Análise de Grafos na Informática Forense

Mapeando uma Rede Criminosa Simulada Através da Rede de Comunicações da Enron

1st Vinícius Ramalho de Oliveira
Engenharia da Computação
CEFET-MG
Divinópolis, Brasil
ramalhooliveiravini@gmail.com

***Abstract*—The analysis of large-scale communication data, such as emails, remains a central challenge in digital forensics and corporate crime investigations. Identifying leaders and information intermediaries within complex organizations is often hindered by data volume and unstructured nature. This paper proposes a methodological pipeline based on Graph Theory and Social Network Analysis (SNA) to model and dissect a corporate communication network, using the public Enron email dataset (1985-2003) as a case study. The communication network was modeled as a directed graph and analyzed using multiple static centrality metrics (Betweenness, PageRank, and Closeness) and community detection (Louvain algorithm). The static analysis successfully identified distinct actor profiles based on their roles: 'Brokers' (high-betweenness), 'Authorities' (high-PageRank), and 'Disseminators' (high-closeness). Furthermore, the community analysis revealed a core-periphery structure and identified the key internal intermediaries for each major group. Finally, a network disruption simulation revealed that the organization exhibits a high degree of centralization, making it significantly more vulnerable to the targeted removal of authority figures than to the removal of intermediaries. This work demonstrates the efficacy of this SNA pipeline as a tool to map hidden power structures and identify high-priority targets for a forensic investigation.**

***Index Terms*—Informática Forense, Análise de Redes Sociais (SNA), Teoria dos Grafos, Métricas de Centralidade, Detecção de Comunidade, Análise de E-mail**

I. Introdução

A investigação de incidentes corporativos, fraudes e outras condutas ilícitas no âmbito organizacional tem, progressivamente, ampliado seu escopo para além da análise tradicional de documentos isolados, exigindo abordagens capazes de lidar com volumes massivos de dados e com a complexidade das interações humanas mediadas por tecnologia. [1], [2], [9] Em particular, os registros de comunicações eletrônicas (e-mails, mensagens instantâneas, logs de colaboração) constituem uma fonte rica de evidências, embora seu conteúdo textual seja relevante, é a estrutura relacional de quem se comunicou com quem, quando e com que frequência, que frequentemente revela padrões organizacionais, cadeias de comando implícitas e rotas de transmissão de informação que não são evidentes na leitura linear de mensagens. [3], [4], [9] Nesse contexto, a Análise de Redes Sociais (Social Network Analysis — SNA), fundamentada na Teoria dos Grafos, emerge como um arcabouço metodológico privilegiado para modelar, quantificar e interpretar tais estruturas relacionais. [11], [13], [14]

A SNA oferece um conjunto articulado de métricas e algoritmos que permitem identificar atores centrais (por exemplo, agentes com elevado grau de intermediação/betweenness), detectar agentes com alto potencial de influência (PageRank) e mapear a proximidade estrutural entre nós (closeness), além de revelar subestruturas coesas por meio de algoritmos de detecção de comunidades. [11]–[14] Esses instrumentos permitem transformar metadados de comunicação em indicadores acionáveis para a perícia digital: a identificação de pontos de falha informacional, a descoberta de "corredores" de disseminação de informação entre departamentos, e o reconhecimento de núcleos organizacionais que podem corresponder a equipes formais ou a grupos informais de coordenação. [3], [5], [8] Complementarmente, ao se fazer uma análise temporal das métricas de rede possibilita a construção de narrativas dinâmicas (por exemplo, o surgimento de picos de intermediação antes de eventos críticos, ou a desconexão gradual de determinados executivos) que são essenciais para correlacionar comportamento comunicacional com eventos externos (crises, auditorias, decisões estratégicas). [7]

No presente trabalho, foram adotados o conjunto de dados públicos de e-mails da Enron como estudo de caso representativo para avaliar a aplicabilidade e a robustez de um pipeline analítico orientado à investigação forense de redes corporativas. [17] Embora o corpus da Enron seja frequentemente empregado em pesquisas de mineração de texto e de redes por sua disponibilização e riqueza histórica, nossa abordagem enfatiza a utilização rigorosa de metadados e de técnicas de SNA para a construção de um quadro interpretativo que priorize reprodutibilidade, transparência e validade investigativa. [9], [10] Para tanto, foi descrito etapas explícitas de pré-processamento (normalização de identificadores, tratamento de duplicatas, janela temporal), modelagem do grafo dirigido de comunicações, cálculo de medidas de centralidade múltiplas e aplicação de métodos de detecção de comunidades (com ênfase no algoritmo de Louvain para robustez e escalabilidade). [14], [15]

Além da apresentação dos procedimentos técnicos, o artigo discute critérios de interpretação dos resultados no contexto forense: como distinguir entre um "hub" operacional e um "broker" investigativo; e quais métricas oferecem maior sensibilidade para a identificação de atores que servem como pontos cruciais de coordenação ou encobrimento. [5], [6], [11]

Também foram abordados aspectos práticos e éticos relevantes em trabalhos desse tipo, incluindo limitações inerentes ao uso de metadados (ausência de conteúdo semântico, possibilidade de remetentes/recipientes múltiplos em mensagens agregadas), riscos de inferência indevida e a necessidade de validação cruzada com outras fontes de prova quando disponível. [9], [10] A reprodução dos experimentos foi priorizada por meio da disponibilização dos scripts de processamento e de análise, permitindo que pesquisadores e peritos verifiquem e ampliem os achados. [16], [17]

II. Referêncial Teórico

A análise de redes de comunicação em investigações forenses fundamenta-se na aplicação da Teoria dos Grafos, que provê um arcabouço matemático rigoroso para a representação, quantificação e interpretação de estruturas relacionais complexas. Formalmente, um grafo é definido como $G = (V, E)$, em que V representa o conjunto de *vértices* (nós) e E o conjunto de *arestas* (ligações) que os conectam. No contexto deste estudo, cada vértice $v \in V$ corresponde a um ator da comunicação (por exemplo, um remetente ou destinatário de e-mails), enquanto cada aresta $e \in E$ representa uma interação observada, isto é, o envio de uma mensagem de um indivíduo a outro.

Ao modelar comunicações corporativas sob a forma de grafos direcionados e ponderados, é possível abstrair a complexidade textual dos e-mails e concentrar-se na estrutura subjacente das interações humanas, revelando padrões ocultos de influência, mediação e isolamento. Conforme discutido pioneiramente por Sparrow (1991) [1], a aplicação da Análise de Redes Sociais (Social Network Analysis – SNA) a contextos investigativos possibilita transcender a leitura individual de mensagens e focar na topologia das relações para identificar hierarquias informais, fluxos de comando e vulnerabilidades estruturais. Desde então, a SNA tem se consolidado como ferramenta metodológica essencial na análise de redes ilícitas e comunicações corporativas suspeitas, conforme demonstrado em estudos como os de Sarvari et al. (2014) [2], que reconstruíram redes criminosas a partir de registros de telefonia, e Ferrara et al. (2014) [3], que aplicaram métricas de grafos para detectar organizações fraudulentas.

No âmbito da informática forense, essa abordagem permite compreender como a estrutura da comunicação reflete a dinâmica organizacional, oferecendo subsídios empíricos para a formulação de hipóteses investigativas. A seguir, são discutidos os principais conceitos e métricas utilizados neste estudo: as medidas de centralidade, que quantificam a importância dos atores na rede; os métodos de detecção de comunidades, que mapeiam a estrutura modular das interações; e, finalmente, os conceitos de robustez estrutural e disrupção, que fundamentam a análise de vulnerabilidade da rede frente a ataques direcionados.

A. Métricas de Centralidade: Quantificando a Importância dos Atores

A análise da centralidade constitui o núcleo da SNA, fornecendo indicadores quantitativos da relevância estrutural de cada ator dentro da rede. Cada métrica de centralidade captura um aspecto distinto da posição de um nó, refletindo papéis estratégicos variados, como liderança, intermediação ou difusão de informação. No contexto forense, tais métricas são particularmente valiosas, pois permitem identificar alvos prioritários e nós críticos cuja remoção pode alterar drasticamente a conectividade da rede.

