os valores das métricas a cada nova remoção simulando um ataque progressivo.

4.3 Algoritmos para simulação dos Ataques

A seguir são detalhados os principais passos das diferentes estratégias utilizadas na remoção dos nós para que fique clara a diferença entre cada uma. São também apresentadas as análises de complexidade de cada algoritmo para que, posteriormente, possa ser feita a análise do seu desempenho.

```
CalculaListaBetweennessNos()                                        O(VElgV)
Resultado: Lista de nós ordenados em ordem decrescente de betweenness
início
    betweenness ← CalculaBetweennessNos;                            O(VElgV)
    OrdenaLista( betweenness );                                      O(VlgV)
fim

CalculaEficiencia(G)                                                O(VElgV)

CalculaListaQuedaEficiencia(G)                                      O(V²ElgV)
Resultado: Lista de nós ordenados em ordem decrescente de queda de
           eficiência
início
    Eg ⇐ CalculaEficiencia(G);                                      O(VElgV)
    para todo i do grafo G faça
        GAux ← cópia do grafo G;                                    O(VE)
        remove nó i de GAux;                                        O(E)
        EAux ← calculaEficiencia(GAux);                             O(VElgV)
        diff ← Eg - EAux;                                           O(1)
        eficiencia ← insere (i, diff) na lista de eficiencia;       O(1)
    fim
    OrdenaLista( eficiencia );                                      O(VlgV)
fim

CalculaListaGrauOrdenado()                                          O(VlgV)
Resultado: Lista de nós ordenados em ordem descrescente de seus graus
início
    para todo i do grafo G faça
        grau ← getGrau(i);                                          O(1)
        grausList ← insere (i, grau) na lista de nós;               O(1)
    fim
    OrdenaLista( grausList );                                       O(VlgV)
fim
```

Algoritmo 1: Funções auxiliares nos algoritmos de ataque

```
SimulaAtaqueGrauInicial()                                   O(V²ElgV)
Resultado: Lista da eficiência da rede após cada remoção
início
    lg ← CalculaListaGrauOrdenado(G);                       O(VlgV)
    enquanto existirem nós conectados em G faça
        i ← pega primeiro nó da lista lg;                   O(1)
        remove nó i do grafo G;                             O(E)
        remove nó i da lista lg;                            O(1)
        armazena calculaEficiencia(G);                      O(VElgV)
    fim
fim
```

Algoritmo 2: Ataque Baseado no Grau Inicial do Nó

```
SimulaAtaqueGrauAtualizado()                                O(V²ElgV)
Resultado: Lista da eficiência da rede após cada remoção
início
    enquanto existirem nós conectados em G faça
        lg ← CalculaListaGrauOrdenado(G);                   O(VlgV)
        i ← pega primeiro nó da lista lg;                   O(1)
        remove nó i do grafo G;                             O(E)
        armazena calculaEficiencia(G);                      O(VElgV)
    fim
fim
```

Algoritmo 3: Ataque Baseado no Grau Atualizado do Nó

```
SimulaAtaqueBetweennessInicial()                            O(V²ElgV)
Resultado: Lista da eficiência da rede após cada remoção
início
    lBi ← CalculaListaBetweennessNos(G);                    O(VElgV)
    enquanto existirem nós conectados em G faça
        i ← pega primeiro nó da lista lBi;                  O(1)
        remove nó i do grafo G;                             O(E)
        remove nó i da lista lBi;                           O(1)
        armazena calculaEficiencia(G);                      O(VElgV)
    fim
fim
```

Algoritmo 4: Ataque Baseado no *Betweenness* Inicial do Nó

A abordagem do algoritmo 4 realiza a remoção baseada nos valores de *betweenness* calculados para os nós enquanto a rede estava em seu estágio inicial. Esses valores são ordenados em ordem decrescente para que o nó com maior *betweenness* seja o primeiro da lista, pois é ele que deve ser removido primeiro. A cada iteração, após o isolamento

do nó que deve ser removido, é recalculada a eficiência da rede remanescente e esse valor é salvo para que posteriormente possa ser gerado um gráfico do comportamento da rede diante dos ataques.

```
SimulaAtaqueBetweennessAtualizado()                          O(V²ElgV)
Resultado: Lista da eficiência da rede após cada remoção
início
    enquanto existirem nós conectados em G faça
        lBa ← CalculaListaBetweennessNos(G);                 O(VElgV)
        i ← pega primeiro nó da lista lBa;                   O(1)
        remove nó i do grafo G;                              O(E)
        armazena calculaEficiencia(G);                       O(VElgV)
    fim
fim
```

