Centro Federal de Educação Tecnológica de Minas Gerais
Pós-Graduação em Modelagem Matemática e Computacional

COMPLEXIDADE DAS RELAÇÕES EM REDES DE CRIMINALIDADE

Alex Sander de Oliveira Toledo

Orientador: Prof. Dr. Antônio Paulo Baeta Scarpelli
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadores: Prof. Dr. Allbens Atman Picardi Faria e Profa. Dra. Laura C. Carpi
Centro Federal de Educação Tecnológica de Minas Gerais

Belo Horizonte
Maio de 2024

Alex Sander de Oliveira Toledo

COMPLEXIDADE DAS RELAÇÕES EM REDES DE CRIMINALIDADE

Tese apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional.

Linha de pesquisa: Métodos Matemáticos Aplicados.

Orientador: Prof. Dr. Antônio Paulo Baeta Scarpelli
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadores: Prof. Dr. Allbens Atman Picardi Faria e Profa. Dra. Laura C. Carpi
Centro Federal de Educação Tecnológica de Minas Gerais

Centro Federal de Educação Tecnológica de Minas Gerais
Pós-Graduação em Modelagem Matemática e Computacional
Belo Horizonte
Maio de 2024

Toledo, Alex Sander de Oliveira
T649c Complexidade das relações em redes de criminalidade / Alex Sander de Oliveira Toledo. – 2024.
101 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Orientador: Antônio Paulo Baeta Scarpelli.
Coorientadores: Allbens Atman Picardi Faria e Laura C. Carpi.
Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Complexidade (Filosofia) – Teses. 2. Criminalidade – Análise de redes –Teses. 3. Análise de redes (Ciências sociais) – Modelos – Teses. I. Scarpelli, Antônio Paulo Baeta. II. Faria, Allbens Atman Picardi. III. Carpi, Laura Corina. IV. Centro Federal de Educação Tecnológica de Minas Gerais. V. Título.

CDD 005.754

Elaboração da ficha catalográfica pela bibliotecária Jane Marangon Duarte, CRB 6ª 1592 / Cefet/MG

SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

"COMPLEXIDADE DAS RELAÇÕES EM REDES DE CRIMINALIDADE".

Tese de Doutorado apresentada por **Alex Sander de Oliveira Toledo,** em 16 de maio de 2024, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Antônio Paulo Baêta Scarpelli (Orientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Alibens Atman Picardi Faria (Coorientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Profª. Drª. Laura Corina Carpi (Coorientadora)
Universidade Federal de Minas Gerais

Prof. Dr. Tiago Alves Schieber de Jesus
Universidade Federal de Minas Gerais

Prof. Dr. Hélio Hiroshi Hamada
Instituto Brasileiro de Segurança Pública

Profª. Drª. Izabela Marques de Oliveira
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Arthur Rodrigo Bosco de Magalhães
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Thiago Gomes de Mattos
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida à impressão,

Prof. Dr. Thiago de Souza Rodrigues
Coordenador do Programa de Pós-Graduação Stricto Sensu em Modelagem Matemática e Computacional

A minha esposa Patrícia, minha alma gêmea, em minha lucides e minha loucura, sempre ao meu lado.

Agradecimentos

Desejo exprimir os meus agradecimentos a todos aqueles que, de alguma forma, permitiram que esta tese se concretizasse.

Agradeço ao Prof. Dr. Antônio Paulo Baêta Scarpeli pela orientação desta tese, pela inspiração, pelo apoio e pela sua disposição e capacidade de fomentar um diálogo de saberes com um espírito libertário, nada fácil e ainda menos comum em meio a uma academia constituída, em termos formais, por relações hierárquicas. Aprendi muito nestes anos e agradeço muito por esse aprendizado.

Ao amigo e querido mestre Prof. Dr. Allbens Atman Picardi Faria, por ter acreditado em mim e nas minhas capacidades. Agradeço ainda o trato simples, correto e científico, com que sempre abordou as nossas reuniões de trabalho, sem nunca ter permitido que o desalento se instalasse, mesmo quando as coisas não corriam bem.

À Dra. Laura C. Carpi, quero de uma forma imensa, expressar o meu muito obrigado... por tudo. Pela forma amiga e generosa com que sempre me incentivou e ajudou, e pelo estímulo sentido após cada conversa, que me faziam "carregar baterias" e seguir em frente.

A la Dra. Cristina Masoller, por haberme acogido tan lejos de casa en un entorno nuevo y desafiante. Tu dominio de la enseñanza y tu habilidad para explicar conceptos complejos de manera clara y cautivadora fueron impresionantes. Cada día fue un viaje de descubrimiento y aprendizaje, y nunca me sentí intimidado al hacer preguntas o pedir ayuda, ya que siempre estuvistes dispuesta a tenderme la mano y orientarme.

"La libertad, Sancho, es uno de los más preciosos dones que a los hombres dieron los cielos; con ella no pueden igualarse los tesoros que encierran la tierra y el mar: por la libertad, así como por la honra, se puede y debe aventurar la vida."
(Miguel de Cervantes, El Quijote)

Resumo

A compreensão de que a sociedade funciona como redes interligadas levou a mudanças significativas nas técnicas e abordagens científicas em diversas disciplinas. Em particular, teve um impacto profundo na criminologia, oferecendo benefícios substanciais em termos de alocação de recursos e até salvando vidas. O surgimento de conhecimentos teóricos avançados de Ciência de Redes e Aprendizado de Máquina abriu novos caminhos para o desenvolvimento de modelos de análise de redes, considerando a relevância topológica e dinâmica dos indivíduos dentro das redes criminosas. Nos últimos anos, as abordagens baseadas em redes melhoraram através da criação de representações interconectadas, permitindo análises de rede mais abrangentes e integradas. Aproveitando vastos bancos de dados de segurança pública do Estado de Minas Gerais, Brasil, este estudo propõe modelos para identificar e classificar agentes-chave com base em redes complexas e mineração de *outlier* por meio de capital humano, social e misto. Várias estratégias para identificar agentes-chave que perturbam a estrutura da rede são comparadas e o desempenho das estratégias de identificação é avaliado. Os resultados demonstram que a definição proposta de capital social e Outilier Score 1 (OS1) é a abordagem mais eficiente. O modelo de identificação dos principais intervenientes sugerido para perturbar as redes criminosas permite que as autoridades responsáveis pela aplicação da lei identifiquem, visualizem e avaliem as atividades ilegais, aumentando assim a probabilidade de desmantelamento bem sucedido destas redes.

Palavras-chave: Complexidade. Outlier. Redes de criminalidade. Modelagem.

Abstract

The understanding that society functions as interconnected networks has led to significant changes in scientific techniques and approaches across multiple disciplines. In particular, it has profoundly impacted criminology, offering substantial benefits regarding resource allocation and even saving lives. The emergence of advanced theoretical knowledge of Network Science and Machine Learning has opened new paths for developing network analysis models, considering the topological and dynamic relevance of individuals within criminal networks. In recent years, network-based approaches have been improved by creating interconnected representations, allowing for more comprehensive and integrated network analysis. Taking advantage of a vast public security database from the State of Minas Gerais, Brazil, this study proposes models to identify and classify key agents using methods based on complex networks and outliers mining through human and social capital and mixed. Various strategies for identifying key actors to disrupt network structure are compared, and the performance of the identification strategies is evaluated. The results demonstrate that the definition of the social capital proposal and Outilier Score 1 (OS1) is the most efficient approach. The model for identifying key actors suggested for disrupting criminal networks allows law enforcement agencies to identify, visualize, and evaluate illegal activities, thereby increasing the likelihood of successfully dismantling these networks.

Keywords: Complexity. Outlier. Crime networks. Modeling.

Lista de Figuras

Lista de Tabelas

Lista de Algoritmos

Lista de Abreviaturas e Siglas

AUC	Area Under Curve
COV	Robust Covariance
IA	Inteligência Artificial
IsF	Isolation Forest
LOF	Local Outilier Factor
ML	Machine Learning
MG	Minas Gerais
OS	Outlier Score
PMMG	Polícia Militar de Minas Gerais
RA	Rede Alvo
SGBD	Sistema Gerenciador de Banco de Dados
SGD	Stochastic Gradient Descent
SNA	Social Network Analysis
SVM	Support Vector Machine
TIC	Tecnologias de Informação e Comunicação

Sumário

1 Introdução

Nos dias atuais, a sociedade enfrenta desafios significativos decorrentes de mudanças populacionais, instabilidade econômica e financeira, divisões políticas e sociais, bem como desigualdades de poder e conflitos sociais, sendo a criminalidade um dos pontos críticos dessa realidade (1). Nos últimos anos, cientistas sociais têm estudado pontos importantes em sistemas sociais complexos e dado conselhos importantes. (2, 3, 4).

A resposta a esses desafios está intrinsecamente ligada às características da complexidade social. Os sistemas sociais complexos são caracterizados por múltiplas camadas ontológicas e conexões multidirecionais, que permeiam tanto do nível micro para o macroscópico, quanto do macro para o microscópico (5). Essas camadas podem ser representadas por opiniões individuais e públicas, movimentos sociais e políticos, identificação local ou global, preferências, atitudes e humor individuais e coletivos.

Para lidar com esses problemas, é imperativo explorar e compreender a inovação tecnológica, que desempenha um papel fundamental e ainda pouco explorado. O desenvolvimento tecnológico abre canais importantes de informação e introduz novos padrões de comportamento, influenciando substancialmente os princípios das organizações, cujos produtos moldam o comportamento social e se tornam cada vez mais determinantes na construção da história. Consequentemente, o comportamento social sofre mudanças rápidas e desafiadoras à adaptabilidade individual.

O amplo acesso à internet e às redes sociais tem gerado impactos significativos na frequência, no alcance e na forma de comunicação humana. Essa comunicação eletrônica desempenha um papel crucial na organização social das redes de criminalidade, que são definidas como sistemas nos quais conexões, informações, estilos e perfis são compartilhados com o objetivo de cometer ações criminosas (6). Esse acesso generalizado foi possibilitado pela introdução das Tecnologias de Informação e Comunicação (TIC), que agora têm o papel de desvendar as leis que regem a complexidade da sociedade (1).

A compreensão das relações sociais em redes de criminalidade é de extrema importância, uma vez que isso nos permite entender as dinâmicas e os padrões subjacentes a esse fenômeno social complexo (7, 8). Conforme destacado por estudos recentes (9, 10), a análise de redes de criminalidade é fundamental para identificar estruturas organizacionais, identificar agentes-chave e compreender como a informação e a influência se propagam dentro dessas redes.

Os agentes-chave são aqueles que desempenham papéis relevantes e influentes em uma rede de criminalidade, atuando como nós de conexão ou influenciadores dentro da rede. Esses agentes têm um grande impacto na estrutura e no funcionamento da rede de criminalidade, sendo capazes de afetar a disseminação de informações e a coordenação de atividades criminosas. A identificação e análise desses agentes são fundamentais para compreender a dinâmica das redes de criminalidade e desenvolver estratégias eficazes para a sua disrupção.

Em (9) é enfatizada a necessidade de abordagens interdisciplinares para a compreensão da criminalidade em redes, enquanto (11) destaca a importância da aplicação de técnicas avançadas de análise de redes para detectar e interromper atividades criminais. Portanto, explorar a complexidade das relações sociais em redes de criminalidade contribui para o desenvolvimento de estratégias eficazes de prevenção e combate ao crime, além de fornecer percepções valiosas para a segurança pública e as políticas de justiça criminal.

A utilização das TIC no entendimento e combate às redes de criminalidade não é apenas uma ferramenta, mas uma necessidade emergente. As TIC permitem uma análise mais aprofundada e precisa das complexas interações que ocorrem dentro dessas redes. Através da coleta e análise de grandes volumes de dados, é possível identificar padrões, tendências e conexões que poderiam passar despercebidos em uma avaliação tradicional. Além disso, a integração dessas tecnologias com a inteligência artificial e aprendizado de máquina promete transformar a maneira como as forças de segurança abordam e combatem o crime organizado. Entretanto, essa integração também traz desafios éticos e de privacidade, que devem ser considerados e equilibrados com os benefícios potenciais. A evolução da tecnologia, quando empregada com responsabilidade, tem o potencial de ser uma aliada inestimável na luta contra a criminalidade e na construção de uma sociedade mais segura e justa.

Nesse sentido, melhor entender as relações intrínsecas da sociedade e dos comportamentos coletivos emergentes tem sido um tema de interesse crescente no campo da ciência social computacional (12). Estudos computacionais têm explorado a influência de fatores sociais e comportamentais na formação e na dinâmica das redes sociais (13). A análise de redes sociais (SNA, do inglês *Social Network Analysis*) e modelos baseados em redes têm se mostrado ferramentas poderosas para investigar a origem, a evolução e a organização das redes de criminalidade.

A modelagem computacional (14) e a ciência social computacional (12) têm proporcionado avanços significativos nesse campo. Diversos estudos têm utilizado abordagens interdisciplinares, envolvendo cientistas sociais, cognitivos, da computação, matemáticos e físicos, para entender as complexas interações sociais e comportamentais que influenciam a formação de redes de criminalidade. Essa colaboração entre diferentes disciplinas tem sido fundamental para o desenvolvimento de modelos inovadores baseados em redes e para a exploração de grandes volumes de dados disponíveis.

No contexto específico da criminalidade, estudos têm investigado as condições sociais e comportamentais que favorecem o surgimento de redes de criminalidade (15). A análise de redes de criminalidade tem permitido identificar estruturas organizacionais, agentes-chave e padrões de cooperação, contribuindo para uma compreensão mais profunda dos mecanismos subjacentes aos comportamentos criminais (9).

A modelagem computacional tem sido aplicada para acompanhar o surgimento e a evolução dessas redes, fornecendo percepções valiosas para o desenvolvimento de estratégias de prevenção e combate ao crime (16). A análise computacional das redes sociais complexas tem

proporcionado avanços significativos no entendimento das redes de criminalidade. Trabalhos relevantes incluem a ciência das redes complexas e sua aplicação em diversos contextos sociais (17), a teoria dos laços fracos e sua influência na disseminação de informações e comportamentos (18) e a SNA e suas aplicações na compreensão da estrutura e dinâmica das redes (19, 20, 21, 22).

Além disso, estudos têm se dedicado a investigar as características específicas das redes de criminalidade, como a formação de coalizões e a estrutura hierárquica de coalizões criminosas, utilizando teoria de jogos (23) e a modelagem de redes de criminalidade com base em redes complexas (24, 25). Esses estudos contribuem para a compreensão das dinâmicas e mecanismos por trás da formação e da evolução das redes de criminalidade.

A chamada "ciência social computacional" é uma abordagem multidisciplinar que estuda a sociedade como sistema complexo. Nessa abordagem, cientistas sociais, cognitivos, da computação, matemáticos e físicos colaboram para criar modelos inovadores e fundamentados (26). Utilizando da teoria de redes e da ciência social computacional busca-se criar modelos para compreender a organização das redes de criminalidade.

A eficácia da ciência social computacional está atrelada à sua capacidade de sintetizar vastas quantidades de dados e compreender os padrões subjacentes nas redes criminais. A combinação de técnicas quantitativas avançadas com percepções qualitativas dos campos sociais e comportamentais permite um entendimento mais holístico do fenômeno do crime. Como resultado, as autoridades podem desenvolver estratégias de combate ao crime mais direcionadas e eficazes. Além disso, a ciência social computacional ajuda na identificação precoce de tendências emergentes no panorama criminal, permitindo intervenções mais tempestivas. Essas abordagens interdisciplinares também ressaltam a importância da colaboração entre diferentes setores da sociedade, incluindo acadêmicos, profissionais da justiça e tomadores de decisão, para enfrentar coletivamente os desafios impostos pela criminalidade na era digital. Ao combinar a profundidade da análise social tradicional com o alcance e a precisão das técnicas computacionais, tem-se uma ferramenta poderosa para desvendar os intricados meandros das redes criminais e formular respostas adequadas.

A melhor ciência social, estruturada em modelos computacionais, tem se mostrado uma abordagem valiosa para explicar fenômenos sociais complexos. Nos últimos anos, houve um aumento significativo no uso de métodos quantitativos e modelagem computacional na pesquisa em ciências sociais, permitindo a avaliação e verificação de modelos por meio de abordagens baseadas em dados (15).

A ciência social computacional, apoiada pelas TIC, desempenha um papel importante ao permitir a manipulação de grandes conjuntos de dados de redes, possibilitando a análise e a compreensão das estruturas e dinâmicas das redes sociais (26). Essa abordagem interdisciplinar, envolvendo cientistas sociais, computacionais, matemáticos e físicos, contribui para a compreensão dos fenômenos sociais complexos e fornece uma base sólida para a tomada de decisões informadas em diversos campos (27).

A compreensão da complexidade social e da adaptação humana aos desafios do ambiente

requer ferramentas avançadas que permitam a construção de mundos virtuais para simular o comportamento social. Os modelos computacionais, especialmente a modelagem de redes complexas, desempenham um papel fundamental nesse processo (28).

A modelagem computacional é uma ferramenta poderosa para capturar a complexa dinâmica do comportamento em redes sociais interdependentes (29). Ao criar modelos computacionais, podemos comparar, entender e prever comportamentos que ocorrem no mundo real. Esses modelos permitem a interação entre indivíduos e entidades do mundo real, utilizando fluxos de dados genéricos para entrada e saída (14, 30).

A adaptação humana aos desafios do meio ambiente é resultado da complexidade social (15). Para conhecer e compreender esta complexidade, precisamos de ferramentas poderosas para construir mundos virtuais que simulem a complexidade social, que permitam analisar, experimentar e testar o comportamento social. Modelos computacionais oferecem essa possibilidade, com o uso da modelagem computacional.

O uso das TIC tem gerado e agregado quantidades massivas de dados provenientes de bancos de dados, internet e redes sociais. Esses dados incorporam uma ampla gama de atividades individuais, incluindo atividades criminosas, o que possibilita uma abordagem científica inovadora para a análise social. Os conjuntos de dados volumosos, conhecidos como *Big Data*, podem fornecer percepções que ajudam a explicar fenômenos e processos complexos das relações sociais na criminalidade. Cada indivíduo deixa rastros nesses conjuntos de dados, e a identificação desses rastros pode contribuir para a compreensão das relações sociais e da sociedade em geral (31), bem como para o estudo dos padrões de comunicação humana (32).

Apesar do vasto potencial desses dados, o desenvolvimento de teorias nessa área tem sido lento, principalmente devido à escassez de informações que precisam ser coletadas em meio à imensa massa de dados disponíveis. No entanto, diferentes tecnologias estão sendo desenvolvidas para categorizar e avaliar de forma eficiente as informações individuais. É aí que as TIC entram em cena, permitindo o acesso, a análise e a construção de grandes bases de dados para abordar o desafiador problema da complexidade social das redes de criminalidade. Por meio do uso de métodos estatísticos e da capacidade de armazenar e processar grandes quantidades de informações, as TIC possibilitam a rápida e eficiente incorporação de dados individuais.

Neste cenário, se destaca o problema: **a capacidade de desenvolver modelos que, por meio da vasta quantidade de dados disponíveis, possam identificar os agentes-chave e simular a disrupção de redes de criminalidade**. A identificação precisa desses agentes é crucial, pois eles frequentemente atuam como nodos de conexão ou influenciadores dentro da rede, possuindo um impacto significativo na sua estrutura e operação. Além disso, entender como essas redes operam e se adaptam após a remoção ou desativação de um ou mais desses agentes-chave é fundamental para a criação de estratégias de intervenção mais eficientes. Desse modo, a integração entre a coleta de dados, análise avançada e a modelagem computacional se apresenta como um caminho promissor para avançar na luta contra a criminalidade organizada, otimizando os esforços de instituições de segurança pública e justiça.

O surgimento e o desenvolvimento de ferramentas de análise, representação e visualização de dados são de extrema importância para a abordagem computacional das redes de criminalidade. Essas ferramentas permitem estabelecer relações entre teoria, fatos e pesquisa, fornecendo uma base sólida para a compreensão e a análise dessas redes.

As ferramentas de análise do comportamento social, em particular a ciência social computacional e a modelagem computacional, são poderosas para a compreensão da complexidade dos sistemas sociais. Estas permitem a construção de mundos sociais virtuais, nos quais é possível alimentar, experimentar, analisar e testar dados, proporcionando uma abordagem inovadora na compreensão dos fenômenos sociais.

A análise dos dados coletados pode revelar informações valiosas sobre a estrutura, os padrões de interação e outros aspectos relevantes das redes de criminalidade, contribuindo para o desenvolvimento de estratégias mais eficazes no combate ao crime.

Dessa forma, como hipótese, postula-se que **há uma correlação entre redes de criminalidade que possibilitam estabelecer padrões para a identificação de agentes-chave**.

Combinando percepções de várias disciplinas, este estudo visa obter uma imagem mais holística e aprofundada das redes de criminalidade. Entender as dinâmicas intrincadas, interações e padrões dentro dessas redes é vital para direcionar esforços eficazes de combate ao crime. Ao desenvolver modelos que destilam e decodificam essas complexas relações, a pesquisa se posiciona não apenas para identificar os agentes-chave, mas também para prever possíveis movimentos e reações dentro dessas redes após intervenções estratégicas. Tal análise multifacetada é essencial para fortalecer as medidas preventivas, de repressão e de desmantelamento dessas redes, garantindo uma resposta mais adaptativa e antecipada às ameaças que representam para a sociedade.

Assim, tem-se como objetivo geral: **desenvolver modelos que permitam a análise de redes de criminalidade, identificando os agentes-chave que as compõem**.

Esses indivíduos podem incluir líderes, comunidades, padrões de interação e outros fatores relevantes para a compreensão da estrutura e do funcionamento das redes de criminalidade.

Embora a análise holística das redes de criminalidade forme o cerne desta pesquisa, é fundamental subdividir o estudo em objetivos específicos para uma abordagem mais direcionada e sistemática. Estes objetivos específicos facilitarão a abordagem dos múltiplos aspectos das redes de criminalidade e proporcionarão um entendimento granular das suas variáveis constituintes. Através deste detalhamento, busca-se decifrar os mecanismos subjacentes, prever tendências emergentes e, crucialmente, conceber estratégias eficazes de intervenção.

Os objetivos específicos servem como pilares de sustentação para o objetivo geral, garantindo uma investigação meticulosa e abrangente do fenômeno em estudo, sendo:

1. A criação de modelos que permitam identificar redes de criminalidade. Esses modelos devem ser capazes de analisar e identificar os padrões e as estruturas das redes criminais, levando em consideração os dados disponíveis e as características específicas dessas redes;

2. A criação de simulações que promovam a disrupção de redes de criminalidade. Por meio de modelos, busca-se identificar os elementos cruciais que influenciam a estrutura e o funcionamento das redes criminais, buscando sua disrupção;

3. Desenvolvimento de ferramentas computacionais que facilitem a análise e o estudo das redes de criminalidade. Isso inclui o desenvolvimento de algoritmos, técnicas de visualização e outras ferramentas que possam auxiliar na compreensão e na interpretação dos dados relacionados às redes criminais, contribuindo assim para a investigação e o combate ao crime.

Este trabalho está organizado em capítulos: o presente Capítulo 1,introdução; Capítulo 2 apresenta a fundamentação teórica; Capítulo 3 descreve trabalhos relacionados a este estudo; Capítulo 4 descreve a metodologia, com a coleta e tratamento dos dados, os modelos propostos e a avaliação destes modelos; Capítulo 5 apresenta os resultados, a discussão e perspectivas do estudo; Capítulo 6 apresenta as conclusões.

2 Fundamentação Teórica

Este capítulo explora a interseção de diversas disciplinas para compreender a dinâmica das redes criminais.

Na vanguarda do combate ao crime, a interseção entre redes complexas, aprendizado de máquina e detecção de *outliers* oferece perspectivas inovadoras e promissoras. As redes complexas fornecem uma representação abrangente das interações entre entidades criminosas, revelando padrões e estruturas subjacentes que escapam à observação superficial. A aplicação de algoritmos de aprendizado de máquina nesse contexto permite a análise eficiente e escalável dessas redes, identificando padrões emergentes e comportamentos anômalos que podem indicar atividades criminosas. A detecção de *outliers* desempenha um papel crucial ao destacar nós incomuns que podem ser indicativos de atividades ilícitas ou pontos de vulnerabilidade na rede. Ao integrar esses elementos, é possível não apenas compreender melhor a dinâmica da criminalidade, mas também desenvolver estratégias proativas de interrupção das redes de criminalidade, interrompendo fluxos de informação, identificando agentes-chave e desarticulando operações criminosas com maior eficácia. A disrupção de redes de criminalidade, portanto, torna-se uma meta alcançável com a aplicação sinérgica de análises de redes complexas e técnicas avançadas de aprendizado de máquina na identificação e neutralização de ameaças criminosas.

2.1 Perspectivas de combate ao crime

Implementar novas tecnologias de análise de dados e redes para o trabalho policial é um grande desafio para conter o crescimento das organizações criminosas (1, 2). As ferramentas para isso devem estar inseridas nos contextos institucionais, organizacionais e práticos do trabalho policial, de forma a permitir um policiamento baseado em práticas sociotécnicas para desorganizar as redes de criminalidade e reduzir a prática de crimes (3). Entre os crimes mais desafiadores encontram-se os crimes de homicídio, tráfico de drogas, roubo e tráfico de armas, que requerem intervenções proativas para atingir as organizações criminosas que os apoiam (1). No mundo todo, agências governamentais, acadêmicos e agentes da lei estão constantemente buscando maneiras mais eficazes de quebrar ou controlar essas estruturas ilegais. Tecnologias têm sido desenvolvidas para estudar e simular as relações humanas, combinando o mundo real e a computação em modelos que representem como essas relações são estabelecidas (2, 3, 4, 9, 23, 24, 33, 34). Isso permite padronizar, com agilidade e precisão, comportamentos que podem ser utilizados como pontos de referência na busca de formas de identificar agentes-chave e causar a disrupção de redes de criminalidade (10, 11, 13, 16, 34, 35, 36, 37).

Desde a década de 1990, a abordagem de rede para estudar organizações criminosas aumentou (38). Técnicas de SNA têm sido usadas com sucesso para entender melhor as interações entre criminosos (9, 33, 34, 39, 40). Entretanto, criar sociedades humanas é uma tarefa desafiadora,

já que as pessoas têm comportamentos distintos quando se juntam, resultando em comportamentos variados e muitas vezes inesperados, especialmente em organizações criminosas. (9, 41, 42). Além disso, a aplicação da abordagem em rede ao estudo dos fenômenos criminais vem ampliando as definições de crime organizado (38, 43, 44, 45, 46). Apesar de técnicas de teoria de redes estarem sendo usadas com sucesso para explicar fenômenos criminais, principalmente na investigação de interações entre atores (39, 40), o estudo nesta área ainda permanece em análise mais descritiva (1). Contudo, alguns estudos recentes surgem para quebrar este paradigma (10, 33, 34, 36, 37)

Um melhor entendimento das redes de criminalidade é crucial para as agências de aplicação da lei. A identificação precisa de indivíduos que desempenham funções importantes nessas organizações poderá permitir uma disrupção ou prevenção mais eficientes das atividades criminosas.

2.2 Sistemas complexos

Os sistemas complexos têm sido objeto de estudo e pesquisa em várias disciplinas, incluindo física, biologia, ciências sociais e computação (47). Esses sistemas são caracterizados por uma estrutura intrincada e interações não lineares entre seus componentes, resultando em comportamentos emergentes e imprevisíveis (48, 49).

A teoria dos sistemas complexos fornece um arcabouço conceitual para compreender e analisar a natureza e o comportamento de sistemas complexos. Em (50), são discutidos os princípios fundamentais da teoria dos sistemas complexos e é definido um sistema complexo como "um sistema composto por um grande número de partes interagentes que exibem comportamentos coletivos não triviais emergentes dessas interações".

Os sistemas complexos exibem características distintivas que os diferenciam dos sistemas simples. Em (51), são enfatizadas algumas dessas características, como não linearidade, interconectividade, auto-organização, adaptação e emergência. A não linearidade refere-se às relações não proporcionais entre causa e efeito, levando a comportamentos imprevisíveis. A interconectividade destaca as múltiplas interações entre os componentes do sistema. A auto-organização refere-se à capacidade do sistema de se reorganizar espontaneamente sem uma entidade central controladora. A adaptação implica na capacidade do sistema de ajustar seu comportamento em resposta a mudanças no ambiente. A emergência se refere ao surgimento de novos padrões e propriedades que não podem ser reduzidos às características de seus componentes individuais.

Os sistemas complexos têm aplicações em várias áreas, desde a modelagem de ecossistemas até a SNA. Em (52), são exploradas as aplicações dos sistemas complexos em diferentes domínios e é discutido como a teoria dos Sistemas Complexos pode ser aplicada para entender fenômenos complexos, como a dinâmica populacional, o comportamento coletivo de animais, a propagação de doenças e a SNA.

Métodos de análise de sistemas complexos requerem abordagens e técnicas específicas para lidar com a complexidade e a imprevisibilidade desses sistemas, considerando a interação

entre agentes individuais e seu ambiente. (53, 54).

2.2.1 Redes complexas

Redes complexas têm sido amplamente estudadas em diversos campos, incluindo ciência da computação, física e policial (9, 10, 11, 33, 34, 47). A aplicação de redes complexas pode revelar padrões de interação humana, permitindo a análise de redes de criminalidade para revelar agentes-chave e contribuir na disrupção destas redes.

A teoria de redes complexas fornece um conjunto de ferramentas e conceitos para analisar a estrutura e a dinâmica dessas redes. Em (53) oferece-se uma visão abrangente sobre a teoria de redes complexas. Ele define uma rede como "um conjunto de nós interconectados por ligações" e enfatiza a importância de analisar as propriedades globais da rede, além das características locais dos nós individuais.

Redes complexas exibem características distintivas que as diferenciam de redes simples. Em (55) discute-se algumas dessas características, destacando a presença de distribuições de grau, onde o número de conexões de cada nó segue uma distribuição estatística específica. Além disso, enfatiza a presença de clusters e comunidades, em que conjuntos de nós estão fortemente interconectados entre si. Essas características têm implicações importantes na análise e na modelagem de redes complexas.

A análise de redes complexas requer métodos específicos para investigar sua estrutura e dinâmica. Em (28, 56) apresentam-se uma variedade de ferramentas para a análise de redes complexas, incluindo medidas de centralidade, identificação de comunidades, modelos de crescimento de rede e modelagem de propagação de informações. Além disso, em (53) discute-se métodos estatísticos para analisar a estrutura da rede, como modelos nulos para comparar a rede real com redes aleatórias e técnicas de detecção de modularidade para identificar comunidades.

A teoria das redes complexas tem suas raízes na teoria dos grafos e na sociometria, mas foi amplamente expandida para incluir modelos matemáticos e computacionais que capturam as propriedades emergentes de sistemas complexos. Em particular, a análise de redes sociais criminosas tem se beneficiado significativamente desses avanços. De acordo com (57), as redes sociais são estruturas formadas por atores (nós) e as relações (arestas) entre eles, e a análise dessas redes permite entender como a posição de um ator dentro da rede influencia seu comportamento e acesso a recursos.

Uma característica fundamental das redes complexas é a presença de heterogeneidade estrutural, onde certos nós possuem um grau de conectividade significativamente maior que outros, formando hubs ou nós centrais (58). Essa propriedade é observada em redes de criminalidade, onde alguns indivíduos desempenham papéis críticos, facilitando atividades ilícitas e a coordenação entre diferentes partes da rede (9).

Além disso, a teoria dos pequenos mundos e a conectividade robusta em redes sociais oferecem perpectivas sobre como as redes de criminalidade mantêm a coesão e a resiliência mesmo quando alguns de seus membros são removidos (59). Redes de criminalidade frequente-

mente exibem uma alta modularidade, com comunidades ou subgrupos fortemente conectados internamente, mas com conexões esparsas entre diferentes comunidades (60). Essa estrutura modular pode ser explorada para desmantelar redes, focalizando intervenções em pontos de interconexão entre comunidades.

A aplicação de algoritmos de detecção de comunidades, como o algoritmo de Louvain, pode ajudar a identificar subestruturas dentro das redes de criminalidade (61). Essas subestruturas ou módulos podem corresponder a diferentes facções ou células operacionais dentro de uma organização criminosa. Identificar e entender essas comunidades é crucial para desenvolver estratégias de intervenção que minimizem a reorganização e resiliência da rede após a remoção de atores chave (11).

A centralidade de intermediários (betweenness centrality) é outra medida importante na análise de redes complexas, especialmente em redes de criminalidade. Nós com alta centralidade de intermediários agem como pontes críticas na rede, facilitando a comunicação e a transferência de recursos entre diferentes partes da rede (62). A remoção ou neutralização desses nós pode causar fragmentação significativa na rede, dificultando a coordenação das atividades criminosas (9).

Modelos dinâmicos de propagação em redes complexas, como modelos SIR (Susceptible-Infected-Recovered), têm sido adaptados para estudar a disseminação de informações e comportamentos em redes sociais (63). Em redes de criminalidade, esses modelos podem ser usados para simular a propagação de informações sobre atividades policiais ou estratégias de disrupção, permitindo prever o impacto de diferentes intervenções na rede (64).

Redes multiplex são uma classe especial de redes complexas em que diferentes tipos de interações ou relações ocorrem simultaneamente entre os mesmos conjuntos de nós (65, 66). Essas redes são compostas por várias camadas, cada uma representando um tipo específico de relação entre os nós (67, 68). A análise de redes multiplex tem se tornado cada vez mais relevante em diversos campos, como ciências sociais, policiais, biológicas, econômicas e da computação, devido à capacidade de capturar a complexidade das interações entre os elementos do sistema (9, 67, 69).

Para entender as redes multiplex, é essencial compreender os conceitos básicos das redes complexas. Uma rede complexa consiste em um conjunto de nós interconectados por arestas, que representam as relações entre eles (53). No entanto, nas redes multiplex, múltiplas camadas coexistem, cada uma representando um tipo de relação (67). Por exemplo, uma rede social pode ser representada por uma camada que captura as amizades entre indivíduos, outra camada que representa as interações profissionais e uma terceira camada que reflete os interesses em comum.

Uma das principais abordagens para analisar redes multiplex é a teoria dos grafos ponderados, que considera não apenas a estrutura da rede, mas também os atributos associados a cada relação (66, 70, 71). Esses atributos podem ser qualitativos, como o tipo de interação, ou quantitativos, como a intensidade da conexão. A incorporação desses atributos permite uma compreensão mais refinada das dinâmicas e propriedades do sistema em estudo.

A análise de redes multiplex para compreender a estrutura e a dinâmica das relações entre indivíduos envolvidos em atividades criminosas ainda é escassa (72). Um dos principais desafios na análise de redes de criminalidade é a natureza oculta e complexa das atividades criminosas (9). Em um contexto bem definido, as redes multiplex podem representar diferentes tipos de interações, como associações criminais, transferência de recursos ilícitos, comunicação entre criminosos e padrões de mobilidade (4, 45). A abordagem multiplex permite capturar as múltiplas facetas das interações entre os atores envolvidos, fornecendo percepções valiosas para investigadores e formuladores de políticas de segurança (72).

A integração dessas abordagens teóricas e metodológicas oferece uma visão holística da estrutura e dinâmica das redes de criminalidade. A análise detalhada das interações complexas dentro dessas redes pode fornecer subsídios valiosos para a formulação de políticas de segurança mais eficazes e para a implementação de estratégias de intervenção mais precisas e direcionadas (9).

2.3 Aprendizado de máquina

Aprendizado de máquina ou *Machine Learning* (ML), do inglês, é um campo da Inteligência Artificial (IA) que se concentra no desenvolvimento de algoritmos e modelos que permitem aos computadores aprenderem e realizarem tarefas sem serem explicitamente programados (73, 74). É um campo interdisciplinar que combina estatística, matemática, ciência da computação e teoria da informação para extrair conhecimentos de conjuntos de dados (75).

Dentro do campo do ML, existem várias subdivisões importantes que abordam diferentes tipos de problemas e técnicas de aprendizado (73, 76). As principais subdivisões incluem:

1. Aprendizado supervisionado: envolve o treinamento de modelos usando dados rotulados, em que o algoritmo é alimentado com pares de entrada e saída correspondentes. O objetivo é fazer com que o modelo aprenda a mapear as entradas para as saídas corretas;

2. Aprendizado não supervisionado: os dados não possuem rótulos e o objetivo é encontrar padrões e estruturas intrínsecas nos dados. Os algoritmos de aprendizado não supervisionado podem ser utilizados para agrupar dados similares em clusters, reduzir a dimensionalidade dos dados, encontrar regras de associação entre as variáveis ou detectar *outliers*;

3. Aprendizado por reforço: envolve um agente que interage com um ambiente e aprende a tomar ações que maximizam uma recompensa numérica. O agente recebe *feedback* em forma de recompensa ou penalidade com base nas ações que realiza. Ele aprende a melhor sequência de ações para maximizar a recompensa ao longo do tempo.

