Ellen Leal dos Santos

# Análise de redes sociais criminais: identificação automática das funções dos criminosos

Relatório Final de Projeto de Seis Meses de Iniciação Científica Voluntária

Orientador: Prof. Dr. Luciano Antonio Digiampietri

São Paulo
2025

Ellen Leal dos Santos

# Análise de redes sociais criminais: identificação automática das funções dos criminosos

Relatório de trabalho de Iniciação Científica voltada à análise de redes sociais aplicada ao contexto criminal, com o objetivo de realizar a identificação automática das funções dos integrantes de uma quadrilha criminosa. Este trabalho foi desenvolvido voluntariamente ao longo do primeiro semestre de 2025 sob a orientação do Professor Doutor Luciano Antonio Digiampietri.

Universidade de São Paulo
Escola de Artes, Ciências e Humanidades
Bacharelado em Sistemas de Informação

São Paulo
2025

**Resumo:** O avanço tecnológico tem promovido transformações sociais profundas, impactando também o ambiente criminal, com a migração de atividades ilícitas para o meio virtual. Esse cenário exige que os órgãos responsáveis pela aplicação da lei adaptem seus métodos de investigação para aumentar sua efetividade. Este projeto de pesquisa propõe o uso da Análise de Redes Sociais como ferramenta de análise criminal, associada a métricas de áreas correlatas e métodos da ciência de dados para desenvolver ferramentas que automatizam a identificação de funções desempenhadas por criminosos em redes organizadas. Para validar o método, foram coletados dados criminais recentes do estado de São Paulo e analisados os resultados de uma operação policial, criando um grafo que possibilitou a análise da organização criminal com indicação da posição hierárquica e função de cada criminoso. A pesquisa buscou oferecer uma abordagem inovadora para auxiliar no enfrentamento do crime no contexto digital, auxiliando na identificação de alvos estratégicos e na alocação eficiente de recursos investigativos. Como resultado, obteve-se uma acurácia de 82,3% na identificação da função de cada criminoso.

**Palavras-chave:** Análise de Redes Sociais Criminais;  Migração Criminal; Crime Organizado

# 1. Introdução

Com o avanço da tecnologia e sua crescente presença no cotidiano da sociedade, novos desafios emergem para os órgãos de segurança pública, especialmente no combate à criminalidade organizada. A digitalização das interações sociais também refletiu no comportamento criminoso, que passou a explorar os ambientes virtuais para praticar crimes complexos, articulados e difíceis de rastrear. Diante desse cenário, torna-se essencial o desenvolvimento de novas estratégias investigativas que combinem técnicas computacionais, estatísticas e sociais para entender e neutralizar essas ameaças.

A Análise de Redes Sociais (ARS) tem se mostrado uma ferramenta promissora no mapeamento de estruturas criminosas. Através da modelagem das relações entre indivíduos, como contatos, interações ou transações, é possível identificar padrões comportamentais, graus de influência e posições hierárquicas em organizações criminosas. Quando associada a métodos de ciência de dados e aprendizado de máquina, a ARS permite automatizar parte do processo investigativo, auxiliando na identificação de funções específicas desempenhadas por cada integrante da rede.

Este projeto desenvolveu uma solução automatizada para a identificação das funções de criminosos em redes organizadas, a partir da análise de dados reais obtidos na Operação Anteros, realizada pela polícia do Estado de São Paulo. A operação desmantelou um grupo que praticava o crime de estelionato amoroso digital, uma prática crescente que utiliza redes sociais para estabelecer relacionamentos fictícios com vítimas e, assim, obter vantagens financeiras de forma ilícita. Esse tipo geralmente envolve uma rede articulada, com diferentes papéis e responsabilidades entre os envolvidos.

A proposta considera o problema sob a ótica da classificação multiclasse, onde, a partir das características de cada indivíduo na rede, busca-se prever qual papel ele desempenha. O uso de técnicas de aprendizado supervisionado, combinado à representação gráfica das relações

criminais, visa oferecer uma solução capaz de apoiar os investigadores na priorização de alvos estratégicos e na compreensão das dinâmicas internas das organizações criminosas.

## 2. Contextualização

O presente trabalho foca em um tipo específico de crime, o estelionato amoroso digital. O estelionato amoroso é um crime cibernético que explora a vulnerabilidade emocional das vítimas, utilizando redes sociais e plataformas digitais para estabelecer vínculos afetivos com o objetivo de obter vantagens financeiras ilícitas. Apesar de existirem diferentes configurações e papéis na atuação de redes criminosas focadas neste tipo de estelionato, é possível identificar os seguintes papéis (ou funções) dos criminosos, especialmente quando considera-se grupos criminosos com ação internacional (SANTOS, 2024):

**Liderança:** No topo da hierarquia encontra-se a liderança. Os líderes são responsáveis por estruturar os golpes e movimentar as maiores quantias de dinheiro, garantindo o funcionamento da organização criminosa.