Algoritmo 5: Ataque Baseado no *Betweenness* Atualizado do Nó

O algoritmo 5 é semelhante ao 4 mas a diferença é que a remoção é feita baseada no valor atualizado de *betweenness* de cada nó. Ou seja, antes de cada remoção, os valores de *betweenness* são recalculados para refletir o estado corrente da rede e os nós são ordenados novamente de acordo com os valores obtidos.

```
SimulaAtaqueEficienciaInicial()                              O(V²ElgV)
Resultado: Lista da eficiência da rede após cada remoção
início
    lQE ← CalculaListaQuedaEficiencia(G);                    O(V²ElgV)
    enquanto existirem nós conectados em G faça
        i ← pega primeiro nó da lista lQE;                   O(1)
        remove nó i do grafo G;                              O(E)
        remove nó i da lista lQE;                            O(1)
        armazena calculaEficiencia(G);                       O(VElgV)
    fim
fim
```

Algoritmo 6: Ataque Baseado na Queda na Eficiência do Grafo Inicial

A diferença do algoritmo 6 em relação ao 4 é somente no momento de recuperação da lista ordenada de nós a serem removidos. No algoritmo 6, a importância do nó é dada pela queda na eficiência da rede causada pela remoção do nó enquanto que no outro, ela é dada pelo valor de *betweenness* do nó.

O algoritmo 7 também é semelhante ao utilizado anteriormente na simulação utilizando uma métrica atualizada só que desta vez, a métrica é a queda na eficiência da rede causada pela remoção do nó.

A tabela 4.1 mostra o tempo de execução gasto para simular um ataque e calcular o impacto na rede causado pelo isolamento dos nós importantes. Nela são exibidos

```
SimulaAtaqueEficienciaAtualizado()                                  O(V^3 ElgV)
Resultado: Lista da eficiência da rede após cada remoção
início
    enquanto existirem nós conectados em G faça
        lQEa ← CalculaListaQuedaEficiencia(G);                      O(V^2 ElgV)
        i ← pega primeiro nó da lista lQEa;                         O(1)
        remove nó i do grafo G;                                     O(E)
        armazena calculaEficiencia(G);                              O(VElgV)
    fim
fim
```

Algoritmo 7: Ataque Baseado na Queda na Eficiência do Grafo Atualizado

os custos temporais das diferentes estratégias utilizadas na escolha do nó mais importante a ser removido sendo que B_I corresponde a abordagem utilizando o valor inicial de *betweenness*, B_A corresponde à utilização de valores atualizados da métrica *betweenness*, ΔE_I representa a utilização dos valores calculados inicialmente para a queda na eficiência enquanto a rede ainda estava completa e, finalmente, ΔE_A corresponde aos valores calculados para a queda da eficiência da rede atualizada levando em consideração os nós removidos.

Tabela 4.1. Custo para simulação dos ataques

Remoção baseada em:	B_I	B_A	ΔE_I	ΔE_A
c1	61 horas	64 horas	-	-
c2	3 seg	7 seg	43 seg	5 min
c3	2 seg	5 seg	25 seg	3 min
c4	1 seg	2 seg	13 seg	1 min
c6	1 seg	2 seg	7 seg	32 seg
c7	1 seg	2 seg	7seg	32 seg

É possível notar claramente que o tempo de execução das estratégias baseadas em *betweenness* é muito menor do que aquelas baseadas na queda de eficiência. O custo computacional para calcular a queda na eficiência é tão alto que tornou inviável a realização desse cálculo para o maior componente. O programa que realizava o cálculo da queda na eficiência, baseada no grafo inicial, demorou mais de 3 semanas e, por isso, foi abortado, pois na prática este tempo é inconcebível.

Então, levando em consideração os tempos de execução, o algoritmo de remoção de nós baseado na queda de eficiência tende a ser desconsiderado por ser muito demorado. Resta avaliar o impacto da retirada dos nós no que diz respeito à desestruturação da rede.

Diante dessa diferença tão grande nos tempos de execução, é necessário avaliar melhor os algoritmos para verificar o porquê da diferença de execução sendo que com exceção do algoritmo baseado na eficiência atualizada, todos os demais têm custo de execução O(V^2ElgV) sendo que V é o número de vértices da rede.