A diferença fundamental entre aprendizado supervisionado e não supervisionado reside na presença ou ausência de rótulos nos dados de treinamento. No aprendizado supervisionado, os dados são rotulados e o objetivo é fazer previsões com base nesses rótulos (73). No aprendizado

não supervisionado, não há rótulos disponíveis, e o objetivo é encontrar estruturas, padrões ou grupos nos dados (76),

Ambas as abordagens têm suas próprias aplicações e são utilizadas de acordo com a natureza do problema e a disponibilidade de dados rotulados (74). O aprendizado supervisionado é comumente usado em tarefas de previsão e classificação, enquanto o aprendizado não supervisionado é útil para descobrir *outliers*, agrupar dados e reduzir a dimensionalidade em conjuntos de dados não rotulados.

Importante ressaltar que a escolha da técnica de ML e do pré-processamento de dados mais adequados depende do contexto e dos objetivos específicos, para que desempenhe papéis fundamentais para resultados mais precisos e confiáveis.

Além disso, os modelos de avaliação desempenham um papel fundamental na construção e otimização de sistemas de ML. Eles permitem medir o desempenho e a eficácia dos modelos, fornecendo uma base objetiva para avaliar seu desempenho em diferentes cenários. A escolha adequada de métricas de avaliação é essencial para compreender como um modelo está se comportando e para comparar diferentes abordagens. Além disso, os modelos de avaliação permitem identificar possíveis problemas, como *overfitting*[1] ou *underfitting*[2], e orientar a seleção de hiperparâmetros e técnicas de pré-processamento de dados. Em suma, as técnicas de avaliação são essenciais para o desenvolvimento de modelos de ML confiáveis e bem sucedidos, fornecendo uma maneira de medir, monitorar e melhorar seu desempenho ao longo do tempo.

2.3.1 Medidas de avaliação para modelos de aprendizado de máquina

As medidas de avaliação são fundamentais para analisar o desempenho e a qualidade dos modelos de ML. Elas fornecem uma maneira objetiva de quantificar o quão bem o modelo está realizando suas tarefas.

Os modelos de ML supervisionados são treinados usando um conjunto de dados rotulados, em que cada amostra possui uma variável de entrada e sua correspondente variável de saída (rótulo). As medidas de avaliação comuns para esses modelos incluem (77):

1. Matriz de confusão: é uma tabela que mostra a contagem de verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN) obtidos pelo modelo. Com base nesses valores, várias medidas adicionais podem ser calculadas, como precisão, *recall* e *F1-score*;

2. *Precision-Recall*: a precisão é a proporção de verdadeiros positivos em relação à soma dos verdadeiros positivos e falsos positivos. Ela mede a taxa de acertos positivos em relação ao total de previsões positivas. Já o *recall* é a proporção de verdadeiros positivos em relação à

[1] *Overfitting* ocorre quando um modelo de aprendizado de máquina se ajusta excessivamente aos dados de treinamento, resultando em uma baixa capacidade de generalização para novos exemplos (73).

[2] *Underfitting* ocorre quando um modelo de aprendizado de máquina não é capaz de capturar os padrões e complexidades dos dados de treinamento, resultando em um desempenho insuficiente tanto nos dados de treinamento quanto em novos exemplos (73).

soma dos verdadeiros positivos e falsos negativos. Ele mede a taxa de captura de todos os casos positivos corretamente;

3. *F1-score*: é uma métrica que combina a precisão e o *recall* em uma única medida. É calculado como a média harmônica entre essas duas métricas, proporcionando um equilíbrio entre elas. O F1-score é particularmente útil quando há um desequilíbrio significativo nas classes de saída;

4. *Curva ROC e Área sob a Curva (AUC-ROC)*: a curva ROC (*Receiver Operating Characteristic*) é uma representação gráfica da taxa de verdadeiros positivos em função da taxa de falsos positivos. A AUC-ROC é a área sob essa curva e fornece uma medida agregada do desempenho do modelo. Quanto maior a AUC-ROC, melhor é o desempenho do modelo em discriminar entre as classes.

Os modelos de ML não supervisionados não dependem de rótulos de saída, pois são usados principalmente para descobrir padrões e estruturas ocultas nos dados. As medidas de avaliação comuns para esses modelos incluem (78):

1. Coeficiente de silhueta: é uma medida que quantifica a coesão e a separação dos grupos formados pelo algoritmo de agrupamento. Ele varia entre -1 e 1, em que valores mais próximos de 1 indicam que as amostras estão bem agrupadas, enquanto valores próximos de -1 indicam que as amostras estão mais próximas de *clusters* vizinhos;

2. Índice de Davies-Bouldin: mede a separação entre os grupos e a dispersão dentro de cada grupo. Quanto menor o valor do índice, melhor é a separação e a compactação dos grupos;

3. Coeficiente de Calinski-Harabasz: é uma medida que compara a dispersão entre os grupos com a dispersão dentro dos grupos. Valores mais altos do coeficiente indicam uma melhor separação entre os grupos.

As medidas de avaliação mencionadas anteriormente para modelos não supervisionados não são especificamente projetadas para detecção de *outliers* (77).

As medidas de avaliação são ferramentas essenciais para analisar a qualidade e o desempenho dos modelos de ML, tanto supervisionados quanto não supervisionados. Essas medidas fornecem uma maneira objetiva de quantificar a eficácia dos modelos em tarefas específicas, tomada de decisões informadas sobre o ajuste e a otimização dos modelos. É importante selecionar as medidas de avaliação adequadas, levando em consideração o tipo de modelo, a natureza dos dados e o objetivo da tarefa. Assim, a avaliação dos modelos deste trabalho serão as descritas na seção 4.4.

2.4 Detecção de *outlier*

Um *outlier* pode ser definido como "um valor que está fora do intervalo esperado de um conjunto de dados" (79). Existem várias definições e abordagens para identificar *outlier*, dependendo do contexto e dos objetivos da análise (80). As técnicas de detecção de *outlier* (81, 82) incluem abordagens baseadas em distância, densidade, agrupamento e estatísticas:

1. as abordagens baseadas em distância avaliam a proximidade de uma observação em relação às demais no conjunto de dados. O algoritmo de ML *SGDOneClassSVM* (83) se baseia na abordagem de distância para detecção de *outlier*. Ao utilizar a técnica de descida de gradiente estocástica, o *SGDOneClassSVM* otimiza os pesos do modelo para encontrar a melhor fronteira de decisão que separa os exemplos normais de possíveis *outliers*. A separação é baseada na distância das observações em relação a essa fronteira de decisão. Os *Outliers Scores* (84) baseiam-se também na abordagem de distância para mireração de *outlier*. Partindo de grafo totalmente conectado e não direcionado, em que os nós são os elementos do conjunto de dados e os links têm pesos que são as distâncias entre os nós, pontuações de *outliers* são definidas pela análise da estrutura do grafo. Em particular, usando a divergência de Jensen-Shannon (JS), explicada no capítulo Metodologia, para comparar as distribuições de pesos de diferentes nós;

2. as abordagens baseadas em densidade exploram a distribuição dos dados e identificam *outliers* como pontos de baixa densidade. O algoritmo de ML *Isolation Forest* é um algoritmo de detecção de *outlier* (85, 86). Ele utiliza o conceito de árvores de isolamento para identificar observações anômalas em um conjunto de dados. As árvores de isolamento são construídas de forma aleatória, selecionando-se aleatoriamente um atributo e um valor de corte em cada nó da árvore. Os *outliers* são identificados como instâncias que requerem menos passos para serem isoladas em uma árvore, ou seja, são facilmente separáveis do restante dos dados;

3. as abordagens baseadas em agrupamento identificam *outliers* como pontos que não pertencem a nenhum grupo ou que pertencem a grupos pequenos e pouco densos. O algoritmo de ML *Local Outlier Factor* (87) é uma técnica amplamente utilizada que calcula o fator de *outlier* local de cada ponto com base na densidade relativa de seus vizinhos. O *Local Outlier Factor* atribui um escore a cada ponto, permitindo a identificação de *outliers* com base em sua falta de aderência ao comportamento local;

4. as abordagens estatísticas utilizam técnicas matemáticas e estatísticas para detectar *outliers* com base em medidas como média, desvio padrão e distribuição dos dados. O algoritmo de ML *Robust Covariance* se baseia na abordagem estatística, utilizando a estimativa robusta da matriz de covariância para detectar *outliers* (88). Essa abordagem visa mitigar os efeitos prejudiciais dos valores discrepantes na análise estatística tradicional.

A mineração de *outliers* é uma tarefa importante na análise de dados, cujo objetivo é identificar padrões anômalos, incomuns ou divergentes que se destacam em relação ao restante dos dados (79, 80). Esses *outliers* podem fornecer percepções valiosos e identificar comportamentos atípicos em uma variedade de domínios, como finanças, saúde, detecção de fraudes, manutenção preditiva e identificação de indivíduos chave em grandes bases de dados (89, 90).

No entanto, a interpretação dos resultados dos modelos não supervisionados pode ser um desafio, pois eles não fornecem uma medida direta de probabilidade de anomalia (76). Isto porque modelos não supervisionados fazem uso de dados não rotulados (Fig. 1).

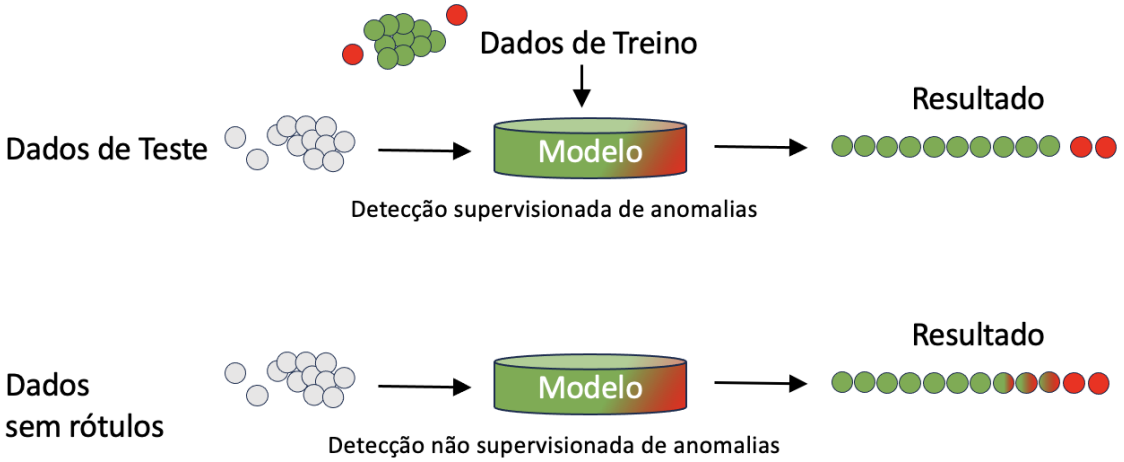


Figura 1 – Detecção de anomalias - as técnicas de detecção de *outliers* baseadas em modelos supervisionados utilizam dados de treinamento rotulados. Já os modelos não supervisionados utilizam dados não rotulados. O resultado de modelos supervisionados é a classificação dos dados. O de modelos não supervisionados é a classificação dos dados e a atribuição de um valor a cada elemento.

A detecção de *outliers* é uma área de pesquisa ativa e em constante evolução, com várias abordagens e técnicas disponíveis, que muito se beneficia do avanço dos algoritmos de ML, especialmente aqueles não supervisionados, (91, 92, 93).

2.5 Disrupção de redes de criminalidade

A disrupção de redes de criminalidade é um desafio complexo que requer abordagens inovadoras e multidisciplinares. Para lidar com essa questão é preciso levar em consideração as estratégias de disrupção, que podem ser categorizadas em duas técnicas principais: a abordagem do capital humano e a abordagem do capital social. A combinação das duas dá origem a uma terceira: a abordagem mista. Essas estratégias podem ser aplicadas a redes de criminalidade que são construídas a partir de dados oriundos das agências de aplicação da lei. As principais referências em relação a este campo de pesquisa estão resumidas na Tabela 1.

O capital social refere-se à rede de relações sociais, normas e confiança que existem em uma comunidade ou grupo. Ele representa o valor que as pessoas obtêm por meio de suas

Tabela 1 – Referências importantes sobre técnicas de disrupção de redes de criminalidade.

Abordagem	**Conceitos Chave**	**Referências**
Capital Social	*Degree centrality*	(94)
	Betweenness centrality	(94)
	Closeness Centrality	(95)
	Clustering Coefficient	(36)
	Multiplexity	(33)
Capital Humano	Substituibilidade	(96)
	Cadeia de valor	(94)
	Análise de roteiro criminal	(97)
		(98)
	Crime e acumulação de capital humano	(99)
Misto	Combinação de capital humano e social	(11)
		(100)
		(101)

conexões sociais, cooperação e interações com os outros membros do grupo. O capital social está intimamente relacionado às técnicas de SNA, uma vez que essas técnicas fornecem uma maneira de visualizar, medir e compreender as relações sociais presentes em uma determinada comunidade ou grupo. A SNA permite identificar e explorar as estruturas de conexões e interações entre os indivíduos, que são a base do capital social.

O capital humano refere-se ao conjunto de habilidades, conhecimentos, experiências e atributos pessoais que um indivíduo possui e que podem ser usados para alcançar objetivos pessoais e sociais. No contexto de redes de criminalidade, o capital humano refere-se ao conjunto de conhecimentos, habilidades, experiências e atributos pessoais relacionados às atividades ilícitas. É o potencial humano que pode ser mobilizado para construir e manter ações ilícitas.

No contexto da disrupção de redes de criminalidade, a resiliência pode ser entendida como a capacidade de uma comunidade, grupo ou indivíduo se adaptar, resistir e se recuperar diante dos desafios e impactos da disrupção (102). É a habilidade de enfrentar e superar as adversidades, fortalecendo o tecido social, promovendo a segurança e contribuindo para a prevenção da desarticulação das redes criminosas. As principais referências em relação a este campo de pesquisa estão resumidas na Tabela 2.

Embora a resiliência seja frequentemente vista como uma qualidade desejável em uma comunidade, também é importante entender sua importância na perspectiva do crime e do criminoso.

A resiliência pode dificultar a aplicação da lei. Criminosos e redes de criminalidade

Tabela 2 – Referências importantes sobre resiliência de redes de criminalidade.

Resiliência	**Descrição**	**Referências**
Conceitos chave	*Definição*	(102)
	Características	(103)
	Compensação de eficiência/segurança	(104)
	Redundância	(105)
	Não redundância	(11)
Estudos de caso	Mercados de drogas ilegais	(96)
	Terrorismo	(106)
	Guanges de rua	(103)
	Tráfico de pessoas	(107)
	Cultivo de maconha	(11)
	Máfia	(108)

resilientes podem se adaptar rapidamente a mudanças nas táticas de repressão, evitando detecção e esquivando-se de investigações.

A resiliência também está relacionada à capacidade de se recuperar rapidamente após a desarticulação de uma rede ou a prisão de agentes-chave. Criminosos resilientes podem se reorganizar, estabelecer novas conexões e retomar suas atividades criminosas em curto prazo. Portanto, é fundamental que as estratégias de disrupção considerem a resiliência dos criminosos, adaptando-se e inovando constantemente para superar essas barreiras.

2.6 Desassortatividade em redes de criminalidade

Organizações em rede são flexíveis e dinâmicas, capazes de limitar a exposição a riscos, permitindo alianças que dependem de papéis e do nível de confiança dos criminosos (13). Essas características adaptativas intrínsecas tornam as redes criminosas um desafio para identificar e desmontar (11). Se considerarmos a estrutura das redes sociais em geral, os indivíduos com papéis essenciais apresentam um maior grau de centralidade do que aqueles cujos papéis são menos significativos (9). Na verdade, essa suposição é válida para redes sociais, que são principalmente redes assortativas (41, 42).

A assortatividade refere-se ao padrão de que nós semelhantes tendem a se conectar uns aos outros com mais frequência do que o esperado ao acaso (41, 42). Em outras palavras, nós que compartilham características semelhantes têm maior probabilidade de se conectarem na rede. Essas características podem ser diversas, dependendo do tipo de rede que está sendo analisada. Por exemplo, em uma rede social, as características podem ser interesses, idade, gênero ou status

socioeconômico. Em uma rede de colaboração científica, as características podem ser o número de publicações, área de pesquisa ou instituição de afiliação.

Por outro lado, a desassortatividade ocorre quando nós com características diferentes têm maior probabilidade de se conectar entre si (41, 42). Nesse caso, a rede tende a formar conexões entre nós que possuem atributos complementares ou opostos.

Tanto a assortatividade quanto a desassortatividade têm implicações importantes na estrutura e dinâmica das redes. Redes mais assortativas tendem a formar comunidades mais homogêneas, nas quais nós com características semelhantes agrupam-se em sub-redes densamente conectadas (41, 42). Já redes mais desassortativas tendem a ser mais resilientes a falhas e propagação de informações, pois nós com diferentes características podem funcionar como pontes que conectam diferentes partes da rede. Essas características de assortatividade e desassortatividade são úteis para entender a dinâmica e a evolução de diferentes tipos de redes e podem ter implicações em várias aplicações práticas.

As redes de criminalidade frequentemente exibem características desassortativas (34), o que significa que há uma tendência para que nós com diferentes níveis de importância ou influência estejam conectados entre si. Esse padrão é crucial para o estudo das redes de criminalidade por várias razões fundamentais. (1, 9, 11, 16, 34, 36):

- diversidade criminosa: a desassortatividade sugere que a rede de criminalidade é composta por indivíduos envolvidos em diferentes tipos de crimes. Pode haver especialização entre diferentes grupos de criminosos que se envolvem em atividades criminosas específicas, como tráfico de drogas, roubo, extorsão, entre outros;
- integração na rede: a conexão entre indivíduos com diferentes perfis criminosos pode facilitar a integração e a comunicação entre grupos criminosos distintos. Isso pode levar à colaboração entre esses grupos em algumas atividades criminosas ou permitir a coordenação de ações em escala maior;
- dificuldade para detecção: a desassortatividade pode tornar a identificação de padrões criminosos mais desafiadora para as autoridades. Como diferentes tipos de criminosos estão interligados, pode ser mais difícil rastrear e entender a complexa dinâmica da rede;
- propagação de informações: a existência de conexões entre diferentes grupos criminosos pode facilitar a propagação de informações, táticas e recursos dentro da rede criminal. Isso pode levar a uma maior disseminação de métodos criminosos eficazes e ao compartilhamento de conhecimento sobre a aplicação da lei;
- resiliência: a desassortatividade pode tornar a rede criminal mais resiliente a perturbações ou ações de aplicação da lei, uma vez que a desconexão de um grupo específico de criminosos pode não fragmentar totalmente a rede.

A natureza desassortativa das redes de criminalidade reflete a sua capacidade de adaptabilidade e resiliência. Ao contrário de redes aleatórias onde os nós têm conexões igualmente

distribuídas, nas redes desassortativas, os nós centrais tendem a estar conectados a nós periféricos. Isso espelha a forma como os líderes criminosos muitas vezes estão ligados a indivíduos menos proeminentes que desempenham papéis fundamentais em operações ilegais. Compreender essas conexões é vital para desmantelar redes de criminalidade de forma eficaz.

Além disso, a análise das redes desassortativas permite identificar pontos críticos de controle e vulnerabilidade. Nesses tipos de redes, a remoção de nós altamente conectados pode ter um impacto significativo na estrutura e na eficácia da rede como um todo. Identificar e neutralizar esses nós centrais pode interromper o fluxo de informações, recursos e influência dentro da rede criminosa, minando sua capacidade de operar de forma eficiente.

É essencial compreender a estrutura e as características das redes de criminalidade para enfrentar adequadamente o crime e desenvolver estratégias eficientes de aplicação da lei. A análise das propriedades das redes de criminalidade, como a desassortatividade, pode fornecer informações valiosas para a formulação de políticas de segurança pública e para o desenvolvimento de medidas preventivas contra o crime, impondo novas abordagens na identificação de agentes-chave e disrupção destas redes.

3 Trabalhos Relacionados

A evolução da discussão sobre a identificação de agentes-chave e a disrupção de redes de criminalidade está sendo influenciada pelos avanços tecnológicos, particularmente com a análise de dados e a ciência de rede. Pesquisadores e agências de aplicação da lei reconheceram o potencial de abordagens baseadas em dados para entender a estrutura, a dinâmica e a vulnerabilidades das redes de criminalidade. Além disso, a colaboração entre diferentes setores, como a academia, o setor privado e as organizações internacionais, desempenham um papel fundamental no aprimoramento dessas estratégias. O compartilhamento de conhecimentos e recursos entre essas entidades permite uma visão mais abrangente e uma abordagem mais eficaz para combater a criminalidade organizada em escala global. Essa cooperação multidisciplinar traz uma compreensão mais profunda dos métodos de operação dos grupos criminosos e contribui para o desenvolvimento de estratégias mais eficientes de prevenção e repressão criminal.

Em (4), os autores analisam uma complexa rede *multlayers* para identificar os atores centrais envolvidos em uma operação da máfia siciliana. Construíram uma rede *multiplex* criminal única a partir operação Montagna em 2007. Os pesquisadores utilizam medidas da análise de redes *multlayers* para caracterizar os atores na rede com base em suas conexões locais e relevância em cada camada específica. Além disso, medidas de similaridade entre camadas foram usadas para estudar as relações entre as diferentes camadas. Através do estudo da conectividade dos atores e da correlação entre as camadas, os pesquisadores demonstram que uma visualização completa da estrutura e das atividades de uma organização criminosa só pode ser obtida considerando as camadas como uma rede de *multlayers* completa, e não como redes de camada única separadas. O artigo discute a importância de adotar uma abordagem de rede *multlayers* na análise de redes criminais. Ao considerar *multlayers*, é possível obter informações que não estão presentes em cada camada isoladamente, o que fornece uma compreensão mais precisa da estrutura da organização criminosa e da posição estratégica de seus atores. Além disso, o estudo destaca a relevância de coletar dados *multiplex* que incluam diferentes tipos de arestas em uma rede criminosa. A identificação de agentes-chave em uma rede criminosa pode ser de grande importância para os esforços das agências de aplicação da lei em reduzir as capacidades da organização criminosa. As agências podem considerar agentes-chave em diferentes camadas para tomar decisões sobre alvos de vigilância, inteligência e prisão. Os pesquisadores também abordam a questão dos dados ausentes em redes criminais, que são geralmente incompletos e inconsistentes. Eles sugerem que a aplicação de algoritmos de previsão de conexões em uma rede *multlayers* pode ajudar a lidar com esse problema, fornecendo resultados confiáveis e ajudando a compreender melhor a topologia e os limites da rede.

Em (9) é explorada a análise da estrutura de uma rede envolvida em uma operação de tráfico de drogas. O estudo concentrou-se em uma operação de tráfico de drogas específica e utilizou uma combinação de dados e métodos de análise de redes sociais para identificar

os agentes-chave e entender a estrutura da rede. A pesquisa foi motivada pela importância de compreender as dinâmicas complexas das redes de tráfico de drogas, a fim de definir políticas e intervenções mais eficazes. Uma das contribuições chave do trabalho foi a utilização de múltiplos tipos de conexões dentro da rede para analisar a estrutura da operação de tráfico de drogas. Os pesquisadores usaram dados de várias fontes, incluindo registros de prisões, interceptações de comunicações e informações de inteligência, para construir uma representação abrangente da rede. Para analisar a estrutura da rede, os autores utilizaram métricas comumente usadas na análise de redes sociais, como centralidade, densidade e assortatividade. Eles também empregaram técnicas avançadas, como a identificação de comunidades e a análise de papéis estruturais, para identificar subgrupos dentro da rede e entender as funções desempenhadas pelos diferentes atores. Os resultados da análise revelaram a presença de redes dentro da rede principal de tráfico de drogas. Os autores identificaram diferentes tipos de conexões, como conexões financeiras, de comunicação e de transporte, que representavam diferentes aspectos da operação. Essa abordagem permitiu uma compreensão mais completa da estrutura da rede e das interações entre os atores envolvidos. Além disso, os pesquisadores conseguiram identificar os agentes-chave dentro da rede, ou seja, aqueles que desempenhavam um papel central e influente na operação de tráfico de drogas. Esses agentes-chave eram frequentemente responsáveis por conectar diferentes partes da rede e desempenhavam um papel crucial na coordenação das atividades ilícitas. Ao combinar diferentes tipos de conexões e analisar a estrutura da rede de forma holística, os autores puderam obter percepções mais profundos sobre a operação de tráfico de drogas estudada.

Em (10) é explorada a eficácia de diferentes estratégias de disrupção em duas redes reais da máfia, utilizando ferramentas de SNA. O objetivo é investigar como a remoção estratégica de agentes-chave com base em seu capital social e capital humano pode afetar a estrutura dessas redes criminais. As redes criminais apresentam desafios significativos para as agências de aplicação da lei devido à sua resistência à disrupção. Nesse estudo, os pesquisadores aplicam métodos e ferramentas da SNA para analisar a eficácia de sete estratégias de intervenção em duas redes da máfia, com base em um conjunto de dados do mundo real obtido a partir da operação anti-máfia *Montagna* concluída em 2007. O conceito de crime organizado como uma rede, em oposição a uma estrutura hierárquica, tem ganhado destaque na literatura criminológica. A SNA tem sido uma ferramenta fundamental nesse contexto, permitindo a análise dos padrões de interações individuais e revelando as conexões e papéis críticos dentro das redes criminais. Existem dois tipos principais de capital que podem ser considerados em uma rede social: o capital social e o capital humano. O capital social refere-se às conexões entre os indivíduos em uma rede, enquanto o capital humano diz respeito aos atributos pessoais e recursos individuais. As agências de aplicação da lei podem tentar interromper as redes criminais visando tanto o capital social quanto o capital humano. Para avaliar a eficácia das estratégias de disrupção, foram realizadas simulações de remoção de atores com base em diferentes medidas de capital social e capital humano. As estratégias que visam o capital social consideram medidas de centralidade, como a centralidade de grau, de intermediação e de proximidade. As estratégias

que visam o capital humano se concentram em papéis específicos dentro da estrutura da máfia. Os resultados das simulações mostram que a remoção de atores com base no capital social se mostrou a estratégia mais eficaz, levando à disrupção total da rede criminal com o menor número de etapas. Por outro lado, as estratégias baseadas no capital humano foram menos eficazes. É importante mencionar que o estudo apresenta algumas limitações. Os dados de redes criminais são geralmente incompletos, incorretos e inconsistentes devido à natureza ilegal dessas redes. Além disso, a definição dos limites da rede e a transformação dos dados em um grafo dependem das considerações e experiência dos analistas. Ainda assim, o estudo oferece uma abordagem promissora para a disrupção de atividades criminais por meio da análise de redes sociais e pode servir como base para futuras pesquisas e desenvolvimento de estratégias de prevenção e disrupção mais eficazes.

Em (11) é investigada a eficácia das estratégias de disrupção de redes criminais e seus impactos na desarticulação dessas organizações. A disrupção de redes de criminalidade é uma estratégia comumente empregada pelas agências de aplicação da lei para combater o crime organizado. A disrupção de redes envolve a identificação e prisão de agentes-chave da rede, com o objetivo de desmantelar a organização criminosa. Os pesquisadores utilizarm técnicas de modelagem e simulação baseadas em dados reais para analisar a eficácia dessa estratégia. Eles construíram um modelo computacional que representava uma rede criminal fictícia, com base em princípios de teoria de redes complexas e comportamento humano. Ao simular a disrupção de redes criminais em seu modelo, os autores descobriram que essa estratégia não era tão eficaz como se esperava. Embora a prisão de agentes-chave da rede tivesse um impacto inicial na desarticulação da organização, o sistema se adaptava rapidamente, reorganizando-se e mantendo sua funcionalidade. Os resultados mostraram que, apesar das prisões realizadas, a rede criminal continuava a operar, embora com algumas modificações em sua estrutura. Isso sugere que a disrupção de redes criminais pode ter apenas um efeito temporário na redução da atividade criminosa, pois a rede é capaz de se adaptar e se reconfigurar. Os pesquisadores também analisam a influência de diferentes fatores na eficácia da estratégia de disrupção de redes criminais. Eles consideraram variáveis como a posição dos agentes-chave na rede, o tamanho da rede e a conectividade entre os membros. Esses fatores demonstraram ter impacto na resiliência da rede à disrupção. Além disso, o estudo abordou o fenômeno conhecido como “efeito de substituição”. Esse efeito ocorre quando a retirada de um agente-chave da rede resulta em outros membros assumindo suas funções e preenchendo a lacuna deixada pela prisão. Os resultados da simulação mostraram que o efeito de substituição era comum e contribuía para a resiliência da rede.

Em (16) é explorada a dinâmica de ataque, defesa e contágio dentro das redes sociais, com foco nas implicações econômicas e sociais. O estudo aborda a importância das redes em vários contextos, como redes sociais, econômicas e tecnológicas. As interações entre os nós de uma rede podem ser influenciadas por ataques, defesas e a propagação de influência de um nó para outro. Os autores utilizaram modelos teóricos e simulações computacionais para examinar os efeitos de ataques e defesas em redes complexas. Eles consideraram diferentes estratégias de

ataque e defesa, bem como diferentes padrões de contágio. Os resultados do estudo revelaram que a eficácia de ataques e defesas depende fortemente da estrutura da rede. Em redes altamente conectadas, onde os nós têm muitos vizinhos, os ataques são menos eficazes devido à capacidade de propagação da influência. Por outro lado, em redes menos conectadas, os ataques podem ser mais eficazes na disrupção da propagação de influência. Os pesquisadores também identificaram que a estratégia de defesa adotada pelos nós influencia a dinâmica da rede. Eles descobriram que a defesa concentrada em nós altamente influentes pode ter um efeito significativo na redução do contágio e na disrupção da propagação de influência em toda a rede. Além disso, o estudo analisou o papel do contágio na propagação de influência. Os autores mostraram que o contágio pode desempenhar um papel crucial na disseminação de informações e na influência mútua entre os nós da rede. A presença de contágio pode facilitar a propagação de ataques ou defesas, ampliando seu impacto em toda a rede.

Em (36) são explorados os aspectos estratégicos de disrupção e defesa em redes complexas, com foco nas implicações econômicas e teóricas dessas ações. O estudo aborda a importância das redes na economia e em várias áreas de pesquisa. As interações entre os nós de uma rede podem ser afetadas por estratégias de disrupção, que visam perturbar ou desmantelar a rede, e estratégias de defesa, que buscam proteger e manter a integridade da rede. Os autores utilizaram uma abordagem teórica baseada em modelos matemáticos e de teoria de jogos para analisar as estratégias de disrupção e defesa em redes complexas. Eles consideram diferentes formas de poder e influência dos nós da rede, bem como diferentes objetivos e motivações dos atores. Os resultados do estudo mostram que tanto as estratégias de disrupção quanto as de defesa podem ter consequências significativas na dinâmica da rede. A disrupção estratégica pode enfraquecer ou mesmo desmantelar a rede, alterando suas propriedades estruturais e causando impactos econômicos substanciais. Por outro lado, a defesa estratégica pode fortalecer a rede, tornando-a mais resiliente e capaz de resistir a ataques externos. Os autores mostram que, em certos casos, as estratégias defensivas podem até mesmo resultar em uma reorganização da rede, levando a uma estrutura mais eficiente e estável. O estudo também analisa a interação entre as estratégias de disrupção e defesa, destacando que essas ações podem desencadear respostas e contra-respostas dos atores na rede. Isso pode levar a um ciclo contínuo de disrupção e defesa, em que os atores buscam constantemente enfraquecer ou proteger sua posição na rede. Além disso, os autores consideraram a influência das características da rede, como a densidade e a centralidade dos nós, nas estratégias de disrupção e defesa. Eles mostram que a eficácia dessas estratégias pode variar de acordo com a estrutura e as propriedades da rede em questão.

Em (37), discute-se a aplicação da SNA para compreender e desarticular redes ilícitas, incluindo organizações criminosas e terroristas. Essas redes apresentam uma estrutura complexa e operam fora dos limites da lei, representando um desafio significativo para as agências de aplicação da lei. A SNA é uma abordagem poderosa que permite examinar a estrutura e o funcionamento dessas redes, revelando informações valiosas sobre os papéis, posições e interações dos indivíduos envolvidos. Uma revisão da literatura existente revela que a SNA tem sido amplamente utilizada

para analisar e entender redes de criminalidade. Ela oferece uma maneira sistemática de examinar as conexões entre os membros dessas redes, bem como identificar pontos fracos e vulnerabilidades que podem ser explorados para interromper suas atividades. Ao medir a centralidade dos nós e outras métricas de rede, a SNA permite identificar os agentes-chave e sua importância na estrutura da rede. Essas informações são essenciais para desenvolver estratégias eficazes de desarticulação. Um dos desafios enfrentados ao aplicar a SNA em redes de criminalidade é a coleta de dados confiáveis e completos. Muitas vezes, os dados disponíveis são obtidos por meio de investigações policiais, limitando sua abrangência e precisão. Além disso, a natureza clandestina dessas redes dificulta ainda mais a obtenção de informações precisas. No entanto, avanços na coleta de dados, incluindo o uso de dados eletrônicos e técnicas de mineração de dados, têm contribuído para melhorar a disponibilidade e a qualidade dos dados utilizados na SNA. Outro aspecto importante a ser considerado é a resiliência das redes de criminalidade. Essas redes são adaptáveis e têm a capacidade de se reorganizar após sofrerem perturbações. Compreender a resiliência é fundamental para desenvolver estratégias eficazes de desarticulação. A literatura existente discute diferentes abordagens para desarticular redes de criminalidade, incluindo a remoção de agentes-chave, interromper as comunicações e influenciar as dinâmicas sociais dentro da rede. A combinação de abordagens baseadas no capital humano e no capital social tem se mostrado promissora para enfraquecer essas redes e mitigar sua resiliência. O estudo ressalta que a SNA é apenas uma ferramenta na luta contra as redes ilícitas. Ela deve ser complementada por outras estratégias e abordagens, como a inteligência policial e a cooperação internacional. Além disso, a ética e a privacidade devem ser consideradas ao lidar com dados sensíveis e pessoais.

As descobertas destes estudos têm implicações significativas para a aplicação da lei e as políticas de combate ao crime, pois fornecem uma compreensão mais detalhada das redes envolvidas e dos agentes-chave que podem ser alvos de intervenções estratégicas para disrupção das redes de criminalidade.

4 Metodologia

Redes complexas e detecção de *outlier* são duas áreas de estudo que convergem para ciência social computacional e desempenham papéis importantes na análise e na interpretação de dados complexos. As redes complexas fornecem um arcabouço teórico para modelar e compreender as interações entre os elementos de um sistema (55), enquanto a detecção de *outlier* busca identificar pontos de dados anômalos ou incomuns (82).

As redes complexas, por meio da teoria de grafos, fornecem uma representação abstrata para modelar uma ampla variedade de sistemas, nos quais os elementos são representados por nós e as interações entre eles são representadas por arestas, com a topologia da rede, a presença de *hubs* e a distribuição dos graus dos nós podendo afetar a detecção de *outliers* (53). As medidas de SNA são amplamente utilizadas em redes complexas para identificar nós importantes ou influentes (10, 36, 53). Essas medidas também podem ser utilizadas na detecção de *outlier*, identificando nós que se destacam em termos de valores. Assim, pode-se explorar diferentes medidas, como *degree*, *betweenness*, *closeness*, *clustering* (56) e *multiplexity* (68), na detecção de *outlier* em redes complexas

Em ML, os modelos não supervisionados podem desempenhar um papel crucial no contexto de redes complexas e detecção de *outlier* (76). As redes complexas, com suas estruturas intricadas e interações não lineares, muitas vezes requerem técnicas sofisticadas para identificar e analisar *outliers* de forma eficaz (73). ML oferece ferramentas e algoritmos poderosos que podem ser utilizados para enfrentar esse desafio. A detecção de *outlier* em redes complexas envolve identificar comportamentos, padrões ou nós incomuns ou anômalos dentro da rede (84, 82, 34). Modelos de ML podem aprender com as características e propriedades de redes complexas para detectar *outliers* que se desviam do comportamento esperado (75, 80). Esses modelos podem descobrir anomalias sutis e ocultas que podem não ser aparentes por meio de abordagens tradicionais, como SNA (74).