**Fake Lovers:** Os "*fake lovers*" desempenham um papel central no crime, sendo os responsáveis por interagir diretamente com as vítimas. Por meio de perfis falsos em redes sociais, eles estabeleciam relações afetivas fictícias, conquistando a confiança das vítimas para posteriormente manipular suas emoções e extorquir dinheiro sob pretextos variados.

**Agentes de Lavagem:** Os agentes de lavagem cuidam da etapa financeira do golpe, ocultando a origem ilícita dos valores obtidos. Esses membros realizam transações complexas para dificultar o rastreamento dos recursos, essencial para a continuidade das atividades

criminosas.

**Operadores e Oficiais:** Atuam como intermediários em diversas fases do golpe. Três funções principais são:

- **Diplomatas:** Entram em contato com as vítimas para informar que presentes enviados por seus supostos amantes estavam retidos devido a problemas diplomáticos, solicitando pagamentos para liberar as encomendas.
- **Agentes da Alfândega:** Simulam situações em que as vítimas deveriam pagar taxas alfandegárias para liberar bagagens ou presentes valiosos supostamente enviados por seus amantes.
- **Transportadores:** Alegam que o transporte dos presentes requer custos adicionais, como pagamento de fretes ou escoltas armadas, devido ao valor elevado dos itens.

**Recrutadores:** Os recrutadores são responsáveis por atrair correntistas para o esquema. Esses correntistas cedem suas contas bancárias para o recebimento de valores provenientes dos golpes, em troca de uma comissão. Os operadores também atuam como ponte entre os correntistas e os demais membros da organização.

**Correntistas:** Os correntistas são peças-chave na logística financeira do crime, fornecendo suas contas bancárias para o depósito dos valores obtidos ilegalmente. Em troca, recebem uma porcentagem do montante movimentado, permitindo que o dinheiro circule sem expor diretamente os líderes e outros membros operacionais.

## 3. Objetivo

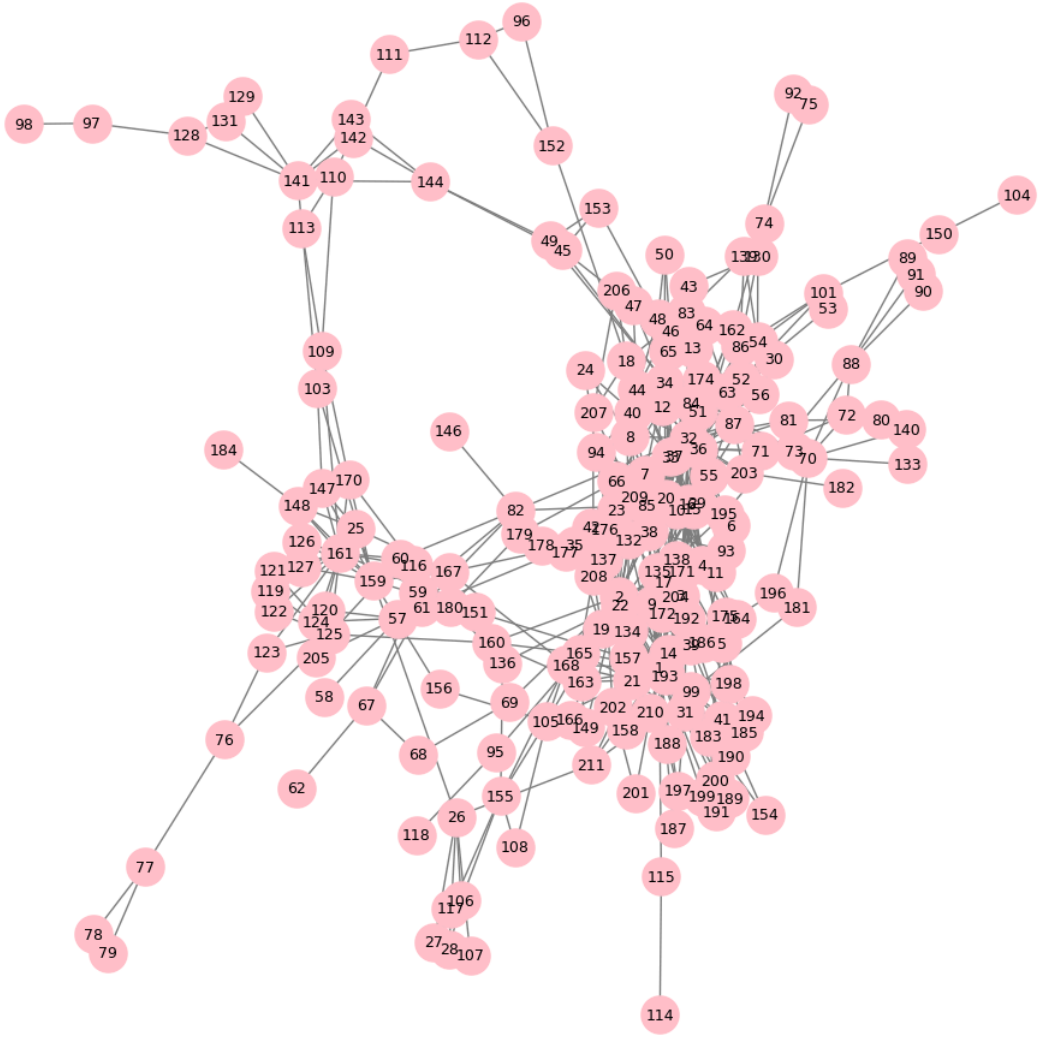
O objetivo do presente trabalho é especificar, desenvolver e avaliar uma solução automática para identificação da função de criminosos dentro de organizações criminosas com base em informações de investigações já concluídas.