Comparando-se os algoritmos baseados no *betweenness* inicial e na queda na eficiência da rede inicial, é possível notar que a única diferença entre os dois é que enquanto um chama a função CalculaListaQuedaEficiencia que é $O(V^2ElgV)$, o outro chama a função CalculaListaBetweennessNos(G) que tem custo $O(VElgV)$ que é claramente menos custosa. No entanto, os demais passos dos algoritmos também são $O(V^2ElgV)$, então por que o tempo de execução do algoritmo 6 não foi proporcional ao tempo do algoritmo 4 sendo que esses algoritmos são muito similares aos demais e seu maior custo é dado pela execução da função calculaEficiencia dentro de um loop de tamanho V exatamente como nas demais. A sutil diferença na função CalculaListaQuedaEficiencia é o fato dela copiar sempre o grafo inicial. Além do próprio custo de copiar todo o grafo, o fato do cálculo ser feito sempre no grafo completo faz com que o custo efetivo da execução da função calculaEficiencia seja sempre alto e da ordem de O(V^3) enquanto que nos demais algoritmos, o grafo está sofrendo modificações além da remoção de arestas são removidos os nós o que faz com que o custo de computação dos caminhos mínimos seja menor. Como a remoção dos nós é feita baseada na conectividade desse nó, a rede se desconecta muito facilmente fazendo com que o custo de execução da função calculaEficiencia seja linear. Esse comportamento ficará mais claro com a análise dos gráficos da seção seguinte que mostram o impacto da remoção dos nós.

4.4 Tolerância das Redes aos Ataques

Após explicado o funcionamento dos algoritmos que fizeram a simulação de diferentes estratégias de ataque às redes, serão analisados, nesta seção, os dados que refletem o impacto causado nelas pelas remoções.

Em outras análises da literatura, é possível ver o grau de conectividade de um indivíduo sendo utilizado como um indício de sua importância na rede. Isso ocorre porque um nó central, com muitas ligações, tem alta probabilidade de exercer um papel importante na sua rede social. No entanto, vários outros trabalhos já mostraram que essa métrica é menos eficiente que a métrica *betweenness* e, por isso, ela não será contemplada aqui.

A seguir são exibidos alguns gráficos mostrando o impacto na eficiência da rede causada pela remoção de nós utilizando as diferentes estratégias citadas. Na figura 4.1,

alguns gráficos apresentam uma queda claramente mais acentuada e, por isso, mostram ser estratégias mais eficazes.