A abordagem inovadora apresentada neste estudo está em utilizar modelos de detecção de *outlier*, por meio de métodos de distância, com o modelos *Outilier Scores* (84) e *SGDOneClass-SVM* (83), de densidade, com o modelo *Isolation Forest* (86), de agrupamento, com o modelo *Local Outlier Factor* (87) e estatístico, com o modelo *Robust Covariance* (88), para identificar *outliers* em redes complexas. Esses modelos analisam as propriedades estruturais das redes e dos indivíduos e detectam padrões ou grupos que diferem dos demais.

Ao integrar os modelos de detecção de *outlier* na análise de redes complexas, podemos aprimorar a detecção de agentes-chave e conseguir uma melhor disrupção das redes. Esses modelos podem aproveitar a estrutura de redes complexas, os padrões de conectividade e as características dos nós para fornecer uma detecção precisa e robusta de *outlier*. A utilização deste cenário (Fig. 2) proporciona novas oportunidades para identificar anomalias em redes complexas, fortalecendo a compreensão e a disrupção das redes de criminalidade.

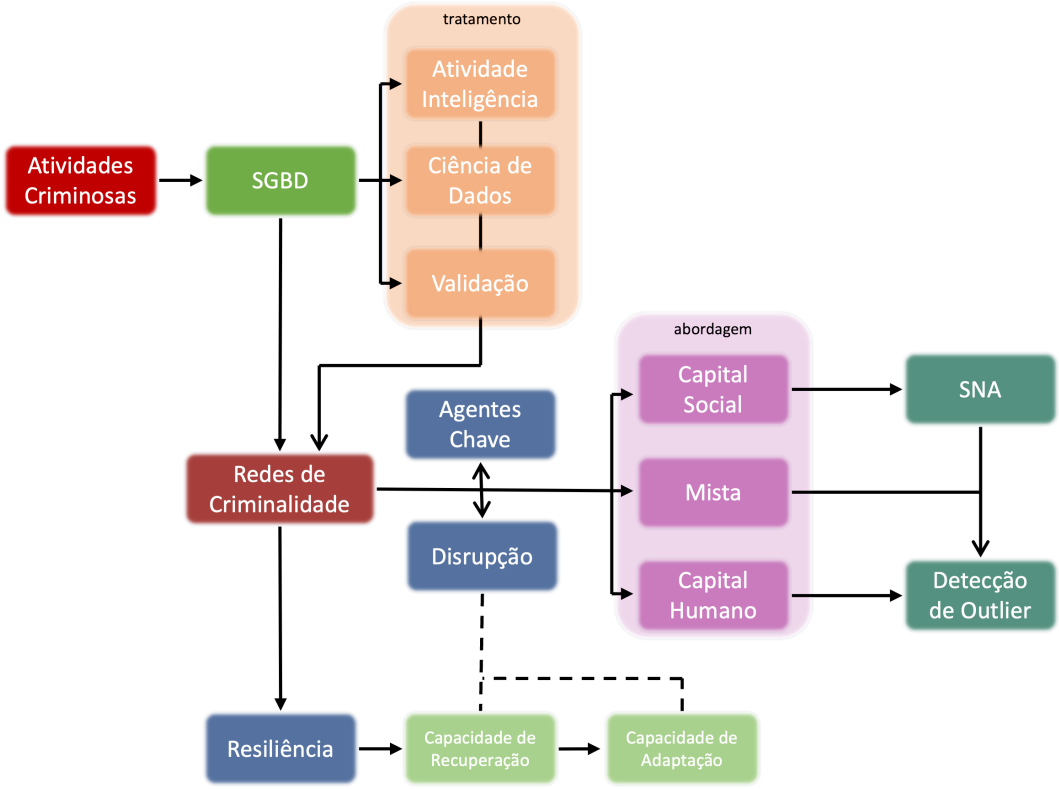


Figura 2 – Cenário conceitual da pesquisa - as atividades criminonas são registradas em bancos de dados. Estes dados permitem a construção de vínculos para visualização e analise de redes de criminalidade. Estes dados recebem tratamentos para melhor entendimento e armazenamento em datasets que serão usados na pesquisa. Este tratamento iniciou com os agentes de inteligência, passando nesta pesquisa por tratamento e validação específicos para comporem os *datasets*. O uso adequado de ferramentas permite a visualização e a análise das redes de criminalidade, que vão ser a base para identificação de agentes-chave e disrupção das redes. Naturalmente, as redes de criminalidade estabelecem mecanismos de resiliência a disrupção, tornando-as mais ou menos vulneráveis, de acordo a capacidade de adaptação e recuperação das abordagens de disrupção. Através de três abordagens, capital social, capital humano e mista, serão utilizadas ferramentas de SNA e modelos de detecção de *outilier* para identificar os agentes-chave e buscar a disrupção mais eficiente.

Neste estudo, todos os algoritmos foram desenvolvidos na linguagem Python (109).

4.1 Dados

Neste estudo os dados foram coletados em dois Sistemas Gerenciadores de Banco de Dados (SGBD) reais de crimes, mantidos pela Polícia Militar de Minas Gerais (PMMG), o "Cadastro de Vívulos de Ocorrências" (CVO) e o "Registro de Eventos de Defesa Social" (REDS), desde suas respectivas implantações, até o ano de 2019, quando foram cedidas cópias dos dois bancos de dados para estudos. Nestes bancos de dados foram armazenadas as atividades criminosas monitoradas pela PMMG no período.

O CVO foi implantado no ano de 2006 com o objetivo de ser um repositório de dados

de vínculos entre criminosos e seus comparsas, contendo 152.973 registros de indivíduos relacionados a diversos crimes. O CVO foi implandado para o trabalho de Inteligência Policial e alimentado durante estes anos por agentes de inteligência em todo estado de Minas Gerais. Ligações entre criminosos correspondem à participação real em crimes, em organização criminosa, interceptação de ligações telefônicas e operações de inteligência deflagradas. O CVO (Fig. 3) está dividido em 7.609 redes de criminalidade, 34.505 nós (indivíduos) e 46.239 arestas (ligações).

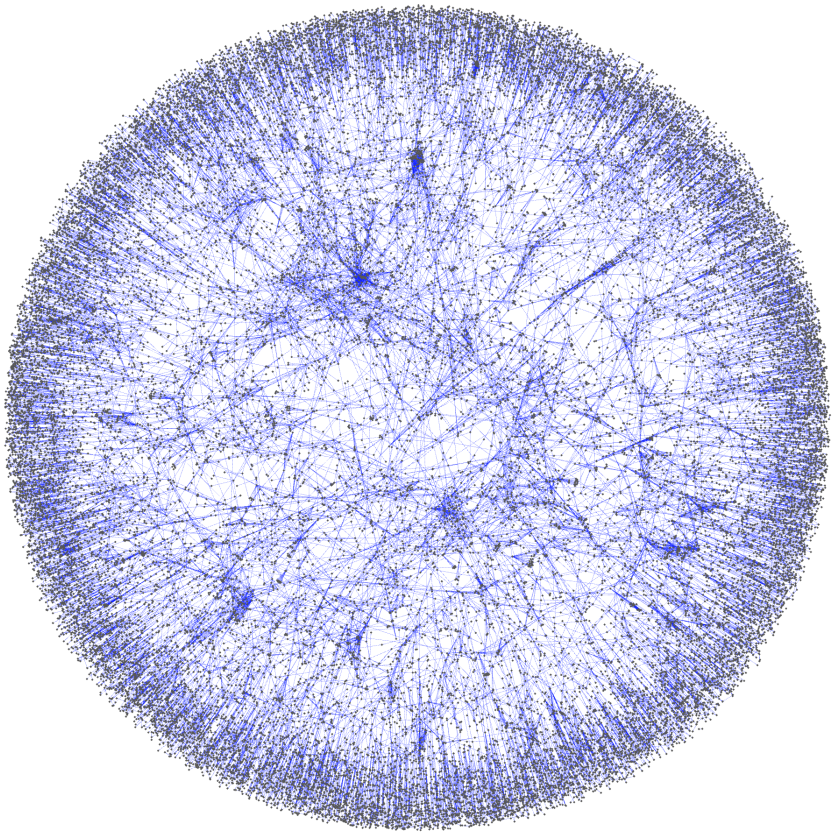
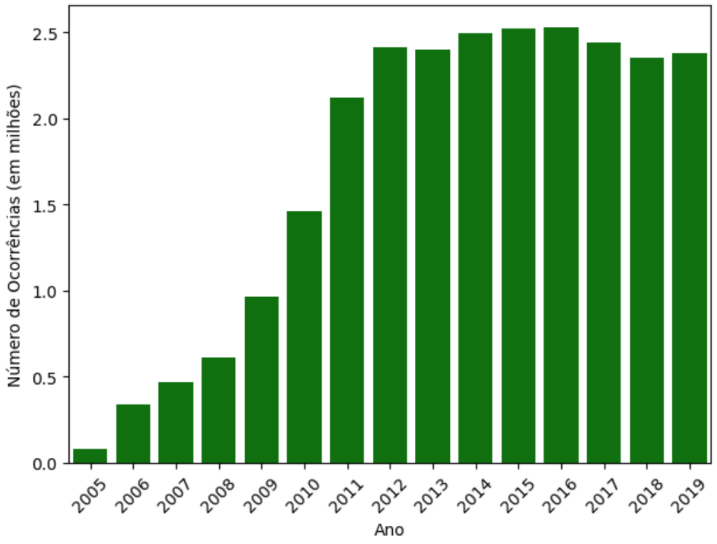


Figura 3 – Componentes de redes de criminalidade contendo 34.505 nós (indivíduos) e 46.239 arestas (ligações) dos diferentes tipos de conexões presentes no CVO. Agentes de inteligência alimentaram esse banco de dados coletando informações reais.

O REDS foi implantado no ano de 2005 com o objetivo de ser um repositório de dados de todas as ocorrências registradas no estado de Minas Gerais, contendo 25.566.017 registros de ocorrências até o ano de 2019 (Fig. 4). Este banco de dados contém dados de crimes, criminosos, vítimas e testemunhas, dentre outros, fundamentais para este estudo.

Este estudo destaca uma das redes de criminalidade que foi alvo de uma operação de inteligência em Belo Horizonte, capital de Minas Gerais e o centro econômico da região, em 2013. O objetivo dessa operação de inteligência foi identificar suspeitos de roubos a joalherias. Naquela época, foi necessário realizar um trabalho árduo de três meses e utilizar recursos humanos e materiais consideráveis para desmantelar uma rede de criminalidade composta por 58 indivíduos e 77 conexões. Apesar do excelente resultado, essa operação gerou um custo significativo para o Estado.



Ano	Número de Ocorrências
2005	82036
2006	336617
2007	470404
2008	610759
2009	961855
2010	1458486
2011	2117781
2012	2411843
2013	2399629
2014	2497135
2015	2521626
2016	2530302
2017	2440276
2018	2351045
2019	2376223

Figura 4 – Desde sua criação, no ano de 2005, até o ano de 2019, foram registradas 25.566.017 ocorrências no REDS. Na imagem podemos visualizar a evolução destes registros. Inicialmente o REDS foi implantado em Belo Horizonte e região metropolitana, capilarizando nos anos seguintes por todo estado de Minas gerais.

Por meio da utilização de algoritmos de mineração, análise e visualização de gráficos, neste estudo foi possível desvendar toda a rede criminosa que causou alarme na região em 2013, gerando manchetes na mídia regional e nacional. Com o auxílio desses algoritmos, além de identificar todos os indivíduos envolvidos na operação de 2013, também foi possível identificar indivíduos adicionais e conexões extras entre eles. Para fins deste estudo, consideramos a rede de criminalidade representada na Fig. 5, chamada de Rede Alvo (RA), que conta com 92 nós e 132 arestas. A RA é uma versão expandida da rede de criminalidade de 2013, resultado dos estudos e algoritimos aqui desenvolvidos. O trabalho que anteriormente levou três meses para ser concluído em 2013, agora pode ser realizado em menos de um minuto utilizando algoritmos e modelos desenvolvidos, o que demonstra eficácia na agilização da identificação de redes de criminalidade e na redução dos custos associados a esse modelo de operação de inteligência em segurança pública.

4.1.1 Tratamento dos dados

Os bancos de dados do CVO e REDS são mantidos na plataforma Microsoft-SQL©. Estes bancos de dados contêm centenas de tabelas com dados de crimes, criminosos, vítimas, testemunhas e vínculos, dentre outros. Todos os dados necessários para este estudo foram filtrados destes bancos e salvos em dois *datasets* primários:

- *dataset redes*: onde estão armazenados os dados de nós (*nodes*) e ligações (*edges*) de todas as 7.609 redes de criminalidade;
- *dataset individuos*: onde estão armazenados os dados com as variáveis (características) de

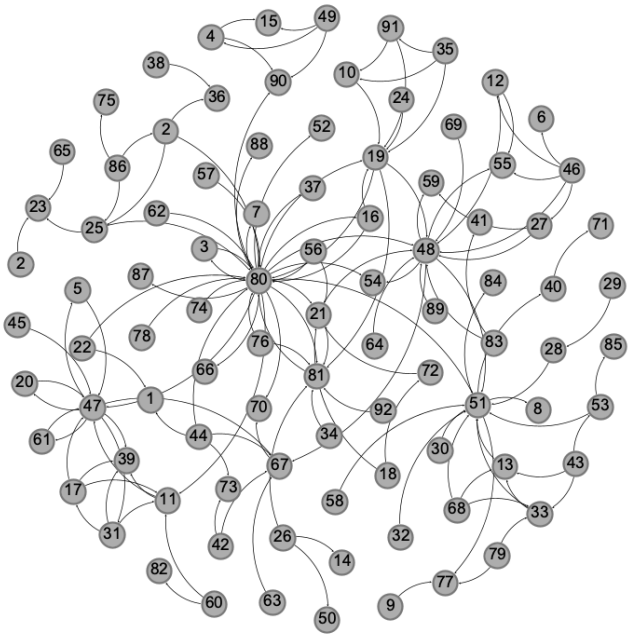


Figura 5 – A RA, uma das examinadas neste estudo, consiste em 92 indivíduos, incluindo criminosos e seus comparsas, com um total de 132 conexões extraídas dos bancos de dados deste estudo. Essa rede desempenha um papel fundamental nas medições, análises, modelagens e simulações realizadas neste trabalho.

todos os 152.973 indivíduos.

As etapas de limpeza e tratamento de dados são fundamentais para garantir a qualidade e a integridade dos conjuntos de dados utilizados em análises e modelos preditivos (110). Essas etapas podem variar dependendo do projeto e do tipo de dados. Neste estudo foram seguidas as seguintes etapas:

1. **Entendimento dos dados**: antes de começar o processo de limpeza e tratamento, foi importante entender os dados. Isso incluiu a compreensão dos tipos de variáveis, dos formatos de dados e do significado de cada campo. Também foi importante identificar quaisquer problemas potenciais nos dados, como valores ausentes, erros de formatação, etc;

2. **Remoção de variáveis para anonimizar os dados**: todas as variáveis que poderiam identificar os indivíduos foram removidas, visando atender a Lei Geral de Proteção de Dados (LGPD) e o requisito de seção e uso dos bancos de dados por parte da PMMG;

3. **Remoção de dados duplicados**: as duplicatas nos dados foram removidas. Dados duplicados podem levar a resultados enviesados e distorcidos em análises;

4. **Tratamento de valores ausentes**: é comum encontrar valores ausentes nos conjuntos de dados. O *dataset redes* não tinha dados ausentes. Os dados ausentes no *dataset indivíduos* foram preenchidos através de pesquisas e filtros recorrentes nas tabelas dos bancos de

dados. Assim não houve necessidade de remover linhas ou colunas, tampouco usar técnicas de imputação de dados (como média e mediana);

5. **Tratamento de erros de digitação e formatação**: foram encontrados erros de digitação e formatação nos dados, como dados escritos de forma inconsistente, datas em formatos diferentes, etc. Estes erros foram corrigidos para garantir a consistência dos dados. Esta foi a etapa mais exaustiva e que demandou mais tempo no tratamento de dados;

6. **Padronização e normalização**: dependendo das características dos dados, pode ser necessário padronizá-los ou normalizá-los. Isso foi feito para garantir que todas as variáveis estivessem na mesma escala e tivessem a mesma importância durante a análise ou modelagem;

7. **Engenharia de recursos**: foi necessário criar novas variáveis a partir das variáveis existentes para melhorar a representação dos dados. Isso envolveu a aplicação de técnicas de engenharia de recursos, como a criação de variáveis *dummy*[1], agregações, transformações matemáticas, etc;

8. **Validação dos dados tratados**: após realizar todas as etapas anteriores, foi importante validar os dados tratados para garantir que eles estivessem em conformidade com as expectativas e os requisitos do projeto. Isso envolveu a verificação de estatísticas básicas, a execução de verificações de consistência e a realização de comparações com fontes de dados originais.

Essas foram as principais etapas envolvidas na limpeza e tratamento de dados. No entanto, é importante ressaltar que cada projeto pode ter requisitos específicos, e o processo pode variar de acordo com as necessidades e características dos dados envolvidos.

Despois do tratamento de dados, o *dataset redes* ficou formatado como uma lista de adjacência [2], contendo duas colunas (origem e destino) do tipo inteiro. Este *dataset* contem os nós e as ligações para os grafos das redes de criminalidade. O Algoritmo 1 demonstra computacionalmente a divisão deste *dataset* em 7.609 componentes (redes de criminalidade).

[1] Variáveis dummy, também conhecidas como variáveis indicadoras, são variáveis binárias que são criadas a partir de variáveis categóricas. Elas são usadas para representar a presença ou ausência de uma determinada categoria. Cada categoria da variável original é mapeada para uma nova variável dummy, onde o valor é 1 se a observação pertence à categoria correspondente e 0 caso contrário. Essas variáveis são úteis em análises estatísticas e modelagem, permitindo que os modelos matemáticos capturem a influência das diferentes categorias de forma adequada. Ao utilizar variáveis dummy, é importante ter em mente a "dummy variable trap" para evitar problemas de multicolinearidade, garantindo a correta representação das categorias no modelo.(110)

[2] Uma lista de adjacências é uma estrutura de dados frequentemente utilizada para representar grafos. Nessa representação, cada vértice do grafo é mapeado para uma lista de seus vizinhos adjacentes. Essa lista pode ser implementada de várias maneiras, como uma lista encadeada, um array ou um dicionário, dependendo das necessidades e características do grafo. Cada elemento da lista de adjacências contém as informações dos vértices vizinhos, permitindo um acesso eficiente às conexões de cada vértice. Essa representação é especialmente útil para grafos esparsos, onde o número de arestas é menor em relação ao número total de vértices. A lista de adjacências é amplamente utilizada em algoritmos de busca, caminhos mais curtos e outros problemas que envolvem a análise e a manipulação de grafos.(53)

Algoritmo 1: Divisão da rede em componentes conectados.

Entrada : Arquivo CSV com lista de adjacência: `dataset_vinculos.csv`

```
adjacency_list ← read_csv('dados/dataset_vinculos.csv');
graph ← edgelist(adjacency_list, source='Source', target='Target');
connected_components ← connected_components(graph);
para i, component en enumerate(connected_components) faça
    component_graph ← graph.subgraph(component);
    component_adjacency_list ← edgelist(component_graph);
    component_adjacency_list.to_csv(component_i + 1.csv', index=False);
fim
```

O *dataset indivíduos* ficou formatado como uma matriz, com o conjunto de variáveis descrito na Tab. 3. Este *dataset* contem dados de 34.505 indivíduos que estão inseridos nas redes de criminalidade.

Tabela 3 – Dicionário de dados do *dataset indivíduos contendo 18 variáveis.*

Variável	**Tipo**	**Descrição**
alcunha	categórico	0 = sem apelido criminoso; 1 = com apelido criminoso
sexo	categórico	0 = feminino; 1 = masculino
pele	categórico	0 = amarela; 1 = branca; 2 = negra; 3 = parda; 4 = outra/ignorada
cabelo	categórico	0 = anelados; 1 = crespos; 2 = lisos; 3 = ignorado
estatura	categórico	$0 \leq$ 1,50m; 1 = 1,50 - 1,60m; 2 = 1,61 - 1,70m; 3 = 1,71 - 1,80; $4 \geq$ 1,80m; 5 = ignorado
tatuagem	categórico	0 = sem tatuagem presidiária; 1 = com tatuagem presidiária
idade	numerico	idade do indivíduo
arma	categórico	0 = sem uso de arma; 1 = com uso de arma
preso	categórico	0 = não foi preso; 1 = foi preso
condenado	categórico	0 = não foi condenado; 1 = foi condenado
faccao	categórico	0 = não pertence; 1 = pertence
estupro	categórico	0 = não praticou; 1 = praticou
extorsao	categórico	0 = não praticou; 1 = praticou
sequestro	categórico	0 = não praticou; 1 = praticou
roubo	categórico	0 = não praticou; 1 = praticou
homicidio	categórico	0 = não praticou; 1 = praticou
trafico arma	categórico	0 = não praticou; 1 = praticou
trafico droga	categórico	0 = não praticou; 1 = praticou

As variáveis do *dataset indivíduos* foram selecionadas por atenderem a Lei Geral de Proteção de Dados (LGPD) e o requisito de seção e uso dos bancos de dados por parte da PMMG. Existem centenas de tipos crimes categorizados nos bancos de dados. Neste estudo foi feita a seleção dos tipos de crime monitorados pela PMMG no "Indicador de Crimes Violentos" (111): estupro, extorsão, sequestro, roubo, homicídio; e também do tráfico de armas e tráfico de drogas (o primeiro por ser um delito que movimenta e mantém o aparato logístico para o cometimento

dos demais crimes e o segundo por ser a principal fonte de arrecadação financeira para redes de criminalidade (1)).

4.1.2 Seleção de variáveis

A seleção de variáveis, *feature selection*, é uma etapa importante no processo de análise de dados, que visa identificar um subconjunto relevante e informativo de características a serem utilizadas em um modelo de ML (112, 113). Embora as técnicas de seleção de variáveis sejam frequentemente aplicadas a algoritmos supervisionados, onde há rótulos associados a cada exemplo de treinamento, também é possível adaptar essas técnicas para algoritmos não supervisionados (78).

Algoritmos não supervisionados não têm rótulos para orientar a seleção de características, tornando o processo mais desafiador. No entanto, existem abordagens que podem ser utilizadas nesse contexto. Uma estratégia comum (114, 115) é utilizar medidas que trazem à tona a importância das variáveis, como a diferença média absoluta, desvio padrão, variância, intervalo interquartil ou correlação, para determinar quais características são mais relevantes. Essas medidas podem ser aplicadas a conjuntos de dados não rotulados, avaliando a dispersão dos dados ou a dependência entre as características.

Outra abordagem é a redução de dimensionalidade, que visa projetar os dados em um espaço de menor dimensão, enquanto preserva as informações mais relevantes . Técnicas como Análise de Componentes Principais, do inglês *Principal Component Analysis* (PCA) (114, 115) podem ser aplicadas para reduzir a dimensionalidade dos dados antes de aplicar um algoritmo não supervisionado. Esta técnica identifica combinações lineares ou não lineares de características que capturam a maior parte da variabilidade dos dados.

A abordagem das medidas de diferença média absoluta, desvio padrão, variância e intervalo interquartil é aplicada para selecionar as variáveis relacionadas aos atributos pessoais do indivíduo, relacionada ao capital humano (Fig. 6).

O mapa de correlação (Fig. 7) , é uma estratégia útil para visualizar as correlações entre variáveis, porém aqui não foi suficiente para destacar a importância das variavéis analisadas. Embora o mapa de correlação possa destacar as relações lineares entre as variáveis, ele não captura necessariamente todas as complexidades e interações presentes nos dados. Além disso, a correlação não implica causalidade, ou seja, uma alta correlação entre duas variáveis não significa necessariamente que uma seja a causa da outra.

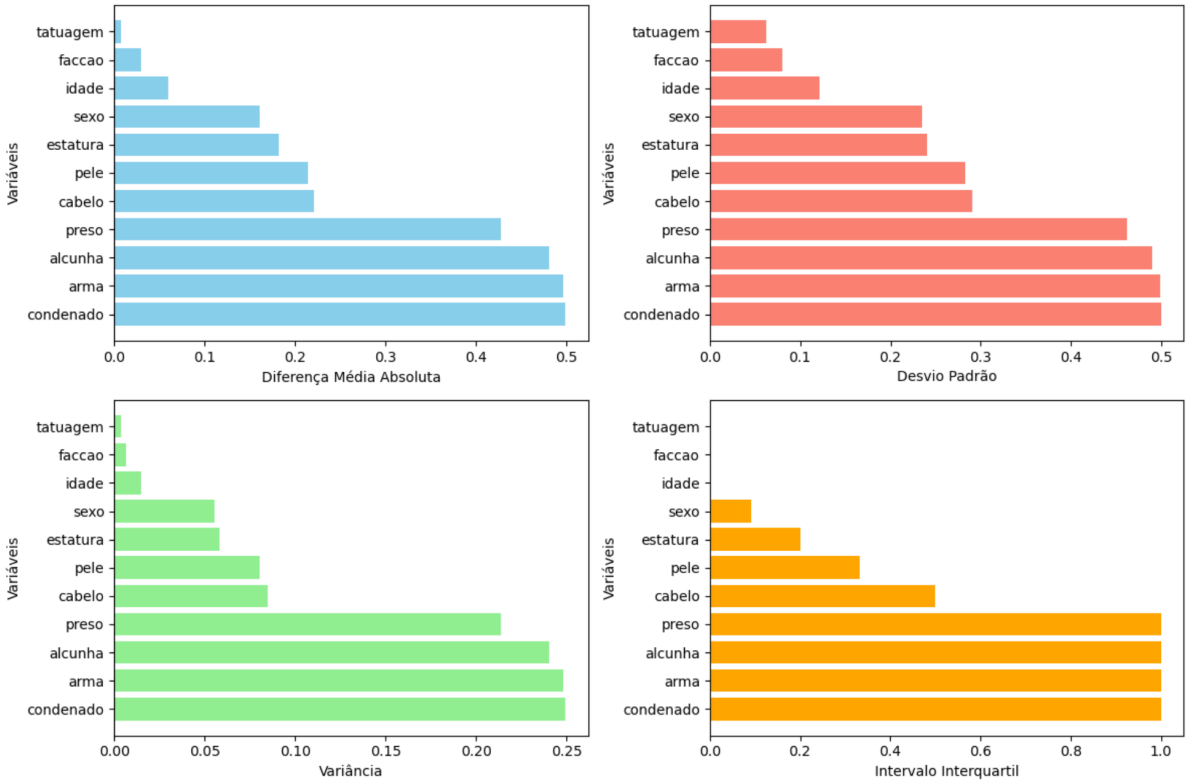


Figura 6 – Medidas para seleção de variáveis.

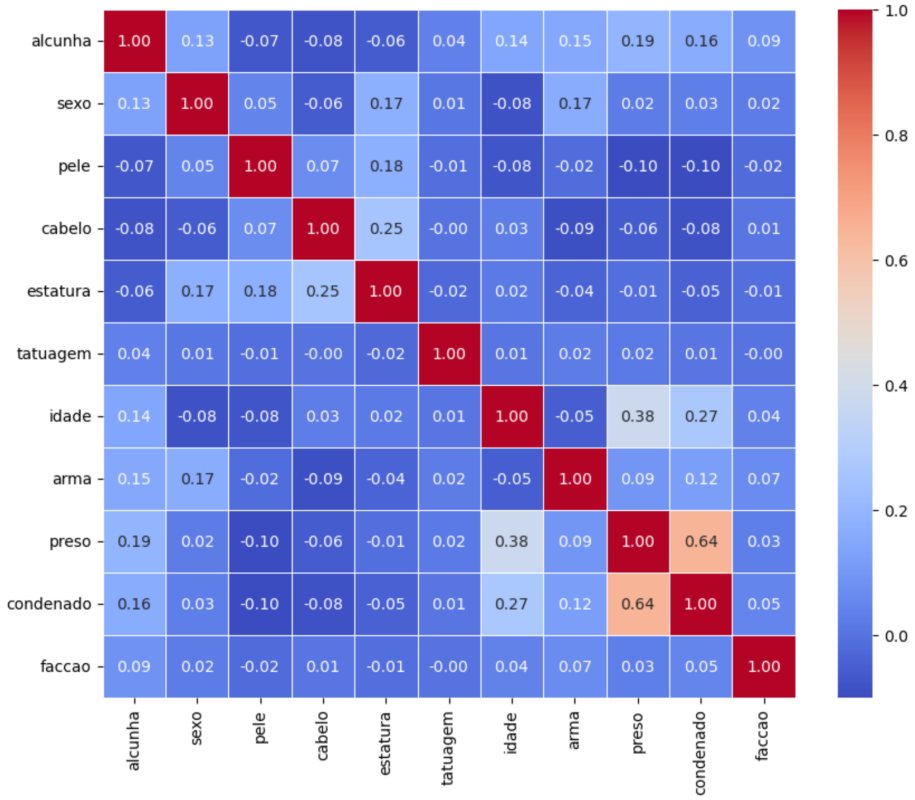


Figura 7 – Mapa de correlação para seleção de variáveis.

O PCA também é utilizado como uma técnica de seleção de variáveis, em conjunto com outras abordagens, sequindo os seguintes passos (52):

1. Normalização dos dados: os dados devem ser normalizados para que todas as variáveis tenham a mesma escala;

2. Aplicação do PCA: com o conjunto de dados normalizado é realizado o PCA, obtendo os componentes principais e seus respectivos autovalores;

3. Avaliação da importância dos componentes: os autovalores associados aos componentes principais são avaliados. Quanto maior o autovalor, maior é a importância desse componente principal na explicação da variabilidade dos dados. Isso pode ser medido pela proporção de variância explicada (PVE) de cada componente principal.

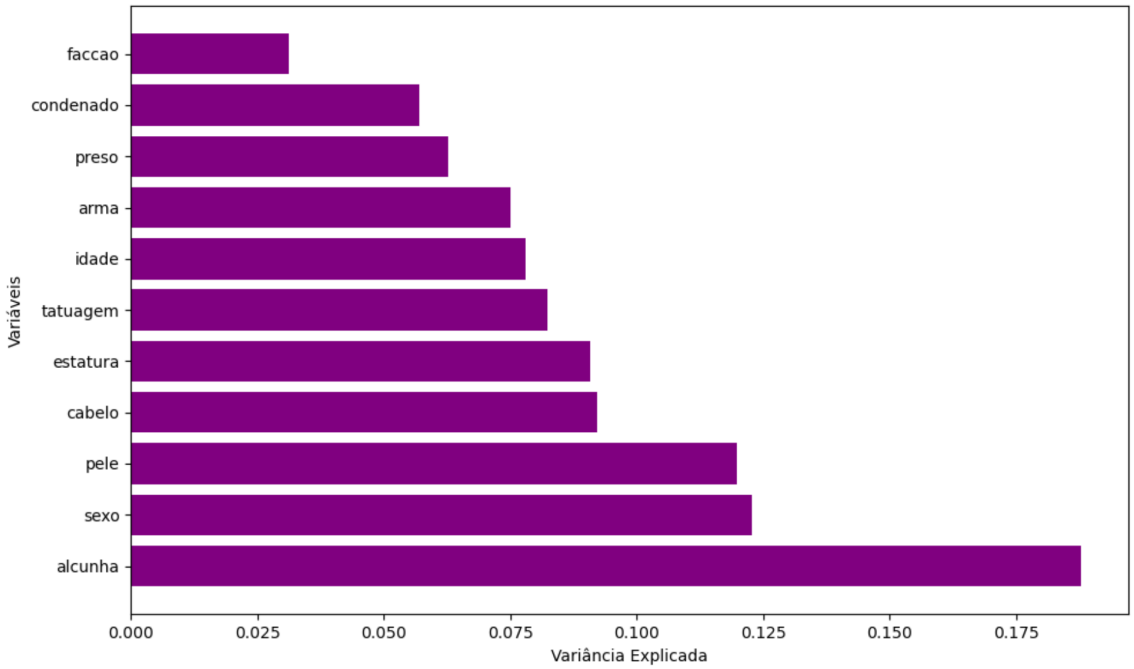


Figura 8 – Variância explicada por variável através do PCA, para selecionar os atributos pessoais mais importantes, relacionados ao capital humano.

Com base na avaliação da PVE é então escolhido um número específico de componentes principais com maior importância. Esses componentes principais selecionados representam os recursos mais relevantes para indicar as variáveis relacionadas aos atributos pessoais do indivíduo, referentes ao capital humano (Fig. 8).

Pelos resultados das medidas de seleção (Fig. 7) e do PCA (Fig. 8), o *dataset indivíduos* ficou com o conjunto de variáveis descrito na Tab. 4. Por meio desta seleção, foi retirada a variável "faccao".

É importante ressaltar que, ao aplicar técnicas de seleção de características em algoritmos não supervisionados, são tomadas decisões baseadas apenas nas características do conjunto de

Tabela 4 – Dicionário de dados do *dataset indivíduos contendo 17 variáveis.*

Variável	**Tipo**	**Descrição**
alcunha	categórico	0 = sem apelido criminoso; 1 = com apelido criminoso
sexo	categórico	0 = feminino; 1 = masculino
pele	categórico	0 = amarela; 1 = branca; 2 = negra; 3 = parda; 4 = outra/ignorada
cabelo	categórico	0 = anelados; 1 = crespos; 2 = lisos; 3 = ignorado
estatura	categórico	0 $\leq$ 1,50m; 1 = 1,50 - 1,60m; 2 = 1,61 - 1,70m; 3 = 1,71 - 1,80; 4 $\geq$ 1,80m; 5 = ignorado
tatuagem	categórico	0 = sem tatuagem presidiária; 1 = com tatuagem presidiária
idade	numerico	idade do indivíduo
arma	categórico	0 = sem uso de arma; 1 = com uso de arma
preso	categórico	0 = não foi preso; 1 = foi preso
condenado	categórico	0 = não foi condenado; 1 = foi condenado
estupro	categórico	0 = não praticou; 1 = praticou
extorsao	categórico	0 = não praticou; 1 = praticou
sequestro	categórico	0 = não praticou; 1 = praticou
roubo	categórico	0 = não praticou; 1 = praticou
homicidio	categórico	0 = não praticou; 1 = praticou
trafico arma	categórico	0 = não praticou; 1 = praticou
trafico droga	categórico	0 = não praticou; 1 = praticou

dados, sem o conhecimento dos rótulos ou da tarefa específica. Portanto, é fundamental avaliar cuidadosamente se a seleção de características é apropriada para o problema em questão e se pode melhorar os resultados do algoritmo não supervisionado.

4.2 Métodos de redes complexas

As redes complexas referem-se a um conjunto de técnicas e métodos que são aplicados ao estudo de sistemas compostos por elementos interconectados, chamados de nós ou vértices, que estão ligados entre si por meio de conexões, chamadas de arestas ou ligações. O foco aqui é explorar e compreender a estrutura e os padrões complexos que existem em redes de criminalidade.

Entre as abordagens mais fascinantes e úteis para entender essas estruturas está a SNA (21). A SNA, por sua vez, é uma subárea específica da análise de redes complexas, focada no estudo das interações sociais entre indivíduos ou organizações. Ela busca entender como as pessoas se relacionam, como as informações fluem através dessas relações e como a estrutura das redes influencia o comportamento coletivo.

A SNA utiliza métodos para identificar e medir diferentes propriedades das redes. Essas medidas fornecem percepções sobre a importância dos nós individuais dentro de uma rede, a densidade de conexões, a existência de subgrupos distintos e outros padrões emergentes (53). Cinco medidas são usados: *degree*, *betweenness*, *closeness*, *clustering* (53, 56) e *multiplexity* (68).

Degree centrality é usada para avaliar a importância de um nó com base no número de conexões que ele possui na rede. Essa medida calcula a quantidade de arestas incidentes em um nó específico em relação ao número máximo possível de conexões para esse nó na rede. É definida como:

$$D_C(i) = \frac{k(i)}{n-1}, \tag{1}$$

em que $k(i)$ é o grau do nó i e n é o número de nós da rede.

Betweenness centrality é utilizada para identificar a importância relativa de um nó na comunicação e fluxo de informações em uma rede. Essa medida é baseada na ideia de que certos nós podem atuar como intermediários chave na comunicação entre outros nós na rede. É definida como:

$$B_C(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}, \tag{2}$$

em que $\sigma_{st}(i)$ é o número de caminhos mais curtos entre s e t que passam pelo nó i; σ_{st} é o número total de caminhos mais curtos entre o nó de origem s e o nó de destino t.

Closeness Centrality é uma medida de centralidade usada para quantificar a importância de um nó com base na sua proximidade dos demais nós da rede. Especificamente, é definida como o inverso da soma das distâncias mais curtas de um nó para todos os outros nós na rede. A equação de Closeness Centrality é dada por

$$C_C(i) = \frac{N-1}{\sum_u d(i,u)}, \tag{3}$$

em que N é o número de nós, $d(i,u)$ é a distância mais curta entre o nó i e o nó u. A soma no denominador é feita para todos os nós u na rede, excluindo o próprio nó i.

