UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FíSICA

Estudo sobre a topologia das redes criminais

Por
Bruno Requião da Cunha

Agosto de 2017.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FíSICA

Tese de Doutorado

Estudo sobre a topologia das redes criminais

Por
Bruno Requião da Cunha

Tese doutoral submetida como requisito parcial
para a obtenção do título de Doutor em Física.

Orientador: Prof. Dr. Sebastián Gonçalves

Agosto de 2017.

Dedico este trabalho a todos os que lutam incansavelmente pelo bem. Principalmente aos herois sem rosto, desconhecidos, que operam longe dos holofotes e do suave bem-estar dos gabinetes ar-condicionados. A estes abnegados, que não raro arriscam suas vidas para servir e proteger aos inocentes. Aos fortes na linha avançada. Que este estudo auxilie-vos de alguma forma a combater o bom combate. Também aos que vieram antes, meus avós e antepassados. Aos que se foram, mas que também ficam pois o futuro emerge do passado. E pelos que virão, que sejamos exitosos em construir uma sociedade mais justa e segura.

"AD VTRVMQVE PARATVS!"

— PVBLIVS VERGILIVS MARO

Agradeço à minha esposa Fernanda pela paciência, carinho, companheirismo, cumplicidade, amor e principalmente por me incentivar a voar, mas me mantendo sempre os pés no chão. Aos meus pais Marco Antônio e Mônica por terem sempre acreditado em mim apesar das muitas curvas e obstáculos que ocorreram até aqui e pelo seu incondicional e sempre presente amor.

Sou grato também aos Policias Federais Luiz Walmocyr dos Santos Júnior, Jean Fernando Passold e Jair Soares Fonseca Filho pela parceria no bom combate, pelas discussões e por acreditarem no projeto. Também aos Policiais Federais Mauro Lima Silveira e Fernando Schwengber Casarin pelo suporte, visão e apoio institucional. A todos pelo profissionalismo e espírito de corpo, e à Polícia Federal brasileira, casa que me acolheu.

Gratulo, por termo, ao meu orientador, Professor Sebastián, e ao Doutor Juan Carlos González-Avella por apostarem em um campo da ciência ainda bastante inexplorado, pelo profissionalismo e pela parcimônia na indicação do rumo acadêmico-científico.

Enfim, remerceio a todos que de alguma maneira contribuíram para este estudo.

> “I am only one, but I am one. I cannot do everything, but I can do something. And because I cannot do everything, I will not refuse to do the something that I can do"
>
> — Edward Everett Hale, Essentials of English Composition

Artigos publicados

Esta tese é parcialmente fundamentada nas seguintes publicações:

- da Cunha, B. R., González-Avella, J. C., & Gonçalves, S. (2015).
  Fast fragmentation of networks using module-based attacks.
  Plos one, 10(11), e0142824, doi: 10.1371/journal.pone.0142824.

- da Cunha, B.R., & Gonçalves, S. (2017).
  Performance of attack strategies on modular networks.
  J Complex Netw, doi:10.1093/comnet/cnx015.

- da Cunha, B.R., & Gonçalves, S. (2017).
  Web of crime in Brazil is controllable and sensitive to module-based attacks.
  arXiv, 1706.03153.

- da Cunha, B.R., & Gonçalves, S. (2017).
  Patterns of criminal activities inside deep web's clandestine networks.
  Unpublished.

“DIMIDIVM FACTI QVI COEPIT HABET: SAPERE AVDE!"

— QVINTVS HORATIVS FLACCVS, EPISTVLARVM LIBER PRIMVS

Resumo

Nesta tese investigam-se três pontos ligados a fragilidades topológicas de grafos e suas aplicações a redes complexas reais e, em especial, a redes de relacionamentos criminais. Na primeira etapa, apresenta-se *in abstracto* um método inédito e eficiente de fragmentação de redes complexas por módulos. O procedimento identifica em primeiro lugar comunidades topológicas por meio da qual a rede pode ser representada usando algoritmos heurísticos de extração de comunidades. Então, somente os nós que participam de ligações inter-comunitárias são removidos em ordem decrescente de sua centralidade de intermediação. Ilustra-se o método pela aplicação a uma variedade de redes reais nas áreas social, de infraestrutura, e biológica. Mostra-se que a abordagem por módulos supera ataques direcionados a vértices baseados somente no ordenamento de índices de centralidade, com ganhos de eficiência fortemente relacionados à modularidade da rede.

No segundo momento, introduzem-se os conceitos de robustez e fragilidade de redes generalizadas para avaliar o quanto um determinado sistema se comporta frente a ataques incompletos. Ainda, avalia-se o desempenho (relação entre robustez e custo computacional) de diversos ataques sequenciais e simultâneos a redes modulares por meio de uma medida empírica que chamamos de *performance*. Mostra-se por meio de redes artificiais de referência e de redes reais que para sistemas altamente modulares a estratégia de fragmentação por módulos apresenta um desempenho até 10 vezes superior aos demais ataques.

Na última etapa, explora-se com maior profundidade a natureza subjacente de redes reais de relacionamentos criminais. Apresenta-se uma rede única e sem precedentes construída pela Polícia Federal Brasileira consistindo de mais de 35.000 relacionamentos entre 24.000 indivíduos. Os dados foram coletados entre abril e agosto de 2013 e consistem em informações fornecidas diretamente pelos investigadores responsáveis de cada caso. O sistema apresenta características típicas de redes sociais, porém é bem mais “escuro"que o comportamento típico, com baixos níveis tanto de densidade de arestas quanto de eficiência de rede. Além do mais, o sistema é extremamente modular o que implica ser possível desmantelar toda a rede de crimes federais brasileiros com a remoção de aproximadamente 2% dos indivíduos escolhidos conforme a prescrição do método modular. Também, a rede é controlável no sentido da teoria matemática de controle, significando que com acesso a aproximadamente 20% dos nós é possível, em tese, levar qualquer variável dinâmica de um estado inicial a um estado final arbitrário em um tempo finito.

Exibi-se também uma análise topológica e de fragilidades de uma segunda rede criminal relacionada a investigações da Polícia Federal. Trata-se de um fórum online destinado à prática de crimes cibernéticos na chamada camada profunda da internet (*deep web*). Após a coleta dos dados foi possível construir uma rede de relacionamentos com quase 10.000 indivíduos. Comparou-se, então, a estratégia usada de fato pela Polícia Federal durante a Operação Darknet com a previsão teórica de ataques topológicos à rede criminal e mostrou-se que ataques dirigidos por grau teriam fragmentado o sistema de maneira quase 15 vezes mais eficiente. Por outro lado, esta rede não é modular apesar de novamente

apresentar uma arquitetura mais “escura"que o usual.