1) Centralidade de Intermediação (Betweenness): A centralidade de intermediação mede a frequência com que um nó $v \in V$ aparece nos caminhos mais curtos entre outros pares de nós da rede. Formalmente, ela é definida como:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

onde σ_{st} é o número total de caminhos geodésicos entre os nós s e t, e $\sigma_{st}(v)$ é o número desses caminhos que passam por v. [11]

Um nó com alta intermediação atua como *broker* (intermediário), controlando o fluxo de informação entre subgrupos que, de outra forma, estariam desconectados [5] [6]. Em um cenário de análise forense, esses nós são de extrema relevância: sua remoção pode fragmentar a rede em componentes isolados, interrompendo a circulação de informações sensíveis e revelando estruturas de poder ocultas [1] [9].

2) Centralidade de Autoridade (PageRank): Inspirada na análise de hiperlinks da web, a métrica de PageRank foi originalmente desenvolvida para classificar páginas da internet, mas mostra-se igualmente eficaz em redes de e-mails e comunicações corporativas. Trata-se de uma forma de centralidade de autovetor, em que a importância de um nó é determinada de maneira recursiva com base na importância dos nós que apontam para ele:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in M(v)} \frac{PR(u)}{L(u)} \tag{2}$$

onde $PR(v)$ é o PageRank do nó v, d é o fator de amortecimento (geralmente 0,85), N é o número total de nós no grafo, $M(v)$ é o conjunto de nós que enviam mensagens a v, $PR(u)$ é o PageRank do nó u, e $L(u)$ é o número de saídas do nó u. [12]

Em uma rede de e-mails, um nó com alto PageRank corresponde frequentemente a um líder hierárquico ou ponto de convergência informacional, ou seja, indivíduos que recebem grande volume de comunicações relevantes de atores também influentes, mas que, por sua posição de autoridade, podem enviar poucas mensagens em resposta. [3]

3) Centralidade de Proximidade (Closeness): A centralidade de proximidade avalia a eficiência de um nó na disseminação de informação através da rede. Ela é definida como o inverso da soma das distâncias geodésicas de um nó v para todos os outros nós alcançáveis:

$$C_C(v) = \frac{1}{\sum_{t \in V} d(v,t)} \tag{3}$$

onde $d(v,t)$ representa a distância mínima entre v e t. [13]

Um ator com alta centralidade de proximidade é considerado um difusor eficiente isto é, está estrategicamente posicionado para propagar rapidamente informações (ordens, alertas ou desinformação) a toda a organização. Em redes corporativas, tais nós geralmente correspondem a coordenadores intermediários ou gestores operacionais, cuja posição lhes permite atuar como vetores de comunicação transversal. [1]

B. Detecção de Comunidades: Mapeando a Estrutura Organizacional

As redes sociais e corporativas raramente são homogêneas; elas tendem a se auto-organizar em subestruturas coesas, conhecidas como comunidades ou módulos, que refletem divisões funcionais, afinidades informais ou hierarquias de projeto. A detecção de comunidades consiste em identificar tais agrupamentos de nós que apresentam densidade interna de conexões significativamente superior à densidade de conexões externas [8].

Neste trabalho, adotou-se o algoritmo de Louvain, um método heurístico amplamente reconhecido por sua eficiência computacional e escalabilidade em grafos de grande porte. O algoritmo busca maximizar a modularidade (Q), uma métrica que quantifica a qualidade de uma partição da rede:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{4}$$

onde A_{ij} é o elemento da matriz de adjacência, k_i e k_j são os graus dos nós i e j, m é o número total de arestas e $\delta(c_i, c_j)$ é uma função indicadora que vale 1 se i e j pertencem à mesma comunidade, e 0 caso contrário. [15]

Uma alta modularidade indica que a divisão é significativa, refletindo uma estrutura real de colaboração interna ou de divisão funcional dentro da organização. No contexto forense, essa técnica permite mapear a estrutura organizacional de fato, muitas vezes distinta do organograma formal, ou seja, revelando silos informacionais, equipes ocultas e líderes locais, que são responsáveis por coordenar fluxos de comunicação em cada módulo. [14]

Ao integrar a detecção de comunidades com as métricas de centralidade, torna-se possível construir uma visão multinível da rede: identificar não apenas os atores mais influentes globalmente, mas também os líderes regionais e intermediários locais que sustentam a coesão interna de cada grupo. Essa integração metodológica é essencial para investigações forenses, pois permite compreender tanto a macroestrutura (a rede como um todo) quanto as microdinâmicas de poder e comunicação em seu interior. [9]

C. Robustez Estrutural e Análise de Disrupção

A análise da resiliência de uma rede foca em sua capacidade de manter a conectividade funcional diante da remoção de vértices ou arestas. Redes de comunicação humana frequentemente exibem uma topologia "livre de escala" (*scale-free*), caracterizada por uma distribuição de graus que segue uma lei de potência ($P(k) \sim k^{-\gamma}$), resultando na existência de *hubs* altamente conectados [18]. A literatura estabelece que tais topologias, embora robustas a falhas aleatórias, apresentam fragilidade crítica a ataques direcionados (*targeted attacks*), onde nós de alta centralidade são removidos deliberadamente [5], [9], [19]–[22].

Para quantificar o impacto de tais intervenções na coesão da rede, utiliza-se o conceito de componentes conexos. Formalmente, seja $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ o conjunto de todos os subgrafos conexos disjuntos de G, onde cada $C_i = (V_i, E_i)$ é um subgrafo maximal tal que existe um caminho não direcionado entre qualquer par de vértices $u, v \in V_i$. O Maior Componente Conexo (*Giant Connected Component* - GCC) é definido como o componente com a maior cardinalidade de vértices:

$$GCC = \arg\max_{C_i \in \mathcal{C}} |V_i| \tag{5}$$

Em investigações forenses, a eficácia de uma estratégia de desmantelamento é mensurada pela variação do tamanho relativo do GCC em função da remoção de uma fração f de nós da rede. Seja N o número total de nós iniciais e $|V_{GCC}(f)|$ o tamanho do componente gigante após a remoção dos f nós mais centrais. A integridade estrutural $\Phi(f)$ é dada por:

$$\Phi(f) = \frac{|V_{GCC}(f)|}{N} \tag{6}$$

Uma redução abrupta em $\Phi(f)$ indica uma transição de fase onde a rede deixa de ser uma entidade globalmente conectada e se fragmenta em ilhas isoladas, impedindo a coordenação e o fluxo de informação entre diferentes setores da organização [6], [9], [19]–[22].

III. METODOLOGIA

O presente estudo emprega um pipeline experimental implementado integralmente em `Python 3.12`, apoiado por bibliotecas de código aberto para manipulação de dados, construção e análise de grafos, detecção de comunidades e visualização interativa. As decisões de projeto (da modelagem dos dados brutos até a geração de artefatos analíticos reprodutíveis) foram orientadas por requisitos forenses: rastreabilidade, reprodutibilidade e clareza interpretativa [1]. A implementação completa está organizada em scripts modulares e documentados [16].

A. Ferramentas e Bibliotecas

O ambiente de desenvolvimento faz uso das seguintes bibliotecas principais:

- `pandas` — pré-processamento e manipulação tabular dos metadados de e-mail;
- `networkx` — construção e análise topológica de grafos (DiGraph);

- `community` (`python-louvain`) — detecção de comunidades por otimização de modularidade;
- `pyvis` — visualização interativa de subgrafos em páginas HTML.

B. Visão Geral do Pipeline

O pipeline consiste em seis etapas principais: Pré-processamento e filtragem; Modelagem do grafo estático; Cálculo de métricas de centralidade; Detecção de comunidades; Validação estrutural; e Visualização de subgrafos. Cada etapa corresponde a um script ou conjunto de funções documentadas no repositório [16].