Informações dessas investigações foram utilizadas tanto para fornecer dados base para a extração da estrutura da rede criminosa, no formato de uma rede social, bem como para validar a solução proposta, pois os criminosos já encontram-se rotulados de acordo com suas funções na rede.

## 4. Metodologia

A metodologia deste projeto baseou-se em atividades iniciadas em trabalho anterior, no qual foi realizada a coleta e organização de dados da Operação Anteros, conduzida pela Polícia Civil do Estado de São Paulo. Esses dados foram anonimizados e estruturados de forma a representar uma rede social criminal, em que os vértices correspondem aos indivíduos investigados e as arestas representavam vínculos identificados nas investigações, como relações de confiança, transações financeiras e interações comunicacionais. O grafo correspondente foi originalmente construído e analisado na linguagem R, incluindo o cálculo de métricas estruturais de análise de redes sociais, como medidas de centralidade.

Figura 1 - Representação em grafo da rede social construída a partir dos dados da operação Anteros



Neste projeto de Iniciação Científica, foram especificados e desenvolvidos códigos na linguagem Python com o uso da biblioteca NetworkX, o que permitiu maior integração com os modelos de aprendizado de máquina a serem utilizados posteriormente. As métricas extraídas da

rede incluíram:

- **Centralidade de intermediação (*betweenness centrality*):** mede a frequência com que um nó aparece nos caminhos mais curtos entre outros pares de nós. Indica o quanto um indivíduo pode atuar como intermediário ou elo de ligação entre diferentes partes da rede.
- **Centralidade de proximidade (*closeness centrality*):** avalia a média das distâncias geodésicas entre um nó e todos os outros na rede. Quanto mais próximo um nó estiver dos demais, maior será seu valor de centralidade.
- **Grau dos vértices (*degree*):** representa a quantidade de conexões diretas de um nó, ou seja, o número de vizinhos. É uma métrica simples, mas eficaz para identificar indivíduos com muitos contatos.
- **PageRank:** métrica originalmente desenvolvida pelo Google para classificar páginas da web. Avalia a importância de um nó com base na qualidade e quantidade de conexões recebidas.
- **Centralidade de autovalor (*eigenvector centrality*):** similar ao grau, mas considera também a importância dos vizinhos. Um nó tem alta centralidade se estiver conectado a outros nós também importantes.
- **Coeficiente de agrupamento local (*clustering coefficient*):** mede o grau de coesão entre os vizinhos de um nó, ou seja, a tendência de formarem triângulos ou grupos fechados.

Além das medidas tradicionais, foi realizada a extração de características baseadas no perfil dos vizinhos de cada indivíduo. Para isso, foi calculada, para cada nó, a proporção de vizinhos que pertenciam a cada uma das funções criminosas identificadas (como líderes, operadores, agentes de lavagem etc.). Essas variáveis foram expressas com base na quantidade absoluta de vizinhos por função.

Para avaliar a robustez dos modelos de classificação automática, foram gerados diferentes conjuntos de dados com porcentagens variadas de vizinhos sem função conhecida. Essa simulação buscou representar cenários realistas, nos quais apenas parte da rede está rotulada, como costuma ocorrer em investigações reais. Assim, foram construídas versões dos dados em que 100%, 50% ou 0% dos vizinhos estavam com as funções conhecidas, com o objetivo de testar a capacidade dos modelos em realizar inferência mesmo com informações parciais. A avaliação dos modelos considerou a validação cruzada com seis subconjuntos.