Clustering Coefficient é usada para quantificar o grau de agrupamento ou clusterização dos nós em uma rede. Essa medida avalia a probabilidade de dois nós conectados a um nó em comum também estarem conectados entre si. É dada por

$$CC(i) = \frac{2 \cdot E(i)}{k(i) \cdot (k(i)-1)}, \tag{4}$$

em que $E(i)$ é o número de arestas que existem entre os vizinhos diretos do nó i, e $k(i)$ é o número de vizinhos do nó i (ou seja, o número de vizinhos diretos do nó i). Quanto maior o valor de $CC(i)$ para um determinado nó i, maior é o seu coeficiente de agrupamento, indicando que os vizinhos diretos desse nó estão mais interconectados entre si.

Multiplexity se refere à existência de múltiplas camadas ou dimensões em uma rede. Pode-se considerar a existência de diferentes tipos de relações ou interações entre os nós de uma rede, como múltiplos tipos de conexões ou diferentes modos de comunicação. É dada pela equação

$$M(i) = \frac{c}{C}, \tag{5}$$

onde c é o número de camadas das quais o nó participa, e C o número total de camadas da rede. Neste estudo são consideradas 7 camadas, representando os crimes inseridos na Tab. 4.

O Algoritmo 2, tendo como entrada os *datasets redes e indivíduos*, computacionalmente obtém os resultados de cada método de SNA e os salva no *dataset sna*, sendo uma matriz com 34.505 indivíduos e 5 variáveis (Tab. 5).

Algoritmo 2: Cálculo de medidas por rede de criminalidade

```
Entrada : Arquivo CSV do dataset: dataset_redes.csv
para cada component in components faça
    subgraph ← G.subgraph(component);
    subgraph_df ← edgelist(subgraph);
    degree_centrality ← degree_centrality(subgraph);
    betweenness_centrality ← betweenness_centrality(subgraph);
    closeness_centrality ← closeness_centrality(subgraph);
    clustering_coefficient ← clustering(subgraph);
    multiplexity ← multiplexity(subgraf, df_individuos);
    component_df ← pd.DataFrame({'id': list(degree_centrality.keys()),;
                                 'Dc': list(degree_centrality.values()),;
                                 'Bc': list(betweenness_centrality.values()),;
                                 'Cc': list(closeness_centrality.values()),;
                                 'CC': list(clustering_coefficient.values()),;
                                 'M': list(multiplexity)});
    dataset_sna ← dataset_sna.append(component_df);
fim
```

Tabela 5 – Dicionário de dados do *dataset sna* contendo 5 variáveis.

Variável	Tipo	Descrição
Dc	numérico	valor de *Degree Centrality* do nó
Bc	numérico	valor de *Betweennes Centrality* do nó
Cc	numérico	valor de *Closeness Centrality* do nó
CC	numérico	valor de *Clustering Coefficient* do nó
M	numérico	valor de *Multiplexity* do nó

4.3 Modelos de detecção de *outlier*

Modelagem computacional é o processo de desenvolver um modelo que represente um sistema ou um conjunto de observações (116). O modelo é então implementado em um

computador para realizar cálculos mais complexos que não podem ser facilmente realizados manualmente. A modelagem computacional pode incluir técnicas estatísticas, modelos de aprendizado de máquina, e outros métodos para representar dados e sistemas. A ideia por trás da modelagem computacional é simplificar um sistema complexo a um nível que possa ser computacionalmente gerenciado e analisado.

Seis modelos de detecção de *outlier* são apresentados neste estudo, para identificação dos agentes-chave em redes de criminalidade.

4.3.1 *Outlier Scores*

Considerando que um conjunto de N indivíduos (nós) possui vetores associados com M recursos (número de variáveis selecionadas do *dataset*), a distância euclidiana entre dois elementos i e j, cujos vetores de características são $\{f_{i,1} \dots f_{i,M}\}$ e $\{f_{j,1} \dots f_{j,M}\}$ é dada por

$$d_{ij} = \sqrt{\sum_{k=1}^{M} (f_{i,k} - f_{j,k})^2}. \tag{6}$$

Como os recursos são usados apenas nesta primeira etapa, para calcular a matriz d_{ij}, é conveniente normalizar as distâncias de modo que a distância máxima seja 1:

$$d'_{ij} = \frac{d_{ij}}{\max_{i,j}\{d_{i,j}\}}. \tag{7}$$

Usando esta matriz, definimos o primeiro *Outilier Score* do nó i, $\mathrm{OS1}_i$, como a soma de suas distâncias euclidianas normalizadas para todos os outros nós:

$$OS1_i = \sum_{l=1}^{N} d'_{i,l}, \tag{8}$$

($d'_{ii} = 0$). Elementos que possuem altos valores de OS1 possuem, em média, grandes distâncias para outros nós.

Mais informações podem ser obtidas inspecionando não a soma, mas a forma da distribuição das distâncias. Dados dois elementos i e k, se a forma das distribuições de d'_{ij} e d'_{kj} for semelhante, os elementos podem ser considerados semelhantes. Caso contrário, eles são muito diferentes e, portanto, uma distância entre i e k pode ser definida comparando as distribuições dos valores d'_{ij} e d'_{kj}. Diferentes medidas podem ser usadas para comparar as distribuições e aqui foi usada a divergência de Jensen-Shannon (JS).

Especificamente, D_{ik} é a distância entre as distribuições de distância dos elementos i e k, P_i e P_k. Aqui, P_i é a distribuição das distâncias de i para todos os outros nós, $\{d_{i,1} \dots d_{i,j} \dots d_{i,N}\}$, $j \neq i$, e P_k é a distribuição das distâncias de k para todos os outros nós, $\{d_{k,1} \dots d_{k,j} \dots d_{k,N}\}$, $j \neq k$. Para cada par de elementos i e k foi calculada a divergência JS:

$$D_{ij} = JS[P_i,P_k] = H[(P_i + P_j)/2] - H[P_i]/2 - H[P_j]/2, \tag{9}$$

em que $H[P]$ é a entropia de *Shannon*: $H[P] = -\sum_{m=1}^{k} p_m \ln p_m$. O novo conjunto de distâncias também é normalizado tal que $\max_{i,j}\{D_{i,j}\} = 1$. Então, usando a nova matriz de distância, Eq. 9, foi calculado um segundo *Outilier Score* de cada nó i como

$$OS2_i = \sum_{l=1}^{N} D_{i,l}. \tag{10}$$

Todo o procedimento (Fig. 9) pode ser resumido da seguinte forma: comece com o conjunto de N indivíduos e M variáveis, obtenha a $N \times N$ matriz de distâncias euclidianas, Eq. 6, obtenha a distribuição de valores de cada coluna (ou cada linha), calculando o histograma de probabilidade e gerando a $N \times N$ matriz de probabilidade, então calcule a divergência JS para cada par de distribuições usando a Eq. 9, e obtenha a $N \times N$ matriz de distâncias JS.

$$d = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ \dots & \dots & \dots & \dots \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{bmatrix} \Rightarrow \begin{cases} P_1 & = \{p_{1,1} \dots p_{1,k}\}; \sum_{m=1}^{k} p_{1,m} = 1 \\ P_2 & = \{p_{2,1} \dots p_{2,k}\}; \sum_{m=1}^{k} p_{2,m} = 1 \\ \dots & \dots \\ P_N & = \{p_{N,1} \dots p_{N,k}\}; \sum_{m=1}^{k} p_{N,m} = 1 \end{cases} \Rightarrow D = \begin{bmatrix} D_{11} & D_{12} & \dots & D_{1N} \\ D_{21} & D_{22} & \dots & D_{2N} \\ \dots & \dots & \dots & \dots \\ D_{N1} & D_{N2} & \dots & D_{NN} \end{bmatrix}$$

Por fim, somando os valores nas colunas (ou nas linhas) de d e D é calculado o conjunto de *Outlier Scores*, $\{OS1_i\}$ e $\{OS2_i\}$ (Eqs. 8 e 10 e Algoritmo 3).

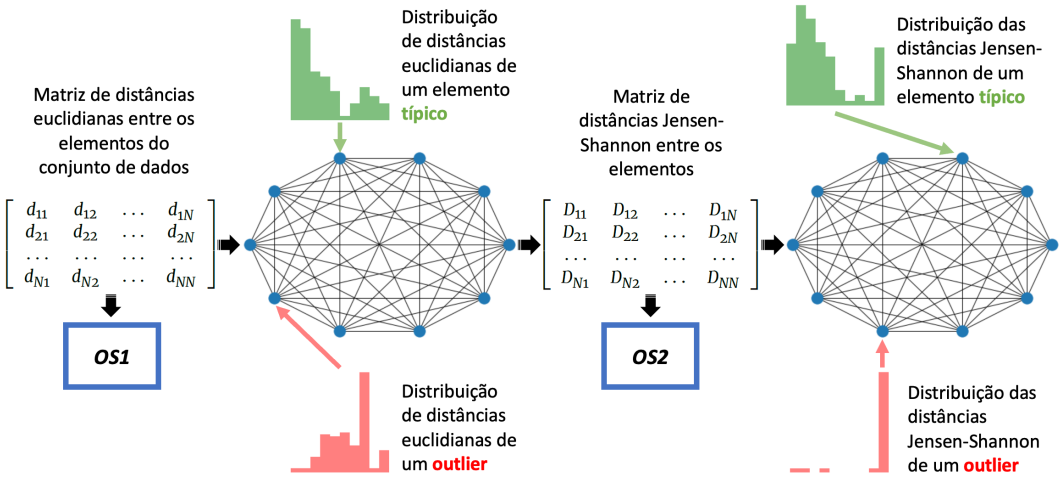


Figura 9 – Representação esquemática do procedimento de atribuição de *Outlier Scores* (OS) a elementos de um banco de dados que são representados por nós de rede. OS1 e OS2 são definidos nas Eqs.(8) e (10).

Algoritmo 3: Detecção de *Outliers Scores*

Entrada : Arquivo CSV do dataset

$data \leftarrow$ `import_data(dataset)`;

$OS1 \leftarrow$ `zeros(len(data), dtype=np.int64)`;

$OS2 \leftarrow$ `zeros(len(data), dtype=np.int64)`;

$matrizfinal \leftarrow$ `limpa_matriz(`$data$`)`;

$m \leftarrow$ `len`(matrizfinal);

$mdisteuc \leftarrow$ `distanciaeuc`(matrizfinal);

$mdisteucnorm \leftarrow mdisteuc/mdisteuc$`.max()`;

$P \leftarrow$ `probabilidade`(mdisteucnorm);

$m \leftarrow$ `len`(P);

$mdistjen \leftarrow$ `distanciajen`(P);

$mdistjennorm \leftarrow mdistjen/mdistjen$`.max()`;

$OS1 \leftarrow mdisteucnorm$`.sum(axis=1)`;

$OS2 \leftarrow mdistjennorm$`.sum(axis=1)`;

4.3.2 *SGDOneClassSVM*

O *SGDOneClassSVM* (SVM) é um algoritmo de ML para detecção de *outilier* que combina os princípios da *One-class Support Vector Machine* ou *One-class SVM* com o *Stochastic Gradient Descent* (SGD) (83, 117, 118).

O primeiro componente, *One-class SVM*, é um algoritmo que é utilizado para detecção de *outlier* (83). Ao contrário do SVM clássico que é um algoritmo de aprendizado supervisionado, o *One-class SVM* é um algoritmo de aprendizado não-supervisionado, o que significa que ele aprende apenas a partir de um conjunto de dados sem rótulos. O objetivo do *One-class SVM* é encontrar uma função que melhor distingue os pontos de dados normais dos anormais (119). Em outras palavras, o algoritmo tenta encontrar uma fronteira de decisão em torno da maioria dos pontos de dados, e qualquer ponto de dados fora dessa fronteira é considerado uma anomalia.

Já o segundo componete, o SGD (118), é um método de otimização iterativo para minimizar uma função objetivo. É particularmente útil quando o conjunto de dados é muito grande para ser processado de uma vez. Em cada iteração, o SGD seleciona aleatoriamente um subconjunto (ou um único exemplo) do conjunto de dados, calcula o gradiente da função objetivo para esse subconjunto e atualiza os parâmetros do modelo.

O *SGDOneClassSVM* combina esses dois conceitos para criar um algoritmo de detecção de anomalias que pode ser treinado de forma eficiente em grandes conjuntos de dados, que funciona da seguinte maneira (83, 119):

1. Primeiro, o algoritmo inicializa um hiperplano aleatório;
2. Em seguida, em cada etapa de treinamento, o algoritmo seleciona um exemplo aleatório do conjunto de dados;

3. O algoritmo então atualiza o hiperplano para se mover na direção que aumenta a margem entre o hiperplano e o exemplo;

4. Este processo é repetido várias vezes, até que o hiperplano não mude significativamente entre as iterações ou até que um número máximo de iterações seja alcançado;

5. Finalmente, o hiperplano resultante é usado como a fronteira de decisão e qualquer ponto de dados que caia fora desta fronteira é considerado uma anomalia.

Ao combinar a *One-class SVM* com o SGD, o *SGDOneClassSVM* pode ser treinado em grandes conjuntos de dados de maneira eficiente, enquanto ainda mantém a capacidade de detectar anomalias que o *One-class SVM* oferece. o Algoritmo 4 desmonstra computacionalmente o modelo.

Algoritmo 4: Detecção de *outlier* com One-Class SVM

Entrada : Contaminação: `best_contamination`, Fit intercept: `fit_intercept`,
Número máximo de iterações: `max_iter`, Tolerância: `tol`,
Embaralhar: `shuffle`, Verbose: `verbose`, Semente aleatória: `random_state`,
Taxa de aprendizado: `learning_rate`, Taxa inicial: `eta0`, Parâmetro power_t: `power_t`,
Continuar treinando: `warm_start`, Média: `average`

```
svm ← SGDOneClassSVM(nu=best_contamination,
 fit_intercept=fit_intercept,
 max_iter=max_iter, tol=tol, shuffle=shuffle, verbose=verbose,
 random_state=random_state, learning_rate='optimal',
 eta0=eta0, power_t=power_t, warm_start=warm_start,
 average=average);
svm.fit(X);
y_pred ← svm.predict(X);
scores_SMV ← svm.decision_function(X);
```

4.3.3 *Isolation Forest*

Isolation Forest (IsF) é um algoritmo de ML eficiente para detecção de *outilier* (86, 120). Ele é baseado em um método bastante diferente dos algoritmos de detecção de anomalias tradicionais (85). Em vez de tentar construir um modelo dos dados normais, *Isolation Forest* tenta isolar os pontos de dados anômalos. A ideia central do *Isolation Forest* é que os pontos de dados anômalos são mais fáceis de "isolar" do que os pontos de dados normais. O *Isolation Forest* funciona da seguinte maneira (120):

1. O algoritmo cria uma "floresta" composta por várias "árvores de isolamento";

2. Cada árvore de isolamento é construída selecionando aleatoriamente um recurso e selecionando aleatoriamente um valor de divisão entre o valor mínimo e o máximo desse recurso;

3. Esse processo é repetido recursivamente até que todos os pontos de dados sejam isolados ou até que um comprimento máximo de caminho seja atingido;

4. Uma vez que a floresta de árvores de isolamento é construída, o algoritmo determina se um ponto de dados é normal ou anômalo calculando a média do comprimento do caminho para isolar esse ponto de dados em todas as árvores na floresta;

5. Pontos de dados anômalos terão, em média, caminhos mais curtos do que os pontos de dados normais;

6. O ponto de dados é identificado como uma anomalia se o comprimento do caminho for inferior a um determinado limite.

Os principais benefícios do *Isolation Forest* são (120):

- Eficiência: A construção da árvore é muito eficiente, pois requer apenas o logaritmo do número de amostras.
- Efetividade: O algoritmo funciona bem em conjuntos de dados de alta dimensão, uma vez que não requer uma estimativa de densidade.
- Menos hiperparâmetros: Ao contrário de muitos algoritmos de detecção de anomalias, *Isolation Forest* tem menos hiperparâmetros para ajustar, tornando-o mais fácil de usar.

Computacionalmente o modelo *Isolation Forest* é definido no Algoritmo 5.

Algoritmo 5: Detecção de *outilier* com Isolation Forest

Entrada : Número de estimadores: `n_estimators`, Tamanho máximo de amostras: `max_samples`,
Contaminação: `best_contamination`, Semente aleatória: `random_state`

```
isf ← IsolationForest(n_estimators=n_estimators,
 max_samples=max_samples,
 contamination=best_contamination, random_state=random_state,
 verbose=0);
isf.fit(X);
y_pred ← isf.predict(X);
scores_isf ← isf.decision_function(X);
```

4.3.4 *Local Outlier Factor*

O *Local Outlier Factor* (LOF) é um algoritmo de ML para detecção de anomalias (87). Ele trabalha em cima do princípio da densidade local, sendo capaz de identificar pontos de dados que possuem uma densidade substancialmente inferior à de seus vizinhos (78). Dessa forma, ao contrário de muitos algoritmos que modelam as "normalidades" dos dados, o LOF é projetado especificamente para fornecer uma pontuação indicando o grau de anormalidade dos dados.

O funcionamento do LOF pode ser detalhado nos seguintes passos (121):

1. Para cada ponto de dados, o LOF calcula a distância para seus vizinhos mais próximos. Essa quantidade de vizinhos é um parâmetro definido pelo usuário, normalmente denominado k;

2. Em seguida, para cada par de pontos de dados, o algoritmo calcula a chamada *Reachability Distance*, que é definida como a distância máxima entre dois pontos para que um seja alcançado pelo outro;

3. O LOF calcula a densidade local de um ponto como a inversa da média das *Reachability Distance* de seus vizinhos k mais próximos;

4. Finalmente, para cada ponto de dados, o LOF é calculado como a média da razão entre a densidade local de um ponto e a densidade local de seus vizinhos.

Uma grande vantagem do LOF é que ele leva em consideração as variações na densidade dos dados (121). Em regiões de alta densidade, um ponto de dados deve ter uma densidade semelhante à dos seus vizinhos para ser considerado normal. Por outro lado, em regiões de baixa densidade, um ponto de dados pode ter uma densidade muito menor que a dos seus vizinhos sem ser considerado uma anomalia. O modelo é computacionalmente descrito no Algoritmo 6.

Algoritmo 6: Detecção de *outlier* com Local Outlier Factor

Entrada : Número de vizinhos: `n_neighbors`, Algoritmo: `algorithm`, Tamanho da folha: `leaf_size`,
Métrica: `metric`, Valor de p: `p`, Parâmetros de métrica: `metric_params`,
Contaminação: `best_contamination`, Novidade: `novelty`

```
lof ← LocalOutlierFactor(n_neighbors=n_neighbors,
 algorithm=algorithm,
 leaf_size=leaf_size, metric=metric, p=p,
 metric_params=metric_params,
 contamination=best_contamination, novelty=True);
lof.fit(X);
y_pred ← lof.predict(X);
scores_Lof ← lof.decision_function(X);
```

4.3.5 *Robust Covariance*

O *Robust Covariance* (COV) é uma técnica estatística para detecção de *outilier* (88) que faz uso do conceito de estimativa de covariância robusta, geralmente representada por *Minimum Covariance Determinant* (MCD) (122). A ideia básica é modelar os dados usando uma distribuição Gaussiana multivariada e, em seguida, identificar como anomalias os pontos de dados que não se ajustam bem a esse modelo.

Estes são os passos de funcionamento do algoritmo (123):

1. O algoritmo começa calculando a estimativa de covariância robusta dos dados. A estimativa de covariância robusta é uma medida da dependência entre variáveis que é resistente a *outliers*. O método comum usado para isso é o MCD, que seleciona o subconjunto de observações com o menor determinante de covariância;

2. Uma vez que a matriz de covariância robusta é calculada, a distribuição dos dados é modelada usando uma distribuição Gaussiana multivariada. A média e a covariância desta distribuição são calculadas usando a matriz de covariância robusta;

3. Finalmente, o algoritmo identifica anomalias com base no quão bem cada ponto de dados se ajusta à distribuição Gaussiana modelada. Pontos de dados que estão a muitos desvios padrão de distância da média (ou seja, pontos de dados que têm uma alta "distância de *Mahalanobis*"[3]) são identificados como anomalias.

Uma grande vantagem do *Robust Covariance* como uma técnica de detecção de *outilier* é que ele leva em consideração as relações entre diferentes características dos dados (122). Ele pode ser mais eficaz do que técnicas univariadas para identificar anomalias em dados multivariados. O Algoritimo 7 descreve computacionalmente o modelo.

Algoritmo 7: Detecção de *outlier* com Elliptic Envelope

Entrada : Contaminação: `best_contamination`, Semente aleatória: `random_state`

```
cov ← EllipticEnvelope(contamination=best_contamination,
 random_state=random_state);
cov.fit(X);
y_pred ← cov.predict(X);
scores_Cov ← cov.decision_function(X);
```

4.3.6 Hiperparâmetros

Hiperparâmetros em ML são os parâmetros que são definidos antes do processo de aplicar os dados ao modelo e determinam a estrutura e o comportamento do algoritmo. Eles

[3] A distância de Mahalanobis é uma medida de distância multivariada que considera as correlações entre as variáveis de um conjunto de dados. A distância de Mahalanobis ajusta a escala de cada variável e a correlação entre elas, proporcionando uma medida mais precisa em cenários onde as variáveis são interdependentes. (123)

desempenham um papel fundamental em algoritmos não supervisionados de detecção de *outliers*, influenciando a forma como o modelo aprende a partir dos dados e, finalmente, o seu desempenho (78).

Pontos importantes sobre os hiperparâmetros em algoritmos não supervisionados de detecção de outliers (123):

- Hiperparâmetros podem influenciar a complexidade do modelo e sua capacidade de se ajustar aos dados;
- Alguns hiperparâmetros têm uma influência direta sobre como o algoritmo lida com outliers;
- A escolha dos hiperparâmetros também pode ajudar a equilibrar o viés e a variância do modelo. Hiperparâmetros que levam a um modelo muito complexo podem resultar em sobreajuste (alta variância), enquanto hiperparâmetros que levam a um modelo muito simples podem resultar em subajuste (alto viés);
- Ajustar os hiperparâmetros corretamente pode melhorar significativamente o desempenho do modelo.

É importante ressaltar que, apesar da importância dos hiperparâmetros, eles devem ser usados com cuidado. Escolhas inadequadas podem levar a resultados indesejáveis, como sobreajuste ou subajuste (93). Portanto, o entendimento do domínio do problema e a experiência em ML são fundamentais ao ajustar hiperparâmetros.

A "taxa de contaminação" é um hiperparâmetro importante usado em algoritmos de detecção de *outliers* (85). Esta taxa é utilizada para definir a proporção de outliers esperada na base de dados. A taxa de contaminação ideal pode variar de acordo com o conjunto de dados e o domínio do problema (93). A definição desse hiperparâmetro pode requerer experimentação e validação. Existem abordagens que podem ser utilizados para definir a taxa de contaminação em modelos não supervisionados para detecção de *outliers*. Aqui estão algumas abordagens mais comuns (93):

- Baseada em limiar: define-se um limite (*threshold*) para considerar um ponto como um outlier. A taxa de contaminação pode ser definida como a proporção de pontos que excedem esse limite. No entanto, a escolha do limiar é muitas vezes subjetiva e pode variar dependendo do conjunto de dados;
- Estimativa percentual: define-se a taxa de contaminação como uma porcentagem fixa do total de pontos no conjunto de dados. Esta abordagem é, portanto, estimada, sem nenhum método para isto;
- Estimação de densidade: uma abordagem comum é utilizar métodos de estimativa de densidade, como o método de mistura de Gaussianas. Nesse caso, a taxa de contaminação representa a proporção de componentes da mistura que são considerados como *outliers*;

- Abordagem iterativa: essa abordagem envolve ajustar o modelo inicialmente com uma taxa de contaminação alta e, em seguida, reduzir gradualmente essa taxa até obter um resultado satisfatório. Isso pode ser feito por meio de um processo iterativo de ajuste do modelo e avaliação dos resultados.

Neste estudo foi utilizada a abordagem iterativa e a estimação de densidade através de mistura de Gaussianas (124) para definir a "taxa de contaminação" do *dataset indivíduos*. O Algoritmo 8 traz a representação computacional para definição da "taxa de contaminação".

Por este algoritimo, a melhor taxa de contaminação a ser utilizada na detecção de *outlier* é de 0,099.

Algoritmo 8: Definição da taxa de contaminação

```
Entrada : Número de componentes: n_components, Tipo de covariância:
          covariance_type
gm ← GaussianMixture(n_components=n_components,
 covariance_type=covariance_type);
gm.fit(X_train);
densities ← gm.score_samples(X_train);
sorted_densities ← np.sort(densities);
contamination ← 0.3;
min_contamination ← 0.001;
contamination_step ← 0.001;
best_score ← float('-inf');
best_contamination ← contamination;
anomaly_cutoff_indices ←
 np.floor(np.arange(min_contamination, contamination +
 contamination_step, contamination_step) × len(X_train)).astype(int);
anomaly_cutoffs ← sorted_densities[anomaly_cutoff_indices];
para i en range(len(anomaly_cutoffs)) faça
    anomaly_cutoff ← anomaly_cutoffs[i];
    contamination− = contamination_step;
    contamination ← max(contamination, min_contamination);
    model ←
     IsolationForest(contamination=contamination, random_state=42);
    model.fit(X_train);
    train_predictions ← model.predict(X_train);
    num_anomalies ← np.count_nonzero(train_predictions == −1);
    se num_anomalies > best_score então
        best_score ← num_anomalies;
        best_contamination ← contamination;
    fim
fim
print("Melhor taxa de contaminação:", best_contamination);
```

4.4 Modelos de disrupção

Neste estudo foram simuladas as intervenções que as agências da lei costumam realizar para interromper e desmantelar redes de criminalidade, removendo um nó e todas as arestas incidentes. Os nós são selecionados de acordo com seu capital humano, social ou misto, seguindo critérios discutidos em detalhes na sequência. Os nós são ordenados de forma decrescente de valores (do maior para o menor), obtidos como resultado de cada modelo de detecção de *outlier*. Cada uma dessas intervenções é simulada por um modelo escolhido, que começa com as redes completas. A cada etapa é excluído um nó de acordo com o modelo específico e medido o número de componentes, a eficiência global média e o tamanho do maior componente conectado, e o próximo nó assume as arestas do nó removido (resiliência da rede). Para cada intervenção, a simulação para quando a rede é totalmente interrompida, ou seja, quando não restam nós conectados.

Os resultados de disrupção vão orientar a escolha do modelo mais eficaz para identificação dos agentes-chave (*outiliers*), por meio do número de componentes, da eficiência global média e do tamanho do maior componente conectado. O maior número de componentes na rede de criminalidade indica uma estrutura mais fragmentada e descentralizada, o que pode criar oportunidades para desestabilizar e interromper a rede de maneira mais eficiente. Um maior número de componentes pode contribuir para a identificação de uma estratégia de disrupção (9, 57, 125):

- Alvos mais acessíveis: com mais componentes na rede, é provável que haja vários grupos menores e isolados de criminosos. Esses grupos menores podem ser alvos mais acessíveis para a aplicação da lei, pois são menos organizados e provavelmente têm menos recursos para resistir à pressão das agências da lei;
- Impacto em cascata: o desmantelamento de um componente conectado pode ter um efeito cascata na rede. Ao interromper um grupo ou um conjunto de conexões importantes, é possível que outros componentes também sejam afetados, enfraquecendo ainda mais a estrutura da rede como um todo;
- Identificação de agentes-chave: um maior número de componentes pode ajudar a identificar nós cruciais na rede. Esses nós chave podem ser indivíduos ou grupos que atuam como pontes entre diferentes componentes ou que têm influência significativa na coordenação das atividades criminosas. Identificar e desativar esses nós pode ter um impacto significativo na desorganização da rede;
- Análise de padrões e comportamentos: com mais componentes, é possível observar melhor os padrões de comportamento e as interações entre os membros da rede. Isso pode fornecer informações sobre como a rede funciona, como as informações são disseminadas e como as decisões são tomadas. Essas informações podem ser úteis para desenvolver estratégias de interrupção mais eficazes;

- Identificação de vulnerabilidades: um maior número de componentes também pode revelar vulnerabilidades na rede. Pode haver componentes que estão altamente dependentes de outros indivíduos para coordenar suas atividades. Identificar essas dependências pode ajudar a direcionar esforços para desestabilizar esses pontos fracos;

- Segmentação de esforços: com mais componentes, é possível segmentar os esforços de interrupção em diferentes áreas da rede. Em vez de se concentrar apenas nos elementos mais proeminentes ou centrais da rede, a abordagem pode ser mais distribuída para afetar várias partes da rede.

A eficiência global média, também conhecida como coeficiente de eficiência global, é dada pela equação da eficiência (126), que calcula a média inversa das distâncias mais curtas entre todos os pares de nós distintos ($i \neq j$) da rede, normalizada pelo número total de pares de nós ($N(N-1)$),

$$E(G) = \frac{1}{N(N-1)} \sum_{\substack{i \neq j \\ i,j \in G}} \frac{1}{d_{ij}}, \tag{11}$$

em que

- $E(G)$ representa a eficiência global média da rede G;

- N é o número total de nós (ou vértices) na rede;

- d_{ij} é a distância mais curta entre o nó i e o nó j. A distância é medida em termos de comprimento do caminho mais curto (ou número de arestas) entre esses nós na rede.

Essa equação é útil para avaliar a eficiência de redes complexas e compreender como a estrutura da rede influencia o fluxo de informações ou recursos (capital) através dela.

O tamanho do maior componente conectado refere-se ao grupo mais extenso de nós (indivíduos ou organizações) interconectados, ou seja, é a maior parte da rede onde a maioria dos membros pode se comunicar ou colaborar diretamente. A importância do tamanho do maior componente conectado na disrupção de uma rede de criminalidade pode ser explicada por diversos fatores (9, 57, 125):

- Eficiência da comunicação: em um grande componente conectado, os membros têm uma maior capacidade de comunicação e compartilhamento de informações. Isso permite que a rede opere de forma mais eficiente, coordenando suas atividades criminosas com maior rapidez e facilidade;

- Resiliência e estabilidade: quanto maior o maior componente conectado, mais resiliente e estável é a rede. A remoção ou prisão de alguns membros pode não afetar significativamente a capacidade da rede de operar. Portanto, a disrupção de uma rede de criminalidade requer mais esforços e recursos quando o maior componente conectado é grande;

- Centralidade e controle: os nós mais importantes ou influentes geralmente estão no maior componente conectado, e eles desempenham um papel fundamental na coordenação das atividades criminosas. Se esses nós-chave forem identificados e alvos de ações de combate ao crime, a rede pode sofrer um golpe significativo;

- Propagação de informações e influência: em redes de criminalidade, a informação e a influência podem se espalhar rapidamente através das conexões. Um grande componente conectado facilita a disseminação de informações e estratégias, tornando a rede mais ágil e adaptável a mudanças;

- Potencial de expansão: redes criminosas com um maior componente conectado têm maior capacidade de expansão e recrutamento de novos membros. Isso pode levar a um crescimento mais rápido da rede, tornando-a mais difícil de controlar se não for interrompida em estágios iniciais.

O número de componentes conectados, a eficiência global média e o tamanho do maior componente conectado são métricas relevantes para a identificação de uma estratégia de disrupção. O ponto de disrupção é atingido quando a rede é dividida em mais componentes e, consequentemente, há um declínio na eficiência global média e no tamanho do maior componente conectado, os dois últimos avaliados pelas derivadas negativas sucessivas das curvas do processo.

Esta análise em conjunto com as estratégias de capital humano, social e misto são essenciais para uma compreensão abrangente da rede e a identificação da estratégia mais eficaz de disrupção.

A abordagem de disrupção pelo capital humano centra em utilizar as variáveis do *dataset indivíduos* (Tab. 4), que contém a cadeia de valores e crimes, representando a acumulação do capital humano (o conhecimento e características) de cada indivíduo.

A disrupção pelo capital social visa posições estratégicas dentro das redes de criminalidade. Para isto são utilizadas as variáveis do *dataset SNA* (Tab. 5), que são as medidas de *Degree Centrality*, *Betweennes Centrality*, *Closeness Centrality*, *Clustering Coefficient* e *Multiplexity*.

A disrupção mista consiste em utilizar o capital humano e o capital social, utilizando as variáveis do *dataset indivíduos* (Tab. 4) e do *dataset SNA* (Tab. 5).

Seis modelos são utilizadas para cada abordagem de disrupção: *Outlier Scores (OS1 e OS2), SGDOneClassSVM, Isolation Forest, Local Outlier Factor e Robust Covariance*. Assim, três abordagens de disrupção são usadas: disrupção pelo capital humano, disrupção pelo capital social e disrupção mista, para um total de seis modelos de detecção de agentes-chave, totalizando então 18 simulações por rede de criminalidade, utilizando as métricas de número de componentes, a eficiência global média e o tamanho do maior componente conectado. O Algoritmo 9 descreve computacionalmente como isto é feito.

Algoritmo 9: Disrupção de redes de criminalidade

Entrada : Arquivos CSV da rede e da estratégia de disrupção:
`edges_crime_network.csv; key.csv`

Saída : passos, número de componentes, tamanho do maior componente e eficiência global

```
df_edges ← read_csv('edges_crime_network.csv', sep=';');
G ← criar_grafo(df_edges);
df_key ← read_csv('key.csv', sep=';');
key ← array(df_key.iloc[:, 0]);
steps ← [];
components ← [];
size_largest_component ← [];
global_efficiency ← [];
total_nodes ← len(G.nodes());
steps.append(total_nodes− len(G.nodes()));
num_components, largest_component ← components_graph(G);
components.append(num_components);
size_largest_component.append(largest_component);
global_efficiency.append(global_efficiency(G));
n ← 0;
H ← G.copy();
para i en range(len(G)) faça
    new_G ← remodel_node(H, key[n]);
    num_components, largest_component ← components_graph(new_G);
    components.append(num_components);
    size_largest_component.append(largest_component);
    global_efficiency.append(global_efficiency(new_G));
    df_edges ← substituir_valores(df_edges, key, n);
    df_edges ← deletar_valores_iguais(df_edges);
    H ← criar_grafo(df_edges);
    n ← n + 1;
    steps.append(total_nodes− len(H.nodes()));
fim
```

5 Resultados

O impacto socioeconômico das redes de criminalidade tem chamado a atenção de pesquisadores para o desenvolvimento de novos modelos para fazer previsões mais precisas para identificar agentes-chave, analisar e interromper redes de criminalidade. A partir dos modelos propostos, foram gerados resultados para reproduzir os fatos observados.

Neste trabalho foram analisadas 14 redes de criminalidade do CVO que contém mais de 50 indivíduos (nós) e são desassortativas, a RA e outras 13.

5.1 Deteção de agentes-chave

Detectar agentes-chave em redes criminosas é um desafio complexo que envolve a análise de dados para identificar indivíduos ou entidades que desempenham papéis cruciais em atividades ilegais. Para abordar essa questão, foram utilizados os modelos de detecção de *outlier*: *Outlier Score 1* (OS1), *Outlier Score 2* (OS2), *Isolation Forest* (IsF), *Local Outlier Factor* (LOF), *Robust Covariance* (COV) e *SGDOneClassSVM* (SVM), em combinação com as abordagens de capital humano, capital social e capital misto.

Os resultados numéricos obtidos por meio desses modelos foram fundamentais para a identificação de agentes-chave em redes criminosas, uma vez que os maiores escores de *outlier* apontam para esses indivíduos.

Os resultados numéricos dos modelos para RA estão nos Apêndices A, B e C.

A detecção de *outlier* é uma técnica fundamental nesse contexto. Ela envolve a análise dos dados para identificar os elementos que se desviam significativamente do comportamento típico da rede. Os maiores escores de *outlier* indicam a presença de agentes-chave. Isso ocorre porque esses indivíduos costumam ter características únicas, se destacando tanto em termos de capital humano quanto social.

5.2 Disrupção de redes de criminalidade

A identificação do modelo mais eficaz para as abordagens propostas e consequentemente a identificação dos agentes-chave será orientada pelo melhor método de disrupção, utilizando o número de componentes, a eficiência global média e o tamanho do maior componente conectado.

A partir da RA as simulações se estenderam por outras 13 redes. A cada etapa, um ator foi removido de acordo com a abordagem do capital humano, social ou misto e os modelos de detecção de *outlier*. Como resultado foram obtidas as métricas de número de componentes conectados, a eficiência global média e o tamanho do maior componente conectado.