C. Pré-processamento e Filtragem de Dados

O pré-processamento tem por objetivo transformar o corpus de e-mails semi-estruturados em um arquivo tabular adequado para modelagem de rede. A rotina principal encontra-se em `PreProcessamento.py`. As etapas executadas são:

1) **Extração de metadados:** aplicação de expressões regulares para extrair os campos `From:`, `To:` e `Date:` do texto bruto de cada mensagem.
2) **Normalização temporal:** conversão de strings de data para `pandas.Timestamp` com `utc=True` para unificação de fusos horários; tratamento de valores inválidos via `errors='coerce'` seguido por remoção das entradas corrompidas.
3) **Filtragem histórica:** seleção do intervalo temporal definido pelo estudo (1º de janeiro de 1985 a 31 de dezembro de 2003), removendo mensagens fora desse período.
4) **Explosão de destinatários:** e-mails com múltiplos destinatários são normalizados de modo que cada linha do arquivo final represente uma interação única (remetente $\rightarrow$ destinatário) através do uso de `DataFrame.explode`. Esta escolha preserva a granularidade de pares de comunicação e simplifica a construção do grafo.
5) **Limpeza final:** padronização, eliminação de linhas nulas e salvamento do artefato limpo em `EnronEmailsTratados.csv` para uso posterior.

[16] [17]

Observação metodológica: a opção por representar uma aresta única entre dois atores (independente da frequência de mensagens) foi deliberada para enfatizar a existência de um canal de comunicação em vez de contabilizar volume. Esta modelagem favorece análises topológicas (conectividade, pontos de articulação) porém implica perda de informação relativa a intensidade; métodos alternativos como a utilização de grafos ponderados com contagem de mensagens são descritos nas limitações e estão disponíveis como uma possível melhoria do trabalho.

D. Modelagem do Grafo Estático

A modelagem foi realizada com um grafo dirigido (`networkx.DiGraph`), criado a partir do arquivo tratado via `nx.from_pandas_edgelist` (ver `AnaliseEstatica.py`) [2]. O grafo resultante contém 37 078 nós e 105 717 arestas únicas na configuração reportada pelo experimento.

Justificativas:

- **Direcionalidade:** preserva informação sobre sentido da comunicação (remetente $\rightarrow$ destinatário), essencial para distinguir papéis de emissor e receptor.
- **Aresta única (não multigrafo):** modela a presença/ausência do canal comunicacional; evita complexidade adicional em algoritmos que nem sempre suportam multigrafos de forma nativa.

E. Análise Estática de Centralidade

Os cálculos de centralidade foram realizados no grafo completo (sem amostragem) conforme implementado em `AnaliseEstatica.py` [16]. As métricas computadas incluem:

- **Grau de entrada e saída (`in/out_degree`):** métricas básicas de atividade receptiva e emissiva.
- **Betweenness centrality:** cálculo exaustivo para identificação de *brokers* e gargalos; implementado via `nx.betweenness_centrality`. Devido à sua complexidade computacional de depender da ordenação dos caminhos mais curtos, é uma operação custosa em grafos de grande porte; o script executa o cálculo completo e armazena os resultados. [11] [5] [6]
- **PageRank:** avaliação da autoridade dos nós com `nx.pagerank` (parâmetro `alpha=0.85`), adequada a grafos direcionados. [12]
- **Closeness centrality:** medida de eficiência de difusão, com tratamento de exceções quando valores não são computáveis em grafos com componentes não alcançáveis. [13]

Todos os resultados foram consolidados em um `DataFrame` e salvos em `centralidade_estatica.csv`. Rankings de Top-10 por métrica também foram exportados para arquivos separados para facilitar inspeção forense e revisão humana (`top10_intermediarios.csv`, `top10_pagerank.csv`, `top10_closeness.csv`).

F. Detecção de Comunidades

A detecção de comunidades foi aplicada sobre a versão não-direcionada do grafo (`G.to_undirected()`) e executada com a função `community_louvain.best_partition` [8] [3]. Para garantir reprodutibilidade do componente heurístico, foi fixado `random_state=42`. O mapeamento nó $\rightarrow$ comunidade foi persistido em `comunidades_estaticas.json`. A partir desse mapeamento, realizou-se:

- Cálculo do líder de comunidade (nó com maior betweenness dentro de cada comunidade).
- Geração de um sumário por comunidade com número de membros e identificador do líder (salvo em `analise_comunidades.csv`).

- Extração das cinco maiores comunidades para inspeção detalhada (`top5_maiores_comunidades.csv`).

Justificativa técnica: a utilização de Louvain balanceia eficiência e qualidade da partição para grafos com dezenas de milhares de nós; a modularidade resultante fornece uma medida quantitativa da significância estrutural das comunidades detectadas. [14] [15]

G. Validação Estrutural

A validação tem dois vetores complementares: verificação programática e inspeção manual assistida por artefatos legíveis. Implementado em `VisualizadorGrafo.py` [16], o procedimento:

- Reconstrói o grafo a partir de `EnronEmailsTratados.csv` e itera sobre todos os nós para extrair predecessores e sucessores.
- Exporta um arquivo de texto (`visualizador_grafo.txt`) que lista, para cada nó, quem lhe enviou e para quem enviou, permitindo comparação 1-a-1 com o dataset tratado e auditoria humana.

Esse processo assegura a rastreabilidade entre os dados tabulares e a topologia do grafo, facilitando a detecção de erros de parsing, incongruências de normalização ou perda de interações.

H. Simulação de Disrupção da Rede

Para validar a hipótese de vulnerabilidade estrutural e comparar estratégias de intervenção, foi realizada uma simulação de ataques direcionados (*Targeted Attacks*). O experimento consistiu na remoção sequencial dos nós classificados no topo dos rankings de *Betweenness* (Intermediários), *PageRank* (Autoridades) e da "Elite Estrutural" identificada qualitativamente. A cada remoção, recalculou-se o tamanho do Maior Componente Conexo (*Giant Component*) da versão não-direcionada da rede para mensurar o nível de desintegração estrutural [5], [6], [9], [19]–[22].

I. Visualização de Subgrafos

Dada a dimensão do grafo completo, a exploração visual foi centrada em subgrafos de primeiro grau para atores de interesse (Top-10 por métrica). O script `PlotSubGrafo.py` automatiza [16]:

1) Leitura do grafo completo e verificação da existência do nó de interesse.
2) Extração dos predecessores e sucessores do nó (vizinhança de 1º grau) e construção do subgrafo correspondente.
3) Renderização interativa com `pyvis`, incluindo configuração de parâmetros (tamanho do nó central, ativação da física, layout) e salvamento em HTML para distribuição e inspeção.

Essas visualizações permitem a navegação interativa pelos pontos de interesse do grafo, sendo possível identificar papéis locais e artefatos de comunicação que não emergem facilmente de tabelas e análises estatísticas, além de fornecer a validação visual das análises [9].

J. Reprodutibilidade, Armazenamento e Documentação

Para apoiar a reprodutibilidade foram adotadas as seguintes práticas:

- Persistência de artefatos intermediários
- Fixação de semente pseudo-aleatória (`random_state=42`) para a etapa de Louvain visando resultados determinísticos.
- Log de mensagens e checkpoints nos scripts para facilitar auditoria e reexecução passo a passo.
- `README.md` que documenta as instruções básicas para compilação e execução da solução elaborada [16].