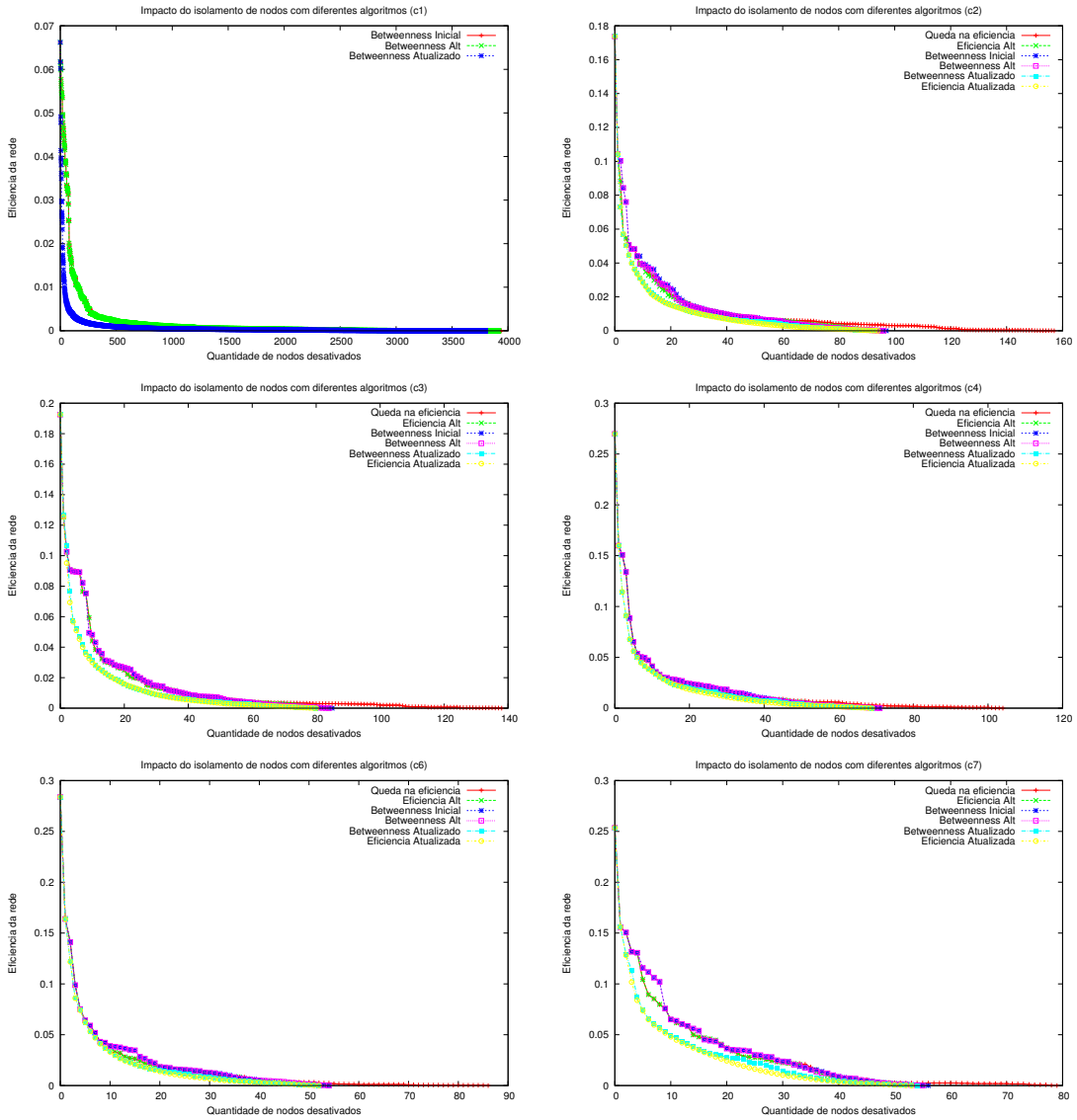

Figura 4.1. Impacto das diferentes estratégias dos ataques

Os gráficos mostram que todas as estratégias conseguem diminuir drasticamente a eficiência da rede com poucas remoções. As curvas mostram uma queda mais acentuada no início, uma vez que a remoção de nós foi feita escolhendo os nós que desempenhavam um papel importante na comunicação da rede. Ou seja, ocorre uma rápida desestruturação das redes ao terem alguns de seus principais elos removidos.

É possível notar também que as estratégias utilizando as métricas atualizadas das redes foram consideravelmente mais eficazes e conseguiram maior queda na eficiência

com menor remoção de nós, como já era esperado. Nota-se também que a diferença no êxito dos métodos é maior nos componentes maiores e as curvas vão se aproximando à medida que o tamanho dos componentes diminui.

Ao contrário do que foi proposto em Latora & Marchiori [2004], não foi possível confirmar que a métrica baseada na queda de eficiência é a métrica mais indicada na identificação dos nós mais importantes das redes aqui estudadas. Na prática, a métrica realmente identificou os nós importantes, pois a remoção deles foi eficaz na desestruturação das redes. E, por definição, essa métrica sempre levará à solução ótima, pois o êxito da remoção de um nó é medido pela queda na métrica eficiência (eq. 2.4). Sendo assim, escolher o nó mais importante pela maior perda de eficiência será sempre a melhor opção. No entanto, o impacto causado pelo ataque baseado na centralidade *betweenness* foi bem próximo dos obtidos com a queda na eficiência. Os resultados mostram que é possível dividir as estratégias em duas classes de equivalência sendo uma baseada nas métricas **iniciais** e outra baseada nas métricas **atualizadas** da queda na eficiência e da centralidade *betweenness*. Pode-se notar que na mesma classe de equivalência, as estratégias utilizando queda na eficiência e *betweenness* tiveram desempenho muito similares e suas curvas são praticamente sobrepostas.

Assim, fica claro que a estratégia de remoção baseada no *betweenness* atualizado é tão eficiente quanto a estratégia utilizando a queda da eficiência do nó mas que o custo computacional do primeiro é bem menor do que o segundo. Esse resultado então confirma que a centralidade *betweenness* é uma métrica sólida para solucionar o alvo dos ataques.

É importante entender porque a remoção baseada na métrica de *betweenness* do nó é eficiente na desestruturação de uma organização. Ela quantifica bem a importância de um indivíduo em sua rede social, pois indica o controle que esse indivíduo exerce sobre a rede em termos de como ele é capaz de afetar sua comunicação. Ou seja, uma posição central estratégica (com *betweenness* alto) permite que ele interrompa o fluxo de informação ao longo da rede. Por sua vez, a *eficiência*, no estudo de redes complexas, foi criada especificamente para desempenhar o papel de medir a habilidade da rede em transmitir informação e sua resposta a perturbações na mesma. Não é por menos que essas métricas se comportam de forma tão similar. Seus conceitos estão fortemente ligados às propriedades topológicas da rede refletindo a forma como os nós estão conectados.