Por termo, demonstra-se que os ataques por arestas estão diretamente relacionados ao aprisionamento enquanto que a ressocialização e/ou morte dos indivíduos é melhor interpretada como a remoção por vértices. Destarte, comprovou-se que de um ponto de vista topológico a ressocialização é de fato mais eficiente em reduzir a criminalidade do que o aprisionamento. Contudo, na rede de crimes federais estudada essa diferença é muito pequena, de tal modo que ambas as políticas poderiam, em tese, ser aplicadas a fim de se combater eficientemente o sistema criminoso.

Abstract

In this thesis we investigate three points connected to topological fragilities of graphs and their applications to real complex networks and, in particular, to networks of criminal relationships. In the first step, we present an unprecedented and efficient method of fragmentation of complex networks by modules. Firstly, the procedure identifies topological communities through which the network can be represented using heuristic communities extraction algorithms. After that, only the nodes that bridge communities are removed in descending order of their betweenness centrality . We illustrate the method by the applying it to a variety of real networks in the social, infrastructure, and biological fields. We show that the modular approach outperforms attacks traditional attacks based only on the ordering of centrality indexes, with efficiency gains strongly related to the modularity of the network.

In the second moment, we introduce the concepts of generalized robustness and fragility of networks to evaluate how much a certain system behaves in the face of incomplete attacks. Also, we evaluate the relation between robustness and computational cost of several sequential and simultaneous attacks to modular networks by means of an empirical measure that we call *performance*. In this sense, we show through artificial and real networks that for highly modular systems the strategy of fragmentation by modules presents a performance up to 10 times superior to traditional attacks.

In the last step, we explore in more depth the underlying nature of real networks of criminal relationships. We present a unique and unprecedented network built by the Brazilian Federal Police consisting of more than 35,000 relationships among 24,000 individuals. The data were collected between April and August 2013 and consist of information provided directly by the investigators responsible for each case. The system has typical characteristics of social networks, but is much "darker"than traditional social networks, with low levels of edge density and network efficiency. Moreover, the network is extremely modular which implies that it is possible to dismantle all the network of Brazilian federal crimes with the removal of approximately 2% of the individuals chosen according to the modular method. Also the network is controllable in the sense of the mathematical control theory, meaning that with access only to 20% of nodes it is possible, In theory, to take any dynamic variable from an initial state to an arbitrary final state in a finite time.

We also show a topological analysis of a second criminal network related to Federal Police investigations. This is an online forum for cybercrime in the so-called deep web. After the data collection, it was possible to build a network of relationships with almost 10,000 individuals. We then compared the strategy actually used by the Federal Police during Operation Darknet with the theoretical prediction of topological attacks on the criminal network and showed that degree-based attacks would have fragmented the system almost 15 times more efficiently. On the other hand, this network is not modular despite presenting a "darker"architecture than usual. As a last result, this particular system is not controllable in practical terms.

We finish the study by showing that edge attacks are directly related to the imprisonment whereas the resocialization and/or death of the individuals is better interpreted as the removal of vertices. Thus,

we prove that from a topological point of view resocialization is in fact more efficient in reducing crime rates than imprisonment. However, in the network of federal crimes studied here this difference is very small, so that both policies could in theory be applied in order to combat effectively the criminal system.

Lista de Figuras

2.1 Representação da rede de distribuição de energia elétrica do oeste dos Estados Unidos (A), uma possível representação modular (B), e a estrutura interna dos nós e arestas dentro de duas comunidades selecionadas, além dos nós conectando dois módulos (C) [56]. 27

2.2 Visualização dos passos do algoritmo *Louvain*. Cada etapa consiste num momento em que a modularidade é otimizada ao se permitir apenas mudanças locais nas comunidades. Na segunda etapa as comunidades extraídas são agregadas para se construir uma nova rede de comunidades. As etapas são repetidas iterativamente até não haver mais ganho na modularidade. A figura foi adaptada do artigo original de Blondel *et al* [59]. 29

3.1 Comparação entre o efeito dos ataques à rede de distribuição de energia dos EUA: por intermediação, por grau, por maior caminho, aleatório e por módulos. (A) Tamanho da maior componente conectada em termos do tamanho original, G, como função da fração de nós retirados, q. (B) Representação modular da rede. (C) Fotos da representação gráfica da rede quando 1%, 2% e 3% dos nós são retirados utilizando os métodos de ataque por centralidade de intermediação (HB) e por módulos (MBA). 36

3.2 Mostra-se a relação entre a fração de nós que conectam diferentes módulos, e a modularidade, Q. Os dados correspondem a dez redes reais: Facebook (FB), Twitter (TW), Google Plus (G +), rede de energia dos Estados Unidos (PG), Estradas Europeias (ER), voos abertos (OF), aeroportos dos Estados Unidos (UA), proteína de levedura (YP), H *pylori* (HP) e C *elegans* (CE). Como esperado, observa-se uma alta correlação (negativa) entre E_{int} e Q, que é precisamente a característica que torna o método de modularidade bem colocado. A extração de comunidades foi realizada utilizando os métodos *Louvain* ou Infomap como detalhado na tabela 3.1. 38

3.3 Tamanho da maior componente conectada em termos do tamanho inicial da rede, G, em função da fração de nós removidos q. Ataque a vértices por módulos (quadrados pretos), ataque a vértices por centralidade de intermediação (círculos vermelhos). (A) Rede de distribuição elétrica do oeste norteamericano. (B) Rede de estradas européias. (C) Voos abertos. (D) Aeroportos norteamericanos. (E) Facebook. (F) Twitter. (G) Google plus. (H) Proteína do levedo. (I) H *pylori*. (J) C *elegans*. As interseções das linhas azuis pontilhadas correspondem ao ponto de dano máximo da rede utilizando o ataque por módulos. 39

3.4 Tamanho da maior componente conectada em termos do tamanho inicial da rede, G, em função da fração de arestas removidas q. Ataque por módulos (quadrados pretos) e por centralidade de intermediação (círculos vermelhos). (A) Rede de distribuição elétrica do oeste norteamericano. (B) Rede de estradas européias. (C) Voos abertos. (D) Aeroportos norteamericanos. (E) Facebook. (F) Twitter. (G) Google plus. (H) Proteína do levedo. (I) H *pylori*. (J) C *elegans*. As interseções das linhas azuis pontilhadas correspondem ao ponto de dano máximo da rede utilizando o ataque por módulos. 40

3.5 Ganho em eficiência do ataque a vértices por módulos se comparado com o ataque por centralidade de intermediação ($\gamma = G_{null}/G$) como função da fração de nós removidos, q. O código das redes é dado por Facebook (FB),Twitter (TW), Google Plus (G+), rede elétrica dos EUA (PG), estradas européias (ER), voos internacionais (OF), aeroportos dos EUA (UA), proteína de levedo (YP), H *pylori* (HP) e C *elegans* (CE). Redes de infraestrutura estão pintadas de vermelho, biológicas de verde e sociais de azul. 41