K. Complexidade Computacional e Recursos

Algumas operações têm custo significativo em grafos de grande escala:

- **Betweenness centrality:** algoritmos exatos têm custo assintótico elevado ($\mathcal{O}(n \cdot m)$ para grafos não ponderados, onde n é número de nós e m número de arestas), por isso recomenda-se executar em máquinas com memória e CPU adequadas ou considerar aproximações caso a máquina utilizada não suporte trabalhar em cima do dataset completo da Enron. [13] [11]
- **PageRank e medidas baseadas em autovetor:** convergem tipicamente em tempo polinomial dependente do número de iterações e esparsidade da matriz de adjacência; são, em geral, mais escaláveis que betweenness exata. [12]
- **Detecção de comunidades (Louvain):** eficiente e escalável na prática, porém com variabilidade heurística que foi controlada via `random_state=42`. [15]

L. Limitações e Extensões

Entre as principais limitações metodológicas identificadas e já consideradas no projeto, estão:

- **Perda de informação por não ponderação:** a modelagem com aresta única descarta informação sobre volume e temporalidade fina; extensão natural é construir grafos ponderados por contagem de mensagens ou por janelas temporais móveis.
- **Ausência de conteúdo semântico:** a SNA sobre metadados não captura intenção ou conteúdo; recomenda-se complementar com análise de texto (NLP) quando autorizada eticamente.
- **Componentes desconectados e closeness:** medidas como closeness podem ser instáveis em presença de nós não alcançáveis; o script já trata exceções, mas comparações intermetodológicas são recomendadas.
- **Escalabilidade:** para corporações maiores, migrar para bibliotecas orientadas a grafos em grande escala (`igraph`, frameworks distribuídos ou grafos em banco de grafos) pode ser necessário.
- **Natureza Estática da Análise:** A metodologia principal deste trabalho foca em um grafo estático (agregado), que representa a rede como uma "foto" de todo o período (1985-2003). Esta abordagem, embora robusta para identificar a estrutura de poder geral, é uma limitação, pois

não captura a dinâmica da rede. A Análise Temporal (proposta como trabalho futuro) seria a extensão crucial, permitindo a análise de ”janelas de tempo” (snapshots) para correlacionar picos de centralidade de atores-chave com eventos externos e detectar mudanças comportamentais [7].

M. Considerações Éticas e Legais

Análises forenses de comunicações implicam gestão cuidadosa de privacidade, consentimento e legislação aplicável. Mesmo em corporações públicas como Enron, recomenda-se [1]:

- Documentar origem dos dados e condições de uso;
- Evitar exposição desnecessária de identificadores pessoais em publicações (ou anonimizar/aggregate quando apropriado);
- Validar interpretações com múltiplas fontes e peritos para reduzir risco de inferências errôneas.

N. Conclusão da Metodologia

O pipeline descrito combina práticas robustas de engenharia de dados com técnicas consolidadas de SNA, favorecendo reprodutibilidade e auditabilidade. Os scripts mencionados encapsulam cada estágio do fluxo experimental e foram projetados para permitir reexecução determinística das etapas de pré-processamento, análise e visualização (`PreProcessamento.py`, `AnaliseEstatica.py`, `VisualizadorGrafo.py` e `PlotSubGrafo.py`) [16].

IV. Resultados e Discussão

A aplicação do pipeline metodológico descrito na Metodologia ao grafo estático da Enron, composto por 37 078 nós e 105 717 arestas, resultou em um modelo relacional capaz de capturar com alta fidelidade a estrutura de poder, comunicação e intermediação da organização. [16], [17] Nesta seção, apresentam-se e discutem-se os resultados obtidos em quatro níveis de granularidade analítica: (A) centrada na identificação dos papéis funcionais dos atores-chave; (B) voltada à estrutura modular e comunitária da rede; (C) focada na análise de robustez e disrupção sob ataques simulados; e (D) validação visual qualitativa dos achados por meio da inspeção de subgrafos.

A. Mapeamento de Papéis por Centralidade

A análise de centralidade revelou que a noção de “importância” em uma rede de comunicação é multifacetada. Cada métrica enfatiza dimensões distintas da influência estrutural, intermediação, autoridade e eficiência de difusão, permitindo a identificação de diferentes tipos de atores estratégicos. [11]–[13] As Tabelas I, II e III (extraídas a partir dos arquivos `top10_intermediarios.csv`, `top10_pagerank.csv` e `top10_closeness.csv` [16]) sumarizam os dez nós mais proeminentes em cada métrica.

A análise cruzada dessas três medidas constitui o cerne da interpretação forense, permitindo distinguir papéis funcionais complementares dentro da rede.

TABLE I
Top 10 Atores por Centralidade de Intermediação (Betweenness)

Ator (E-mail)	In-Degree	Out-Degree	Betweenness
jeff.dasovich@enron.com	571	687	0.017990
tana.jones@enron.com	571	740	0.012796
vince.kaminski@enron.com	354	751	0.011642
sara.shackleton@enron.com	604	695	0.010854
gerald.nemec@enron.com	455	508	0.010679
sally.beck@enron.com	483	426	0.010497
louise.kitchen@enron.com	560	349	0.008975
kenneth.lay@enron.com	693	28	0.008279
jeff.skilling@enron.com	662	56	0.008095
kay.mann@enron.com	317	528	0.007500

a) Os Intermediários (“Brokers”): A Tabela I quantifica o papel de intermediação, identificando os atores que funcionam como ”gargalos” ou ”pontes estratégicas” na rede. Esta métrica não mede o volume total de comunicação (como o *in/out-degree*), mas sim o controle estrutural de um nó sobre o fluxo de informação. Um nó com alta *betweenness* é aquele que se situa com maior frequência nos caminhos mais curtos que conectam outros pares de nós, essencialmente ”controlando” a comunicação entre eles. [5], [6], [11]

A análise revela que jeff.dasovich@enron.com (pontuação 0.0179) e tana.jones@enron.com (0.0127) são os intermediários mais críticos da organização. A sua alta pontuação sugere que eles conectam *clusters* (comunidades) e múltiplas pessoas da organização que, de outra forma, provavelmente estariam estruturalmente isolados. Eles funcionam como os principais ”canais” para a comunicação fluir.

Seguindo-os, vince.kaminski@enron.com (0.0116), sara.shackleton@enron.com (0.0108), gerald.nemec@enron.com (0.0106) esally.beck@enron.com (0.0104) formam um segundo escalão de *brokers* essenciais, todos com pontuações robustas acima de 0.01. É notável o perfil de Vince Kaminski: além de ser o broker #3, ele possui o maior *out-degrees* (751) da lista, indicando que ele não apenas é uma ponte, ou seja, repassa apenas oque recebe, mas também uma fonte ativa de disseminação de informação para múltiplos grupos. Em contraste, kay.mann@enron.com (#10) apresenta um perfil de *broker* mais equilibrado, com *in-degree* de 317 e *out-degree* de 528.

Do ponto de vista forense, este ranking é crucial. Ele demonstra que a função tática de conectar os silos da empresa era, delegada principalmente a esse grupo de 10 pessoas. Para um investigador que busca entender o fluxo de informação dentro da organização, ou para uma ação de ruptura que vise fragmentar a rede de comunicação, os alvos prioritários seriam os atores que ocupam o topo da Tabela I, como Dasovich, Jones e Kaminski, e não necessariamente os líderes hierárquicos formais da organização. [1], [3], [9]

b) As Autoridades (“Líderes”): A Tabela II apresenta os resultados da centralidade de PageRank, uma métrica que quantifica a ”autoridade” ou prestígio de um nó. O PageRank mede a importância de um nó de forma recursiva: a pontuação

TABLE II
TOP 10 ATORES POR AUTORIDADE (PAGERANK)

Ator (E-mail)	In-Degree	Out-Degree	PageRank
klay@enron.com	1294	0	0.013190
jeff.skilling@enron.com	662	56	0.007337
kenneth.lay@enron.com	693	28	0.005955
sara.shackleton@enron.com	604	695	0.004513
tana.jones@enron.com	571	740	0.004187
ebass@enron.com	93	1	0.003865
louise.kitchen@enron.com	560	349	0.003647
jeff.dasovich@enron.com	571	687	0.003368
sally.beck@enron.com	483	426	0.003272
gerald.nemec@enron.com	455	508	0.002879

de um ator é alta se ele recebe comunicações de outros atores que também são importantes. [12] Este resultado serve como a mais poderosa validação de toda a metodologia: o topo da lista é ocupado inequivocamente por klay@enron.com (0.0131), jeff.skilling@enron.com (0.0073) e kenneth.lay@enron.com (0.0059).