Para facilitar as comparações entre as estratégias de disrupção descritas na seção 4.4, foi

feito um gráfico para cada uma das três métricas de resultado, unidas em uma figura: número de componentes, a eficiência global média e o tamanho do maior componente conectado. Para cada gráfico, o eixo x mostra o número de etapas executadas.

Para a RA, a abordagem do capital humano é vista na Fig. 11, a abordagem do capital social na Fig. 12 e a abordagem do capital misto na Fig. 13.

Os resultados mostram melhor eficiência na abordagem do capital **social** com o modelo ***Outilier Score 1*** (OS1), capaz de aumentar o número de componentes (37), diminuir a eficiência da rede (0,02735) e o tamanho do maior componente conectado (9) na etapa de disrupção **6** (Fig. 10). Este resultado da abordagem do capital social com o modelo *Outilier Score 1* (OS1) foi identificado como mais eficaz nas outras 13 redes de criminalidade.

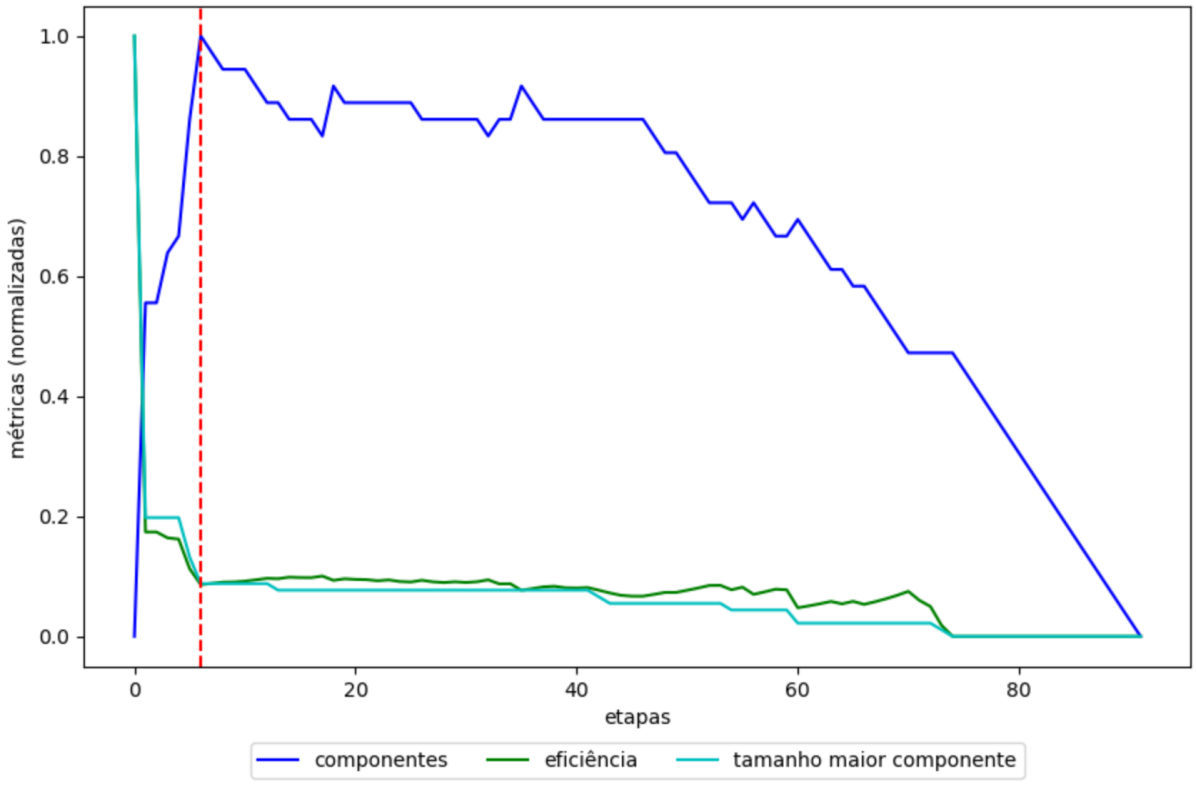


Figura 10 – RA - a disrupção pelo capital social com o modelo OS1, capaz de aumentar o número de componentes, diminuir a eficiência da rede e o tamanho do maior componente conectado com menor número de etapas de disrupção, é o resultado mais eficiente deste estudo. A linha vertical representa a etapa de disrupção.

Conforme discutido na seção 4, toda rede de criminalidade estabelece mecanismos de resiliência a disrupção. Assim, a identificação da etapa de disrupção pela métrica de números de componentes é considerada quando atinge o maior patamar. Na sequência, ocorre o decréscimo desta métrica, o que identifica que o mecanismo de resiliência da rede foi interrompido.

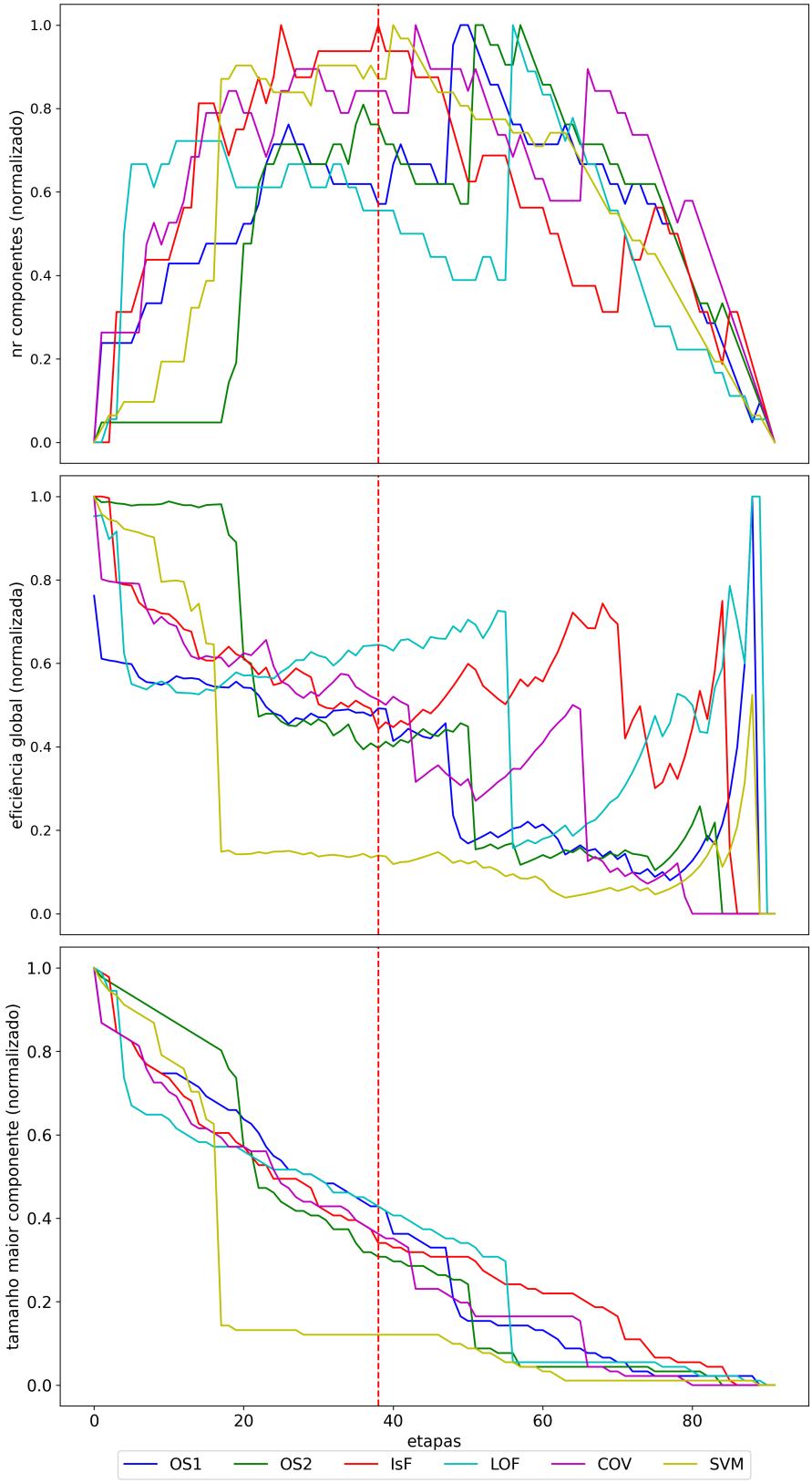


Figura 11 – RA - na disrupção pelo capital humano o modelo mais eficaz é o IsF, aumentando o número de componentes conectados (17), diminuindo a eficiência da rede (0,14043) e o tamanho do maior componente conectado (32), e causando a disrupção da rede na etapa 38, identificada pela linha vertical.

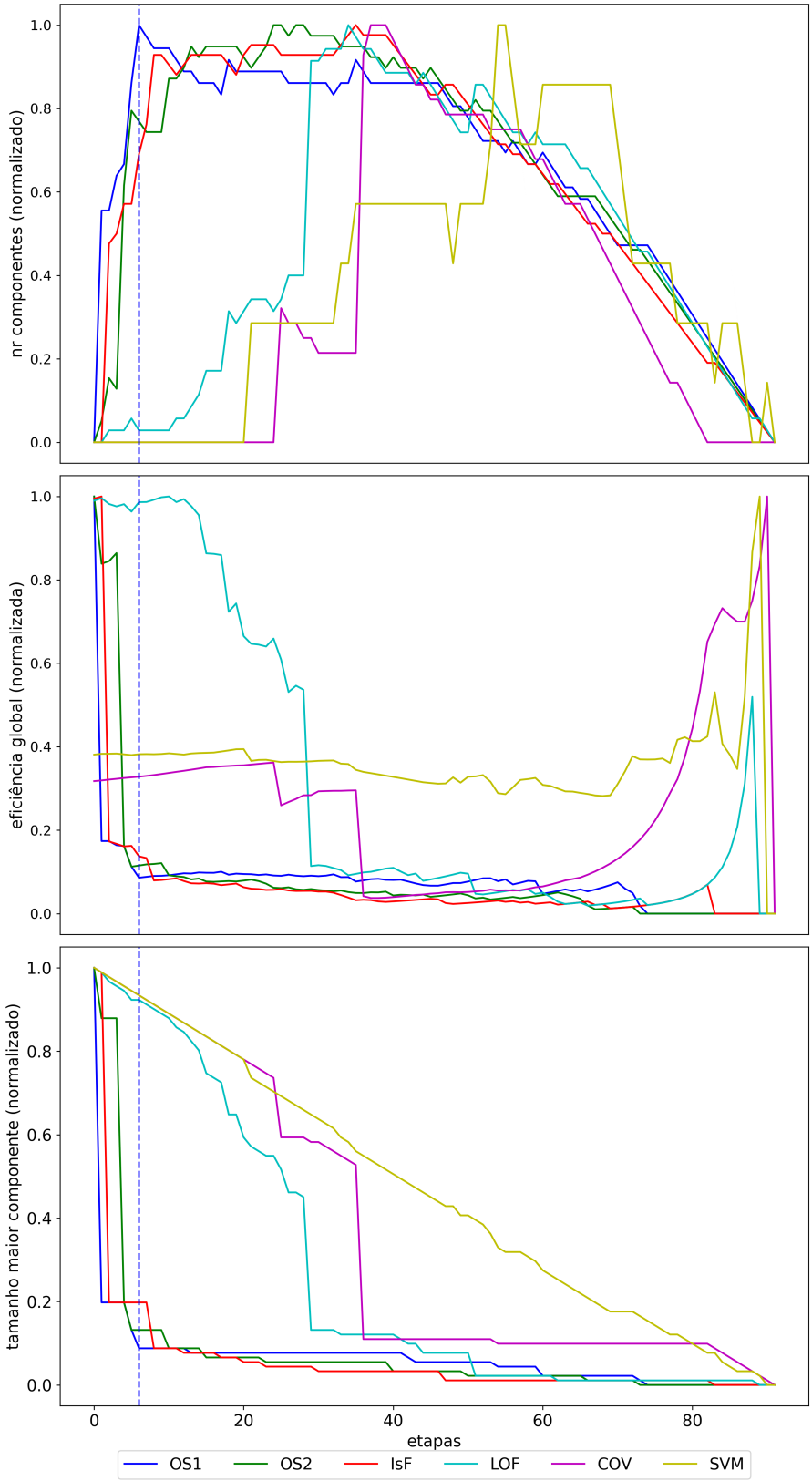


Figura 12 – RA - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (37), diminuindo a eficiência da rede (0,02735) e o tamanho do maior componente conectado (9), e causando a disrupção da rede na etapa 6, identificada pela linha vertical.

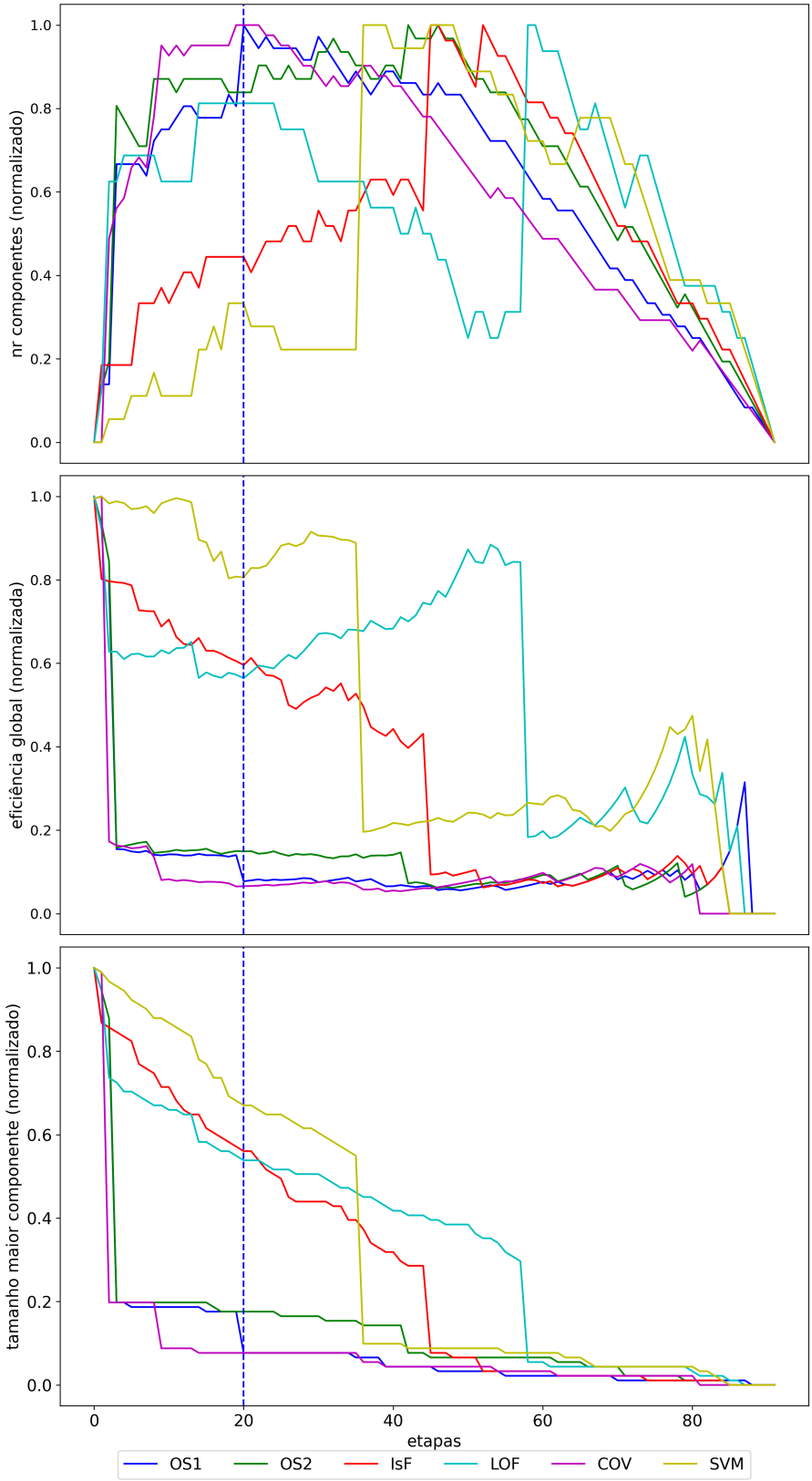


Figura 13 – RA - na disrupção pelo capital misto o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (37), diminuindo a eficiência da rede (0,02462) e o tamanho do maior componente conectado (8), e causando a disrupção da rede na etapa 20, identificada pela linha vertical.

A identificação dos agentes-chave então é orientada pelo melhor modelo de disrupção, OS1, com a abordagem do capital social. A Fig. 14 identifica os agentes-chave da RA. Estes agentes são os retirados nas seis primeiras estapas de disrupção da RA, orientadas por OS1.

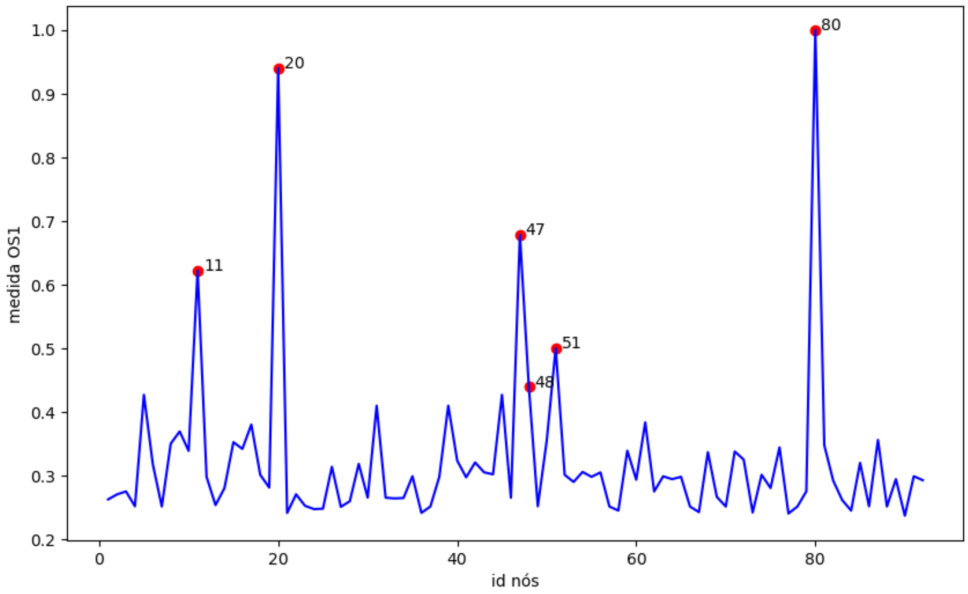


Figura 14 – Identificação dos seis agentes-chave da RA, pela abordagem do capital social com o modelo OS1.

Neste estudo, foi investigada a eficácia de diferentes modelos em prever e entender as redes de criminalidade. Foram realizadas simulações em 14 redes distintas para avaliar a *performance* de cada modelo em capturar os padrões e características inerentes a essas redes.

Além dos resultados gráficos da rede denominada RA, são compartilhados os resultados gráficos de mais uma rede, RA2, que possui uma estrutura composta por 1855 nós e 2449 ligações. Para RA2, a abordagem do capital humano é vista na Fig. 15, a abordagem do capital social na Fig. 16 e a abordagem do capital misto na Fig. 17.

Para todas as 14 redes, o modelo OS1, com uma abordagem baseada no capital social, se mostrou superior em eficiência comparado aos outros modelos avaliados. Ao analisar a RA2, é possível visualizar como este resultado persistiu (Fig. 18). A Fig. 19 identifica os agentes-chave da RA2. Estes agentes são os retirados nas 758 primeiras estapas de disrupção da RA2, orientadas por OS1. O resultado da disrupção das demais 12 redes de criminalidade podem ser visualidados nas figuras do Apêndice D.

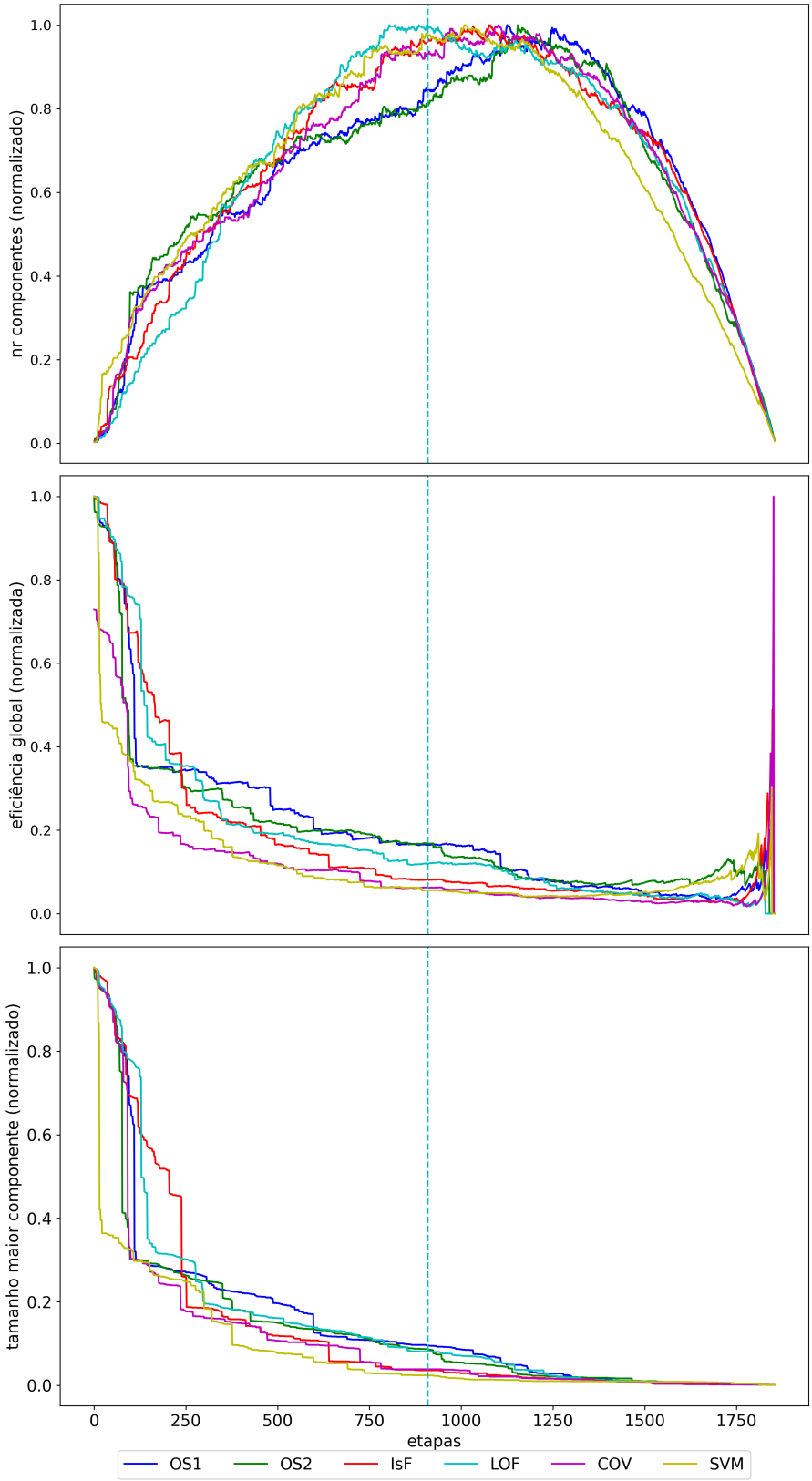


Figura 15 – RA2 - na disrupção pelo capital humano, o modelo mais eficaz é o LOF, aumentando o número de componentes (326), diminuindo a eficiência da rede (0,00872) e o tamanho do maior componente conectado (149), e causando a disrupção da rede na etapa 909, identificada pela linha vertical.

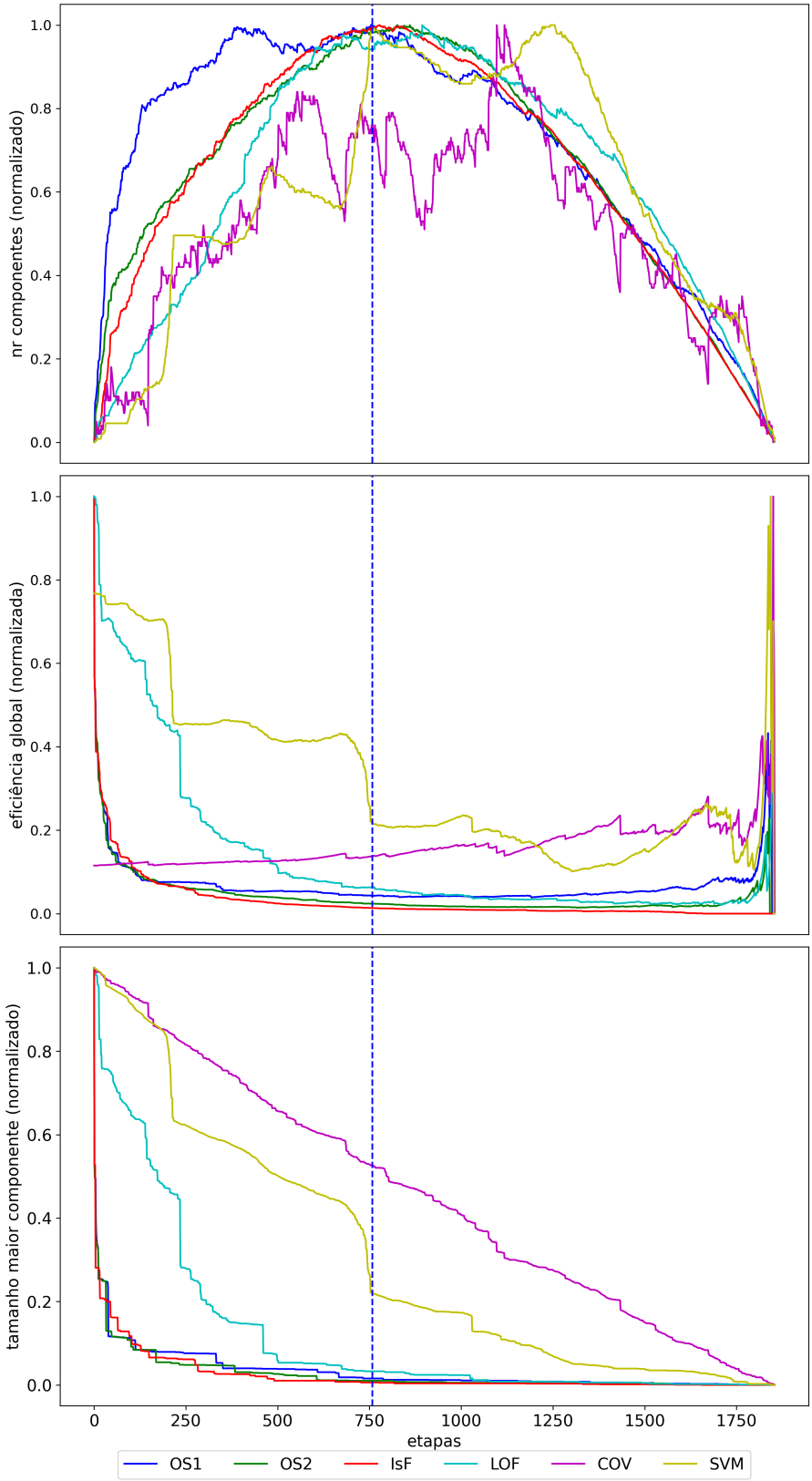


Figura 16 – RA2 - na disrupção pelo capital social, o modelo mais eficaz é o OS1, aumentando o número de componentes (413), diminuindo a eficiência da rede (0,00312) e o tamanho do maior componente conectado (30), e causando a disrupção da rede na etapa 758, identificada pela linha vertical.

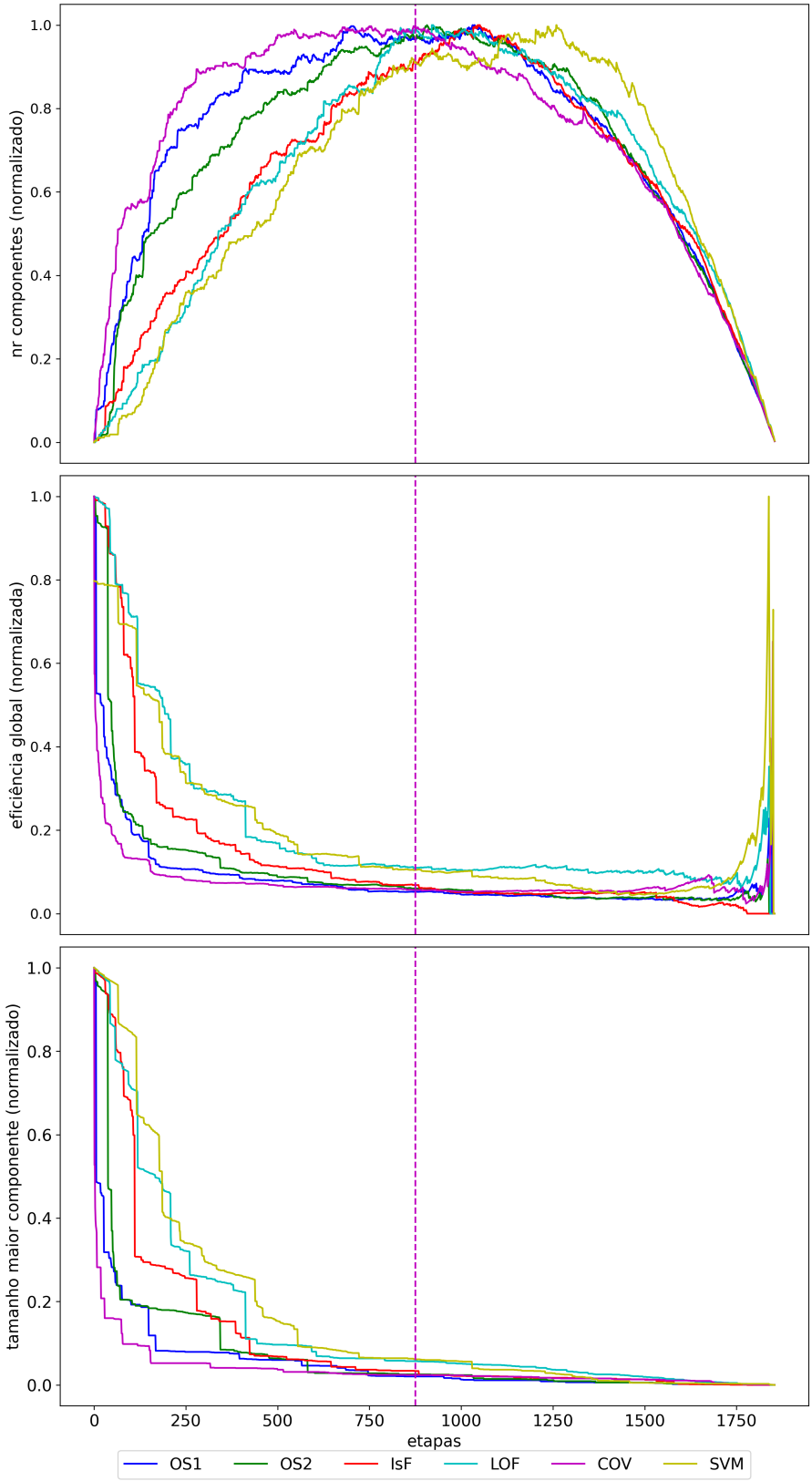


Figura 17 – RA2 - na disrupção pelo capital misto, o modelo mais eficaz é o COV, aumentando o número de componentes (379), diminuindo a eficiência da rede (0,02473) e o tamanho do maior componente conectado (45), e causando a disrupção da rede na etapa 875, identificada pela linha vertical.

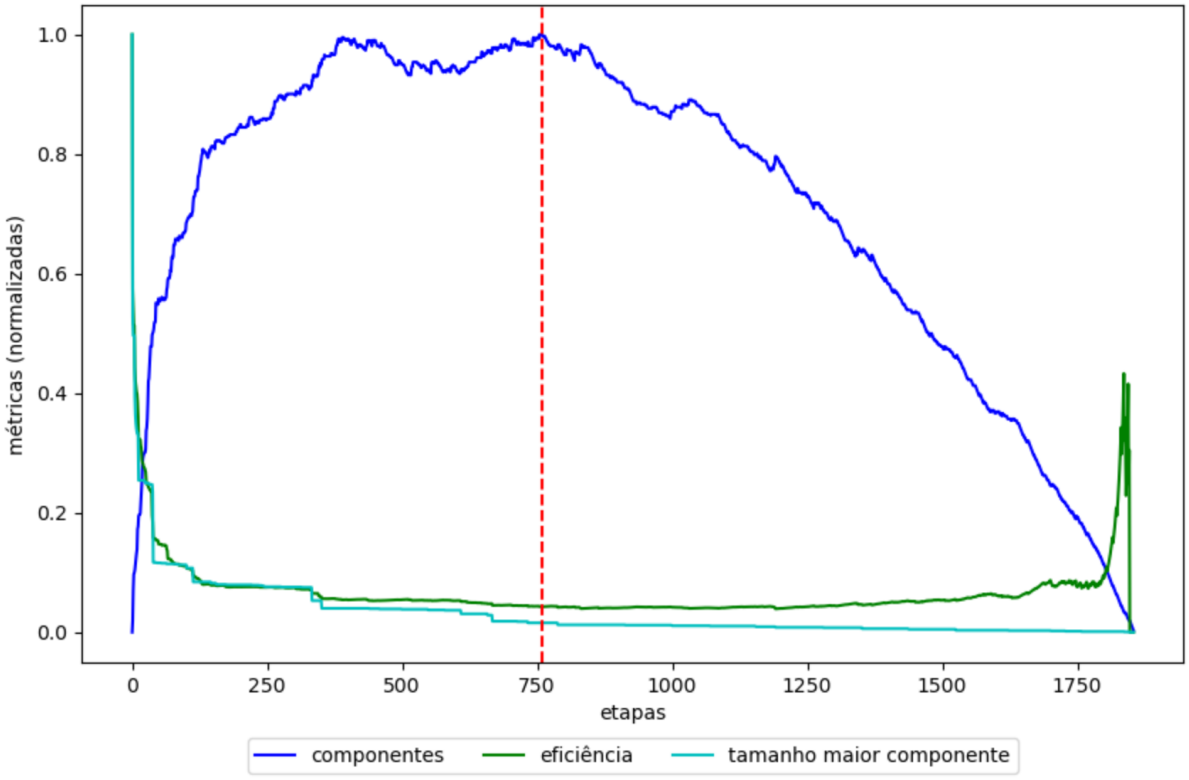


Figura 18 – RA2 - a disrupção pelo capital social com o modelo OS1, capaz de aumentar o número de componentes, diminuir a eficiência da rede e o tamanho do maior componente conectado com menor número de etapas de disrupção, é o resultado mais eficiente deste estudo. A linha vertical representa a etapa de disrupção.

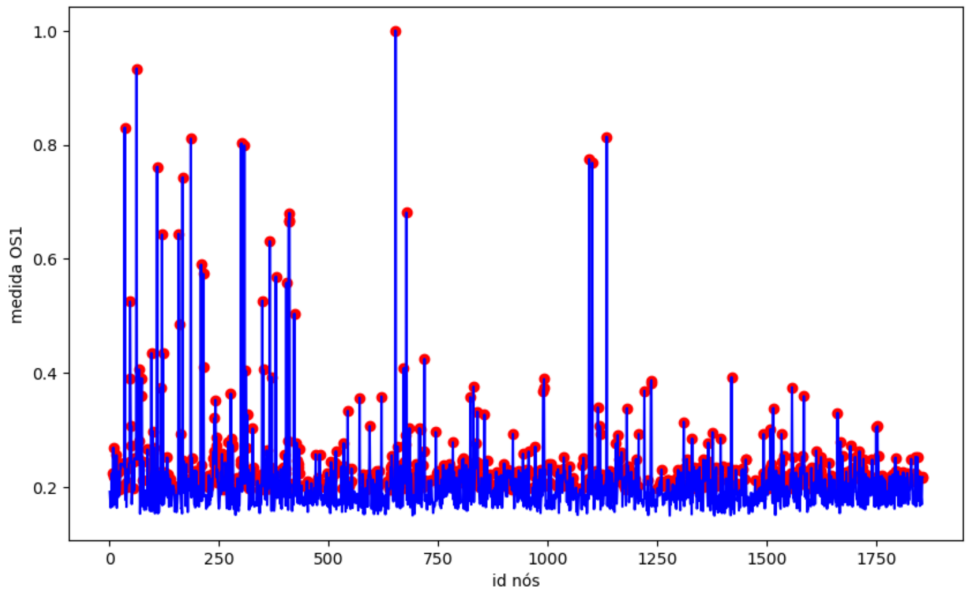


Figura 19 – Identificação dos 758 agentes-chave da RA2, pela abordagem do capital social com o modelo OS1.