Outra questão importante é com relação a diferença no comportamento da rede em função dos ataques baseados nas métricas calculadas no grafo inicial e métricas atualizadas após cada remoção. Essa diferença é causada não só em decorrência da quebra do componente em componentes menores, fazendo com que nós que tinham

métrica de *betweenness* altos passassem a ter *betweenness* baixos mas também em função dessa métrica não refletir conjuntos de nós redundantes. O valor de *betweenness* de nó leva em consideração a equivalência de nós somente de um para um, ou seja, que participam de algum caminho mínimo sem levar em consideração caminhos que sejam ligeiramente maior do que o mínimo. A figura 4.2 exemplifica essa situação ajudando a esclarecer o cenário.

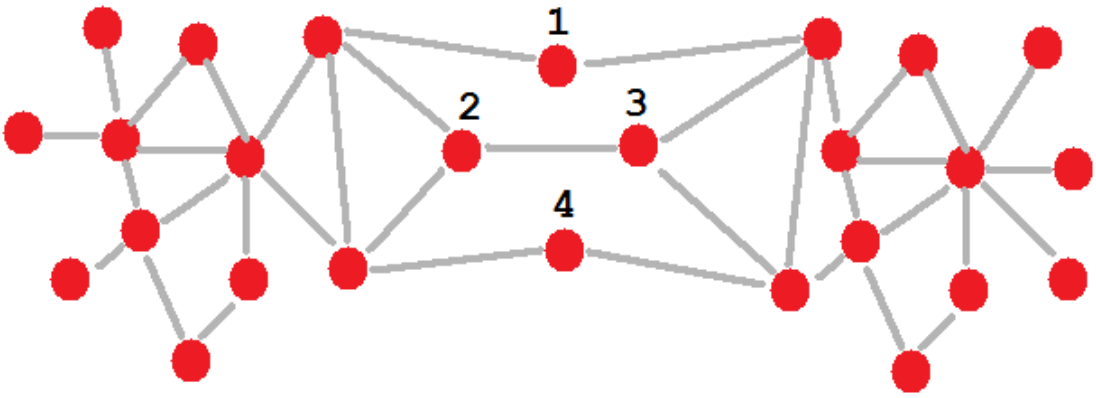

Figura 4.2. Exemplo de rede com nós redundantes

Nesta figura, os nós 1 e 4 são equivalentes e terão o mesmo valor de *betweenness* pois essa métrica leva em consideração todos os caminhos mínimos existentes e não só o primeiro encontrado. Sendo assim ambos terão o mesmo valor de *betweenness*. Por outro lado, os nós 2 e 3, que se analisados em conjunto são equivalentes aos nós 1 e 4, têm valor de *betweenness* muito baixos pois não fazem parte de muitos caminhos da rede. É fácil notar que em um ataque, após a remoção dos nós 1 e 4, o nó que deve ser removido em seguida para desestruturar a rede mais rapidamente deve ser o nó 2 ou 3. No entanto, se as métricas não forem atualizadas, esses nós demorarão mais para serem removidos e fará com que a rede continue firme por mais tempo. Assim, ao atualizar as métricas resolve-se o fato de não se levar em consideração conjuntos de nós equivalentes. Outra maneira de enxergar o mesmo problema é fazendo uma analogia com uma rede de computadores. Considerando um caminho mínimo dessa rede passando pelo roteador X e que tenha como estrutura redundante a X um outro caminho formado pelos roteadores Y e Z (ou seja, um caminho mais longo), a rede não terá Y e Z identificados como roteadores importantes de acordo com a métrica *betweenness*, já que não fazem parte do caminho mínimo. No entanto, se um dos nós Y ou Z forem removidos, o impacto é equivalente a remover X, pois a rede perde a redundância naquele ponto ficando mais vulnerável. Isso mostra a necessidade de

utilizar a métrica atualizada para indicação da importância do nó para evitar desvios na queda de eficiência como pode ser visto no gráfico do componente c3 da figura 4.1.

4.5 Considerações Finais

Após analisar o comportamento da rede em virtude da remoção de nós baseada em diferentes métricas é possível concluir que a melhor forma para quantificar a importância de uma pessoa dentro da rede é medindo o valor de *betweenness* que ela tem dentro da organização. Foi mostrado na seção 4.3 que apesar da métrica baseada na queda da eficiência ser intrinsecamente a solução ótima para identificar a importância do nó, essa métrica é computacionalmente inviável pois não escala para redes grandes. Além do mais, o comportamento da métrica *betweenness* foi muito similar.