3.6 Ganho global de eficiência (η) do ataque por módulos relativo ao ataque por intermediação como função da modularidade Q para remoção de vértices (A) e de arestas (B). O eixo vertical está em escala logarítmica e o eixo horizontal em escala linear. As redes atacadas são Facebook (FB),Twitter (TW), Google Plus (G+), rede elétrica dos EUA (PG), estradas européias (ER), voos internacionais (OF), aeroportos dos EUA (UA), proteína de levedo (YP), H *pylori* (HP) e C *elegans* (CE). 42

3.7 Vinte rodadas distintas do ataque por módulos segundo os algoritmos de *Louvain* e Infomap. Mostram-se os ataques aos nós no caso da rede elétrica dos EUA e a média deles para demonstrar a típica sensibilidade do método proposto à escolha particular do algoritmo de extração de comunidades. 42

3.8 Representação geométrica da robustez generalizada como a razão entre a área sob a curva vermelha e a área máxima de ataque delimitada pelo retângulo com lados $1 - G_{min}$ e 1 conforme a definição equação 3.4. 46

3.9 Aqui se plotam em escala semi-log as performances ($\mathcal{P}$) dos métodos HBA (triângulos vermelhos para cima), HDA (triângulos verdes para baixo), MBA (pontos azuis) e CI (quadrados pretos) em redes de referência LFR com tamanho $N = 10^4$ e modularidade variando entre $0,48 < Q < 0,99$. Uma bola com raio $l = 3$ foi utilizada nas simulações CI após um processo de otimização de raios. 48

3.10 A figura mostra o processo de fragmentação, isto é, o tamanho da maior componente conectada (G) em função da fração de nós removidos (q), das redes de energia elétrica da União Europeia e dos Estados Unidos da América, o sistema de rodovias da União Europeia, o proteoma do levedo, extratos do Google Plus, voos abertos, proteoma do H *pylori* e do C *elegans* além dos aeroportos dos Estados Unidos da América sob os ataques HBA (linhas vermelhas pontilhadas), HDA (linhas verdes pontilhadas), MBA (linhas azuis) e CI (linhas tracejadas pretas). Uma bola com raio $l = 3$ foi utilizada para as simulações CI após um processo de otimização de raios. 49

3.11 A figura retrata os histogramas da performance ($\mathcal{P}$) para cada attack (azul), HBA (vermelho), HDA (verde) e CI (preto) em cada rede real. 49

3.12 Representação da maior componente conectada da rede de crimes federais brasileiros consistindo em 9.887 indivíduos e 91 módulos. As cores representam vértices da mesma comunidade conforme extraídos pelo método de *Louvain*. 52

3.13 Distribuição de grau, p_k, para a rede de crimes federais em escala log-log. A figura mostra a característica típica de redes sociais com saturação de baixa conectividade e corte para alto grau (painel à esquerda). Já no painel à direita tem-se a distribuição de grau reescalonada p_k como função de $k + k_{sat}$ também em escala log-log e respectivo ajuste para lei de potência. 54

3.14 O mapa tipo radar apresenta os seguintes parâmetros para a rede da PF (padrão em cinza) e sua contraparte randomizada mantendo-se o grau médio e a densidade de arestas constantes (padrão em vermelho): diâmetro ($D = 49$ e 15), modularidade ($Q = 0,96$ e $0,52$), fração de controladores ($n_d = 0,21$ e $0,02$), comprimento médio de caminho mais curto ($\lambda = 14,43$ e $6,78$), assortatividade ($r = 0,017$ e $0,001$), coeficiente de agrupamento ($C = 0,391$ e $0,001$). 56

3.15 As figuras mostram as curvas de fragmentação da rede de crimes federais a partir do tamanho relativo da maior componente conectada G em função da fração de nós/arestas removidos q de acordo com os seguintes procedimentos: (esquerda), remoção de vértices por grau (HD - quadrados pretos), por intermediação (HB - triângulos azuis) e por módulos (MBA - círculos vermelhos); (centro), remoção de arestas por intermediação (HB - quadrados pretos), por intermediação adaptativa (HBA - triângulos azuis) e por módulos (MBA - círculos vermelhos); (direita), remoção de vértices por grau adaptativo (HDA - quadrados pretos), por influência coletiva (CI - triângulos verdes para baixo), por intermediação adaptativa (HBA - triângulos azuis) e por módulos (MBA - círculos vermelhos). 57

3.16 Os histogramas mostram a performance dos três melhores ataques sobre a rede de crimes federais. O painel (A) mostra os ataques por vértices MBA (sombreado horizontal vermelho), HB (sombreado azul inclinado) e HD (sombreado dourado), enquanto o painel (B) mostra MBA, HB e HBA (sombreado preto inclinado) para ataque a arestas. 58

3.17 Representação da rede criminal da deep web. Vértices vermelhos e maiores possuem valores mais altos de grau interno. Para facilitar a visualização as arestas foram omitidas e dos vértices mostram-se apenas os 25% mais conectados. 60
3.18 Neste mapa são mostradas características topológicas da rede da deep web (em vermelho) e sua versão aleatorizada (em azul). A figura contém o diâmetro da rede ($D = 46$ e 5), o comprimento médio de caminho mais curto ($\lambda = 2,07$ e $2,52$), densidade da rede ($\delta = 0,0078$ e $0,0078$), eficiência da rede ($E_{ff} = 0,42$ e $0,39$), coeficiente de agrupamento ($C = 0,130$ e $0,016$) e o valor em módulo das correlações de grau ($r_{in-out} = -0,08$ e $-0,02$, $r_{out-in} = -0,15$ e $0,012$, $r_{out-out} = -0,0884$ e $-0,0004$, $r_{in-in} = -0,0212$ e $-0,0007$). O valor máximo para cada quantidade é mostrado na borda da figura e cada raio equivale a um terço desse valor. 62
3.19 Distribuições cumulativas de grau total (a), In (b) e out (c). O gráfico é apresentado em escala log-log. O ajuste para uma lei de potência no caso da distribuição de grau total resulta em um expoente de $\gamma_{all} = 2,10$, para grau-out em $\gamma_{out} = 2.08$, e para grau-in em $\gamma_{in} = 12.74$ com um valor-p do teste de Kolmogorov-Smirnov de $0,14$, $0,16$ e $0,12$ respectivamente. 64
3.20 O histograma mostra a performance $\mathcal{P}$ de cada estratégia de ataque: HDA $(0,1047)$, HDA_{in} $(0,1047)$, HDA_{out} $(0,1047)$, HBA $(0,0011)$, MBA $(0,0411)$, HD $(0,0545)$, HD_{in} $(0,2847)$, HD_{out} $(0,0121)$ e HB $(0,1384)$. 66
3.21 O tamanho relativo da maior componente conectada G (painel à esquerda) da rede da deep web como função da fração de vértices removidos q conforme as principais estratégias HD (linha sólida azul), HD_{in} (linha duplamente tracejada verde), HD_{out} (linha pontilhada magenta), HBA (linha ponto-tracejada vermelha) e HB (linha tracejada longa preta). Os valores em no painel à direita são uma ampliação na região onde ocorreu a ação policial. Este ponto está marcado pelo “x"vermelho e pelo acrônimo “PF"no cruzamento entre as duas linhas pontilhadas. Os policiais obtiveram mandados para 182 alvos, isto é, $1,75\%$ do total de usuários, resultando em um grafo com 1.060 usuários, ou seja, $96,7\%$ da rede original. 66