5.3 Discussão

A metodologia inovadora apresentada pode ajudar na identificação de agentes-chave e na disrupção de redes de criminalidade. Primeiramente, a análise de redes de criminalidade permite que as agências de aplicação da lei identifiquem agentes-chave e conexões importantes em uma rede criminosa. Essa identificação pode ser feita através de ferramentas de redes complexas, que permitem visualizar a estrutura da rede e identificar os nós mais importantes.

Além disso, o estudo apresentado utiliza modelos de detecção de *outilier* para identificar comportamentos nas redes de criminalidade. Esses modelos permitem identificar padrões de comportamento que são incomuns na rede. Esses padrões podem indicar a presença de agentes-chave que merecem atenção especial das agências de aplicação da lei.

Foram construídos modelos computacionais que analisam redes de criminalidade, com base em princípios de teoria de redes complexas e comportamento humano. Estes modelos permitiram simular diferentes estratégias de disrupção da rede de criminalidade e avaliar sua eficácia.

O capital social, que é a capacidade dos indivíduos de se beneficiarem das relações sociais, pode ser um recurso crucial em redes de criminalidade. Ao considerá-lo nos modelos, foi incorporado um componente essencial que pode influenciar o funcionamento e a eficiência de disrupção das redes.

Neste estudo, foram empregados cinco métodos de SNA, a saber: *degree*, *betweenness*, *closeness*, *clustering* e *multiplexity*. Por meio da aplicação do modelo de disrupção delineado no Algoritmo 9, foram derivadas métricas referentes ao número de componentes, à eficiência global média e ao tamanho do maior componente conectado para os métodos de SNA em questão. A avaliação dos resultados obtidos mediante a aplicação da disrupção utilizando as métricas de SNA revelou que o método *degree* obteve o desempenho mais destacado.

Na Figura 20, apresentam-se os resultados da disrupção na RA, evidenciando que o método *degree* ocasiona a disrupção na etapa 20, ao passo que o modelo OS1, já apresentado, promove a disrupção da rede na etapa 6, conforme ilustrado na Figura 12.

Redes de criminalidade são intrinsecamente complexas e possuem um alto grau de interconexão entre seus membros. O modelo OS1, ao considerar o capital social, foi mais sensível às nuances dessas conexões, proporcionando uma compreensão mais aprofundada da estrutura e disrupção das redes.

Redes de criminalidade são dinâmicas e frequentemente se adaptam para enfrentar novos desafios ou contornar obstáculos. O modelo OS1, com sua abordagem baseada no capital social, pode ser mais adaptável e flexível em prever as mudanças e evoluções na rede.

O modelo OS1, assim como o OS2, é isento de parâmetros, o que elimina a necessidade de treinamento, em contraste com os outros quatro modelos, que exigem parâmetros específicos. Isto representa um ponto forte no modelo, que não depende de ajustes em parâmetros.

Os resultados do estudo podem ajudar na identificação de agentes-chave e na disrupção de redes de criminalidade, por meio da colaboração entre diferentes setores como academia,

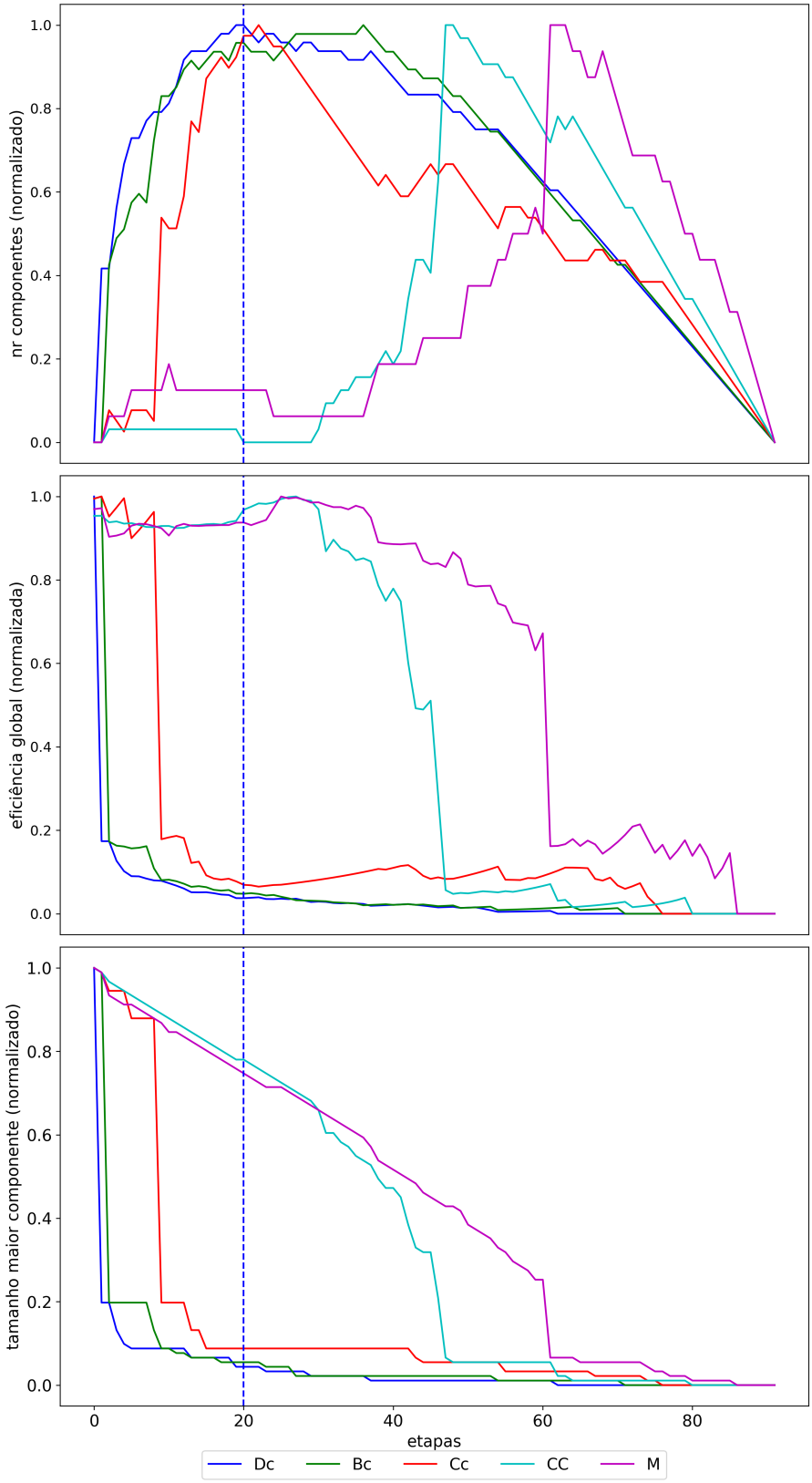


Figura 20 – RA - na disrupção pela SNA o modelo mais eficaz é o Degree, aumentando o número de componentes (49), diminuindo a eficiência da rede (0,01183) e o tamanho do maior componente conectado (5), e causando a disrupção da rede na etapa 20, identificada pela linha vertical.

setor privado e agências de aplicação da lei, que podem desenvolver um papel fundamental no aprimoramento das estratégias de prevenção e combate ao crime organizado. A metodologia e os resultados apresentados podem ser utilizados como ferramenta para promover essa colaboração e permitir que estes diferentes setores trabalhem juntos na identificação de agentes-chave e na disrupção de redes de criminalidade.

Estes resultados reforçam a importância de integrar conceitos como o capital social na modelagem para disrupção e identificação de agentes-chave em redes de criminalidade. Essa integração não só melhora a precisão das previsões, mas também oferece percepções valiosos sobre a natureza e funcionamento dessas redes.

6 Conclusão

O crime organizado é um problema global que afeta a segurança e o bem-estar de milhões de pessoas em todo o mundo. As organizações criminosas são altamente adaptáveis e resilientes, e muitas vezes operam em redes complexas que dificultam a identificação e a desarticulação de suas atividades. Além disso, o crime organizado está cada vez mais se valendo de tecnologias avançadas para facilitar suas operações e escapar da vigilância das agências de aplicação da lei.

Diante desse cenário desafiador, a identificação de agentes-chave e a disrupção de redes de criminalidade se tornam tarefas cada vez mais importantes para as autoridades. Identificar os agentes-chave e as conexões importantes em uma rede criminosa pode ajudar a orientar as ações de agências de aplicação da lei e concentrar os esforços em desmantelar essas redes. No entanto, essa tarefa é extremamente complexa e requer o uso de ferramentas avançadas de análise de dados e ciência de rede.

O estudo apresentado aborda esse problema de forma inovadora, propondo uma metodologia que combina técnicas de SNA, modelos de detecção de *outilier* e simulação computacional para identificar agentes-chave e provocar a disrupção de redes de criminalidade. A metodologia proposta é baseada em princípios de teoria de redes complexas e comportamento humano, e utiliza ferramentas computacionais avançadas para analisar e visualizar a estrutura de redes de criminalidade.

A hipótese apresentada, de que há uma correlação entre os dados das redes de criminalidade que possibilitam estabelecer padrões para a identificação de agentes-chave, foi comprovada através da metodologia proposta. A análise de redes de criminalidade permitiu identificar agentes-chave e conexões importantes em redes de criminalidade, e a simulação computacional permitiu testar diferentes estratégias de disrupção de redes de criminalidade e avaliar sua eficácia.

Além disso, a metodologia proposta na tese foi validada através de uma revisão sistemática da literatura sobre análise de redes de criminalidade. A revisão sistemática mostrou que a metodologia proposta é consistente com as abordagens mais recentes de análise de redes de criminalidade e que pode ser aplicada em diferentes contextos e cenários.

O estudo destaca o cumprimento dos objetivos propostos através do poder da colaboração interdisciplinar e das tecnologias de ponta para lidar com os complexos desafios que as redes de criminalidade representam. Ao combinar percepções de sistemas complexos, redes complexas, ML e detecção de *outliers*, foi desenvolvida uma estrutura robusta para identificar agentes-chave e interromper organizações criminosas.

Este estudo focou em simular intervenções para desmantelar 14 redes de criminalidade, identificando os agentes-chave nesse processo. Foram analisadas três estratégias: baseadas em capital humano, social e uma combinação dos dois, o capital misto. Para isso, foram utilizados seis modelos de detecção de *outilier*.

As estratégias mais eficazes para interromper essas redes focam nos agentes com o maior

capital social. Entre os seis modelos, o OS1 apresentou o melhor desempenho. Por outro lado, abordagens centradas apenas no capital humano não foram tão eficientes. As abordagens que combinam os dois tipos de capital (misto) tiveram uma eficácia moderada.

Contudo, há desafios em identificar um agente-chave cuja remoção resultaria no colapso total da rede, devido à adaptabilidade e dinâmica das redes criminosas. Estas redes tendem a ser extremamente resilientes a interrupções, adaptando-se rapidamente e reorganizando seus líderes.

Uma observação relevante é que há limitações na obtenção e modelagem de dados sobre redes criminosas. Eles podem ser incompletos, devido à natureza clandestina dessas redes, incorretos por erros humanos, ou inconsistentes por conta de informações falsas. Completar o conjunto de dados dessas redes é dificílimo devido à sua natureza dinâmica e indefinida.

Contudo, o êxito alcançado na análise e disrupção da RA destaca-se como uma exceção a esta regra, e essa distinção pode ser creditada em grande parte à dedicação e foco impecável da atividade de inteligência. O diferencial da operação em torno da RA foi que, cientes da relevância e da gravidade da ameaça representada por essa rede específica à época, as equipes de inteligência aprofundaram-se mais intensamente no trabalho de coleta e tratamento dos dados. Enquanto a análise convencional poderia ter se baseado em fontes genéricas ou informações mais acessíveis, neste caso, um esforço extraordinário foi feito para garantir que cada pedaço de informação fosse verificado, validado e, se necessário, correlacionado com outras fontes para aferir sua veracidade.

Pode-se argumentar que o nível de atenção e os recursos destinados à RA foram em razão do reconhecimento de sua importância estratégica naquela conjuntura. Isso não significa que outras redes fossem menos ameaçadoras ou significativas, mas, na avaliação da inteligência de segurança pública à época, a RA demandava uma abordagem mais focada.

As outras redes, apesar de estarem em operação, não foram submetidas a este tratamento especial por parte da inteligência. Este é um lembrete de que a alocação de recursos e a atenção em segurança pública frequentemente precisa fazer escolhas estratégicas baseadas em avaliações de risco e prioridade.

Em retrospecto, o tratamento diferenciado dado à RA provou ser uma decisão acertada. A desarticulação efetiva dessa rede serviu não apenas como um testemunho do que pode ser alcançado quando a coleta e o tratamento de dados são executados de forma impecável, mas também como um marco para futuras operações, demonstrando a importância de uma abordagem estratégica e focada no combate ao crime organizado.

Os resultados ressaltam a importância da abordagem de capital social com o modelo OS1, revelando sua eficácia no desmantelamento de redes de criminalidade por meio de intervenções direcionadas. A visualização dos estágios de interrupção e dos mecanismos de resiliência da rede aprofundam ainda mais a compreensão da dinâmica em jogo. Como as agências de aplicação da lei buscam ferramentas cada vez mais sofisticadas para combater atividades criminosas, este estudo oferece um roteiro para integrar modelos teóricos e aplicações práticas, contribuindo para uma sociedade mais segura e protegida. Este trabalho exemplifica o potencial da pesquisa interdisciplinar para gerar soluções inovadoras para desafios complexos do mundo real.

6.1 Trabalhos futuros

Nos últimos anos, observamos o rápido avanço da inteligência artificial, especialmente no campo do *Deep Learning*. Estes modelos de aprendizado profundo demonstraram eficácia superior em uma variedade de tarefas, desde reconhecimento de imagem a tradução de linguagem. Em paralelo, a criminalidade, cada vez mais sofisticada, expande-se usando redes complexas que se assemelham a sistemas distribuídos. Portanto, a combinação de técnicas avançadas de *Deep Learning* com a análise de redes pode ser uma abordagem promissora para identificar e desmantelar estas redes criminosas.

Os modelos de aprendizado profundo, como as Redes Neurais Recorrentes (RNN) e os *Transformers*, podem capturar e analisar sequências de eventos, relações e interações ao longo do tempo, ajudando a mapear a estrutura e a dinâmica de redes de criminalidade.

O uso de *Deep Learning* para identificar agentes-chave e interromper redes de criminalidade não é apenas uma evolução natural da aplicação da tecnologia, mas uma necessidade diante da complexidade crescente das redes de criminalidade modernas. Ao combinar a capacidade analítica dos modelos de aprendizado profundo com as técnicas tradicionais de análise de redes, podemos esperar melhorar significativamente a capacidade de combater e prevenir o crime em múltiplas frentes.

Referências

1 EGBERT, S.; LEESE, M. **Criminal futures: Predictive policing and everyday police work**. [S.l.]: Taylor & Francis, 2021. Citado 5 vezes nas páginas 14, 20, 21, 31 e 45.

2 WHEELER, A. P.; REUTER, S. Redrawing hot spots of crime in dallas, texas. **Police Quarterly**, SAGE Publications Sage CA: Los Angeles, CA, v. 24, n. 2, p. 159–184, 2021. Citado 2 vezes nas páginas 14 e 20.

3 CAPLAN, J. M. et al. Data-informed and place-based violent crime prevention: the kansas city, missouri risk-based policing initiative. **Police quarterly**, Sage Publications Sage CA: Los Angeles, CA, v. 24, n. 4, p. 438–464, 2021. Citado 2 vezes nas páginas 14 e 20.

4 FICARA, A. et al. Multilayer network analysis: the identification of key actors in a sicilian mafia operation. In: SPRINGER. **International conference on future access enablers of ubiquitous and intelligent infrastructures**. [S.l.], 2021. p. 120–134. Citado 4 vezes nas páginas 14, 20, 24 e 33.

5 HELBING, D.; BALIETTI, S. From social data mining to forecasting socio-economic crises. **The European Physical Journal Special Topics**, Springer, v. 195, p. 3–68, 2011. Citado na página 14.

6 TOLEDO, A. S. O. **Estudo da Complexidade de Redes de Criminalidade**. 2018. Citado na página 14.

7 LI, J. C. et al. Understanding and preventing financial fraud against older citizens in chinese society: Results of a focus group study. **International Journal of Offender Therapy and Comparative Criminology**, SAGE Publications Sage CA: Los Angeles, CA, v. 60, n. 13, p. 1509–1531, 2016. Citado na página 14.

8 WORTLEY, R. et al. What is crime science. **Routledge handbook of crime science**, Routledge, p. 1–29, 2019. Citado na página 14.

9 BRIGHT, D.; BREWER, R.; MORSELLI, C. Using social network analysis to study crime: Navigating the challenges of criminal justice records. **Social Networks**, Elsevier, v. 66, p. 50–64, 2021. Citado 12 vezes nas páginas 14, 15, 20, 21, 22, 23, 24, 30, 31, 33, 60 e 61.

10 FICARA, A. et al. Human and social capital strategies for mafia network disruption. **IEEE Transactions on Information Forensics and Security**, IEEE, v. 18, p. 1926–1936, 2023. Citado 6 vezes nas páginas 14, 20, 21, 22, 34 e 38.

11 DUIJN, P. A.; KASHIRIN, V.; SLOOT, P. M. The relative ineffectiveness of criminal network disruption. **Scientific reports**, Nature Publishing Group UK London, v. 4, n. 1, p. 4238, 2014. Citado 8 vezes nas páginas 15, 20, 22, 23, 29, 30, 31 e 35.

12 SCHELLING, T. C. Dynamic models of segregation. **Journal of mathematical sociology**, Taylor & Francis, v. 1, n. 2, p. 143–186, 1971. Citado na página 15.

13 CAMPANA, P. Explaining criminal networks: Strategies and potential pitfalls. **Methodological Innovations**, SAGE Publications Sage UK: London, England, v. 9, p. 2059799115622748, 2016. Citado 3 vezes nas páginas 15, 20 e 30.

14 BORRILL, P. L.; TESFATSION, L. 11 agent-based modeling: the right mathematics for the social sciences?'. **The Elgar companion to recent economic methodology**, v. 228, 2011. Citado 2 vezes nas páginas 15 e 17.

15 EPSTEIN, J. M. **Generative social science: Studies in agent-based computational modeling**. [S.l.]: Princeton University Press, 2012. Citado 3 vezes nas páginas 15, 16 e 17.

16 GOYAL, S.; VIGIER, A. Attack, defence, and contagion in networks. **The Review of Economic Studies**, Oxford University Press, v. 81, n. 4, p. 1518–1542, 2014. Citado 4 vezes nas páginas 15, 20, 31 e 35.

17 BARABÁSI, A.-L. Network science. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, The Royal Society Publishing, v. 371, n. 1987, p. 20120375, 2013. Citado na página 16.

18 GRANOVETTER, M. The strength of weak ties: A network theory revisited. **Sociological theory**, JSTOR, p. 201–233, 1983. Citado na página 16.

19 CROSS, R.; BORGATTI, S. P.; PARKER, A. Making invisible work visible: Using social network analysis to support strategic collaboration. **California management review**, SAGE Publications Sage CA: Los Angeles, CA, v. 44, n. 2, p. 25–46, 2002. Citado na página 16.

20 CROSS, R. et al. Knowing what we know: Supporting knowledge creation and sharing in social networks. **Networks in the knowledge economy**, Oxford University Press Oxford, p. 208–231, 2003. Citado na página 16.

21 CROSSLEY, N. et al. **Social network analysis for ego-nets: Social network analysis for actor-centred networks**. [S.l.]: Sage, 2015. Citado 2 vezes nas páginas 16 e 48.

22 EDWARDS, G. Mixed-method approaches to social network analysis. NCRM, 2010. Citado na página 16.

23 GREEN, D. P.; MCFALLS, L. H.; SMITH, J. K. Hate crime: An emergent research agenda. **Annual review of sociology**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 27, n. 1, p. 479–504, 2001. Citado 2 vezes nas páginas 16 e 20.

24 SULTAN, H. B.; MAHMOOD, B. M. Analyzing crime networks: A complex network-based approach. **AL-Rafidain Journal of Computer Sciences and Mathematics**, Mosul University, v. 15, n. 1, p. 57–73, 2021. Citado 2 vezes nas páginas 16 e 20.

25 WANG, H. et al. Criminal behavior analysis based on complex networks theory. In: IEEE. **2009 IEEE International Symposium on IT in Medicine & Education**. [S.l.], 2009. v. 1, p. 951–955. Citado na página 16.

26 LAZER, D. et al. Social science. computational social science. **Science (New York, NY)**, v. 323, n. 5915, p. 721–723, 2009. Citado na página 16.

27 GILBERT, N.; TROITZSCH, K. **Simulation for the social scientist**. [S.l.]: McGraw-Hill Education (UK), 2005. Citado na página 16.

28 BARABÁSI, A.-L. **Linked: The new science of networks**. [S.l.]: American Association of Physics Teachers, 2003. Citado 2 vezes nas páginas 17 e 22.

29 SNIJDERS, T. A.; BUNT, G. G. Van de; STEGLICH, C. E. Introduction to stochastic actor-based models for network dynamics. **Social networks**, Elsevier, v. 32, n. 1, p. 44–60, 2010. Citado na página 17.

30 BONABEAU, E. Agent-based modeling: Methods and techniques for simulating human systems. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 99, n. suppl_3, p. 7280–7287, 2002. Citado na página 17.

31 ONNELA, J.-P. et al. Structure and tie strengths in mobile communication networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 104, n. 18, p. 7332–7336, 2007. Citado na página 17.

32 KARSAI, M. et al. Small but slow world: How network topology and burstiness slow down spreading. **Physical Review E**, APS, v. 83, n. 2, p. 025102, 2011. Citado na página 17.

33 TOLEDO, A.; CARPI, L. C.; ATMAN, A. Diversity analysis exposes unexpected key roles in multiplex crime networks. In: SPRINGER. **Complex Networks XI: Proceedings of the 11th Conference on Complex Networks CompleNet 2020**. [S.l.], 2020. p. 371–382. Citado 4 vezes nas páginas 20, 21, 22 e 29.

34 TOLEDO, A. S. et al. Multiplex key roles to disrupt criminal networks. **Social Network Analysis and Mining**, Springer, v. 13, n. 1, p. 1–14, 2023. Citado 5 vezes nas páginas 20, 21, 22, 31 e 38.

35 HOYER, B.; JAEGHER, K. D. Strategic network disruption and defense. **Journal of Public Economic Theory**, Wiley Online Library, v. 18, n. 5, p. 802–830, 2016. Citado na página 20.

36 CAVALLARO, L. et al. Disrupting resilient criminal networks through data analysis: The case of sicilian mafia. **Plos one**, Public Library of Science San Francisco, CA USA, v. 15, n. 8, p. e0236476, 2020. Citado 6 vezes nas páginas 20, 21, 29, 31, 36 e 38.

37 FICARA, A. et al. Covert network construction, disruption, and resilience: A survey. **Mathematics**, MDPI, v. 10, n. 16, p. 2929, 2022. Citado 3 vezes nas páginas 20, 21 e 36.

38 VARESE, F. The structure and the content of criminal connections: The russian mafia in italy. **European sociological review**, Oxford University Press, v. 29, n. 5, p. 899–909, 2013. Citado 2 vezes nas páginas 20 e 21.

39 BOUCHARD, M.; AMIRAULT, J. Advances in research on illicit networks. **Global crime**, Taylor & Francis, v. 14, n. 2-3, p. 119–122, 2013. Citado 2 vezes nas páginas 20 e 21.

40 KREAGER, D. A. et al. Toward a criminology of inmate networks. **Justice Quarterly**, Taylor & Francis, v. 33, n. 6, p. 1000–1028, 2016. Citado 2 vezes nas páginas 20 e 21.

41 CATANZARO, M.; CALDARELLI, G.; PIETRONERO, L. Assortative model for social networks. **Physical review e**, APS, v. 70, n. 3, p. 037101, 2004. Citado 3 vezes nas páginas 21, 30 e 31.

42 CROFT, D. et al. Assortative interactions and social networks in fish. **Oecologia**, Springer, v. 143, p. 211–219, 2005. Citado 3 vezes nas páginas 21, 30 e 31.

43 RIBEIRO, H. V. et al. The dynamical structure of political corruption networks. **Journal of Complex Networks**, Oxford University Press, v. 6, n. 6, p. 989–1003, 2018. Citado na página 21.

44 REN, X.-L. et al. Generalized network dismantling. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 116, n. 14, p. 6554–6559, 2019. Citado na página 21.

45 CUNHA, B. R. da; GONÇALVES, S. Topology, robustness, and structural controllability of the brazilian federal police criminal intelligence network. **Applied network science**, SpringerOpen, v. 3, n. 1, p. 1–20, 2018. Citado 2 vezes nas páginas 21 e 24.

46 ALVES, L. G. et al. Distance to the scaling law: a useful approach for unveiling relationships between crime and urban metrics. **PloS one**, Public Library of Science San Francisco, USA, v. 8, n. 8, p. e69580, 2013. Citado na página 21.

47 PLANK, J. et al. Complex systems, interdisciplinary collaboration, and institutional renewal. **Change: The Magazine of Higher Learning**, Taylor & Francis, v. 43, n. 3, p. 35–43, 2011. Citado 2 vezes nas páginas 21 e 22.

48 THURNER, S.; HANEL, R.; KLIMEK, P. **Introduction to the theory of complex systems**. [S.l.]: Oxford University Press, 2018. Citado na página 21.

49 WOLFRAM, S. Complex systems theory. **Emerging syntheses in science**, p. 183–190, 2018. Citado na página 21.

50 MITCHELL, M. **Complexity: A guided tour**. [S.l.]: Oxford university press, 2009. Citado na página 21.

51 MILLER, J. H.; PAGE, S. E. **Complex adaptive systems: an introduction to computational models of social life: an introduction to computational models of social life**. [S.l.]: Princeton university press, 2009. Citado na página 21.

52 GUANRONG, C.; WANG, X.; SHI, D. Complex systems and networks: Dynamics, controls and applciations. Springer, 2016. Citado 2 vezes nas páginas 21 e 47.

53 NEWMAN, M. **Networks**. [S.l.]: Oxford university press, 2018. Citado 5 vezes nas páginas 22, 23, 38, 43 e 48.

54 CROOKS, A. T.; HEPPENSTALL, A. J. Introduction to agent-based modelling. In: **Agent-based models of geographical systems**. [S.l.]: Springer, 2011. p. 85–105. Citado na página 22.

55 PÓSFAI, M.; BARABASI, A.-L. **Network Science**. [S.l.]: Citeseer, 2016. Citado 2 vezes nas páginas 22 e 38.

56 BLOCH, F.; JACKSON, M. O.; TEBALDI, P. Centrality measures in networks. **Social Choice and Welfare**, Springer, p. 1–41, 2023. Citado 3 vezes nas páginas 22, 38 e 48.

57 WASSERMAN, S.; FAUST, K. Social network analysis: Methods and applications. Cambridge university press, 1994. Citado 3 vezes nas páginas 22, 60 e 61.

58 MELAZO, V. B. Explorando redes complexas: a importância da análise visual perante medidas estatísticas. Universidade Federal de Uberlândia, 2021. Citado na página 22.

59 WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world'networks. **nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. Citado na página 22.

60 GIRVAN, M.; NEWMAN, M. E. Community structure in social and biological networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 99, n. 12, p. 7821–7826, 2002. Citado na página 23.

61 CARVALHO, J. A. d. et al. Detecção e identificação de notícias falsas em redes sociais utilizando abordagem de ciência de redes. Universidade Federal de Uberlândia, 2024. Citado na página 23.

62 FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, JSTOR, p. 35–41, 1977. Citado na página 23.

63 PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic spreading in scale-free networks. **Physical review letters**, APS, v. 86, n. 14, p. 3200, 2001. Citado na página 23.

64 VALENTE, T. W. Network interventions. **science**, American Association for the Advancement of Science, v. 337, n. 6090, p. 49–53, 2012. Citado na página 23.

65 GUO, Q. et al. Levy random walks on multiplex networks. **Scientific reports**, Nature Publishing Group UK London, v. 6, n. 1, p. 37641, 2016. Citado na página 23.

66 BOCCALETTI, S. et al. The structure and dynamics of multilayer networks. **Physics reports**, Elsevier, v. 544, n. 1, p. 1–122, 2014. Citado na página 23.

67 KIVELÄ, M. et al. Multilayer networks. **Journal of complex networks**, Oxford University Press, v. 2, n. 3, p. 203–271, 2014. Citado na página 23.

68 DOMENICO, M. D. et al. The physics of spreading processes in multilayer networks. **Nature Physics**, Nature Publishing Group UK London, v. 12, n. 10, p. 901–906, 2016. Citado 3 vezes nas páginas 23, 38 e 48.

69 GAO, J. et al. Networks formed from interdependent networks. **Nature physics**, Nature Publishing Group UK London, v. 8, n. 1, p. 40–48, 2012. Citado na página 23.

70 BARRAT, A. et al. The architecture of complex weighted networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 101, n. 11, p. 3747–3752, 2004. Citado na página 23.

71 PEIXOTO, T. P. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. **Physical Review E**, APS, v. 92, n. 4, p. 042807, 2015. Citado na página 23.

72 PAPACHRISTOS, A. V. Murder by structure: Dominance relations and the social structure of gang homicide. **American journal of sociology**, The University of Chicago Press, v. 115, n. 1, p. 74–128, 2009. Citado na página 24.

73 AWAD, M.; KHANNA, R. **Efficient learning machines: theories, concepts, and applications for engineers and system designers**. [S.l.]: Springer nature, 2015. Citado 3 vezes nas páginas 24, 25 e 38.

74 HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. **Automated machine learning: methods, systems, challenges**. [S.l.]: Springer Nature, 2019. Citado 3 vezes nas páginas 24, 25 e 38.

75 LYNCH, S. **Python for Scientific Computing and Artificial Intelligence**. [S.l.]: CRC Press, 2023. Citado 2 vezes nas páginas 24 e 38.

76 PATEL, A. A. **Hands-on unsupervised learning using Python: how to build applied machine learning solutions from unlabeled data**. [S.l.]: O'Reilly Media, 2019. Citado 4 vezes nas páginas 24, 25, 28 e 38.

77 AGGARWAL, C. C. **Outlier analysis second edition**. [S.l.]: Springer Switzerland, 2016. Citado 2 vezes nas páginas 25 e 26.

78 AGGARWAL, C. C.; SATHE, S. Theoretical foundations and algorithms for outlier ensembles. **Acm sigkdd explorations newsletter**, ACM New York, NY, USA, v. 17, n. 1, p. 24–47, 2015. Citado 4 vezes nas páginas 26, 45, 56 e 58.

79 CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 41, n. 3, p. 1–58, 2009. Citado 2 vezes nas páginas 27 e 28.

80 WANG, H.; BAH, M. J.; HAMMAD, M. Progress in outlier detection techniques: A survey. **Ieee Access**, IEEE, v. 7, p. 107964–108000, 2019. Citado 3 vezes nas páginas 27, 28 e 38.

81 MANDHARE, H. C.; IDATE, S. A comparative study of cluster based outlier detection, distance based outlier detection and density based outlier detection techniques. In: IEEE. **2017 international conference on intelligent computing and control systems (ICICCS)**. [S.l.], 2017. p. 931–935. Citado na página 27.

82 BOUKERCHE, A.; ZHENG, L.; ALFANDI, O. Outlier detection: Methods, models, and classification. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 53, n. 3, p. 1–37, 2020. Citado 2 vezes nas páginas 27 e 38.

83 SCIKIT-LEARN. **svm.OneClassSVM**. 2023. Disponível em: <https://scikit-learn.org-/stable/modules/generated/sklearn.svm.OneClassSVM.html>. Citado 3 vezes nas páginas 27, 38 e 53.

84 TOLEDO, A. S. et al. Outlier mining in high-dimensional data using the jensen–shannon divergence and graph structure analysis. **Journal of Physics: Complexity**, IOP Publishing, v. 3, n. 4, p. 045011, 2022. Citado 2 vezes nas páginas 27 e 38.

85 LIU, F. T.; TING, K. M.; ZHOU, Z.-H. Isolation forest. In: IEEE. **2008 eighth ieee international conference on data mining**. [S.l.], 2008. p. 413–422. Citado 3 vezes nas páginas 27, 54 e 58.

86 SCIKIT-LEARN. **svm.ensemble.IsolationForest**. 2023. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>. Citado 3 vezes nas páginas 27, 38 e 54.

87 SCIKIT-LEARN. **neighbors.LocalOutlierFactor**. 2023. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html>. Citado 3 vezes nas páginas 27, 38 e 56.

88 SCIKIT-LEARN. **covariance.EllipticEnvelope**. 2023. Disponível em: <https://scikit-learn-.org/stable/modules/generated/sklearn.covariance.EllipticEnvelope.html>. Citado 3 vezes nas páginas 27, 38 e 57.

89 AKOGLU, L.; TONG, H.; KOUTRA, D. Graph based anomaly detection and description: a survey. **Data mining and knowledge discovery**, Springer, v. 29, p. 626–688, 2015. Citado na página 28.

90 PANG, G. et al. Deep learning for anomaly detection: A review. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 2, p. 1–38, 2021. Citado na página 28.

91 BAKAR, Z. A. et al. A comparative study for outlier detection techniques in data mining. In: IEEE. **2006 IEEE conference on cybernetics and intelligent systems**. [S.l.], 2006. p. 1–6. Citado na página 28.

92 JIDIGA, G. R.; SAMMULAL, D. P. Machine learning approach to anomaly detection in cyber security with a case study of spamming attack. **International Journal of Computer Engineering & Technology (IJCET)**, v. 4, n. 3, p. 113–122, 2013. Citado na página 28.

93 GOLDSTEIN, M.; UCHIDA, S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. **PloS one**, Public Library of Science San Francisco, CA USA, v. 11, n. 4, p. e0152173, 2016. Citado 2 vezes nas páginas 28 e 58.

94 GOTTSCHALK, P. Value configurations in organised crime. **Policing & Society**, Taylor & Francis, v. 19, n. 1, p. 47–57, 2009. Citado na página 29.

95 KREBS, V. E. Mapping networks of terrorist cells. **Connections**, v. 24, n. 3, p. 43–52, 2002. Citado na página 29.

96 SPARROW, M. K. The application of network analysis to criminal intelligence: An assessment of the prospects. **Social networks**, Elsevier, v. 13, n. 3, p. 251–274, 1991. Citado 2 vezes nas páginas 29 e 30.

97 CORNISH, D. B. The procedural analysis of offending and its relevance for situational prevention. **Crime prevention studies**, v. 3, n. 1, p. 151–196, 1994. Citado na página 29.

98 DEHGHANNIRI, H.; BORRION, H. Crime scripting: A systematic review. **European Journal of Criminology**, SAGE Publications Sage UK: London, England, v. 18, n. 4, p. 504–525, 2021. Citado na página 29.

99 BROWN, R.; VELÁSQUEZ, A. The effect of violent crime on the human capital accumulation of young adults. **Journal of development economics**, Elsevier, v. 127, p. 1–12, 2017. Citado na página 29.

100 VILLANI, S. et al. A virtuous combination of structural and skill analysis to defeat organized crime. **Socio-Economic Planning Sciences**, Elsevier, v. 65, n. C, p. 51–65, 2019. Citado na página 29.

101 TOSTADO, S. d. l. M.; NÚÑEZ-LÓPEZ, M.; HERNÁNDEZ-VARGAS, E. A. Human trafficking in mexico: Data sources, network analysis and the limits of dismantling strategies. **arXiv preprint arXiv:2206.02971**, 2022. Citado na página 29.

102 NORRIS, F. H. et al. Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. **American journal of community psychology**, Springer, v. 41, p. 127–150, 2008. Citado 2 vezes nas páginas 29 e 30.

103 AYLING, J. Criminal organizations and resilience. **International Journal of Law, Crime and Justice**, Elsevier, v. 37, n. 4, p. 182–196, 2009. Citado na página 30.

104 MORSELLI, C.; GIGUÈRE, C.; PETIT, K. The efficiency/security trade-off in criminal networks. **Social networks**, Elsevier, v. 29, n. 1, p. 143–153, 2007. Citado na página 30.

105 WILLIAMS, P. Transnational criminal networks. **Networks and netwars: the future of terror, crime, and militancy**, Rand Santa Monica, CA, v. 1382, p. 61, 2001. Citado na página 30.