Sendo assim, definida a métrica para quantificar a importância do nó, o algoritmo para identificar a ordem das pessoas-chave da rede está detalhado a seguir:

```
OrdenaPessoasPorImportancia(){
    Enquanto existirem nós conectados em G
        CalculaListaBetweennessNos(G)
        OrdenaListaBetweennessNos(G)
        i <-- pega o primeiro da lista de betweenness
        remove nó i do grafo G
        insere i na lista de pessoas importantes

}
```

Algoritmo 8: Identificação da Ordem das Pessoas-Chave

O algoritmo retorna a lista de pessoas ordenadas em ordem decrescente sendo que o mais importante é o primeiro da lista. A complexidade do algoritmo pode ser calculada da seguinte forma:

- CalculaListaBetweennessNos(G) tem custo O(V^2 + VE)
- OrdenaListaBetweennessNos(G) tem custo O(V lgV)
- os demais comandos têm custo O(1)
- OrdenaPessoasPorImportancia() tem custo O(V^2 + VE) + O(V lgV) = O(V^2 + VE)

Pode-se notar que o custo para calcular a lista de indivíduos-chave é inferior ao custo encontrado na simulação. A complexidade passou de O(V^4) para O(V^2 + VE) que é uma diferença considerável. Isso ocorre porque nesse ponto não é necessário calcular a eficiência da rede para avaliar o impacto da remoção do nó. Já se sabe a métrica que mede a importância do nó, então o custo maior consiste em calcular essa métrica.

Capítulo 5

Conclusão

Neste trabalho, caracterizou-se a base de registros de ocorrências da Polícia Militar de Minas Gerais visando extrair o máximo de informações possíveis dessa base, extrair as redes de infratores contidas nesses dados e obter informações sobre o funcionamento dessa redes e sobre a forma como os indivíduos se relacionam.

O tratamento de deduplicação da base, em si, já constitui uma grande contribuição à sociedade pois permite identificar todas as vezes que uma pessoa apareceu em um registro policial. A comunidade policial considera essa informação essencial para decidir as providências que irá tomar em relação a infratores de delitos menores. Através da eliminação das réplicas e identificação da pessoa de forma única é possível levantar como os indivíduos de um determinada ocorrência estão relacionados a indivíduos de outras ocorrências. A partir desses relacionamentos, surgem as redes de criminosos que podem ser usadas para auxiliar no trabalho investigativo da polícia.

Foi encontrado um número elevado de redes de criminosos sem que nenhuma delas pudesse representar o todo e, por isso, as análises foram realizadas em cima de cada uma das redes de forma separada. Uma característica recorrente encontrada na maioria das redes foi que elas apresentaram um percentual significativo de ocorrências relacionadas ao USO/TRÁFICO DE DROGAS mostrando que as drogas, em geral, levam a outros tipos de delitos como roubo, furto, homicídio, agressões etc. Esses dados comprovam o que já era esperado pois usuários de drogas acabam cometendo outros delitos para sustentarem seus vícios e os traficantes, para cobrarem suas dívidas.

A análise das propriedades topológicas das redes mostrou que elas pertencem à categoria de redes *small-world*. O fato das redes terem sido classificadas como *small-world* indica que o fluxo de informações nas redes funciona bem e que a comunicação é feita de forma eficiente. Apesar de não terem sido classificadas como *scale-free*, mostrou-se que nas redes de criminosos, assim como nas *scale-free*, o controle da rede é feito por

poucos membros o que permite identificar indivíduos importantes e que desempenham papéis-chave no funcionamento das organizações criminosas. Esses indivíduos devem ser os alvos das ações policiais, pois essas redes são altamente sensíveis a ataques e tenderão a ser desestruturadas com a remoção dos nós-chave.

Foram comparados diferentes algoritmos utilizados na identificação de nós importantes. Os experimentos realizados e a análise de complexidade desses algoritmos mostraram que a utilização da métrica *betweenness* é a mais apropriada pois além de obter um resultado eficaz, muito próximo do ótimo, é computacionalmente mais eficiente do que a solução ótima.

5.1 Trabalhos Futuros

Novas oportunidades de pesquisa poderão ser incorporadas ao trabalho realizado visando expandir e melhorar os resultados obtidos. Dentre as opções destacamos:

1. **Análise temporal**: avaliar como as redes evoluem ao longo do tempo, se existe algum padrão nessa evolução. Tentar identificar características nas redes que acabam se conectando ao longo do tempo.

2. **Identificação de elos faltantes**: tentar prever onde estão faltando relacionamentos entre as pessoas para obter uma rede mais completa e próxima da real. Aqui, poderão ser utilizadas técnicas de *link prediction* e *random walk*. Esse item tem uma relação estreita com a análise temporal citada anteriormente.