Lista de Tabelas

3.1 Dados topológicos das redes: tamanho (N), número de arestas (E), grau médio ($\langle k \rangle$), modularidade (Q), tamanho relativo da maior componente conectada (N_{mod}^{max}), fração de arestas ligando comunidades (E_{inter}), ganho global em eficiência do método de ataque por comunidades (η, ver equação (3.2) para definição). Para os quatro parâmetros relacionados à detecção de comunidades mostram-se os valores correspondentes ao caso de maior eficiência entre 10 rodadas de extração para os métodos infomap (I) e *Louvain* (L). Os dados estão representados tanto para ataques a nós quanto para ataques a arestas. 37

3.2 Complexidade temporal dos algoritmos de ataque estudados neste no caso de grafos esparsos: HDA, HBA, MBA e CI. 44

3.3 Dados comparativos (densidade, eficiência, número de vértices e número de arestas) entre a rede de crimes federais e outros sistemas sociais: um subgrafo do Facebook [108, 111], um conjunto de dados criminais da Polícia de Saint Louis nos Estados Unidos da América por volta de 1990 [134], relacionamentos entre usuários do sítio hamsterster.com [135] e a própria rede de crimes federais brasileiros. 53

Lista de Abreviaturas, Siglas e Símbolos

A_{ij}	Elementos da Matriz de Adjacência
C	Coeficiente de Agrupamento
CE	Rede do Proteoma do C *elegans*
CI	Collective Influence Algorithm
CPU	Central Processing Unit
δ	Densidade de Rede
D	Diâmetro de Rede
η	Ganho Global de Eficiência
E	Número de Arestas
ER	Rede de Autoestradas Europeias
FB	Extrato da Rede do Facebook
γ	Ganho de Eficiência
$G+$	Extrato da Rede do Google Plus
HBA	High Betweenness Adaptive Attack
HB	High Betweenness Attack
HDA	High Degree Adaptive Attack
HDA_{in}	High in-Degree Adaptive Attack
HDA_{out}	High out-Degree Adaptive Attack
HD	High Degree Attack

HD_{in}	High in-Degree Attack
HD_{out}	High out-Degree Attack
HP	Rede do Proteoma do H *pylori*
k_i	Grau do vértice i
λ	Comprimento Médio das Geodésicas
LFR	Redes de Lancichinetti-Fortunato-Radicchi
MBA	Module-Based Attack
N	Número de Vértices
N_d	Número de Pontes entre Comunidades
$OrCrim$	Organização Criminosa
OF	Rede de Voos Internacionais
$\mathcal{P}$	Performance de Ataques a Redes
PF	Polícia Federal
PG	Rede de Energia Elétrica dos EUA
$p_k, p(k)$	Distribuição de Grau
Q	Modularidade
R	Robustez de Rede
r	Assortatividade
SBM	Stochastic Block Model
TW	Extrato da Rede do Twitter
UA	Rede de Aeroportos dos EUA
YP	Rede do Proteoma do Levedo

Sumário

1 Introdução **19**

2 Teoria de redes complexas **23**

2.1 Redes modulares 26

2.1.1 O algoritmo *Louvain* 28

2.2 Fragilidades estruturais em redes complexas 30

2.3 Teoria de controle 31

3 Resultados **34**

3.1 Ataques dirigidos por comunidades 34

3.2 *Performance* de métodos de ataque 42

3.3 Redes criminais 50

3.3.1 Crimes federais 51

3.3.2 Darknet 59

4 Conclusão **68**

4.1 Trabalhos futuros 72

Referências Bibliográficas **74**

Capítulo 1

Introdução

Em um mundo cada vez mais conectado e multifacetado, o diálogo entre a física e outras disciplinas encontra-se em importante ascendência. No entanto, há obstáculos a essa interação, que incluem a falta de registros precisos e confiáveis, e a escassez de ferramentas capazes de realizar mineração de dados– tópicos que despertam a atenção da comunidade científica. Todavia, devido à rapidez com que essas ferramentas são desenvolvidas, avança-se em um terreno antes dominado por incertezas e por abordagens pouco científicas. Em breve, a natureza e a evolução de sistemas até então inexplorados pela física tradicional poderão ser testadas nos mínimos detalhes. Esses avanços possibilitam ir além de generalizações, permitindo aos pesquisadores a formulação de perguntas mais profundas sobre a maneira como a natureza em seus mais diversos aspectos se comporta, e como uma certa descoberta influencia trabalhos subsequentes dentro e fora da física. Nesse sentido, as explorações científicas mais frutíferas geralmente provêm da fertilização cruzada entre campos anteriormente desconectados. Esta interação interdisciplinar pode ajudar a identificar áreas de pesquisa ainda pouco exploradas, abrindo espaço para inovações com profundos impactos a longo prazo.

Atualmente, um dos campos de pesquisa mais interdisciplinar é o que cuida da codificação das interações entre os componentes de redes complexas— a chamada ciência de redes [1]. De fato, cercam a todos: dos bilhões de pessoas se relacionando em sociedade, passando pelo número incontável de roteadores se comunicando ao formarem a internet, até as interações metabólicas entre os neurônios nos cérebros de cada ser humano [2, 3]. Esses coletivos extremamente complicados são alguns exemplos de sistemas complexos, isto é, estruturas cujos comportamentos superam às somas das descrições de cada componente individual. Uma das mais relevantes descobertas em sistemas complexos é a de que as redes de distintos campos da natureza, da sociedade e da tecnologia são governadas pelos mesmos princípios e, portanto, as arquiteturas desses sistemas são muito similares [4, 5]. Dessa maneira, podemos utilizar um mesmo conjunto de ferramentas analíticas, computacionais e matemáticas para tratar de sistemas totalmente diferentes uns dos outros [6].