Com os dados estruturados e as características extraídas, foram testados diversos classificadores supervisionados de aprendizado de máquina. Os modelos utilizados foram:

- **DummyClassifier:** utilizado como *baseline* (referência), sempre prediz a classe mais frequente, servindo como controle para comparação com os demais modelos.
- **Logistic Regression (RegLog):** modelo estatístico linear que estima a probabilidade de um nó pertencer a uma determinada classe com base em suas características.
- **Random Forest:** algoritmo baseado em múltiplas árvores de decisão, que combina os resultados das árvores individuais para melhorar a precisão e reduzir o overfitting.
- **MLPClassifier (Rede Neural Multicamadas):** modelo com múltiplas camadas de perceptrons e função de ativação ReLU, capaz de capturar padrões não lineares complexos nos dados.
- **Gaussian Naive Bayes (GNB):** classificador probabilístico que assume independência entre as variáveis e distribuições normais dos atributos.
- **K-Nearest Neighbors (KNN):** classifica um nó com base nas classes mais comuns entre seus vizinhos mais próximos.

- **Decision Tree (DT):** constrói uma árvore de decisões com base em critérios de entropia para segmentar os dados conforme suas características.
- **Support Vector Machines (SVM e SVM2):** classificadores baseados na maximização da margem entre classes. Foram testadas versões com kernel linear e polinomial.

Antes do treinamento, os dados passaram por um pipeline que incluiu seleção de características com base no teste qui-quadrado ($chi^2$), projeção com Truncated SVD, e normalização com MinMaxScaler e StandardScaler, dependendo do classificador utilizado.

A avaliação dos modelos foi realizada com base em duas métricas principais:

- **Acurácia:** proporção de previsões corretas em relação ao total de casos.
- **F1-score Macro:** média harmônica entre precisão e revocação para todas as classes, tratando-as de forma equilibrada independentemente do número de exemplos por classe.

## 5. Resultados

A presente seção é dedicada à apresentação e discussão sobre os resultados obtidos.

### 5.1 Resultados com Vizinhança Totalmente Rotulada

A Tabela 1 apresenta os resultados da classificação, considerando o conhecimento do rótulo de todos os vizinhos de cada um dos elementos do conjunto de teste.

Tabela 1 - Resultados da abordagem utilizando a rotulação completa da vizinhança de todos os nós do conjunto de dados

| Pipeline | F1 Macro | Accuracy |
|---|---|---|
| ['PCA', 'MLP'] | 0,484 | **0,823** |
| ['MinMaxScaler', 'RF'] | 0,483 | **0,823** |
| ['KBest', 'MLP'] | 0,478 | 0,817 |
| ['StandardScaler', 'SVM_Poly'] | 0,569 | 0,812 |
| ['SemPreprocessamento, RF'] | 0,539 | 0,812 |
| ['MinMaxScaler', 'SVM_Poly'] | 0,396 | 0,812 |
| ['StandardScaler', 'SVM'] | 0,383 | 0,812 |
| ['MinMaxScaler', 'MLP'] | 0,515 | 0,807 |
| ['KBest', 'SVM_Poly'] | 0,448 | 0,807 |
| ['SemPreprocessamento, SVM'] | 0,332 | 0,807 |
| ['MinMaxScaler', 'SVM'] | 0,327 | 0,807 |
| ['KBest', 'RF'] | **0,57** | 0,806 |
| ['SemPreprocessamento, MLP'] | 0,542 | 0,806 |
| ['MinMaxScaler', 'RF'] | 0,516 | 0,806 |
| ['MinMaxScaler', 'RegLog'] | 0,403 | 0,801 |
| ['KBest', 'SVM'] | 0,328 | 0,801 |
| ['StandardScaler', 'RegLog'] | 0,452 | 0,796 |
| ['PCA', 'SVM'] | 0,319 | 0,796 |
| ['KBest', 'GNB'] | 0,501 | 0,785 |
| ['SemPreprocessamento, RegLog'] | 0,435 | 0,785 |
| ['StandardScaler', 'KNN'] | 0,342 | 0,785 |
| ['PCA', 'RF'] | 0,486 | 0,779 |
| ['StandardScaler', 'MLP'] | 0,47 | 0,779 |
| ['KBest', 'RegLog'] | 0,41 | 0,779 |
| ['MinMaxScaler', 'KNN'] | 0,334 | 0,779 |
| ['PCA', 'SVM_Poly'] | 0,38 | 0,773 |
| ['PCA', 'RegLog'] | 0,43 | 0,768 |
| ['SemPreprocessamento, SVM_Poly'] | 0,404 | 0,768 |
| ['PCA', 'GNB'] | 0,468 | 0,763 |
| ['dummy'] | 0,216 | 0,763 |
| ['KBest', 'KNN'] | 0,354 | 0,762 |
| ['PCA', 'DT'] | 0,496 | 0,757 |
| ['PCA', 'KNN'] | 0,345 | 0,746 |
| ['KBest', 'DT'] | 0,561 | 0,74 |
| ['SemPreprocessamento, KNN'] | 0,326 | 0,74 |