3. **Identificação de policiais corruptos**: acrescentar os policiais nas análises das redes levando-se em consideração que os relacionamentos entre policiais e infratores são circunstanciais e esporádicos. Assim, uma alteração nesse padrão de comportamento pode indicar alguma relação ilegal dos policiais com os infratores e pode levar à identificação de policiais corruptos.

Julgamos que a ferramenta implementada pode auxiliar bastante o trabalho da polícia. Com isso, pretendemos incluir uma interface para administração das redes e identificação das pessoas-chave. Além disso, incluir uma funcionalidade que permite pesquisar as pessoas pelos seus dados para identificar a quais outras pessoas elas estão relacionadas pode auxiliar na identificação de indivíduos suspeitos. Acreditamos que existe um grande potencial de aplicação dos resultados desta dissertação por parte dos órgãos de segurança pública.

Referências Bibliográficas

Adamic, L. A. (1999). The small world web. In *ECDL '99: Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries*, pp. 443--452, London, UK. Springer-Verlag.

Albert, R. & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47--97.

Albert, R.; Jeong, H. & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406:378–382.

Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M. & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175--308.

Carley, K. M.; Reminga, J. & Kamneva, N. (2003). Destabilizing terrorist networks. In *Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) Conference*, Pittsburgh, PA, USA.

Chen, H.; Zeng, D.; Atabakhsh, H.; Wyzga, W. & Schroeder, J. (2003). Coplink: managing law enforcement data and knowledge. *Communications of the ACM*, 46(1):28--34.

Cormen, T.; Leiserson, C. & Rivest, R. (2009). *Introduction to Algorithms*. The MIT Press, Cambridge, MA, 3ª edição.

Crucitti, P.; Latora, V.; Marchiori, M. & Rapisarda, A. (2003). Efficiency of scale-free networks: error and attack tolerance. *Physica A*, 320:622–642(21).

Crucitti, P.; Latora, V.; Marchiori, M. & Rapisarda, A. (2004). Error and attack tolerance of complex networks. *Physica A Statistical Mechanics and its Applications*, 340:388–394.

Dorogovtsev, S. N. & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, 51(4):1079--1187.

Erdös, P. & Rényi, A. (1959). On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290--297.

Fellman, P. V. & Wright, R. (2004). Modeling terrorist networks - complex systems at mid-range.

Goh, K. I.; Oh, E.; Jeong, H.; Kahng, B. & Kim, D. (2002). Classification of scale-free networks. *Proc Natl Acad Sci U S A*, 99(20):12583--12588.

Klerks, P. & Smeets, E. (2001). The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? recent developments in the netherlands. *Connections*, 24(3):53--65.

Krebs, V. E. (2002). Uncloacking terrorist networks. *First Monday*, 7(4).

Latora, V. & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, 87(19).

Latora, V. & Marchiori, M. (2004). How the science of complex networks can help developing strategies against terrorism. *Chaos Solitons Fractals*, 20(1):69--75.

Marteleto, R. M. (2001). Análise de redes sociais - aplicação nos estudos de transferência da informação. *Ciência da Informação*, 30(1):71--81.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167--256.

Ozgul, F.; Bondy, J. & Aksoy, H. (2007). Mining for offender group detection and story of a police operation. In *AusDM '07: Proceedings of the sixth Australasian conference on Data mining and analytics*, volume 70, pp. 189--193, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Santos, W.; Teixeira, T.; Machado, C.; Jr., W. M.; Silva, A. S. D.; Ferreira, R. & Guedes, D. (2007). A scalable parallel deduplication algorithm. In *Proceedings of the 19th International Symposium on Computer Architecture and High Performance Computing*, pp. 79--86.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440--442.

Xu, J. & Chen, H. (2008). The topology of dark networks. *Communications of the ACM*, 51(10):58--65.

Apêndice A

Categorias das Naturezas das Ocorrências

- *INFRAÇÕES CONTRA A PESSOA*: esse grupo contempla todas as naturezas relacionadas a pessoas e inclui, por exemplo, homicídio, lesão corporal, abandono de incapaz, recusa/dificulta/não assiste idoso, dentre outras;

- *INFRAÇÕES CONTRA O PATRIMÔNIO*: como o próprio nome já diz, contempla as infrações relacionadas ao patrimônio, por exemplo, roubo, furto, extorsão, estelionato, dano, dentre outros;

- *INFRAÇÕES CONTRA OS COSTUMES E FAMÍLIA*: contempla naturezas como estupro, corrupção de menores, entrega de filho menor a pessoa inidônea, jogo de azar, perturbação da tranquilidade etc;