Conforme o conhecimento público histórico, Jeffrey Skilling e Kenneth Lay foram os CEOs da Enron. O fato de o modelo matemático, analisando apenas metadados de comunicação (quem enviou para quem) e sem nenhum conhecimento prévio do organograma da empresa, ter identificado precisamente a alta administração como os líderes reais da organização, valida a eficácia do PageRank como ferramenta forense para a detecção de hierarquia. [17]

Portanto a análise forense detalhada desta tabela revela perfis de liderança distintos e cruciais para uma investigação:

- **O "Sumidouro de Informação" (Information Sink):** O perfil de klay@enron.com, o primeiro colocado para a pontuação de autoridade, é emblemático. Os dados mostram um *in-degree* massivo de 1294 e um *out-degree* de exatamente zero. Logo esse vértice trata-se de um "vértice terminal" da rede, um nó que recebe praticamente toda informação relevante existente, mas nunca a origina. A análise sugere fortemente que klay@enron.com pode e deve ser conta secundária de Kenneth Lay ou de Jeff Skilling, os dois líderes da rede de comunicação , usada especificamente para receber relatórios consolidados de alto nível, enquanto suas contas públicas, como por exemplo kenneth.lay@enron.com, eram usadas para comunicações limitadas e específicas, ou seja, sendo uma conta de armazenamento de informação onde os dois líderes utilizavam para averiguar tudo que estava acontecendo dentro da organização. Portanto, esta conta "sumidouro" é um repositório de inteligência de valor inestimável.
- **A "Autoridade Oculta":** O caso de ebass@enron.com demonstra o poder do PageRank sobre a simples contagem de *in-degree*. Este ator possui um *in-degree* de apenas 93, trivial comparado aos 600-700 de outros executivos no Top 10. No entanto, sua pontuação de PageRank é a sexta maior da rede. A única explicação matemática para isso é que as poucas comunicações que ebass@enron.com recebeu vieram de atores de altíssima autoridade, como Skilling ou Lay, conferindo-lhe um prestígio estrutural desproporcional. Ele é uma "autoridade oculta", um conselheiro ou figura-chave que seria completamente ignorado por uma análise de volume. [12]
- **A Liderança Híbrida:** Atores como sara.shackleton@enron.com, a quarta no ranking, e tana.jones@enron.com, a quinta colocada, também aparecem nesta lista. Seus perfis são mais equilibrados tendo *in-degree* e *out-degree* relativamente altos, e sua presença aqui, combinada com suas altas pontuações de intermediação (Tabela I), as consolida como parte da "Elite Estrutural", possuindo tanto autoridade hierárquica quanto controle tático do fluxo de informação.

TABLE III
TOP 10 ATORES POR CENTRALIDADE DE PROXIMIDADE (CLOSENESS)

Ator (E-mail)	In-Degree	Out-Degree	Closeness
louise.kitchen@enron.com	560	349	0.136737
john.lavorato@enron.com	391	244	0.132545
sally.beck@enron.com	483	426	0.130950
tim.belden@enron.com	203	114	0.130394
greg.whalley@enron.com	318	105	0.128330
kenneth.lay@enron.com	693	28	0.127991
david.delainey@enron.com	131	235	0.127424
jeff.skilling@enron.com	662	56	0.126445
tana.jones@enron.com	571	740	0.126282
elizabeth.sager@enron.com	271	302	0.126244

c) Os Difusores ("Núcleo Operacional"): A Tabela III identifica os atores mais eficientes na disseminação de informação. A Centralidade de Proximidade não mede o controle (Betweenness) nem a autoridade (PageRank), mas sim a "velocidade" topológica, calculada a partir da distância geodésica média de um nó a todos os outros nós alcançáveis. Atores com alta *closeness* estão posicionados no centro operacional da rede, sendo os mais capazes de transmitir uma mensagem a toda a organização no menor tempo possível. [7], [13]

Os resultados são liderados por louise.kitchen@enron.com (0.1367) e john.lavorato@enron.com (0.1325), com sally.beck@enron.com (0.1309) logo em seguida. Do ponto de vista forense, este grupo representa o "núcleo de difusão" da rede, os gerentes operacionais de médio a alto escalão que servem como "nós de difusão" entre a alta administração e as equipes táticas.

Observa-se também que os CEOs kenneth.lay@enron.com e jeff.skilling@enron.com, bem como pessoas com alto grau de intermediação como sally.beck@enron.com e tana.jones@enron.com, também estão presentes nesta lista. Isso demonstra que, além de seus papéis de comando e intermediação, eles também estavam posicionados de forma ideal para a rápida disseminação de informações.

Simultaneamente, esta lista revela os "operadores-puros": atores como john.lavorato@enron.com, tim.belden@enron.com, greg.whalley@enron.com e david.delainey@enron.com, que não figuravam no topo das listas de Betweenness ou PageRank. Eles são os

gerentes táticos cuja principal função estrutural é a eficiência operacional e a comunicação rápida, completando o mapa da estrutura de poder da rede.

d) A Elite Estrutural da Rede: Finalmente, a sobreposição de atores nas Tabelas I, II e III fornece a evidência empírica mais significativa para uma análise forense. Após a identificação dos papéis funcionais de ”Brokers”, ”Autoridades” e ”Difusores”, é possível agora sintetizar os achados para identificar a elite da rede. [5], [9]

Observamos uma convergência notável: atores como tana.jones@enron.com, sally.beck@enron.com, kenneth.lay@enron.com e jeff.skilling@enron.com aparecem consistentemente nos rankings das três métricas. Este perfil híbrido é raro na rede e indica um papel multifuncional que transcende a simples especialização.

Do ponto de vista forense, este é o grupo de maior periculosidade. Diferente dos ”brokers puros” (como vince.kaminski@enron.com, que é vital para o fluxo mas não é uma autoridade de topo) ou das ”autoridades puras” (como klay@enron.com, que é um destino final de informação mas não um intermediário), esta ”elite estrutural” possui os três atributos do poder informacional. Eles exercem simultaneamente:

- **O Comando (Alta Autoridade):** Sendo destinatários de comunicações importantes, eles estão no topo da hierarquia de tomada de decisão.
- **A Coordenação (Alta Intermediação):** Sendo ”brokers”, eles têm o controle tático para conectar os diversos silos e grupos que comandam.
- **A Disseminação (Alta Proximidade):** Estando no núcleo operacional, eles possuem os meios mais eficientes para disseminar suas decisões pela organização.

Esse pequeno conjunto de nós, portanto, não representa apenas ”pessoas importantes”; ele representa o verdadeiro ”círculo interno” da organização, o nexo de controle estratégico e tático capaz de formular, coordenar e disseminar a informação através de toda a rede.

B. Estrutura Organizacional e Comunidades

A etapa de detecção de comunidades, realizada pelo algoritmo de Louvain, particionou o grafo em *1.697 comunidades* distintas, revelando uma estrutura de comunicação altamente modular, mas com forte tendência a um padrão de núcleo periféria. [14], [15] As cinco maiores comunidades (Tabela IV) concentram a maior parte dos atores e representam o “núcleo corporativo” da empresa. [16], [17]

TABLE IV
TOP 5 MAIORES COMUNIDADES DA REDE (POR Nº DE MEMBROS)

Líder Intermediário Interno	ID	Nº de Membros
louise.kitchen@enron.com	2	3187
jeff.dasovich@enron.com	5	3053
gerald.nemec@enron.com	3	2994
john.arnold@enron.com	0	2847
kevin.hyatt@enron.com	4	2764

A inspeção desses agrupamentos evidencia dois achados principais. Primeiro, há uma forte correlação entre o tamanho da comunidade e a densidade interna de comunicação, o que sugere uma estrutura departamentalizada com lideranças locais bem definidas. Segundo, a análise dos “líderes intermediários internos”, como os nós de maior *betweenness* dentro de cada comunidade, confirma a consistência dos resultados de centralidade global.