| | | |
|---|---|---|
| ['SemPreprocessamento, GNB'] | 0,467 | 0,735 |
| ['MinMaxScaler', 'GNB'] | 0,467 | 0,735 |
| ['StandardScaler', 'GNB'] | 0,467 | 0,735 |
| ['MinMaxScaler', 'DT'] | 0,509 | 0,713 |
| ['SemPreprocessamento, DT'] | 0,453 | 0,697 |
| ['StandardScaler', 'DT'] | 0,472 | 0,685 |

Na abordagem com conhecimento completo sobre as funções dos vizinhos (isto é, pessoas com as quais cada indivíduo estava conectado na rede criminosa) os resultados mostraram que, apesar de alguns pipelines alcançarem altas taxas de acurácia, houve variação considerável no F1-score macro, refletindo um desequilíbrio na capacidade dos modelos em classificar corretamente todas as classes.

Os melhores desempenhos em na F1-score macro, que é a métrica mais adequada para cenários com classes desbalanceadas, foram obtidos com:

- Random Forest combinado com seleção de características via KBest, atingindo 0,57 de F1-score macro e 80,6% de acurácia.
- Rede Neural MLP sem pré-processamento, com 0,542 de F1-score macro e 80,6% de acurácia.
- Random Forest sem pré-processamento, obtendo 0,539 de F1-score macro e 81,2% de acurácia.
- SVM com kernel polinomial e StandardScaler, atingindo o maior F1-score macro entre os SVMs (0,569) e também com 81,2% de acurácia.

Por outro lado, modelos como o SVM linear e o SVM polinomial com MinMaxScaler apresentaram alta acurácia (81,2%), mas baixos F1-scores macro (0,383 e 0,396,

respectivamente), indicando possível viés para as classes mais frequentes.

O classificador Dummy, utilizado como baseline, alcançou F1-score macro de 0,216 e 76,3% de acurácia, evidenciando que todos os modelos testados superaram significativamente a predição aleatória.

Esses resultados indicam que, mesmo com informação completa da vizinhança, o problema apresenta desafios relacionados ao balanceamento das classes e à variabilidade no desempenho entre os pipelines testados. O uso de métodos como Random Forest e MLP, principalmente sem pré-processamento ou com seleção via KBest, se mostrou bem robusto na tarefa de classificação.

### 5.2 Resultados com Vizinhança 50% Rotulada

A Tabela 2 apresenta os resultados da classificação, considerando o conhecimento do rótulo de 50% dos vizinhos de cada um dos elementos do conjunto de teste.

Tabela 2 - Resultados utilizando a abordagem com a rotulação de 50% da vizinhança de todos os nós do conjunto de dados

| Pipeline | F1 Macro | Accuracy |
|---|---|---|
| ['StandardScaler', 'MLP'] | 0,479 | 0,807 |
| ['MinMaxScaler', 'SVM_Poly'] | 0,364 | 0,807 |
| ['SemPreprocessamento, RF'] | 0,387 | 0,801 |
| ['MinMaxScaler', 'SVM'] | 0,32 | 0,801 |
| ['SemPreprocessamento, SVM_Poly'] | 0,496 | 0,796 |
| ['MinMaxScaler', 'RF'] | 0,443 | 0,796 |
| ['MinMaxScaler', 'RF'] | 0,404 | 0,796 |
| ['KBest', 'MLP'] | 0,37 | 0,796 |
| ['StandardScaler', 'SVM'] | 0,33 | 0,796 |
| ['PCA', 'GNB'] | 0,441 | 0,791 |
| ['PCA', 'MLP'] | 0,414 | 0,791 |
| ['StandardScaler', 'SVM_Poly'] | 0,469 | 0,79 |
| ['SemPreprocessamento, MLP'] | 0,362 | 0,79 |
| ['PCA', 'SVM'] | 0,317 | 0,79 |