- *INFRAÇÕES CONTRA A INCOLUMIDADE PÚBLICA*: contempla naturezas como incêndio, inundação, desabamento, uso de gás tóxico, corrupção/poluição de água potável, dentre outras;

- *INFRAÇÕES CONTRA ORGANIZAÇÃO DO TRABALHO*: inclui naturezas como atentar contra a liberdade de trabalho, paralisar trabalho com violência ou perturbação, frustar direito assegurado na lei trabalhista, dentre outras;

- *INFRAÇÕES CONTRA A ADMINISTRAÇÃO E FÉ PÚBLICA*: contempla as ocorrências de moeda falsa, falsificação de documento, emprego irregular de verbas/rendas públicas, abandono de função, desacato, dentre outras;

- *INFRAÇÕES CONTRA SENTIMENTO RELIGIOSO/MORTOS*: contempla impedir/perturbar cerimônia funerária, ultrajar/impedir/perturbar culto religioso, extrair orgão/parte do corpo humano para transplante, dentre outras;
- *INFRAÇÕES REFERENTES A SUBSTÂNCIAS ENTORPECENTES*: inclui tráfico de substância entorpecente, associação para o tráfico, adquirir/guardar/trazer droga para uso próprio, dentre outras;
- *INFRAÇÕES REFERENTES A ELEIÇÕES*: inclui inscrever-se fraudulentamente como eleitor, perturbar/impedir de qualquer forma o alistamento, violar ou tentar violar o sigilo do voto, difamar alguém na propaganda eleitoral, boca de urna, dentre outras;
- *INFRAÇÕES CONTIDAS EM LEIS EXTRAVAGANTES*: contempla naturezas relacionadas a lei de proteção ao consumidor, crime resultante de preconceito de raça/cor, abuso de autoridade, crimes de sonegação fiscal, crimes contra a segurança nacional, dentre outros;
- *INFRAÇÕES COMUNS AO MEIO AMBIENTE E DE ATIVIDADES POTENCIALMENTE POLUIDORAS*: contempla destruir material proibido para caça/apanha/pesca, realizar atividade sem licenciamento ambiental, pichar/grafitar/poluir edificação/monumento urbano, dentre outras;
- *INFRAÇÕES AMBIENTAIS REFERENTES A FAUNA E PESCA*: inclui modificar/destruir ninho/abrigo/criadouro, caçar/apanhar/utilizar animal silvestre, promover rinhas de animais, pescar com uso de tarrafas, dentre outras;
- *INFRAÇÕES AMBIENTAIS RELATIVAS A FLORA*: inclui explorar vegetação em área preservada por lei, provocar queimada sem autorização, comercializar motoserra sem autorização, dentre outros;
- *INFRAÇÕES REFERENTES AO TRÂNSITO*: inclui veículo abandonado, acidentes de trânsito, adulterar sinal/identificador de veículo automotor, dentre outros;
- *NATUREZAS REFERENTES A EXPLOSÃO E INCÊNDIO*: contempla naturezas específicas como explosão de caminhão tanque, explosão em fábrica de explosivos, incêndio em caldeira, incêndio em cilindros, dentre outras;
- *NATUREZAS REFERENTES A PREVENÇÃO*: inclui prevenção contra incêndio e pânico em bailes, eventos esportivos, prevenção de perigo de afogamento, contaminação química, desabamento, inundação, dentre outros;

- *NATUREZAS REFERENTES A DEFESA CIVIL*: contempla fatos relacionados a defesa civil como abastecimento d'água, acidente com produtos perigosos, catástrofes naturais como enchentes, vendavais, dentre outras;
- *NATUREZAS REFERENTES A BUSCA E SALVAMENTO*: inclui as ocorrências relacionadas a busca e salvamento por diferentes motivos como convulsão, afogamento, acidente vascular cerebral, em teleféricos, acidentes, dentre outras;
- *AÇÕES DE DEFESA SOCIAL*: contempla ocorrências relacionadas a atritos verbais, averiguação de disparo de alarme, encontro de feto, remoção de cadáver, dentre outras;
- *OPERAÇÕES DE DEFESA SOCIAL*: contempla as operações realizadas pelos órgão de defesa social visando fiscalizar e manter a ordem pública como operações de fronteira, operações policiais de trânsito, fiscalização de táxi e transporte escolar, dentre outros;
- *COMUNICAÇÕES E SOLICITAÇÕES DE DEFESA SOCIAL*: contempla naturezas relacionadas a extravio de documentos, objetos pessoais e pessoas extraviadas ou desaparecidas.