a) Validação dos Intermediários: A análise cruzada entre a centralidade de intermediação global (Tabela I) e a liderança intracomunitária (Tabela IV) revela um fenômeno de convergência estratégica. Atores como jeff.dasovich@enron.com, gerald.nemec@enron.com e louise.kitchen@enron.com não operam apenas como pontes entre departamentos desconexos; eles estão, simultaneamente, enraizados no comando tático dos maiores grupos funcionais da empresa. [5], [6], [8]

Do ponto de vista forense, esta dupla atuação de alta centralidade global e local, classifica estes indivíduos como Pontos Únicos de Falha (*Single Points of Failure*) na rede de comunicação. Eles possuem o monopólio da informação que entra e sai de seus departamentos. Em uma investigação, o comprometimento (ou a colaboração) de um desses intermediários, ofereceria ao investigador acesso não apenas a um fluxo de dados específico, mas à infraestrutura de comunicação de setores inteiros da organização, além de causar o colapso da rede de dentro pra fora apenas por sua ausência.

b) Identificação dos Líderes de Silo: Um dos achados mais relevantes da análise de comunidades é a identificação de figuras de poder que permanecem ”invisíveis” às métricas globais de centralidade. Atores como john.arnold@enron.com e kevin.hyatt@enron.com emergem como líderes de comunidades massivas com 2.847 e 2.764 membros, respectivamente. Embora comandem grupos com vários funcionários, estes nomes não figuram no Top 10 de Intermediação (Tabela I) ou Autoridade (Tabela II). [9], [14]

Essa discrepância sugere que eles operam como ”Líderes de Silo”: gestores de subredes densas e autossuficientes, com foco operacional interno e menor necessidade de pontes externas. Para a auditoria forense, isso representa um risco, uma investigação focada apenas nos conectores globais (como Dasovich ou Kitchen) poderia deixar pontos cegos operacionais que estão sob a tutela de Arnold e Hyatt. A detecção de comunidades prova-se, portanto, indispensável para mapear a descentralização operacional onde práticas ilícitas podem ocorrer longe da supervisão central ou dos principais corredores de informação.

c) Estrutura Núcleo-Periferia: A síntese dos resultados de centralidade e detecção de comunidades revela que a topologia da rede Enron não é distribuída uniformemente, mas sim caracterizada por uma arquitetura *Core-Periphery* (Núcleo-Periferia) altamente estratificada. O ”Núcleo” é constituído pela interseção da Elite Estrutural com os ”Intermediários Pontos Únicos de Falha” identificados na análise mesoscópica. Este núcleo mantém a coesão das cinco maiores comunidades (”silos”), enquanto a ”Periferia” é composta por milhares de nós com baixo grau de conexões, dependentes

inteiramente desses líderes para o acesso à rede global. [9], [15]

C. Simulação de Disrupção e Robustez da Rede

Para validar empiricamente a vulnerabilidade estrutural da organização, realizou-se uma simulação de ataques direcionados, monitorando a degradação do Componente Gigante da rede em função da remoção de nós-chave [17], [19], [21]. A Figura 1 apresenta a comparação entre as estratégias de remoção de Autoridades (Decapitação), Intermediários (Fragmentação) e da Elite Estrutural.

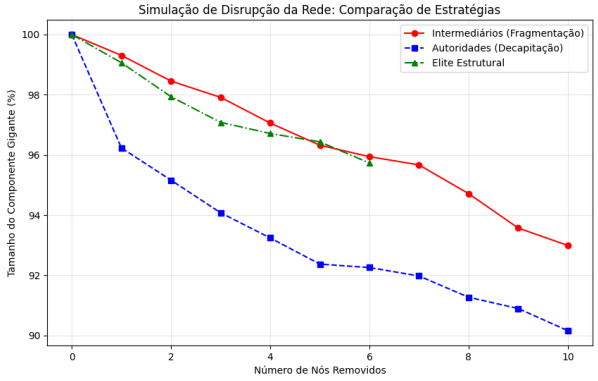

Fig. 1. Análise de Robustez da Rede sob Ataques Direcionados. O eixo Y representa a porcentagem de nós remanescentes no maior componente conectado após a remoção sequencial de N nós (eixo X).

A análise das curvas revela um padrão distinto de fragilidade corporativa [19], [20]:

1) **Dominância da Decapitação (Linha Azul):** A estratégia de remoção de Autoridades apresentou o maior impacto sistêmico durante todo o experimento, provocando a redução mais acentuada no tamanho da rede (de 100% para ≈90%) [17], [19]. Isso confirma que a topologia da Enron é fortemente centralizada [18]. Os líderes executivos atuam como *hubs* massivos; sua remoção causa o desprendimento imediato de um grande volume de nós periféricos que dependem exclusivamente deles para conexão [11], [19].
2) **Resiliência à Fragmentação (Linha Vermelha):** A remoção de Intermediários causou um declínio mais suave na integridade da rede (terminando em ≈93%) [21], [22]. Isso sugere que, embora os *brokers* sejam vitais para a eficiência do fluxo de informação [5], [11], a estrutura corporativa possui redundância suficiente (caminhos alternativos) para manter os departamentos conectados globalmente, mesmo após a perda de pontes importantes [20], [21].
3) **Comportamento da Elite Estrutural (Linha Verde):** A curva da Elite Estrutural posicionou-se de forma intermediária, mas próxima à tendência dos Intermediários [15], [17]. Isso indica que, apesar de conter figuras de autoridade, a seleção qualitativa deste grupo capturou atores que sustentam a coesão interna, mas cuja remoção não provoca o mesmo isolamento massivo de ”nós-folha” que a remoção puramente algorítmica dos maiores *hubs* de PageRank [16], [19].

D. Validação Visual e Interpretação Estrutural

A etapa final da análise consiste na validação qualitativa dos resultados estatísticos por meio da inspeção visual da topologia local. Enquanto as métricas de centralidade resumem a importância de um ator em um único valor numérico, esta análise permite examinar a morfologia das conexões que geram esse valor, distinguindo nuances que os dados agregados podem ocultar. Essa validação é crucial pois, em redes complexas, nós com indicadores de centralidade idênticos podem ocupar posições estruturais fundamentalmente distintas — por exemplo, diferenciando um ator que centraliza um grupo coeso (alta redundância) de um que conecta grupos desconexos (alta dependência). A inspeção visual atua, portanto, como uma camada de verificação semântica sobre o cálculo matemático.

Para tanto, foram gerados subgrafos egocêntricos para os atores-chave identificados nas etapas anteriores. Um subgrafo egocêntrico de um nó v é definido formalmente como o subconjunto do grafo composto pelo nó v, seus vizinhos imediatos e todas as arestas existentes entre esses vizinhos. Esta abordagem, implementada através da biblioteca `pyvis`, isola o ”microcosmo” comunicacional de cada ator, filtrando a complexidade da rede global de mais de 37 mil nós para focar na dinâmica local de influência. O uso de renderização interativa permite identificar padrões morfológicos específicos, como estruturas em estrela (*hubs* de autoridade), pontes (vetores de intermediação) ou cliques densos (núcleos operacionais), facilitando a interpretação forense da função real que cada alvo desempenha na organização [14], [16].

As Figuras 2, 3 e 4 apresentam os subgrafos de três atores que representam arquétipos distintos de influência na rede Enron. A análise visual destas figuras confirma que atores com diferentes perfis de centralidade (Autoridade, Intermediação e Difusão) ocupam posições topologicamente distintas na rede, corroborando a classificação funcional proposta. Fornecidos como material complementar a essa análise estão os subgrafos dos demais agentes de centralidade, ou seja, os subgrafos de todas as pessoas que aparecem nos ranking de centralidade e autoridade, fornecidos através de arquivos htmnl na pasta `Outputs` no repositório do github. [16]

O subgrafo de kenneth.lay@enron.com (Figura 2) confirma o padrão típico de autoridade hierárquica estabelecido na literatura de redes [12]. A topologia é dominada por um nó central com alto *in-degree* de 693 e quase nenhuma conexão de saída *out-degree* de 28, formando uma estrutura em estrela (“*hub-and-spoke*”), identificada classicamente como o modelo de máxima centralização de grau [13]. Essa configuração reforça o papel de Lay como vértice terminal do fluxo comunicacional, operando como um “sumidouro de informação” (*information sink*), com função predominantemente receptora, característica de altos executivos em redes corporativas que centralizam a tomada de decisão estratégica [1].