| | | |
|---|---|---|
| ['SemPreprocessamento, SVM'] | 0,316 | 0,79 |
| ['KBest', 'SVM'] | 0,316 | 0,79 |
| ['MinMaxScaler', 'MLP'] | 0,443 | 0,785 |
| ['KBest', 'SVM_Poly'] | 0,421 | 0,785 |
| ['KBest', 'GNB'] | 0,505 | 0,78 |
| ['SemPreprocessamento, RegLog'] | 0,427 | 0,779 |
| ['KBest', 'RF'] | 0,391 | 0,779 |
| ['StandardScaler', 'KNN'] | 0,381 | 0,779 |
| ['KBest', 'KNN'] | 0,329 | 0,779 |
| ['MinMaxScaler', 'RegLog'] | 0,307 | 0,774 |
| ['SemPreprocessamento, KNN'] | 0,332 | 0,773 |
| ['PCA', 'KNN'] | 0,329 | 0,773 |
| ['PCA', 'SVM_Poly'] | 0,388 | 0,768 |
| ['PCA', 'RegLog'] | 0,37 | 0,763 |
| ['dummy'] | 0,216 | 0,763 |
| ['MinMaxScaler', 'KNN'] | 0,35 | 0,762 |
| ['PCA', 'RF'] | 0,352 | 0,757 |
| ['KBest', 'RegLog'] | 0,353 | 0,752 |
| ['StandardScaler', 'RegLog'] | 0,331 | 0,752 |
| ['SemPreprocessamento, GNB'] | 0,482 | 0,741 |
| ['MinMaxScaler', 'GNB'] | 0,482 | 0,741 |
| ['StandardScaler', 'GNB'] | 0,482 | 0,741 |
| ['PCA', 'DT'] | 0,411 | 0,735 |
| ['KBest', 'DT'] | 0,366 | 0,73 |
| ['MinMaxScaler', 'DT'] | 0,368 | 0,719 |
| ['SemPreprocessamento, DT'] | 0,324 | 0,702 |
| ['StandardScaler', 'DT'] | 0,355 | 0,685 |

No cenário onde apenas metade das funções dos vizinhos é conhecida, os resultados mostram uma diminuição no desempenho dos modelos em relação ao cenário com 100% de informação. Isso reflete a dificuldade de inferência em redes parcialmente rotuladas, que é uma situação comum em investigações reais.

Os melhores resultados em F1-score macro foram:

- Gaussian Naive Bayes com KBest, alcançando 0,505 de F1-score macro e 78% de acurácia.
- MLP com StandardScaler, com 0,479 de F1-score macro e 80,7% de acurácia.
- SVM com kernel polinomial sem pré-processamento, com 0,496 de F1-score macro e 79,6% de acurácia.
- GNB sem pré-processamento ou com escalonadores (MinMax ou StandardScaler), com F1 próximo a 0,482 e 74,1% de acurácia.

O Random Forest, que foi um dos melhores no cenário de 100% rotulado, apresentou uma queda no desempenho. Com pré-processamento MinMaxScaler, o F1-score macro ficou entre 0,404 e 0,443, dependendo da configuração, embora a acurácia ainda tenha se mantido em torno de 79,6% a 80,1%.

SVMs e KNNs tiveram comportamento parecido ao cenário anterior: mantiveram boa acurácia, mas com F1-score macro reduzido (em torno de 0,32 a 0,38), o que indica um viés para as classes mais frequentes.

O classificador Dummy permaneceu com F1-score macro de 0,216 e 76,3% de acurácia, servindo como baseline. Mesmo com a redução de desempenho geral, todos os modelos testados superaram a predição dele.

Nessa abordagem, os resultados evidenciam que a redução da informação disponível impacta negativamente a capacidade dos modelos de equilibrar corretamente as predições entre as diferentes classes. Entretanto, alguns classificadores probabilísticos, como o GNB com seleção de

atributos, se mostraram mais robustos à perda de informação.

### 5.3 Resultados com Vizinhança 0% Rotulada

A Tabela 3 apresenta os resultados da classificação sem utilizar informações dos rótulos dos vizinhos de cada elemento do conjunto de teste.

Tabela 3 - Resultados utilizando a abordagem sem rotulação da vizinhança dos nós do conjunto de dados