106 SUTCLIFFE, K. M. **A Review of:"From Pablo to Osama: Trafficking and Terrorist Networks, Government Bureaucracies, and Competitive Advantage by Michael Kenney" From Pablo to Osama: Trafficking and Terrorist Networks, Government Bureaucracies, and Competitive Advantage. Michael Kenney. University Park, PA: Pennsylvania State University Press, 2007, 293 pages. ISBN 9870271029313**. [S.l.]: Taylor & Francis, 2007. Citado na página 30.

107 MUNRO, P. People smuggling and the resilience of criminal networks in indonesia. **Journal of Policing, Intelligence and Counter Terrorism**, Taylor & Francis, v. 6, n. 1, p. 40–50, 2011. Citado na página 30.

108 AGRESTE, S. et al. Network structure and resilience of mafia syndicates. **Information Sciences**, Elsevier, v. 351, p. 30–47, 2016. Citado na página 30.

109 COMUNNITY, P. **Python**. 2023. Disponível em: <https://www.python.org>. Citado na página 39.

110 MCKINNEY, W. **Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython**. [S.l.]: Novatec Editora, 2018. Citado 2 vezes nas páginas 42 e 43.

111 PMMG. **Plano Estratégico 2020-2023**. [S.l.]: Comando Geral da PMMG, Belo Horizonte, 2. ed., 2021. Citado na página 44.

112 TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. [S.l.]: Pearson Education India, 2016. Citado na página 45.

113 HAN, J.; PEI, J.; TONG, H. **Data mining: concepts and techniques**. [S.l.]: Morgan kaufmann, 2022. Citado na página 45.

114 GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of machine learning research**, v. 3, n. Mar, p. 1157–1182, 2003. Citado na página 45.

115 WU, X. et al. Top 10 algorithms in data mining. **Knowledge and information systems**, Springer, v. 14, p. 1–37, 2008. Citado na página 45.

116 BOCCARA, N.; BOCCARA, N. **Modeling complex systems**. [S.l.]: Springer, 2010. v. 1. Citado na página 50.

117 SCHÖLKOPF, B. et al. Estimating the support of a high-dimensional distribution. **Neural computation**, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 13, n. 7, p. 1443–1471, 2001. Citado na página 53.

118 BOTTOU, L. Large-scale machine learning with stochastic gradient descent. In: SPRINGER. **Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers**. [S.l.], 2010. p. 177–186. Citado na página 53.

119 HADSELL, R.; CHOPRA, S.; LECUN, Y. Dimensionality reduction by learning an invariant mapping. In: IEEE. **2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)**. [S.l.], 2006. v. 2, p. 1735–1742. Citado na página 53.

120 LIU, F. T.; TING, K. M.; ZHOU, Z.-H. Isolation-based anomaly detection. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM New York, NY, USA, v. 6, n. 1, p. 1–39, 2012. Citado 2 vezes nas páginas 54 e 55.

121 BREUNIG, M. M. et al. Lof: identifying density-based local outliers. In: **Proceedings of the 2000 ACM SIGMOD international conference on Management of data**. [S.l.: s.n.], 2000. p. 93–104. Citado na página 56.

122 ROUSSEEUW, P. J. Multivariate estimation with high breakdown point. **Mathematical statistics and applications**, v. 8, n. 283-297, p. 37, 1985. Citado na página 57.

123 ROUSSEEUW, P. J.; DRIESSEN, K. V. A fast algorithm for the minimum covariance determinant estimator. **Technometrics**, Taylor & Francis, v. 41, n. 3, p. 212–223, 1999. Citado 2 vezes nas páginas 57 e 58.

124 BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4. Citado na página 59.

125 ECK, J.; WEISBURD, D. L. Crime places in crime theory. **Crime and place: Crime prevention studies**, v. 4, 2015. Citado 2 vezes nas páginas 60 e 61.

126 LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. **Physical review letters**, APS, v. 87, n. 19, p. 198701, 2001. Citado na página 61.

Apêndices

APÊNDICE A – Medidas do Capital Humano - RA

id	OS1	OS2	IsF	LOF	COV	SVM
1	0.3445820867190598	0.5132473995519444	0.09782608695652174	0.41303447826086957	0.30434782608695654	0.836956521739130
2	0.4103050713723425	0.5329082048804502	0.5978260869565217	0.31521739130434784	0.5	0.9130434782608695
3	0.3374416246565866	0.5422839565731631	0.06521739130434782	0.9347826086956522	0.1956521739130435	0.619565217391304
4	0.3853621062444304	0.5288890744588135	0.3695652173913043	0.6630434782608695	0.32608695652173914	0.10869565217391304
5	0.3872994302406887	0.5128275697130312	0.4673913043478261	0.6086956521739131	0.5760869565217391	0.29347826086956524
6	0.3785363370790418	0.4925564020666973	0.5869565217391305	0.2717391304347826	0.42391304347826086	0.42391304347826086
7	0.3878144170150324	0.4930967232685667	0.45652173913043476	0.391304347826087	0.6086956521739131	0.5108695652173914
8	0.3577096388905291	0.4439317508611948	0.10869565217391304	0.6413043478260869	0.43478260869565216	0.7608695652173914
9	0.35536630498880081	0.5069207533561273	0.15217391304347827	0.358695652173913	0.17391304347826086	0.2826086956521739
10	0.4893463992873274	0.5110106961975089	0.9021739130434783	0.42391304347826086	0.8369565217391305	0.08695652173913043
11	0.4234970031354333	0.4978124497990768	0.75	0.07608695652173914	0.9239130434782609	0.8478260869565217
12	0.390571734395138	0.7631822190874064	0.34782608695652173	0.5108695652173914	0.5217391304347826	0.44565217391304346
13	0.6696882597906336	0.8140484701346988	0.8695652173913043	0.21739130434782608	0.8586956521739131	0.15217391304347827
14	0.3518642331392193	0.5369725085316731	0.043478260869565216	0.8478260869565217	0.1195652173913043	0.6739130434782609
15	0.3694078177271999	0.7312368506498	0.21739130434782608	0.4891304347826087	0.14130434782608695	0.3804347826086957
16	0.3819549024410497	0.682413467795131	0.2608695652173913	0.7934782608695652	0.2391304347826087	0.7934782608695652
17	0.5287974841645071	0.4998786922854764	0.9565217391304348	0.13043478260869565	0.9673913043478261	0.043478260869565216
18	0.4041015964406261	0.5798644324060511	0.22826086956521738	0.8586956521739131	0.3804347826086957	0.2391304347826087
19	0.3791119630222868	0.6630463719859349	0.010869565217391304	0.9891304347826086	0.15217391304347827	0.34782608695652173
20	0.4138992588363179	0.5041088263279506	0.40217391304347827	0.8695652173913043	0.44565217391304346	0.05434782608695652
21	0.5555665929334382	0.5934571951868796	0.9456521739130435	0.20652173913043478	0.8804347826086957	1.0
22	0.369407817271999	0.733123685064989	0.20652173913043478	0.4782608695652174	0.13043478260869565	0.3695652173913043
23	0.4579827349151042	0.5679247745600222	0.5652173913043478	0.3695652173913043	0.06521739130434782	0.8586956521739131
24	0.4860725342880928	0.6094124889385	0.11956521739130435	1.0	0.021739130434782608	0.16304347826086957
25	0.4417363429447128	0.6571648128891042	0.9130434782608695	0.17391304347826086	0.45652173913043476	0.9347826086956522
26	0.4618535607131468	0.5703227917574084	0.1956521739130435	0.7391304347826086	0.7391304347826086	0.14130434782608695
27	0.4837680461887781	0.493177780269249	0.2717391304347826	0.75	0.07608695652173914	0.4891304347826087
28	0.384622833826543	0.595252882249583	0.44565217391304346	0.3369565217391304	0.6304347826086957	0.6086956521739131
29	0.35536630498880081	0.5069207533561273	0.14130434782608695	0.34782608695652173	0.16304347826086957	0.2717391304347826
30	0.3688659813696363	0.686784912654348	0.13043478260869565	0.8152173913043478	0.4673913043478261	0.5434782608695652
31	0.442855762527034	0.5005249069952125	0.9239130434782609	0.03260869565217391	0.9456521739130435	0.7282608695652174
32	0.4938591229170188	0.7582272094023716	0.5543478260869565	0.6195652173913043	0.782608695652174	0.45652173913043476
33	0.3491473414603732	0.4563850672773969	0.18478260869565216	0.6521739130434783	0.2826086956521739	0.8804347826086957
34	0.3665318351054344	0.5016796505382675	0.16304347826086957	0.9239130434782609	0.18478260869565216	0.40217391304347827
35	0.3650051562704427	0.7426203637024631	0.25	0.7173913043478261	0.2608695652173913	0.21739130434782608
36	0.4666883352563641	0.6210956353278797	0.782608695652174	0.25	0.7173913043478261	0.13043478260869565
37	0.4323310994527001	0.4705217137415431	0.5760869565217391	0.532608695652174	0.6847826086956522	0.06521739130434782
38	0.4462625422933299	0.5182144942246213	0.7282608695652174	0.1956521739130435	0.6739130434782609	0.1956521739130435
39	0.6041926722499749	0.4319203074365276	1.0	0.18478260869565216	0.9782608695652174	0.03260869565217391
40	0.4624319486049305	0.4946859617004399	0.7717391304347826	0.5652173913043478	0.8152173913043478	0.11956521739130435
41	0.377082818577254	0.5247417235902571	0.6521739130434783	0.30434782608695654	0.4130434782608957	0.4673913043478261
42	0.3778479827061649	0.6215339554985552	0.717391304347826	0.15217391304347827	0.6521739130434783	0.25
43	0.4099339652389882	0.4419634867146152	0.6630434782608695	0.282608695652139	0.6413043478260869	0.8913043478260869
44	0.3924883417538096	0.5008125367107679	0.7934782608695652	0.08695652173913043	0.7065217391304348	0.5543478260869565
45	0.4405221977812263	0.4395572269999957	0.5434782608695652	0.8804347826086957	0.7608695652173914	0.07608695652173914
46	0.5043339980827055	0.470196376473585 8	0.7391304347826086	0.8913043478260869	0.8478260869565217	0.391304347826087
47	0.4215335860414435	0.5106850176500888	0.8586956521739131	0.010869565217391304	0.9347826086956522	0.8695652173913043
48	1.0	0.6474830510890315	0.9782608695652174	0.043478260869565216	1.0	0.010869565217391304
49	0.3518642331392193	0.5369725085316731	0.03260869565217391	0.8369565217391305	0.10869565217391304	0.6630434782608695
50	0.4419401033504994	0.5030529143492945	0.2826086956521739	0.5217391304347826	0.05434782608695652	0.7391304347826086
51	0.4033855918391829	0.6598923820748936	0.2391304347826087	0.9673913043478261	0.5434782608695652	0.5760869565217391
52	0.3822550231955878	0.4713421468101024	0.4239130434782086	0.2391304347826087	0.4782608695652174	0.9456521739130435
53	0.3860005517041793	0.4762322921794908	0.6739130434782609	0.7282608695652174	0.5978260869565217	0.8043478260869565
54	0.3699039884648889	0.4462699094122325	0.2934782086956524	0.6304347826086957	0.5652173913043478	0.5
55	0.5036067809339307	0.6665862282899612	0.6304347826086957	0.6739130434782609	0.8043478260869565	0.41304347826086957
56	0.4579300387875582	0.6038250991262275	0.6956521739130435	0.9021739130434783	0.7282608695652174	0.2608695652173913
57	0.5254919420723396	0.8913514573872304	0.8478260869565217	0.5760869565217391	0.08695652173913043	0.9565217391304348
58	0.3896933582976876	0.5604319826382779	0.6086956521739131	0.3804347826086957	0.5543478260869565	0.5326086956521741
59	0.436205839619663	0.5824170128223185	0.5108695652173914	0.7717391304347826	0.6956521739130435	0.09782608695652174
60	0.3635785356365581	0.5048150520871 1	0.532608695652174	0.1413043478260869	0.6630434782608695	0.6304347826086957
61	0.4305228096256412	0.5150701086387863	0.8260869565217391	0.7065217391304348	0.7717391304347826	0.31521739130434784
62	0.8076807191197215	0.7220180747954938	0.9891304347826086	0.10869565217391304	0.9891304347826086	0.021739130434782608
63	0.3482864895515718	0.5132030287731661	0.3260869565217391	0.5978260869565217	0.22826086956521738	0.5217391304347826
64	0.4268687395720961	1.0	0.8152173913043478	0.06521739130434782	0.9130434782608695	0.7065217391304348
65	0.3697637771286677	0.4614986315297532	0.391304347826087	0.5543478260869565	0.358695652173913	0.75
66	0.4922145417966609	0.7931347131326463	0.6195652173913043	0.6847826086956522	0.043478260869565216	0.9239130434782609
67	0.5490299834446901	0.6623583110444867	0.9347826086956522	0.11956521739130435	0.8695652173913043	0.9891304347826086
68	0.3649821645146871	0.4605715236134938	0.41304347826086957	0.4347826086956516	0.33695652173913043	0.717391304347826
69	0.3852936091239937	0.668033175679946	0.4891304347826087	0.5869565217391305	0.3695652173913043	0.33695652173913043
70	0.6308441894121475	0.5573380333169351	0.967391304347826	0.16304347826086957	0.8913043478260869	0.9782608695652174
71	0.3518642331392193	0.5369725085316731	0.021739130434782608	0.8260869565217391	0.09782608695652174	0.6521739130434783
72	0.3905717343951 38	0.7631822190874064	0.3369565217391304	0.5	0.5108695652173914	0.43478260869565216
73	0.4934426414030636	0.5188681951130686	0.8913043478260869	0.09782608695652174	0.9565217391304348	0.22826086956521738
74	0.3745696387360422	0.4755903349317549	0.3804347826086957	0.5434782608695652	0.4891304347826087	0.5652173913043478
75	0.4330689864622602	0.539035784826496	0.31521739130434784	0.782608695652174	0.03260869565217391	0.8152173913043478
76	0.5030193613825352	0.4840575618581642	0.7065217391304348	0.45652173913043476	0.21739130434782608	0.2065217391304348
77	0.4546370441039639	0.6462633725112092	0.8804347826086957	0.4673913043478261	0.391304347826087	0.967391304347826
78	0.410174315835126	0.4321563394805945	0.6413043478260869	0.22826086956521738	0.5869565217391305	0.7717391304347826
79	0.354605773868706	0.4328742683010927	0.17391304347826086	0.9130434782608695	0.31521739130434784	0.9021739130434783
80	0.3445820867190598	0.5132473995519444	0.0869565217391304	0.40217391304347827	0.29347826086956524	0.8260869565217391
81	0.3434045757189707	0.4650580165966191	0.05434782608695652	0.9565217391304348	0.20652173913043478	0.6847826086956522
82	0.4334440733634472	0.611751108095636	0.4782608695652174	0.7608695652173914	0.532608695652174	0.4782608695652174
83	0.3604286044148874	0.5127090060988878	0.5	0.44565217391304346	0.34782608695652173	0.358695652173913
84	0.3713714319823823	0.5510849725723602	0.5217391304347826	0.2934782608695524	0.40217391304347827	0.6413043478260869
85	0.4949395599765118	0.6686424711544973	0.8369565217391305	0.6956521739130435	0.8260869565217391	0.3043478260869654
86	0.4268687395720961	1.0	0.8043478260869565	0.05434782608695652	0.9021739130434783	0.6956521739130435
87	0.4303307773850281	0.6182393351919186	0.3043478260869654	0.9456521739130435	0.25	0.1739130434782086
88	0.384622833826543	0.595252882249583 6	0.43478260869565216	0.32608695652173914	0.6195652173913043	0.5978260869565217
89	0.4054410613093537	0.4567029686864207	0.7608695652173914	0.021739130434782608	0.7934782608695652	0.32608695652173914
90	0.3958628890477925	0.451667101710117	0.6847826086956522	0.2608695652173913	0.75	0.782608695652174
91	0.4645500564152561	0.509897275454781	0.07608695652173914	0.9782608695652174	0.010869565217391304	0.5869565217391305
92	0.3778099722409845	0.4925723675630262	0.358695652173913	0.8043478260869565	0.2717391304347826	0.18478260869565216

APÊNDICE B – Medidas do Capital Social - RA

id	OS1	OS2	IsF	LOF	COV	SVM
1	0.262263484546854	0.27179250253603 17	0.05434782608695652	0.05434782608695652	0.7717391304347826	0.2717391304347826
2	0.27014270696885 78	0.30860374813870 54	0.7934782608695652	0.7282608695652174	0.3152173913043478 4	0.1413043478260895
3	0.27473639702946 91	0.31495993855727 58	0.17391304347826086	0.5108695652173914	0.1847826086956521 6	0.3369565217391304 3
4	0.25139891512673 63	0.29270241017964 12	0.5978260869565217	0.9021739130434783	0.5978260869565217	0.6630434782608695
5	0.42658286555822 722	0.71821530129436 47	0.9021739130434783	0.9239130434782609	0.0326086956521739 1	0.3804347826086957
6	0.31731992725981 38	0.28658467883420 98	0.8369565217391305	0.967391304347826	0.9782608695652174	0.7717391304347826
7	0.25105902623764 07	0.26841588315687 829	0.25	0.18478260869565216	0.2826086956521739	0.5978260869565217
8	0.34995966721984 95	0.31449851955565 16	0.6195652173913043	0.7608695652173914	0.7282608695652174	0.0652173913043782
9	0.36901934586255 15	0.29952815726605 4	0.8586956521739131	0.9456521739130435	1.0	0.0543478260869565 2
10	0.33851407551644 99	0.33102882523526 08	0.7282608695652174	0.2608695652173913	0.6521739130434783	0.9130434782608695
11	0.62203310420626 19	1.0	0.9782608695652174	0.6630434782608695	0.0543478260869565 2	0.0326086956521739 1
12	0.29774164036139 46	0.33049579379323 46	0.41304347826086957	0.043478260869565216	0.4456521739130434 6	0.7173913043478261
13	0.25329417114224 44	0.31840433256138 18	0.6086956521739131	0.8260869565217391	0.532608695652174	0.6413043478260869
14	0.28004183625261 81	0.29061385259434 84	0.15217391304347827	1.0	0.8695652173913043	0.1630434782608695 7
15	0.35218811568336 8	0.26319024797043 51	0.3043478260869565 4	0.8043478260869565	0.8586956521739131	0.3913043478260867
16	0.34161207609820 96	0.33064463029339 29	0.7608695652173914	0.010869565217391304	0.1521739130434782 7	0.9565217391304348
17	0.37988076716713 6	0.33868442364371 17	0.4782608695652174	0.3478260869565217 3	0.0652173913043478 2	0.7608695652173914
18	0.30078025806228	0.37095458491692 58	0.33695652173913043	0.31521739130434784	0.4021739130434782 7	0.7282608695652174
19	0.28085152627515 47	0.34826441353707 4	0.6521739130434783	0.8478260869565217	0.3586956521739313	0.1739130434782608 6
20	0.94024986021429 64	0.95294866092468 08	1.0	0.5217391304347826	0.043478260869565216	0.1195652173913043 5
21	0.24092820495254 87	0.44367337540389 97	0.6413043478260869	0.7391304347826086	0.3043478260869565 4	0.6304347826086957
22	0.27022065542206 61	0.29337318184685 77	0.2608695652173913	0.6847826086956522	0.2934782608695652 4	0.2173913043478260 8
23	0.25216349381745 72	0.30998108175170 59	0.5108695652173914	0.782608695652174	0.6630434782608695	0.3586956521739313
24	0.24701053046073 16	0.32779907436579 52	0.11956521739130435	0.5869565217391305	0.6956521739130435	0.5
25	0.24766862604233 75	0.45112631455049 88	0.8260869565217391	0.9782608695652174	0.1956521739130435	0.6195652173913043
26	0.31355974394565 19	0.42549744927955 032	0.9130434782608695	0.7717391304347826	0.4130434782608695 7	0.4239130434782686
27	0.25059122198289 22	0.30817395878888 08	0.5652173913043478	0.9347826086956522	0.5	0.6739130434782609
28	0.25935385129443 59	0.32725095851888 54	0.8478260869565217	0.4239130434782686	0.6847826086956522	0.1304347826086956 5
29	0.31804581775006 38	0.29623692466726 89	0.358695652173913	0.44565217391304346	0.9239130434782609	0.8260869565217391
30	0.26493805572396 12	0.25800933366941 49	0.08695652173913043	0.08695652173913043	0.5543478260869565	0.3043478260869565 4
31	0.40949833646947 97	0.58744384399367 37	0.717391304347826	0.20652173913043478	0.08695652173913043	0.9891304347826086
32	0.26493805572396 12	0.25800933366941 49	0.07608695652173914	0.07608695652173914	0.5434782608695652	0.2934782608695652 4
33	0.26380947351873 36	0.32904150129772 12	0.5543478260869565	0.8586956521739131	0.7173913043478261	0.3478260869565217 3
34	0.26447736429936 07	0.26286589339435 978	0.09782608695652174	0.09782608695652174	0.4891304347826087	0.3152173913043478 4
35	0.29862383226293 41	0.32549624193557 83	0.391304347826087	0.33695652173913043	0.4782608695652174	0.6956521739130435
36	0.24113021270331 65	0.32779907436579 52	0.2826086956521739	0.6086956521739131	0.7608695652173914	0.4782608695652174
37	0.25105902623764 07	0.26841588315687 829	0.23913043478260087	0.17391304347826086	0.2717391304347826	0.5869565217391305
38	0.29782587823399 3	0.29391489545825 26	0.032608695652173913	0.25	0.9021739130434783	0.2391304347826087
39	0.40949833646947 97	0.58744384399367 37	0.7065217391304348	0.1956521739130435	0.0760869565217391 4	0.9782608695652174
40	0.32304994848587 393	0.36763779905796 94	0.32608695652173914	0.1086956521739130 4	0.3804347826086957	0.4130434782608697
41	0.29698433515361 19	0.29391489545825 26	0.043478260869565216	0.23913043478260087	0.8804347826086957	0.2608695652173913
42	0.32032484210628 7	0.28285057428277 98	0.5	0.40217391304347827	0.967391304347826	0.8043478260869565
43	0.30481838426335 88	0.40913190951038 9	0.8804347826086957	0.6521739130434783	0.9456521739130435	0.9021739130434783
44	0.30162249187889 96	0.33808644136304 19	0.8695652173913043	0.7934782608695652	0.3695652173913043	0.782608695652174
45	0.42658286558222 722	0.71821530129436 47	0.8913043478260869	0.9130434782608695	0.0217391304347826 08	0.3695652173913043
46	0.26478573964153 74	0.29389395458136 7	0.6304347826086957	0.8369565217391305	0.7934782608695652	0.1521739130434782 7
47	0.67815068996307 89	0.99783713850836 62	0.9673913043478260	0.6413043478260869	0.0108695652173913 04	0.0217391304347826 08
48	0.43903807153643 1	0.47253765538031 255	0.9456521739130435	0.8152173913043478	0.6086956521739131	0.0434782608695652 16
49	0.25139891512673 63	0.29270241017964 12	0.5869565217391305	0.8913043478260869	0.5869565217391305	0.6521739130434783
50	0.35474344178484 39	0.28979211816927 63	0.9565217391304348	0.8804347826086957	0.8478260869565217	0.1086956521739130 4
51	0.49973915326146	0.90806168101685 4	0.9239130434782609	0.4565217391304347 6	0.7391304347826086	0.0978260869565174
52	0.30092638075469 47	0.29197514557255 32	0.532608695652174	0.6304347826086957	0.3478260869565173	0.8913043478260869
53	0.28965148111383 89	0.29770704556220 91	0.6739130434782609	0.532608695652174	0.8369565217391305	0.7934782608695652
54	0.30537199900026 32	0.29827384579455 35	0.42391304347826086	0.358695652173913	0.1195652173913043 5	0.75
55	0.29774164036139 46	0.33049579379323 46	0.40217391304347827	0.03260869565217391	0.4347826086955216	0.7065217391304348
56	0.30465945978376 11	0.29827384579455 35	0.4673913043478261	0.3804347826086957	0.1304347826086955	0.7391304347826086
57	0.25105902623764 07	0.26841588315687 829	0.22826086956521738	0.16304347826086957	0.2608695652173913	0.5760869565217391
58	0.24470295854014 24	0.32779907436579 52	0.14130434782608695	0.5652173913043478	0.5760869565217391	0.5217391304347826
59	0.33872664800228 12	0.33362809224415 495	0.6847826086956522	0.2717391304347826	0.5108695652173914	0.9347826086956522
60	0.29315671936777 51	0.28658236592323 67	0.2934782608695652 4	0.9891304347826086	0.1086956521739130 4	0.2065217391304347 8
61	0.38324721344988 28	0.41300812319569 32	0.8043478260869565	0.21739130434782608	0.0978260869565174	1.0
62	0.27473639702946 91	0.31495993855727 58	0.16304347826086957	0.5	0.1739130434782608 6	0.3260869565217391 4
63	0.29851278170512 57	0.29391489545825 26	0.010869565217391304	0.2826086956521739	0.9130434782608695	0.2282608695652173 8
64	0.29405460382708 75	0.26788830786570 79	0.44565217391304346	0.4891304347826087	0.8152173913043478	0.8478260869565217
65	0.29775276964726 75	0.29391489545825 26	0.021739130434782608	0.22826086956521738	0.8913043478260869	0.25
66	0.25105902623764 07	0.26841588315687 829	0.21739130434782608	0.15217391304347827	0.25	0.5652173913043478
67	0.24225428478510 63	0.34935231498295 9	0.3695652173913043	0.6739130434782609	0.8260869565217391	0.4673913043478261
68	0.33649158231139 08	0.33102882523526 08	0.7717391304347826	0.30434782608695654	0.5217391304347826	0.9239130434782609
69	0.26616847491405 01	0.26286589339435 978	0.06521739130434782	0.06521739130434782	0.6304347826086957	0.2826086956521739
70	0.25105902623764 07	0.26841588315687 829	0.20652173913043478	0.14130434782608695	0.2391304347826087	0.5543478260869565
71	0.33753508792482 41	0.33362809224415 495	0.6956521739130435	0.2934782608695652 4	0.4565217391304346	0.9456521739130435
72	0.32478685867687 78	0.36274761859994 24	0.31521739130434784	0.391304347826087	0.391304347826087	0.4021739130434782 7
73	0.24166611190371 25	0.32779907436579 52	0.2717391304347826	0.5760869565217391	0.782608695652174	0.4891304347826087
74	0.30092638075469 47	0.29197514557255 32	0.5217391304347826	0.6195652173913043	0.3369565217391343	0.8804347826086957
75	0.28008501077117 8	0.36582214428563 88	0.75	0.717391304347826	0.9891304347826086	0.4456521739130434 6
76	0.34402662015831 61	0.33427506045086 11	0.7391304347826086	0.021739130434782608	0.1413043478260869 5	0.967391304347826
77	0.24014234597249 36	0.34935231498295 9	0.5760869565217391	0.9565217391304348	0.7065217391304348	0.4347826086956521 6
78	0.25105902623764 07	0.26841588315687 829	0.1956521739130435	0.13043478260869565	0.2282608695652173 8	0.5434782608695652
79	0.27501491468229 22	0.34136936589008 61	0.782608695652174	0.75	0.9565217391304348	0.4565217391304346
80	1.0	0.95126885819990 88	0.9891304347826086	0.6956521739130435	0.6195652173913043	0.0108695652173913 04
81	0.34748887762711 33	0.32142366314241 21	0.9347826086956522	0.3695652173913043	0.1630434782608695 7	0.0869565217391304 3
82	0.29175469588050 53	0.25608751822937 97	0.4891304347826087	0.5434782608695652	0.4239130434782686	0.8695652173913043
83	0.26128662422930 87	0.27179250255306 17	0.10869565217391304	0.8695652173913043	0.75	0.1956521739130435
84	0.24470295854014 24	0.32779907436579 52	0.13043478260869565	0.5543478260869565	0.5652173913043478	0.5108695652173914
85	0.31970586045119 68	0.29776340309654 17	0.3478260869565173	0.4347826086956522	0.9347826086956522	0.8152173913043478
86	0.25133974506439 07	0.31430281009355 1	0.5434782608695652	0.5978260869565217	0.6413043478260869	0.1847826086956521 6
87	0.35567106781053	0.35122183024038 74	0.8152173913043478	0.4130434782608695 7	0.3260869565217391 4	0.0760869565217391 4
88	0.25105902623764 07	0.26841588315687 829	0.18478260869565216	0.11956521739130435	0.2173913043478260 8	0.532608695652174
89	0.29405460382708 75	0.26788830786570 79	0.4347826086955216	0.4782608695652174	0.8043478260869565	0.8369565217391305
90	0.23662737897186 17	0.39237228059736 27	0.6630434782608695	0.7065217391304348	0.2065217391304347 8	0.6086956521739131
91	0.29862383226293 41	0.32549624193557 83	0.380434782608695 7	0.32608695652173914	0.4673913043478261	0.6847826086956522
92	0.29245338663482 97	0.25836886631019 39	0.45652173913043476	0.4673913043478261	0.6739130434782609	0.8586956521739131

APÊNDICE C – Medidas do Capital Misto - RA

id	OS1	OS2	IsF	LOF	COV	SVM
1	0.3883343675425676	0.3475587663714927	0.021739130434782608	0.44565217391304346	0.20652173913043478	0.6086956521739131
2	0.4389558373494304	0.359253374612018	0.6086956521739131	0.2826086956521739	0.5869565217391305	0.8369565217391305
3	0.3844269621104508	0.41737781898145 26	0.010869565217391304	0.9347826086956522	0.16304347826086957	0.5760869565217391
4	0.4151905839982448	0.3610246670518249	0.21739130434782608	0.5217391304347826	0.1956521739130435	0.5434782608695652
5	0.4772988509938529	0.3949580096893177	0.45652173913043476	0.8804347826086957	0.43478260869565216	0.9347826086956522
6	0.4306247260143743	0.3359137025978163	0.5	0.1847826086956216	0.4891304347826087	0.44565217391304346
7	0.4187893346607322	0.3256574921362487	0.22826086956521738	0.42391304347826086	0.5	0.391304347826087
8	0.427855065826155	0.3302184845026102	0.32608695652173914	0.7391304347826086	0.5978260869565217	0.6739130434782609
9	0.4327387146180504	0.3306962666505123	0.34782608695652173	0.8913043478260869	0.5543478260869565	0.9782608695652174
10	0.5278805547254678	0.4273386413015279	0.9347826086956522	0.5652173913043478	0.8152173913043478	0.20652173913043478
11	0.5853537648527906	0.9511197915824656	0.9130434782608695	0.06521739130434782	0.967391304347826	0.13043478260869565
12	0.4352385880767013	0.512812373516258	0.3804347826086957	0.8043478260869565	0.5434782608695652	0.2826086956521739
13	0.6667145200675555	0.507844286035308	0.8260869565217391	0.3369565217391304 3	0.8586956521739131	0.33695652173913043
14	0.3989693994476183	0.3461524483849675	0.03260869565217391	0.8478260869565217	0.09782608695652174	0.9021739130434783
15	0.4339655386702046	0.3940941878353615	0.3586956521739313	0.43478260869565216	0.2608695652173913	0.7608695652173914
16	0.4427374447947091	0.3764417667126462	0.5543478260869565	0.8369565217391305	0.6956521739130435	0.8695652173913043
17	0.5714710030488441	0.3449370279154808	0.967391304347826	0.14130434782608695	0.5217391304347826	0.09782608695652174
18	0.4462372988891511	0.3630433446034902	0.18478260869565216	0.6304347826086957	0.4130434782086957	0.40217391304347827
19	0.4167489920740333	0.4028044721313666	0.06521739130434782	1.0	0.13043478260869565	0.8586956521739131
20	0.7509587754815592	0.6859244390528857	0.8586956521739131	0.9239130434782609	1.0	0.8913043478260869
21	0.5594878162891797	0.5116139889852909	0.8913043478260869	0.15217391304347827	0.8695652173913043	0.021739130434782608
22	0.4083099701882534	0.4234138193234582	0.14130434782608695	0.2608695652173913	0.15217391304347827	0.967391304347826
23	0.4764134151367614	0.399688725591492	0.5108695652173914	0.16304347826086957	0.010869565217391304	0.9565217391304348
24	0.5018997756663717	0.386637446847801	0.2391304347826087	0.6847826086956522	0.021739130434782608	0.8260869565217391
25	0.4599928517921551	0.4060792523869998	0.8152173913043478	0.2934782608956524	0.42391304347826086	0.5978260869565217
26	0.4944090980405429	0.4252148485451607	0.6413043478260869	0.782608695652174	0.9456521739130435	0.18478260869565216
27	0.498803878481178	0.4028109569739028	0.5434782608695652	0.75	0.07608695652173914	0.43478260869565216
28	0.417966126182131	0.3272397658065195	0.5652173913043478	0.4891304347826087	0.8478260869565217	0.7934782608695652
29	0.4165470691061693	0.3290873540789171	0.09782608695652174	0.7717391304347826	0.2826086956521739	1.0
30	0.4082176219705959	0.3491963134692837	0.043478260869565216	0.9130434782608695	0.3695652173913043	0.4891304347826087
31	0.5124219173516432	0.6116976893594654	0.9239130434782609	0.08695652173913043	0.45652173913043476	0.10869565217391304
32	0.5150606377871831	0.421905562559159	0.42391304347826086	0.6956521739130435	0.7391304347826086	0.17391304347826086
33	0.3883325553361559	0.3330244674169074	0.07608695652173914	0.5434782608695652	0.14130434782608695	0.358695652173913
34	0.4059234467910276	0.382525612251372	0.05434782608695652	0.4673913043478261	0.21739130434782608	0.7391304347826086
35	0.4145329565735165	0.446602234340527	0.33695652173913043	0.6413043478260869	0.33695652173913043	0.75
36	0.4829414786236758	0.4199085154614935	0.7608695652173914	0.4130434782608695 7	0.6195652173913043	0.5108695652173914
37	0.4568582618414517	0.4220183336336368	0.4673913043478261	0.5869565217391305	0.5760869565217391	0.4673913043478261
38	0.4801724772224617	0.3635987330618535	0.6739130434782609	0.1956521739130435	0.4782608695652174	0.4782608695652174
39	0.6443133789311151	0.4659030792601501	0.9782608695652174	0.3478260869565173	0.8369565217391305	0.06521739130434782
40	0.4990560976538854	0.3911067325925293	0.7717391304347826	0.7065217391304348	0.782608695652174	0.1956521739130435
41	0.4238599935869104	0.3585333533738009	0.4130434782086957	0.11956521739130435	0.2717391304347826	0.9130434782608695
42	0.4320642472798232	0.320430847936046	0.6521739130434783	0.10869565217391304	0.40217391304347827	0.532608695652174
43	0.4507357328033872	0.3192782739322377	0.75	0.30434782608695654	0.5652173913043478	0.3695652173913043
44	0.4340868242187971	0.3379628846857847	0.7391304347826086	0.20652173913043478	0.6304347826086957	0.5217391304347826
45	0.5159621378688637	0.3322234757782339	0.532608695652174	0.9782608695652174	0.6521739130434783	0.41304347826086957
46	0.519687645536342	0.3584734421215509	0.6847826086956522	0.9673913043478261	0.8043478260869565	0.15217391304347827
47	0.6127260154230324	1.0	0.9456521739130435	0.010869565217391304	0.9782608695652174	0.0869565217391304 3
48	1.0	0.6043001378075772	1.0	0.21739130434782608	0.9239130434782609	0.010869565217391304
49	0.3882949627850062	0.3729949843574929	0.1086956521739104	0.45652173913043476	0.06521739130434782	0.7717391304347826
50	0.4887716502064863	0.3421401756161446	0.5760869565217391	0.4782608695652174	0.9347826086956522	0.6956521739130435
51	0.5158504271710 06	0.377457442321521	0.4456521739130436	0.9891304347826086	0.9130434782608695	0.3043478260869564
52	0.4292948263407121	0.3306428364752968	0.40217391304347827	0.5543478260869565	0.7065217391304348	0.6521739130434783
53	0.4299974306872175	0.3435803124719231	0.5869565217391305	0.8152173913043478	0.75	0.9891304347826086
54	0.4222816392508683	0.3666742297918333	0.29347826086956524	0.5	0.532608695652174	0.6847826086956522
55	0.5292473421549719	0.3659903267611156	0.6956521739130435	0.6195652173913043	0.7717391304347826	0.16303437826086957
56	0.4912359681255668	0.5591113039091575	0.717391304347826	0.9021739130434783	0.717391304347826	0.22826086956521738
57	0.5370474557538504	0.778840527758723	0.8043478260869565	0.717391304347826	0.10869565217391304	0.6303478260869657
58	0.4183790092556877	0.469844771361019	0.4891304347826087	0.358695652173913	0.358695652173913	0.45652173913043476
59	0.4850575128915008	0.5161860839633794	0.6303478260869657	0.8260869565217391	0.7934782608695652	0.2608695652173913
60	0.4096836749987862	0.3380581133058872	0.2608695652173913	0.13043478260869565	0.2934782608956524	0.31521739130434784
61	0.4952073198283105	0.3346502856678573	0.7826086956521774	0.5760869565217391	0.7608695652173914	0.3260869565217391 4
62	0.7981498668332672	0.5506488547568291	0.9891304347826086	0.09782608695652174	0.8804347826086957	0.03260869565217391
63	0.4008674395042194	0.3573137003487258	0.13043478260869565	0.07608695652173914	0.08695652173913043	0.8043478260869565
64	0.4667060491516992	0.3914892944379905	0.9021739130434783	0.021739130434782608	0.2391304347826087	0.14130434782608695
65	0.4179399023182166	0.3395692156089755	0.2717391304347826	0.40217391304347827	0.25	0.782608695652174
66	0.5076290344054337	0.7147759895392448	0.5978260869565217	0.5978260869565217	0.05434782608695652	0.717391304347826
67	0.5567028486241887	0.5184786857385763	0.8478260869565217	0.22826086956521738	0.8913043478260869	0.05434782608695652
68	0.4280979006621944	0.3471523006528318	0.6630434782608695	0.3695652173913043	0.6413043478260869	0.5
69	0.4220521367381057	0.3824747170125079	0.1956521739130435	0.6630434782608695	0.32608695652173914	0.5543478260869565
70	0.6321471964797074	0.4697280665596164	0.9565217391304348	0.25	0.9021739130434783	0.043478260869565216
71	0.4199884635 06376	0.4003255803215332	0.3043478260869654	0.532608695652174	0.3152173913043478 4	0.7282608695652174
72	0.4423950619778866	0.5560608929846 55	0.3695652173913043	0.7608695652173914	0.6086956521739131	0.271739130434782 6
73	0.5085750779637821	0.3689639015271 72	0.8695652173913043	0.30434782608956516	0.3043478260869654	0.07608695652173914
74	0.4227704268620754	0.3421994956151781	0.17391304347826086	0.32608695652173914	0.4456521739130446	0.3478260869565173
75	0.4622103107454795	0.3175675208818076	0.25	0.6739130434782609	0.043478260869565216	0.7065217391304348
76	0.5410981222365309	0.3860569479203942	0.7934782608695652	0.6521739130434783	0.6739130434782609	0.847826086956521 7
77	0.4716323425324618	0.3411484353565238	0.6195652173913043	0.7282608695652174	0.6630434782608695	0.6630434782608695
78	0.4367314518651617	0.3617000437194855	0.4782608695652174	0.2391304347826087	0.3804347826086957	0.42391304347826086
79	0.3968917071842395	0.3303226469153387	0.08695652173913043	0.5108695652173914	0.17391304347826086	0.5652173913043478
80	0.7565834886392883	0.7829315299362589	0.5217391304347826	0.3804347826086957	0.9891304347826086	0.6195652173913043
81	0.4105725196372877	0.3534074494243602	0.15217391304347827	0.8586956521739131	0.9565217391304348	0.8152173913043478
82	0.4687292856812057	0.3402502935810391	0.3152173913043784	0.9456521739130435	0.7282608695652174	0.2391304347826087
83	0.3995640401781755	0.3432673534724333	0.16304347826086957	0.05434782608695652	0.11956521739130435	0.9239130434782609
84	0.4028293617015889	0.3437889627830748	0.2826086956521739	0.17391304347826086	0.2282608695652138	0.6413043478260869
85	0.5297812231 96943	0.3691418023559396	0.8369565217391305	0.9565217391304348	0.8260869565217391	0.21739130434782608
86	0.4528138214305464	0.3831659476442045	0.8804347826086957	0.032608695652173 91	0.18478260869565216	0.11956521739130435
87	0.4835874932730643	0.3541771173714114	0.4347826086956521 6	0.7934782608695652	0.4673913043478261	0.3804347826086957
88	0.417914777241829	0.3220261902931285	0.391304347826087	0.2717391304347826	0.5108695652173914	0.8804347826086957
89	0.4475977818747585	0.3493975977554928	0.7065217391304348	0.391304347826087	0.6847826086956522	0.2934782086956524
90	0.4185389776109543	0.3170019798785701	0.7282608695652174	0.31521739130434784	0.34782608695652173	0.25
91	0.4951279349754607	0.3428358907435582	0.11956521739130435	0.8695652173913043	0.03260869565217391	0.5869565217391305
92	0.4248411252562955	0.3391771237688567	0.20652173913043478	0.6086956521739131	0.391304347826087	0.9456521739130435