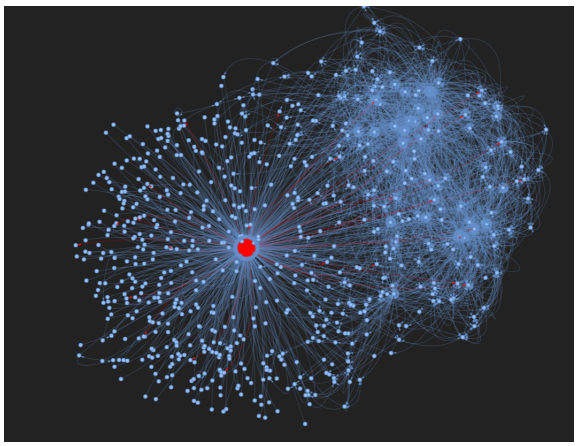

Fig. 2. Subgrafo do Ator de Autoridade — Kenneth Lay - O nó vermelho central e as arestas vermelhas representam Kenneth, enquanto os nós azuis e as arestas azuis são de seus vizinhos

O subgrafo de jeff.dasovich@enron.com (Figura 3)apresenta uma topologia visual de alta complexidade e densidade, diferindo do padrão clássico de ”ponte simples” [11], [13], [18], diversas arestas saindo do nó central sem uma densidade de malha tão grande. A imagem revela uma estrutura maciça e coesa, onde o nó central está imerso em uma teia densa de conexões e se conecta individualmente com a grande maioria dos outros nós dessa red [14], [15]. Isso reflete visualmente sua dupla função na rede, uma vez que ele não é apenas um intermediário isolado que conecta grupos distantes [5], [11], mas também atua como o hub centralizador de uma vasta comunidade operacional (a Comunidade 4) [15], [18]. A alta densidade de arestas entre seus vizinhos e a alta densidade de arestas saindo Dasovich sugere que ele gerenciava um ecossistema de comunicação altamente ativo e interdependente, validando seu papel como um intermediário importante que possuía tanto o papel de ser uma das pontes que conecta a rede, quanto o controle e a liderança de um núcleo interno robusto [6], [9], [10].

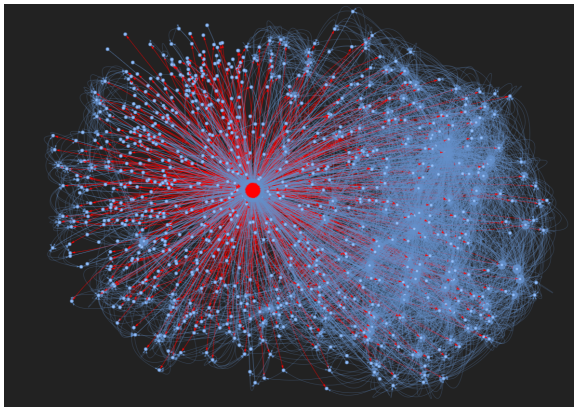

Fig. 3. Subgrafo do Ator Intermediário — Jeff Dasovich - O nó vermelho central e as arestas vermelhas representam Jeff, enquanto os nós azuis e as arestas azuis são de seus vizinhos

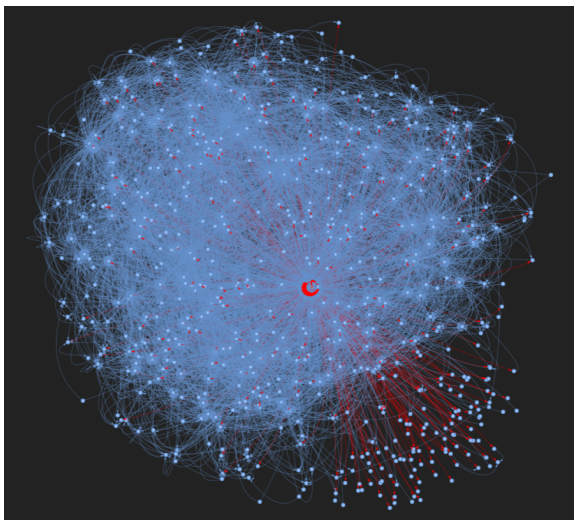

Fig. 4. Subgrafo do Ator Difusor — Louise Kitchen - O nó vermelho central e as arestas vermelhas representam Louise, enquanto os nós azuis e as arestas azuis são de seus vizinhos

O subgrafo de louise.kitchen@enron.com (Figura 4) revela uma topologia visualmente distinta, caracterizada por uma estrutura de malha muito densa [11], [14], [15]. Ao contrário da estrutura radial e centralizada observada nas autoridades, a vizinhança de Kitchen apresenta uma elevada conectividade, onde os atores conectados a ela também se comunicam intensamente entre si, formando múltiplos triângulos e ciclos de comunicação [14], [15]. Visualmente, isso se manifesta como uma ”teia” coesa e saturada de arestas, o que valida empiricamente sua classificação como a principal ”difusora” da rede [3], [12]. Esta densidade local indica que Kitchen não atuava como um gargalo, mas sim como o motor de um ”microcosmo de eficiência”, onde a redundância de conexões permitia que a informação fluísse horizontalmente e verticalmente com extrema rapidez [7], [18].

E. Síntese dos Achados e Implicações Forenses

A integração das diferentes camadas analíticas como a centralidade global, a estrutura modular, a análise de robustez e a morfologia local, revela que a governança informacional da Enron operava sob uma hierarquia híbrida complexa, caracterizada por uma dependência estrutural de sua liderança e uma dependência operacional de seus intermediários. [11], [14], [19]

Os resultados mapearam uma ”Elite Estrutural”: kenneth.lay, jeff.skilling, tana.jones e sally.beck. Esses atores destacam-se por exibir, simultaneamente, os maiores índices de autoridade (PageRank) e de capacidade de difusão. [12], [16] Enquanto Lay e Skilling representam o vértice decisório (o ”sumidouro” de informações), Jones e Beck atuam como o elo executivo dessa elite, convertendo as diretrizes estratégicas em fluxo operacional. [5] A análise de centralidade confirma que este grupo detém o controle dos canais mais curtos de comunicação para qualquer ponto da rede. [11], [13]

Contudo, a análise integrada evidencia que a eficiência tática da rede não residia exclusivamente no topo. Enquanto as métricas de centralidade destacaram uma camada crítica de intermediários de alto *betweenness*, como jeff.dasovich e louise.kitchen, a detecção de comunidades revelou a importância de líderes de silos operacionais, como `john.arnold`, que operavam fora dos rankings globais. [3], [11], [14] Embora estes atores não possuam a autoridade formal máxima da Elite, eles detêm o monopólio das conexões transversais e da coesão local. [5], [6] São eles os responsáveis por traduzir as ordens da Elite para os diversos departamentos operacionais, que de outra forma estariam isolados em suas comunidades. [1], [9]

A simulação de disrupção (Fig. 1) trouxe uma qualificação fundamental para a estratégia de intervenção física. [19], [21] Ao contrário de redes criminosas descentralizadas, a Enron mostrou-se vulnerável à estratégia de Decapitação. [19] A remoção das Autoridades de topo provocou a maior degradação imediata da rede, isolando a base operacional da estrutura de comando. [21]

Entretanto, sob a ótica da coleta de evidências, o papel dos intermediários ganha uma nova dimensão. [3], [10] Embora sua remoção cause menor fragmentação imediata do que a remoção dos líderes, eles constituem ativos de inteligência superiores. [3], [6] Devido à sua posição topológica privilegiada (conectando múltiplos módulos), atores como Dasovich e Kitchen possuem uma "visão de rede" mais ampla do que os próprios executivos de silos isolados. [3], [10]

Em termos forenses, isso sugere uma abordagem tática bifrontal. Para o desmantelamento imediato da atividade (disrupção), a intervenção deve priorizar a prisão da Elite Estrutural e dos demais líderes presentes na Tabela II, pois eles são os pontos de falha estrutural que mantêm a rede unida. [16], [19] Porém, para a instrução probatória e mapeamento da hierarquia oculta, os intermediários identificados devem ser tratados como alvos prioritários para monitoramento ou colaboração premiada, pois servem como vetores de informação que expõem tanto a Elite acima deles quanto os núcleos operacionais abaixo. [3], [5], [10]

V. Conclusão e Próximas Etapas

Este trabalho demonstrou a eficácia da aplicação de um *pipeline* metodológico baseado em Teoria dos Grafos e Análise de Redes Sociais (SNA) para fins de inteligência forense. Utilizando o *dataset* da Enron como estudo de caso [17], foi possível não apenas reconstruir a arquitetura de comunicação da empresa, mas também dissecar sua estrutura de poder oculta, indo além da análise superficial de volumes de mensagens [1], [2].