Estruturas em rede possuem diversas vantagens e outras tantas desvantagens: ao mesmo tempo em que a conectividade agiliza o transporte de informações, dados e riquezas, a falha de um elemento pode causar um efeito em cascata que desmantela o sistema como um todo [7, 8]. Apesar de parecerem

aleatórias e imprevisíveis, essas falhas são quantificáveis e seguem leis bem definidas, o que as tornam passíveis de intervenção do ponto de vista matemático [9–11]. Assim, descobrir quais são os fatores que aumentam a robustez de redes complexas e quais as tornam mais vulneráveis é de suma importância prática. O que leva naturalmente à procura pelo conjunto mínimo de elementos da rede que, se removidos, causariam uma demolição completa do sistema em pequenos pedaços desconectados. Uma aplicação direta desse conceito é identificar quais indivíduos são mais importantes em uma classe muito particular de redes sociais: aquelas voltadas para a prática de crimes. Muitas agências de aplicação da Lei vêm utilizando ferramentas de redes para enfrentar grupos criminosos no mundo inteiro. Apesar deste trabalho ser em grande parte sigiloso, vários casos bem documentado são públicos, a exemplo da captura do ditador Saddam Hussein na Operação Red Dawn do Exército Norte Americano [12], da identificação dos responsáveis pelos bombardeios das linhas de trem de Madrid em 2004 e da criação do conceito de guerra centrada em redes pela Academia Militar de West Point nos Estados Unidos da América [13].

O termo “crime"não possui uma definição universalmente aceita. Contudo, a noção de que algumas condutas (homicídio e roubo por exemplo) devem ser proibidas parece existir mundialmente, apesar da descrição precisa do que vem a ser uma ofensa criminal depender da definição legal própria de cada pais[1]— em alguns países que adotam um sistema jurídico de *commom law* não há tal codificação expressa [14]. Verdadeiramente, o crime é um fenômeno social (antes de ser jurídico) que depende de como as pessoas e a sociedade concebem o delito baseado em normas societais [15]. Todavia, independente da definição legal, social ou consuetudinária do termo, alguns aspectos abstratos como a clandestinidade e a comunicação entre coautores estão sempre presentes. Assim, grupos de criminosos precisam enfrentar um constante conflito entre segurança e eficiência de comunicação, e isso afeta diretamente a arquitetura de suas estruturas de rede [16].

De fato, o crime está em toda a parte, mas se trata de um fenômeno longe de ser distribuído uniformemente no espaço ou no tempo [17–19]. Isto é evidenciado pelo surgimento de gangues e grupos organizados. Essa formação de padrões abre naturalmente as portas para uma análise matemática e quantitativa. Do ponto de vista das ciências naturais, o crime deve então ser encarado como um fenômeno complexo, em que respostas não-lineares, comportamentos coletivos, estruturas de rede e auto-organização dão origem a comportamentos inesperados que são difíceis de prever e controlar [20]. Talvez por isso que grande parte das ações policiais são ora aleatórias, conforme *notitiae criminis* alcançam os setores competentes; ora focadas em líderes ou cabeças de grandes organizações em estratégias que se assemelham vagamente a ataques por conectividade, quando resultado de levantamentos prévios de inteligência. Entretanto, recentemente foi mostrado que a estrutura social de atividades criminais é de fato muito resiliente a essas ações tradicionais [21–23]. Há, portanto, uma falta de direcionamento estratégico na luta contra a criminalidade e, muitas vezes, os órgãos policiais agem de maneira reativa, sempre um passo atrás das empreitadas criminosas [24–27].

[1]No Brasil, a doutrina admite como crime todo o fato típico, ilícito e culpável e as definições legais podem ser encontradas no Decreto-Lei 2.848/40 (Código Penal), na Lei 13.260/16 (Terrorismo), na Lei 9.613/98 (Lavagem de Dinheiro e Capitais), nas Leis 12.850/13 e 12.694/12 (Orgnizações Criminosas), na Lei 11.343/06 (Tráfico de Drogas), na Lei 8.069/90 (Estatuto da Criança e do Adolescente) e outras tantas.

Daí, a necessidade de se abordar a aplicação da lei num quadro pró-ativo e o crime como um sistema complexo coletivo.

A literatura sociológica suporta a adoção de métodos de teoria de redes para se estudar organizações criminosas [28–31]. Por exemplo, o modelo de facilitação social [32, 33] informa que a tendência de um indivíduo a cometer delitos é fortemente elevada por sua participação em gangues, organizações ou outras estruturas coletivas. Nesta mesma esteira, os experimentos de Asch mostraram [34] que as pessoas tendem a se conformar publicamente (influência normativa) ou endossar respostas coletivas (influência informacional) mesmo quando evidentemente erradas. Alguns anos mais tarde, os conhecidos experimentos de Milgram [35] sobre obediência a autoridade revelaram que sob certas circunstâncias pessoas comuns podem inflingir dor e sofrimento a outros, mesmo sendo moralmente contrárias. Em suma, esses estudos comportamentais mostram que quando uma pessoa faz parte de uma rede social, parte de sua individualidade é perdida e o grupo começa a se comportar como um todo. Por conseguinte, abalar a estrutura de rede de uma organização criminal deve, em tese, diminuir as taxas de crime já que os processos de facilitação social, conformidade e autoridade seriam ou barrados ou no mínimo atenuados. Todavia, muitas das contribuições sociológicas são baseadas em aspectos qualitativos e não na importância quantitativa de cada indivíduo em manter o sistema funcionando como um todo [36, 37]. Além disso, alguns pesquisadores já ilustraram recentemente os benefícios de se aplicar métodos de física estatística ao estudo da estrutura e fragilidade do fenômeno criminal [38]. Além do mais, ao abordar o crime do ponto de vista macroscópico, os relacionamentos se tornam cada vez mais importantes e grandes estruturas emergem resultando na existência de redes de crime organizado. Nesse tipo de sistema, o fluxo de dados é prioridade e, como mostram estudos anteriores [13], o sigilo das informações é conseguido por meio da compartimentação, isto é as informações importantes são isoladas em diferentes compartimentos ou células da organização. Essa estratégia evita que a rede fique exposta e é encontrada de fato em diversos grupos criminosos reais [39]. Esses estudos mostram que, a fim de se adaptar aos ataques pelas agências de segurança, as redes criminosas tendem a se organizar de forma não-redundante no que tange ao fluxo interno de informações, protegendo assim os seus membros e a sua estrutura funcional [40]. Como apresentado a seguir, este fenômeno de compartimentação corresponde matematicamente ao conceito de modularidade ou comunidades topológicas.

Apesar dessas considerações, o estudo de redes criminais ainda apresenta um poderoso inconveniente: essas estruturas são em grande parte desconhecidas devido à dificuldade de acesso e coleta de dados confiáveis. Contudo, recentemente a Polícia Federal brasileira obteve dados de inteligência sobre algumas redes de crime organizado, introduzindo informações restritas que podem lançar luz sobre a natureza desses sistemas. Tais registros foram disponibilizados exclusivamente para o presente estudo de maneira inédita[2]. Assim, será possível responder perguntas como: será que redes criminais compartilham das mesmas propriedades topológicas que outros tipos de redes reais? Como esses sistemas alcançam eficiência apesar de sofrerem vigilância e ameaça constantes das autoridades?

[2]Como apresentado nos capítulos apropriados, todos os dados foram criptografados pelo setor competente da Polícia Federal a fim de se resguardar o sigilo das informações e dos indivíduos.