| Pipeline | F1 Macro | Accuracy |
|---|---|---|
| ['MinMaxScaler', 'SVM'] | 0,34 | 0,812 |
| ['SemPreprocessamento, MLP'] | 0,377 | 0,807 |
| ['StandardScaler', 'SVM'] | 0,337 | 0,807 |
| ['MinMaxScaler', 'SVM_Poly'] | 0,327 | 0,807 |
| ['KBest', 'SVM_Poly'] | 0,418 | 0,801 |
| ['PCA', 'MLP'] | 0,408 | 0,801 |
| ['KBest', 'MLP'] | 0,368 | 0,801 |
| ['PCA', 'GNB'] | 0,406 | 0,796 |
| ['PCA', 'SVM_Poly'] | 0,344 | 0,791 |
| ['PCA', 'SVM'] | 0,317 | 0,791 |
| ['SemPreprocessamento, SVM'] | 0,313 | 0,791 |
| ['StandardScaler', 'SVM_Poly'] | 0,419 | 0,79 |
| ['KBest', 'KNN'] | 0,366 | 0,79 |
| ['MinMaxScaler', 'MLP'] | 0,307 | 0,79 |
| ['KBest', 'SVM'] | 0,308 | 0,785 |
| ['MinMaxScaler', 'RegLog'] | 0,298 | 0,785 |
| ['KBest', 'RF'] | 0,441 | 0,784 |
| ['MinMaxScaler', 'RF'] | 0,427 | 0,779 |
| ['MinMaxScaler', 'RF'] | 0,424 | 0,779 |
| ['SemPreprocessamento, RF'] | 0,392 | 0,779 |
| ['SemPreprocessamento, SVM_Poly'] | 0,366 | 0,779 |
| ['SemPreprocessamento, KNN'] | 0,364 | 0,774 |
| ['PCA', 'KNN'] | 0,362 | 0,774 |
| ['PCA', 'RegLog'] | 0,312 | 0,774 |
| ['SemPreprocessamento, RegLog'] | 0,31 | 0,774 |
| ['KBest', 'RegLog'] | 0,31 | 0,774 |
| ['KBest', 'GNB'] | 0,455 | 0,763 |

| | | |
|---|---|---|
| ['dummy'] | 0,216 | 0,763 |
| ['StandardScaler', 'MLP'] | 0,404 | 0,74 |
| ['PCA', 'RF'] | 0,396 | 0,735 |
| ['StandardScaler', 'RegLog'] | 0,351 | 0,735 |
| ['SemPreprocessamento, GNB'] | 0,434 | 0,719 |
| ['MinMaxScaler', 'GNB'] | 0,434 | 0,719 |
| ['StandardScaler', 'GNB'] | 0,434 | 0,719 |
| ['PCA', 'DT'] | 0,388 | 0,712 |
| ['StandardScaler', 'KNN'] | 0,285 | 0,712 |
| ['KBest', 'DT'] | 0,42 | 0,702 |
| ['SemPreprocessamento, DT'] | 0,379 | 0,702 |
| ['MinMaxScaler', 'KNN'] | 0,288 | 0,701 |
| ['MinMaxScaler', 'DT'] | 0,389 | 0,696 |
| ['StandardScaler', 'DT'] | 0,378 | 0,696 |

No caso em que nenhum vizinho tem a função conhecida, os modelos precisam se basear apenas nas características individuais dos nós, sem qualquer informação relacional. Isso representa um caso extremo, próximo ao pior cenário possível em redes criminais parcialmente conhecidas, o que é frequentemente observado na vida real.

Mesmo nessas condições, alguns modelos apresentaram desempenho superior ao baseline (Dummy), que obteve F1-score macro de 0,216 e 76,3% de acurácia.

Os melhores resultados foram:

- Gaussian Naive Bayes com seleção KBest, com 0,455 de F1-score macro e 76,3% de acurácia.
- Random Forest com KBest, atingindo 0,441 de F1-score macro e 78,4% de acurácia.
- Random Forest com MinMaxScaler, com F1 de 0,424 e 0,427 e 77,9% de acurácia.
- MLP com PCA, alcançando 0,408 de F1-score macro e 80,1% de acurácia.

- SVM com kernel polinomial e StandardScaler, com 0,419 de F1-score macro e 79% de acurácia.

Por outro lado, classificadores como SVM linear, KNN e Regressão Logística apresentaram desempenho mais baixo em F1-score macro (entre 0,28 e 0,34), apesar de manterem acurácia acima de 77%, o que indica um desequilíbrio nas predições entre as classes.

Os classificadores GNB (Gaussian Naive Bayes), mesmo sem pré-processamento ou com escalonadores, mantiveram estabilidade, obtendo F1 em torno de 0,434 e acurácia de 71,9%, o que mostra resistência do método probabilístico mesmo com ausência de dados relacionais.

Esses resultados confirmam que a falta de informações reduz significativamente a capacidade dos modelos de distinguir entre as diferentes funções criminosas, mas métodos como Random Forest, MLP com PCA e GNB com seleção de atributos ainda conseguem reconhecer parte dos padrões relevantes.