Os resultados evidenciaram um padrão de estrutura núcleo-periferia, onde a aplicação integrada de métricas de centralidade permitiu classificar os atores em papéis funcionais distintos: a Elite Estrutural (comando), os intermediários importantes (coordenação/brokers) e o Núcleo Operacional (difusão). Essa interpretação é compatível com trabalhos clássicos de centralidade [11]–[13]. A detecção de comunidades, por sua vez, revelou a organização departamental de fato e identificou líderes locais ("Líderes de Silo") que operavam fora do radar das métricas globais, em consonância com os métodos de Girvan–Newman [14] e Louvain [15].

A principal contribuição deste estudo reside na demonstração empírica de que a topologia da rede corporativa investigada difere de redes criminosas resilientes. Os achados indicam que o desmantelamento eficaz de organizações hierarquizadas como a Enron é melhor alcançado pela estratégia de decapitação da liderança, uma vez que a simulação comprovou que a rede é estruturalmente dependente de seus *hubs* centrais de autoridade, para manter a coesão global [9].

Como trabalhos futuros e extensões desta pesquisa, é proposto:

- **Análise Temporal Dinâmica:** A implementação de uma análise de janelas de tempo deslizantes (*sliding windows*) para monitorar a evolução das métricas de centralidade. Isso permitiria correlacionar picos de atividade de intermediários específicos com eventos externos e detectar mudanças comportamentais indicativas, como a súbita desconexão de atores-chave. [7].
- **Modelagem de Grafos Ponderados:** A refatoração do modelo para considerar o peso das arestas (frequência de e-mails) poderia refinar a distinção entre canais de comunicação esporádicos e rotas de comando intensivas, oferecendo uma visão mais granular da hierarquia. [3], [8], [10].
- **Enriquecimento Semântico (NLP):** A integração de técnicas de Processamento de Linguagem Natural para analisar o conteúdo dos e-mails em conjunto com os metadados. A correlação entre a topologia da rede e a análise de sentimento ou modelagem de tópicos permitiria identificar não apenas quem controla o fluxo de informação, mas qual o teor dessa informação. [3], [10].
- **Predição de Links Ocultos:** A aplicação de algoritmos de *Link Prediction* para inferir conexões que foram intencionalmente omitidas ou realizadas por canais alternativos, aumentando a robustez da investigação contra táticas de contra-forense. [4].

Em suma, este estudo reafirma que a análise de metadados, quando tratada com rigor matemático, constitui uma ferramenta indispensável para a moderna investigação criminal, transformando dados brutos em inteligência acionável. [1], [9].

References

[1] M. K. Sparrow, "The application of network analysis to criminal intelligence: An assessment of the prospects," *Social Networks*, vol. 13, pp. 251–274, Sep. 1991. doi:10.1016/0378-8733(91)90008-H. [scholar.harvard.edu]

[2] H. Sarvari, E. Abozinadah, A. Mbaziira, and D. McCoy, "Constructing and analyzing criminal networks," in *Proc. IEEE Security & Privacy Workshops (SPW)*, May 2014, pp. 84–91. doi:10.1109/SPW.2014.22. [IEEE TCSP]

[3] E. Ferrara, P. De Meo, S. Catanese, and G. Fiumara, "Detecting criminal organizations in mobile phone networks," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5733–5750, 2014. doi:10.1016/j.eswa.2014.03.024. [ScienceDirect]

[4] G. Berlusconi, F. Calderoni, N. Parolini, M. Verani, and C. Piccardi, "Link prediction in criminal networks: A tool for criminal intelligence analysis," *PLoS ONE*, vol. 11, no. 4, p. e0154244, Apr. 2016. doi:10.1371/journal.pone.0154244. [PMC]

[5] R. Grassi, C. Piccardi, F. Calderoni and A. Lazzarini, "Betweenness to assess leaders in criminal networks: New evidence using the dual projection approach," *Criminology and Criminal Justice*, vol. 19, no. 3, pp. 341–363, May 2019. doi:10.1016/j.crimjus.2018.10.013.

[6] P. Magalingam, S. Davis, and A. Rao, "Ranking the importance level of intermediaries to a criminal using a reliance measure," *arXiv:1506.06221*, Jun. 2015. [Online]. Available: https://arxiv.org/abs/1506.06221. [arXiv / IEEE workshop]

[7] L. Falzon, E. Quintane, J. Dunn, and G. Robins, "Embedding time in positions: Temporal measures of centrality for social network analysis," *Social Networks*, vol. 54, pp. 168–178, Mar. 2018. doi:10.1016/j.socnet.2018.02.002. [Online]. Available: https://www.researchgate.net/publication/323581466_Embedding_time_in_positions_Temporal_measures_of_centrality_for_social_network_analysis

[8] A. Bahulkar *et al.*, A. Bahulkar, B. K. Szymanski, N. O. Baycik, and T. C. Sharkey, "Community detection with edge augmentation in criminal networks," in Proc. IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, Aug. 2018, pp. 1168-1175. doi: 10.1109/ASONAM.2018.8508326.

[9] L. Cavallaro *et al.*, L. Cavallaro, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, and A. Liotta, "Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia," PLOS ONE, vol. 15, no. 8, p. e0236476, Aug. 2020. doi:10.1371/journal.pone.0236476.

[10] V. Bellandi *et al.*, V. Bellandi, P. Ceravolo, S. Maghool, and S. Siccardi, "Graph embeddings in criminal investigation: Towards combining precision, generalization and transparency," World Wide Web, vol. 25, pp. 2379–2402, Feb. 2022. doi: 10.1007/s11280-021-01001-2.

[11] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977. doi:10.2307/2786543. [JSTOR]

[12] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107–117, Apr. 1998. doi:10.1016/S0169-7552(98)00110-X. [Elsevier]

[13] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978/79. doi:10.1016/0378-8733(78)90021-7. [ScienceDirect]

[14] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, Feb. 2004. doi:10.1103/PhysRevE.69.026113. [APS]

[15] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, Oct. 2008. doi:10.1088/1742-5468/2008/10/P10008. [IOPscience]

[16] Vinícius Ramalho de Oliveira, "Analise de Grafos na Informatica Forense" *GitHub Repository*, 2025. [Online]. Available: https://github.com/ViniciusRO22/Analise-de-Grafos-na-Informatica-Forense.git

[17] W. Cukierski, "The Enron Email Dataset," Kaggle, 2016. [Online]. Available: https://www.kaggle.com/datasets/wcukierski/enron-email-dataset Accessed: Nov. 19, 2025.

[18] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999. doi:10.1126/science.286.5439.509.

[19] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, Jul. 2000. doi:10.1038/35019019.

[20] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," *Physical Review Letters*, vol. 85, no. 21, pp. 4626–4628, Nov. 2000. doi:10.1103/PhysRevLett.85.4626.

[21] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Physical Review Letters*, vol. 85, no. 25, pp. 5468–5471, Dec. 2000. doi:10.1103/PhysRevLett.85.5468.

[22] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, p. 056109, 2002. doi:10.1103/PhysRevE.65.056109.