Como essas redes são robustas frente a ataques?

Em consequência, esta tese pode ser separada em dois grandes tópicos. No primeiro estuda-se o problema do conjunto mínimo de vértices que devem ser retirados de uma rede em abstrato para que ela se despedace e deixe de funcionar como um todo. Propõe-se então um método de ataque a redes que tem inspiração na estrutura modular do crime organizado e que apresenta enorme vantagem em relação aos métodos tradicionais de fragmentação. No segundo momento, estuda-se duas redes criminais empíricas resultantes da cooperação com a Polícia Federal. Descrevemos suas características topológicas mais marcantes, em especial suas fragilidades, propondo estratégias eficientes de enfrentamento ao crime organizado do ponto de vista da ciências de redes.

Por termo, com esse trabalho espera-se, de alguma maneira, sensibilizar os órgãos componentes do sistema de justiça e segurança criminal para a importância de se aplicar conceitos e ferramentas de teoria de redes e sistemas complexos a operações de inteligência e de aplicação da Lei. Enfim, que os resultados a seguir auxiliem no planejamento de estratégias mais eficientes de combate ao crime, contribuindo para a diminuição dos índices de criminalidade no Brasil e no mundo.

Este trabalho é estruturado da seguinte forma. O próximo capítulo é iniciado apresentando-se um breve apanhado de livro-texto de conceitos já bem estabelecidos na literatura: estrutura topológica de redes complexas, modularidade e fragilidade de redes. Após, discute-se o método de ataque modular a redes complexas desenvolvido durante esta tese com o respectivo estudo sobre o desempenho do algoritmo. No capítulo seguinte, exibem-se as redes reais obtidas em cooperação com a Polícia Federal e exploram-se suas naturezas e fragilidades. Encerra-se o feito com conclusões e discussões sobre os principais pontos desenvolvidos, apontando possibilidades de trabalhos futuros.

Capítulo 2

Teoria de redes complexas

Neste capítulo revisam-se conceitos básicos sobre teoria de grafos e redes como usualmente tratado nos principais livros-texto da área [1, 41–43] Um grafo é simplesmente um par ordenado $G = (N, E)$ que consiste de um conjunto de N de vértices, nodos, nós ou pontos e um conjunto E de arestas, arcos, linhas, *links* ou bordas. Uma aresta é sempre associada a dois vértices de forma não ordenada quando se tem um grafo não dirigido ou ordenada quando se tratam de grafos dirigidos. Quando estes entes matemáticos representam sistemas reais tem-se o que comumente se chama de redes. Contudo, para efeitos práticos os dois termos são utilizados indistintamente.

Uma das principais medidas de rede é o grau ou conectividade de um vértice. Trata-se de quantidade de arestas que determinado nó i tem conectada a ele, k_i. Com efeito, a distribuição de grau p_k indica a probabilidade de que um vértice escolhido ao acaso tenha conectividade k. Outras quantidades estatísticas são muito úteis para a caracterização de um grafo, como o grau médio $\langle k \rangle = \sum_k kp_k$ e o enésimo momento generalizado $\langle k^n \rangle = \sum_k k^n p_k$. Pode-se ainda classificar o quanto os vizinhos de um determinado nó estão ligados entre si, o que é expresso pelo coeficiente de agrupamento $C_i = 2L_i/k_i(k_i - 1)$, onde L_i representa o número de arestas entre os k_i vizinhos do vértice i. Como resultado, tem-se um valor global para toda a rede de $C = \frac{1}{N}\sum_i C_i$ [44, 45].

Uma vez conhecidas as propriedades dos vértices de um grafo, é necessário também caracterizar as arestas do sistema para se ter uma compreensão total das propriedades da rede. Para tanto, costuma-se condensar as propriedades do grafo na chamada matriz de adjacência **A** [46] que é uma matriz quadrada $|N| \times |N|$ tal que:

$$A_{ij} = \begin{cases} 1, & \text{se há uma aresta entre } i \text{ e } j. \\ 0, & \text{se os vértices } i \text{ e } j \text{ não estão conectados.} \end{cases} \tag{2.1}$$

A partir desta definição diversas propriedades estatísticas do grafo podem ser obtidas como por exemplo $k_i = \sum_j A_{ij}$.

Pode-se definir ainda o comprimento de um caminho entre dois vértices da rede como a quantidade de arestas que compõem o trajeto, e o menor caminho λ como aquele que consiste do menor número de bordas [46]. Daí vem a definição de diâmetro da rede como o maior menor caminho possível [46].

Outro ponto importante é a correlação entre propriedades similares da rede. Nesse contexto, assortatividade é a tendência que os vértices de uma dada rede têm de se ligar a outros similares [47]. A fim de aproximar os comportamentos observados em redes reais, geralmente essa similaridade é atribuída à correlação da conectividade dos vértices. Por exemplo, em redes sociais, as pessoas tendem a se conectar com outras cujo grau de conexão é parecido– comportamento conhecido como assortativo. Por outro lado, redes biológicas e tecnológicas geralmente apresentam comportamento desassortativo, conforme vértices com alta conectividade costumam se conectar a outros que possuem poucas conexões. O coeficiente de assortatividade nada mais é que o coeficiente de correlação de Pearson ($-1 < r < 1$) entre os graus de pares de vértices conectados [48]. Valores positivos indicam uma correlação entre nós com valores de conectividade similares, enquanto que coeficientes negativos indicam relações fortes entre vértices com graus distintos. Assim, quando $r > 0$ a rede é dita assortativa, quando $r < 0$ diz-se da rede desassortativa, quando $r = 0$ a rede é não-assortativa. Portanto, r é formalmente descrito por:

$$r^{(\alpha-\beta)} = \frac{\frac{1}{E}\sum_e \left(k_e^{(\alpha)} - \overline{k^{(\alpha)}}\right)\left(j_e^{(\beta)} - \overline{j^{(\beta)}}\right)}{G^{(\alpha)}G^{(\beta)}}, \tag{2.2}$$

onde a soma é sobre todas as arestas E, $\alpha, \beta \in \{in, out\}$ é o tipo de grau, k^α é o grau do nó-fonte (in), j^β é o grau do nó-alvo (out), e $\overline{j}_\alpha = \frac{1}{E}\sum_e j_e^\alpha$ é o grau médio dos vértices no início de cada aresta, $G_\alpha^2 = \frac{1}{E}\sum_e \left(k_e^{(\alpha)} - \overline{k^{(\alpha)}}\right)^2$ é a variância com $\overline{k^{(\beta)}}$ e $G^{(\beta)}$ são definidos similarmente.