## 6. Conclusões

Os resultados obtidos demonstram que o uso de características extraídas de redes sociais criminais pode contribuir significativamente para a classificação de indivíduos com base em suas funções na organização, mesmo em cenários de informação incompleta. Pode se obverar que, quanto maior a proporção de vizinhos rotulados, melhor o desempenho dos modelos, com destaque para Random Forest, MLP e GNB. No entanto, mesmo na ausência total de informações da vizinhança, alguns classificadores conseguiram superar a baseline, indicando a possibilidade de abordagens baseadas em aprendizado supervisionado para auxiliar investigações criminais.

Essa análise reforça a importância de combinar informações estruturais da rede com técnicas de machine learning, visando a apoiar a tomada de decisão em contextos complexos e com dados parciais.

## 7. Bibliografia e Referências Bibliográficas


BOBA, Rachel. Introductory Guide to Crime Analysis and Mapping. Crime Mapping Laboratory - Police Foundation, 2001.

BRASIL. DECRETO-LEI No 2.848, DE 7 DE DEZEMBRO DE 1940. Código Penal. Disponível em https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm

BRASIL. LEI Nº 14.155, DE 27 DE MAIO DE 2021. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tornar mais graves os crimes de violação de dispositivo informático, furto e estelionato cometidos de forma eletrônica ou pela internet; e o Decreto-Lei nº 3.689, de 3 de outubro de 1941 (Código de Processo Penal), para definir a competência em modalidades de estelionato. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14155.htm

BRIGHT, D., BREWER, R., & MORSELLI, C. (2021). Using social network analysis to study crime: Navigating the challenges of criminal justice records. Social Networks, 66, 50–64. https://doi.org/https://doi.org/10.1016/j.socnet.2021.01.006

BURCHER, M., WHELAN, C. Social network analysis as a tool for criminal intelligence: understanding its potential from the perspectives of intelligence analysts. Trends Organ Crim 21, 278–294 (2018). https://doi.org/10.1007/s12117-017-9313-8

CARRINGTON, Peter J. et SCOTT, John. The Sage Handbook of Social Network Analysis. Los Angeles: Sage, 2014.

CUNHA, Bruno Requião da. Criminofísica: a ciência das interações criminais. 1ª ed. Porto Alegre : Buqui, 2020.

D. Weisburd and C. Britt, Statistics in Criminal Justice, DOI 10.1007/978-1-4614-9170-5_1, Springer Science+Business Media New York 2014, p. 2

Divisão de Estatísticas do Departamento de Assuntos Econômicos e Sociais das Nações Unidas (ONU)

GALLO, Fernanda de Almeida. Tutorial de redes e um estudo de caso sobre “redes criminais”. Revista USP, São Paulo, nº 92, p. 74-85, dezembro/fevereiro 2011-2012.

HOBBES, Thomas. O Leviatã. São Paulo: Martin Claret, 2014.

KHAN, Túlio. Anuário Brasileiro De Segurança Pública aponta explosão de estelionatos no país e maior número de estupros da série histórica. Fonte Segura – Fórum Brasileiro de Segurança Pública. Disponível em: https://fontesegura.forumseguranca.org.br/anuario-brasileiro-de-seguranca-publica-aponta-explosao-de-estelionatos-no-pais-e-maior-numero-de-estupros-da-serie-historica/

KHAN, Túlio. Migração dos crimes violentos de rua para crimes digitais. Fonte Segura – Fórum Brasileiro de Segurança Pública. Disponível em https://fontesegura.forumseguranca.org.br/migracao-dos-crimes-violentos-de-rua-para-crimes-digitais/

LE BON, Gustave. Psicologia das Multidões. 3ª ed. São Paulo: Editora WMF Martins Fontes, 2018.

MALDONADO, Santiago Vanegas. 'Love bombing': quando amor em excesso se torna perigoso? Disponível em https://www.bbc.com/portuguese/articles/cv2r7v19npeo

MORENO, J.L. Who Shall Survive? New York: Beacon Press, 1934.

SANTOS, Kerlly B. M.. Análise de redes sociais criminais: o desafio dos crimes digitais. Relatório Final de Iniciação Científica, Escola de Artes, Ciências e Humanidades da USP, 2024.

Secretaria de Segurança Pública do Estado de São Paulo. Dados Mensais. Disponível em https://www.ssp.sp.gov.br/estatistica/dados-mensais

Secretaria de Segurança Pública do Estado de São Paulo. Estatísticas Trimestrais. Disponível em https://www.ssp.sp.gov.br/estatistica/dados-trimestrais

Secretaria de Segurança Pública do Estado de São Paulo. Números sem Mistério. Disponível em https://www.ssp.sp.gov.br/estatistica

UNITED NATIONS. Manual for the Development of a System of Criminal Justice Statistics. Series F nº 89. New York, 2003.

WEISBURD, David, et BRITT, Chester. Statistics in Criminal Justice. New York: Springer, 2014.