APÊNDICE D – Resultados de Disrupção

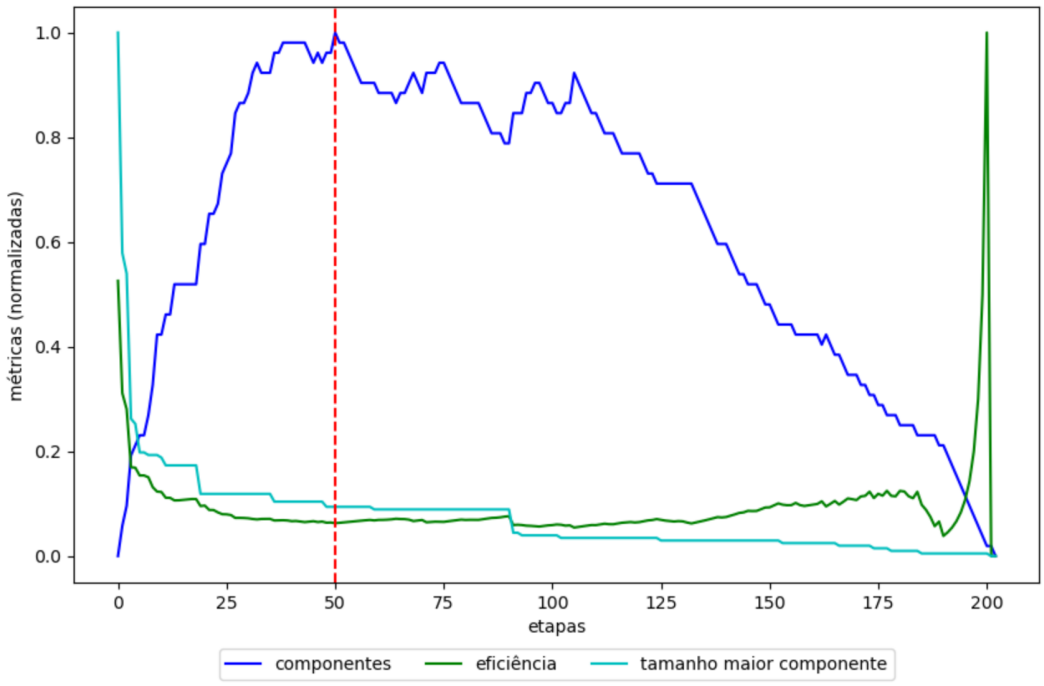


Figura 21 – RA3, com 203 indivíduos e 380 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (53), diminuindo a eficiência da rede (0,02091) e o tamanho do maior componente conectado (20), e causando a disrupção da rede na etapa 50. A linha vertical representa a etapa de disrupção.

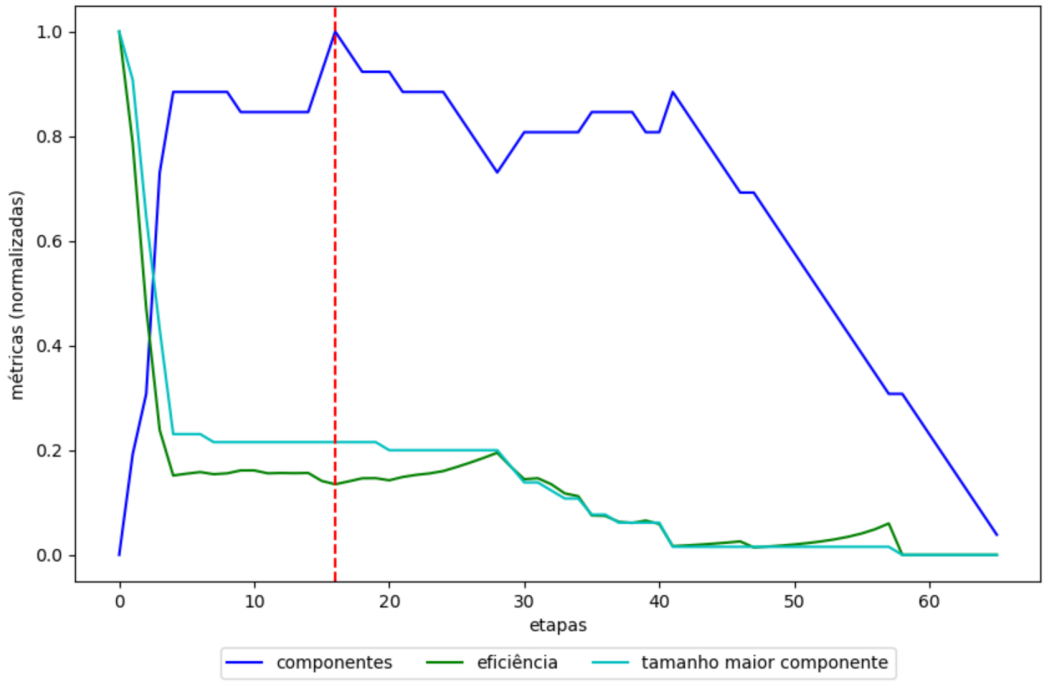


Figura 22 – RA4, com 66 indivíduos e 124 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (27), diminuindo a eficiência da rede (0,04997) e o tamanho do maior componente conectado (15), e causando a disrupção da rede na etapa 16. A linha vertical representa a etapa de disrupção.

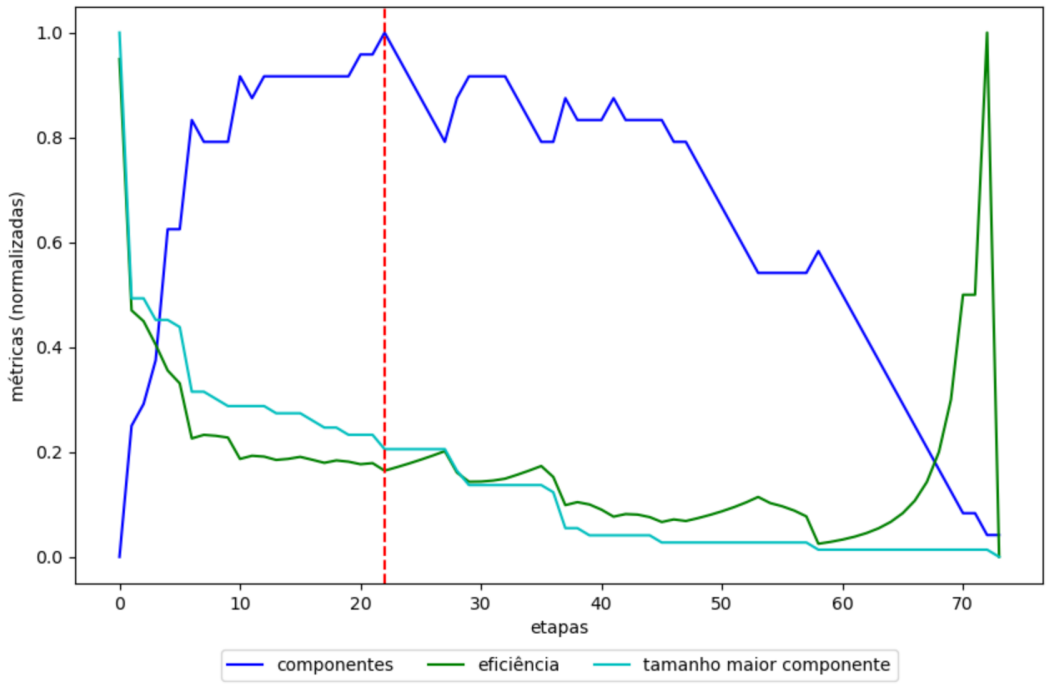


Figura 23 – RA5, com 74 indivíduos e 131 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (25), diminuindo a eficiência da rede (0,05479) e o tamanho do maior componente conectado (16), e causando a disrupção da rede na etapa 22. A linha vertical representa a etapa de disrupção.

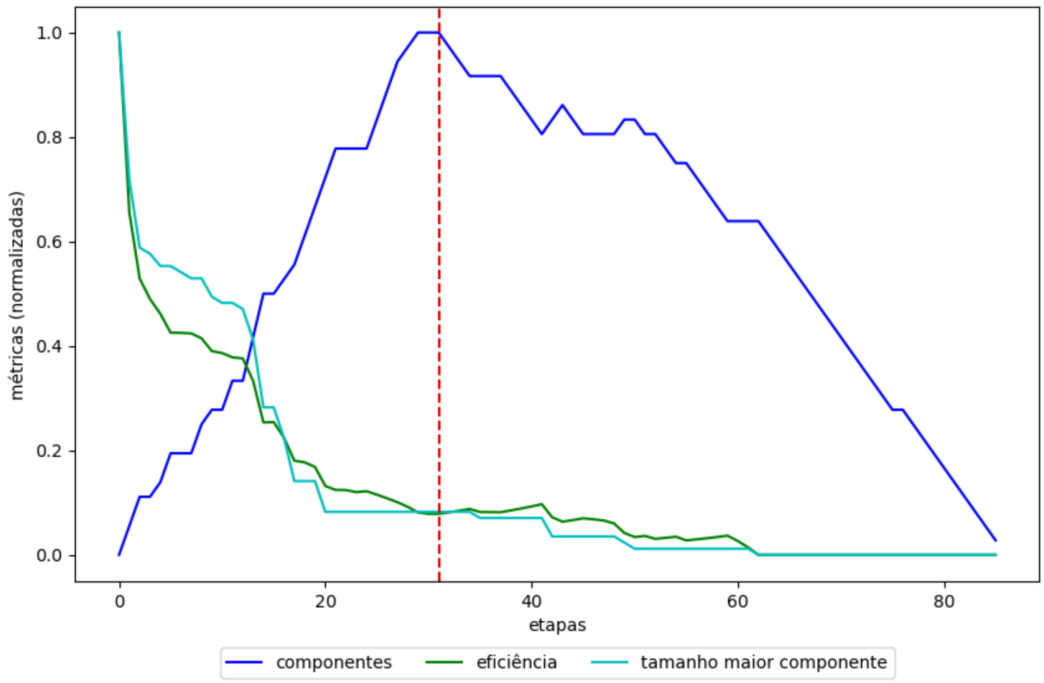


Figura 24 – RA6, com 86 indivíduos e 112 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (37), diminuindo a eficiência da rede (0,01824) e o tamanho do maior componente conectado (8), e causando a disrupção da rede na etapa 31. A linha vertical representa a etapa de disrupção.

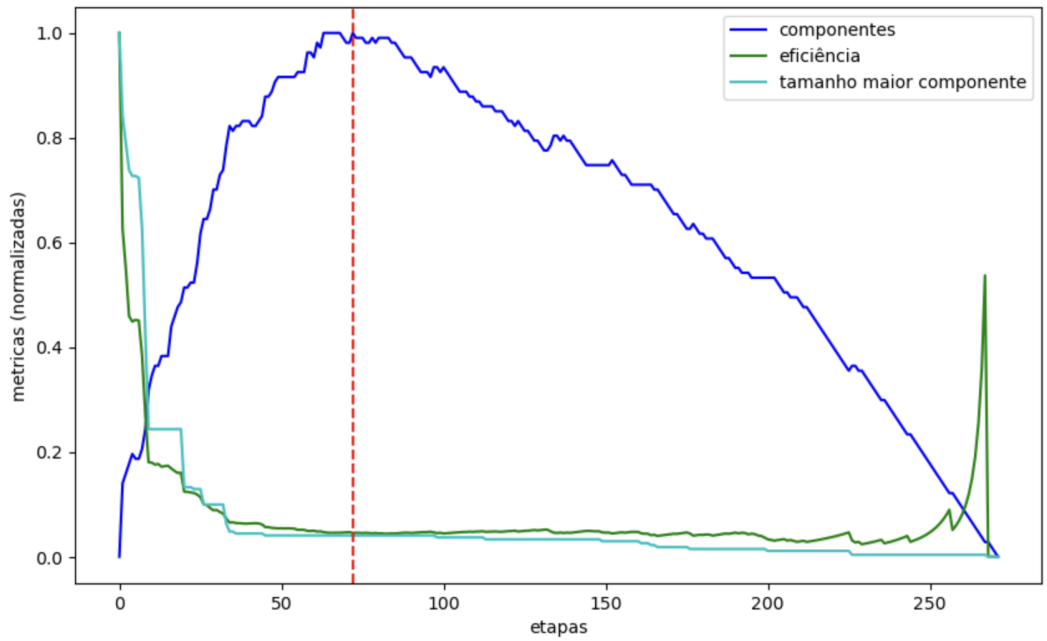


Figura 25 – RA7, com 272 indivíduos e 367 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (108), diminuindo a eficiência da rede (0,00841) e o tamanho do maior componente conectado (12), e causando a disrupção da rede na etapa 72. A linha vertical representa a etapa de disrupção.

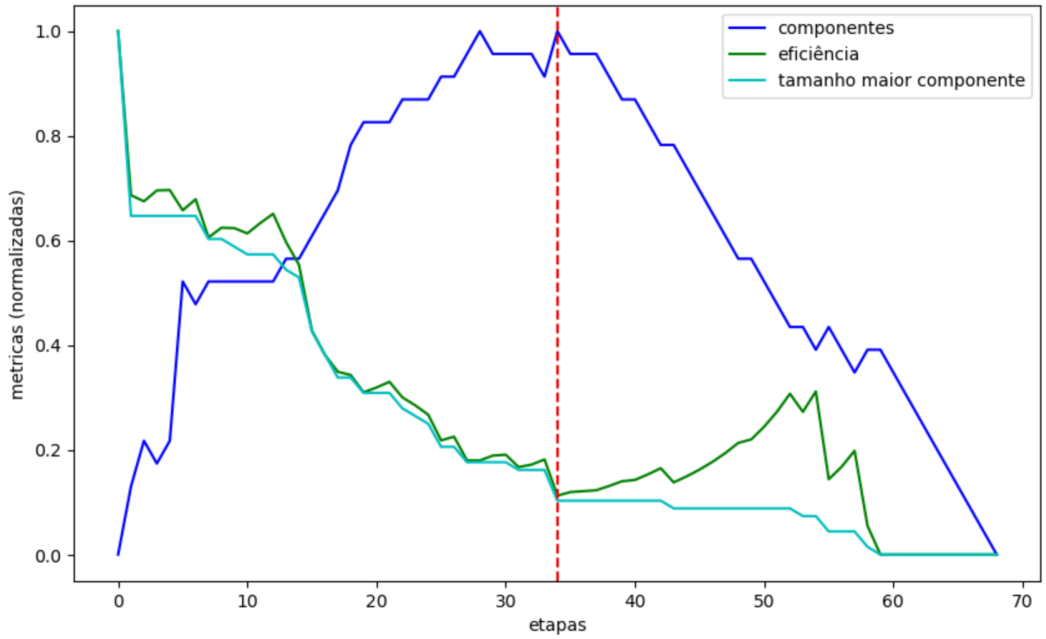


Figura 26 – RA8, com 69 indivíduos e 168 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (24), diminuindo a eficiência da rede (0,03725) e o tamanho do maior componente conectado (8), e causando a disrupção da rede na etapa 34. A linha vertical representa a etapa de disrupção.

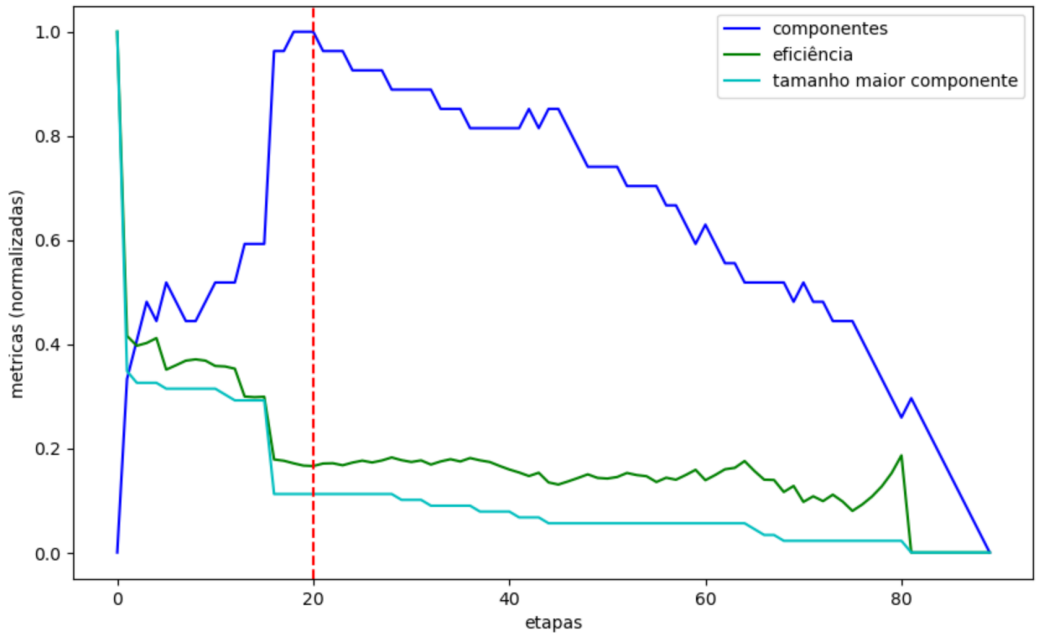


Figura 27 – RA9, com 90 indivíduos e 199 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (28), diminuindo a eficiência da rede (0,04946) e o tamanho do maior componente conectado (11), e causando a disrupção da rede na etapa 20. A linha vertical representa a etapa de disrupção.

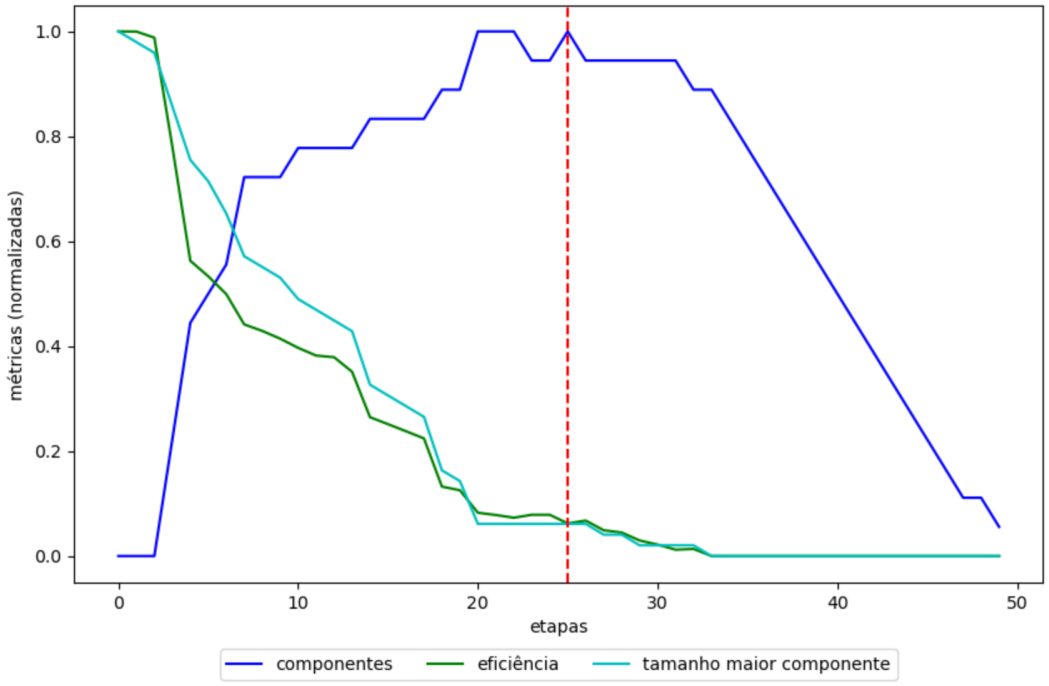


Figura 28 – RA10, com 50 indivíduos e 143 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (19), diminuindo a eficiência da rede (0,03) e o tamanho do maior componente conectado (4), e causando a disrupção da rede na etapa 25. A linha vertical representa a etapa de disrupção.

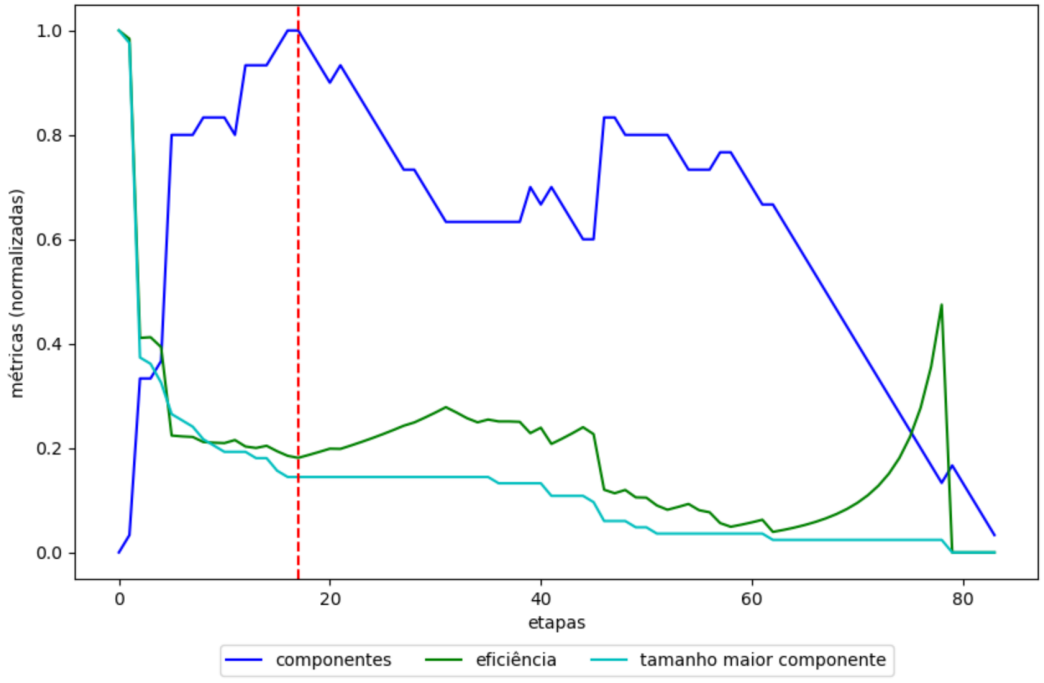


Figura 29 – RA11, com 84 indivíduos e 105 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (31), diminuindo a eficiência da rede (0,04545) e o tamanho do maior componente conectado (13), e causando a disrupção da rede na etapa 17. A linha vertical representa a etapa de disrupção.

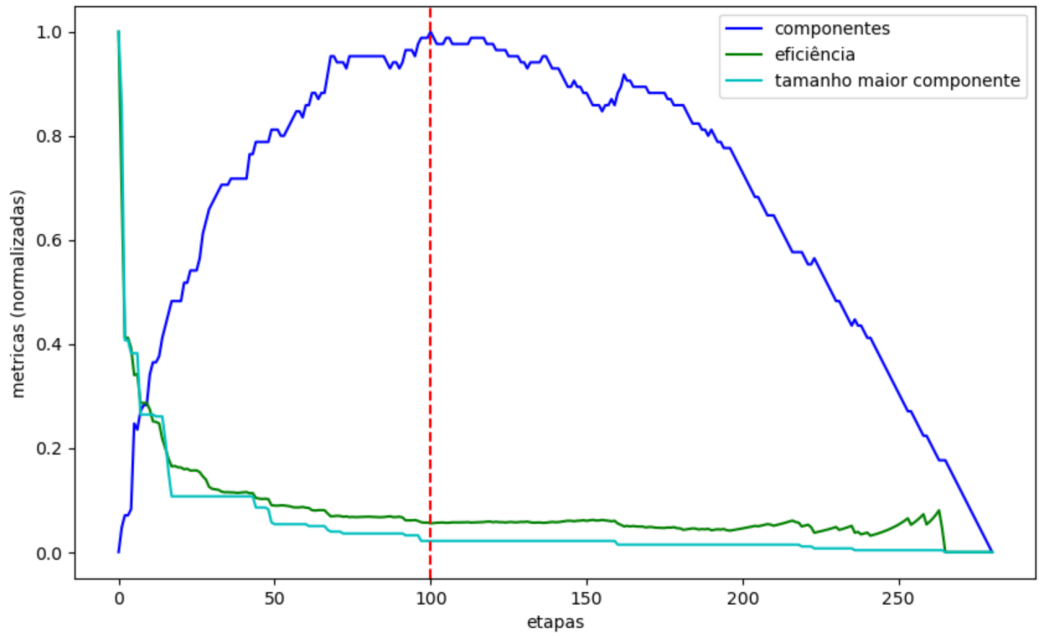


Figura 30 – RA12, com 281 indivíduos e 474 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (87), diminuindo a eficiência da rede (0,00898) e o tamanho do maior componente conectado (7), e causando a disrupção da rede na etapa 100. A linha vertical representa a etapa de disrupção.

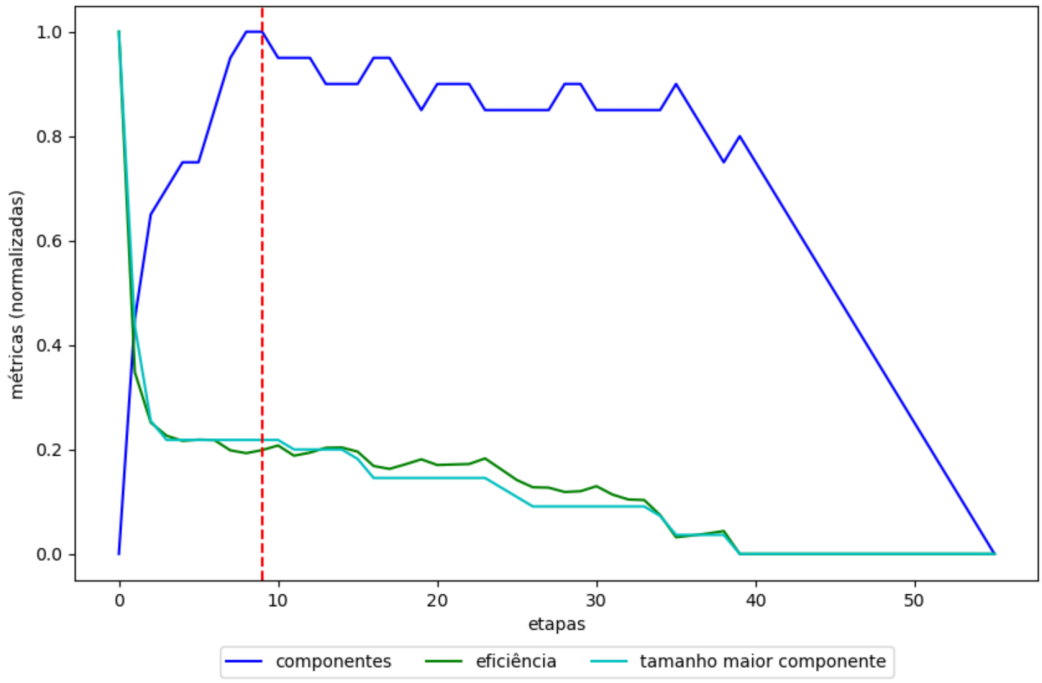


Figura 31 – RA13, com 56 indivíduos e 104 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (21), diminuindo a eficiência da rede (0,07431) e o tamanho do maior componente conectado (13), e causando a disrupção da rede na etapa 9. A linha vertical representa a etapa de disrupção.

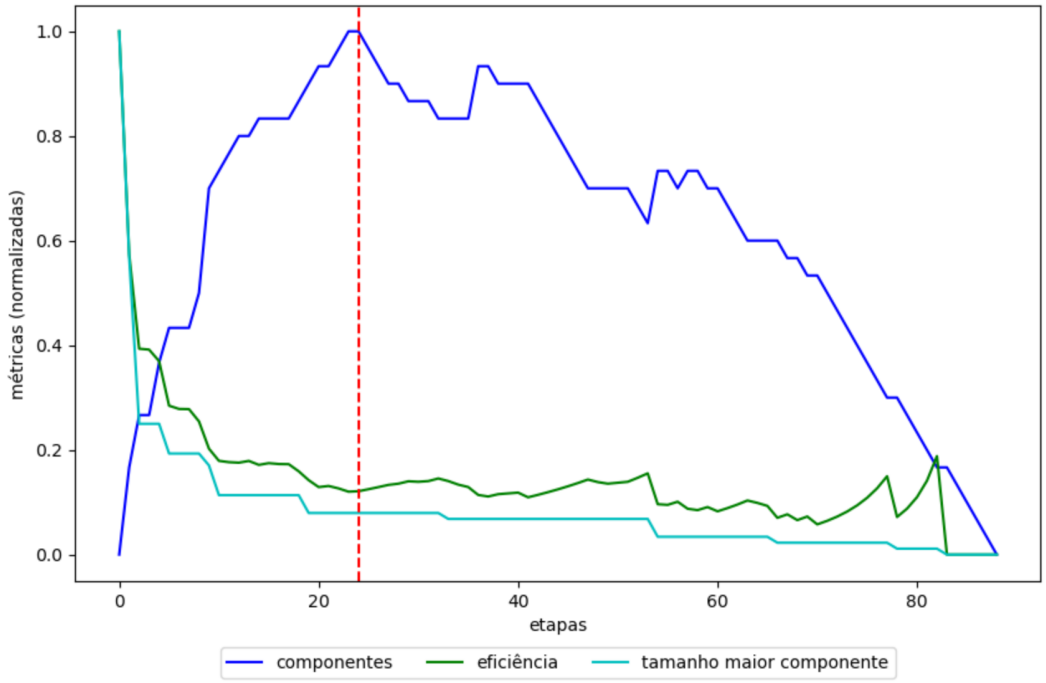


Figura 32 – RA14, com 89 indivíduos e 156 ligações - na disrupção pelo capital social o modelo mais eficaz é o OS1, aumentando o número de componentes conectados (16), diminuindo a eficiência da rede (0,14143) e o tamanho do maior componente conectado (37), e causando a disrupção da rede na etapa 29. A linha vertical representa a etapa de disrupção.

Anexos

ANEXO A – Memorando PMMG

POLÍCIA MILITAR
DE MINAS GERAIS
COMANDO GERAL

Memorando nº 91341 /2019 – CG.

Belo Horizonte, 23 de abril de 2019.

Ao : Diretor de Inteligência e Diretor de Tecnologia e Sistemas da PMMG.
Assunto : Autorização de pesquisa.

1 CONSIDERANDO QUE:

1.1 Deu entrada neste Subcomando-Geral solicitação do nº 095.606-0, Cel QOR Alex Sander de Oliveira Toledo, pertencente ao Centro de Administração de Pessoal (CAP), para acessar dados contidos no "Cadastro de Vínculo de Ocorrências" e no "Registro de Eventos de Defesa Social" da Polícia Militar de Minas Gerais.

1.2 O Oficial encontra-se devidamente matriculado no Centro Federal de Educação Tecnológica de Minas Gerais (CEFET/MG) e está realizando uma pesquisa intitulada "A Complexidade Social das Redes de Criminalidade".

1.3 Para assegurar o caráter ético da pesquisa e viabilizar seu desenvolvimento o referido Oficial necessita ter acesso aos bancos de dados doravante mencionados, haja vista ser necessário transformar os dados coletados em "grafos dinâmicos" que permitam estudar interações em redes de criminalidade, assegurando a preservação da identidade das pessoas contidas nos referidos bancos de dados.

2 DETERMINO:

2.1 Autorizar o acesso e utilização dos dados contidos no "Cadastro de Vínculo de Ocorrências" e no "Registro de Eventos de Defesa Social" da Polícia Militar de Minas Gerais" ao no 095.606-0, Cel QOR Alex Sander de Oliveira Toledo para fins de subsídio à pesquisa pretendida.

2.2 Deverão ser realizados contatos bilaterais entre a DTS, DINT e o Oficial solicitante para adoção das medidas pertinentes, conforme normas em vigor.

(a) MARCELO FERNANDES, CORONEL PM
SUBCOMANDANTE-GERAL

Esse documento foi assinado em 24/05/2019 12:21:08 por MARCELO FERNANDES:77381122634.
Para verificar sua autenticidade escaneie o QrCode ao lado, ou acesse https://intranet.policiamilitar.mg.gov.br/lite/assinador/web/validar e informe o código: 5347FDC25CF