Tendo em vista a caracterização da rede, uma questão que nasce naturalmente é a classificação dos vértices conforme sua importância no sistema. A relevância prática é imediata e vai desde a identificação das pessoas mais relevantes em uma rede social, passando pelos roteadores-chave da internet até os vetores propagadores de epidemias e doenças [6]. Busca-se, pois, uma função real dos vértices de um grafo que forneça um ranking dos nós mais importantes. A essa função dá-se o nome de centralidade [43]. Contudo, essa indexação dos vértices segundo a centralidade depende da definição precisa do termo. Assim, a primeira e mais usada centralidade é a de grau, que nada mais é que a lista de conectividade de cada vértice da rede. Veremos a seguir duas definições que são muito utilizadas para a caracterização de redes reais e serão utilizadas neste trabalho, lembrando que a lista não é exaustiva.

- **Proximidade**: a centralidade de proximidade de um nó consiste no tamanho médio dos menores caminhos entre o vértice e todos os outros da rede [49]. Assim, um nó é central quando ele está de certa maneira "perto" dos outros, ou seja, a centralidade de proximidade de um vértice v é dada por:

  $$C_C(v) = \frac{1}{\sum_i \lambda_{iv}}, \tag{2.3}$$

  onde λ_{iv} é a distância entre os dois vértices i e v.

- **Intermediação**: a chamada centralidade de intermediação de um vértice consiste na fração de quantos menores caminhos de toda a rede que passam por aquele determinado nó [50]. Conceitualmente, a ideia nasceu para quantificar o controle de um ser humano sobre a comunicação

entre outras pessoas numa rede social. Assim, formalmente a centralidade de intermediação de um nó v é obtida pela seguinte equação:

$$C_B(v) = \sum_{i \neq j \neq v} \frac{\lambda(v)_{ij}}{\lambda_{ij}}, \tag{2.4}$$

onde $\lambda(v)_{ij}$ é a quantidade de caminhos entre i e j que passam por v e λ_{ij} é o total de caminhos entre os mesmos vértices.

Quanto à arquitetura das suas distribuições de grau, ou seja quanto à sua topologia, as redes podem ser divididas entre aleatórias e invariante em escala. Redes aleatórias (também chamadas de Ërdos-Rényi— ER [51]) apresentam uma distribuição de grau do tipo Poisson enquanto que em redes invariantes em escala a distribuição segue uma lei de potência (também identificada pelo termo em inglês Scale-Free— SF [52]):

$$p_k \propto \begin{cases} e^{-\langle k \rangle}, & \text{ER.} \\ k^{-\gamma}, & \text{SF.} \end{cases} \tag{2.5}$$

Grafos aleatórios apresentam diversas propriedades que servem de referência, uma das principais delas é o fato de que se os vértices forem removidos um a um aleatoriamente, após uma fração fixa de remoções a rede deixa de ser conexa e perde o que se chama de componente gigante (um subgrafo da rede principal que estatisticamente concentra a maioria dos nós). De acordo com o chamado critério de Molloy-Reed [53], uma rede deixa de apresentar uma componente gigante quando

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2, \tag{2.6}$$

ou seja, para que uma componente gigante exista, cada vértice deve apresentar em média pelo menos duas arestas. Isso implica[1] que quando redes aleatórias são sujeitas a falhas randômicas, elas transicionam de um regime conectado para um regime fragmentado após a eliminação de uma fração $f_c = 1 - \frac{1}{\langle k \rangle}$ de vértices [1]. Por outro lado, redes invariantes em escala apresentam tal comportamento quando [1]:

$$f_c = \begin{cases} 1 - \frac{1}{\frac{\gamma-2}{3-\gamma} k_{min}^{\gamma-2} k_{max}^{3-\gamma} - 1}, & 2 < \gamma < 3, \\ 1 - \frac{1}{\frac{\gamma-2}{\gamma-3} k_{min} - 1}, & \gamma > 3. \end{cases} \tag{2.7}$$

De fato, redes invariantes em escala apresentam poucos vértices muito conectados (*hubs*) e diversos nós com menor conectividade. Assim, remoções aleatórias dificilmente vão atingir esses *hubs* e a rede consegue suportar um nível muito grande de falhas, o que não acontece com sistemas ER, que se quebram facilmente após a falha de poucos nós. Por outro lado, quando as remoções são realizadas a partir dos vértices mais conectados de maneira iterativa redes ER se apresentam robustas já que têm uma distribuição homogênea de grau, enquanto que redes SF perdem sua componente gigante quando uma fração f_c é removida conforme a seguinte equação transcendental:

$$f_c^{\frac{2-\gamma}{1-\gamma}} = 2 + \frac{2-\gamma}{3-\gamma} k_{min} (f_c^{\frac{3-\gamma}{1-\gamma}} - 1). \tag{2.8}$$

[1] Aqui não são levadas em conta correções por tamanho finito.

Destarte, quando γ é suficientemente grande a rede se comporta como um grafo ER, e seu comportamento é similar àquele de falhas aleatórias [1]. Contudo, para outros valores de γ, o limiar crítico diminui e a rede se torna frágil aos ataques dirigidos por grau [1].

Pode-se também medir o quão eficientemente informação pode trocada numa estrutura de rede. Assim, pode-se definir o valor médio da eficiência de rede relacionando-se os menores caminhos existentes no sistema [54]:

$$E_{ff} = \frac{2}{N(N-1)} \sum \frac{1}{\lambda_{ij}} \tag{2.9}$$

Nesse sentido, também costuma-se definir a densidade de arestas de uma rede como sendo a fração de conexões reais da rede em função da quantidade de conexões possíveis para o mesmo número de nós:

$$\delta = \frac{2E}{N(N-1)} \tag{2.10}$$

Como veremos adiante no texto, esses dois pontos estão relacionados com o balanço entre “brilho"da rede, ou seja, o quão perceptível é sua estrutura, e a eficiente comunicação entre elementos quaisquer do sistema [55].

Um último comportamento que vale a pena ser ressaltado é o chamado fenômeno de pequeno mundo [45]. Basicamente, esta característica típica de redes reais assevera que a distância entre dois nós escolhidos ao acaso é pequena no sentido que o menor caminho médio da rede (também se pode utilizar o diâmetro da rede) depende logaritmicamente do tamanho do sistema. Em outras palavras,

$$\langle \lambda \rangle = \log N \tag{2.11}$$

De fato, redes de pequenos mundos reais apresentam pequenos valores de menores caminhos ao mesmo tempo em que o coeficiente de agrupamento (C) é mais significativo que o de simples redes aleatórias. Com efeito, geralmente se usam essas duas medidas (λ,C) para se caracterizar uma rede como de pequeno mundo ao se comparar esses valores com o da rede original randomizada.

2.1 Redes modulares

Redes reais tendem a se agrupar em aglomerados remotamente conectados uns aos outros [57]. A estes grupos concentrados de vértices costuma-se dar o nome de comunidades topológicas ou módulos. Exemplos práticos de estruturas comunitárias são encontrados em diversos sistemas reais: empregados de uma empresa tendem a interagir mais com seus colegas de trabalho do que com funcionários de outras companhias; patologias geralmente estão relacionadas a vizinhanças (ou comunidades) bem definidas das redes celulares; máfias de crime organizado também possuem a tendência a serem bastante conectados para maximizar a eficiência das comunicações ao mesmo tempo em que têm poucas conexões com outros grupos criminosos a fim de se manterem protegidos da ação policial [1,55].

Nesse sentido, a identificação de comunidades é um processo de partição da rede original em subgrafos em número e tamanhos não pré-definidos. A quantidade de maneiras distintas de se separar uma determinada rede em comunidades cresce mais que exponencialmente com o tamanho do sistema,

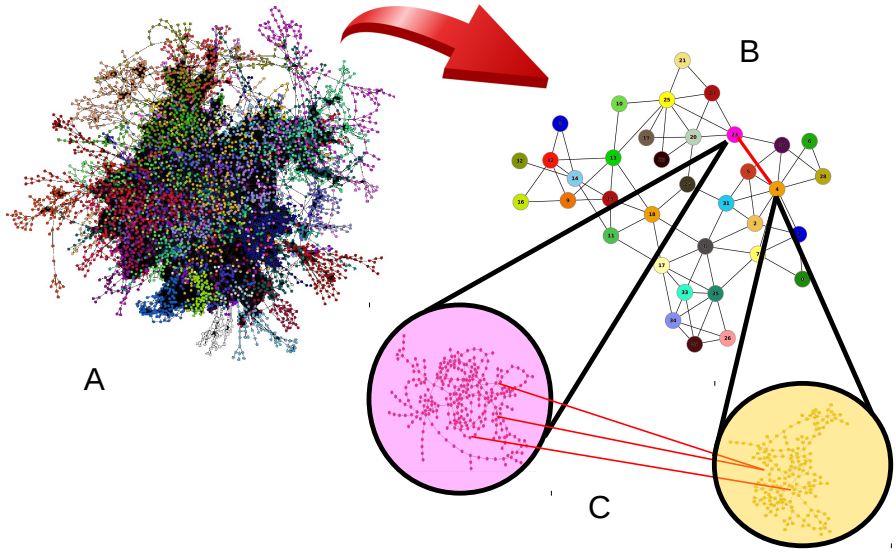

Figura 2.1: Representação da rede de distribuição de energia elétrica do oeste dos Estados Unidos (A), uma possível representação modular (B), e a estrutura interna dos nós e arestas dentro de duas comunidades selecionadas, além dos nós conectando dois módulos (C) [56].

o que torna impossível se inspecionar todas as partições de sistemas reais até se encontrar a configuração ideal que reflita uma propriedade intrínseca da rede. São necessários, então, algoritmos que possam identificar comunidades sem que se vasculhe todas as opções de partições possíveis. Para responder a esta demanda, diversos algoritmos de deteção de comunidades foram desenvolvidos ao longo dos anos. Ora, em uma rede aleatória, a distribuição de grau é uniforme, de modo que tal classe de sistemas não devem apresentar estruturas modulares resultantes de flutuações locais na densidade de conexões. Assim, pode-se definir uma quantidade para medir a qualidade de uma partição pela medida sistemática de desvios da configuração modular de redes aleatórias. Nesse sentido, a modularidade de uma partição de uma rede não ponderada pode ser definida como a densidade de links no interior das comunidades se comparada com a densidade de arestas entre comunidades da seguinte maneira [57, 58]:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{2.12}$$

onde A_{ij} é a matriz de adjacência (tomando os valores 1 quando há uma conexão entre os nós i e j, 0 caso contrário), k_i é o grau de conexão do nó i e c_i representa a comunidade a qual esse vértice pertence. A função δ é tal que $\delta(u, v)$ é 1 se $u = v$ e 0 caso contrário com m sendo o número total de arestas. Dessa maneira Q é um valor escalar entre -1 e 1 que mede a o grau de modularidade da rede. Com efeito, podem-se definir 4 regimes modulares [1]:

- **Partição ótima**: modularidades em torno de $Q = 0,41$ indicam duas comunidades distintas, valores acima disto indicam que a rede possui estrutura modular bem definida.

- **Partição subótima**: uma partição com modularidade próxima a $Q = 0,22$ não identifica corretamente comunidades na rede já que apresenta um valor equivalente ao de sistemas randômicos.

- **Comunidade única**: se todos os vértices pertencerem à mesma comunidade, a subtração na definição da modularidade resulta em $Q = 0$.

- **Modularidade negativa**: à medida que cada nó é designado a uma única comunidade a modularidade vai se ficando negativa e aumentando em módulo.

Assim, como se verá a seguir, para uma determinada rede a estrutura de comunidade com maior modularidade corresponderá à estrutura de partição ótima.

Nesse sentido, a remoção de algumas poucas pontes em redes altamente modulares deve possibilitar a separação de grandes grupos de nós densamente conectados, levando a uma rápida fragmentação de redes complexas como será abordado a seguir[2]. Por exemplo, na figura 2.1 é representada uma possível estrutura de comunidades para a rede elétrica do oeste dos Estados Unidos da América, ilustrando as frágeis conexões entre os aglomerados que são densamente conectados internamente. A figura 2.1A utiliza vértices para representar um gerador, um transformador ou uma subestação, enquanto que as arestas representam uma linha de fornecimento de energia. Cores distintas são usadas para se identificar os módulos nos quais a rede foi particionada. Na figura 2.1B cada comunidade é representada por um nó colorido e as arestas são mostradas sempre que há uma conexão entre os nós de módulos distintos (sem levar em conta a quantidade). Já a figura 2.1C mostra a conexão detalhada entre duas comunidades escolhidas aleatoriamente exibindo as ligações entre elas.

2.1.1 O algoritmo *Louvain*

Muitos algoritmos de identificação de comunidades se baseiam na otimização de alguma função-qualidade, em especial uma classe muito utilizada para grandes redes é a que busca maximizar a modularidade Q. Como já demonstrado em estudos anteriores, o chamado método *Louvain* desenvolvido por Blondel *et al* [59] é, geralmente, o que apresenta menor complexidade computacional e precisão, além de ter melhor escalabilidade, o que o torna propício para uso em grandes redes reais. O método consiste em dois passos interativos como segue (ver figura 2.2):

1. Começa-se no regime de modularidade negativa em que a cada vértice é atribuída uma comunidade distinta. Então, para cada nó i avalia-se o ganho na modularidade ao se colocá-lo na comunidade de um de seus vizinhos j. Move-se o nó i para a comunidade que gera o maior

[2] O termo rápido aqui se refere a uma resposta íngreme da rede à remoção de nós, isto é, quando uma pequena fração de nós removidos resulta na desconexão de uma grande parcela da rede.

ganho positivo na modularidade. O processo é repetido até não haver mais ganho e se alcançar um máximo local de modularidade.

2. A seguir, agrupam-se todos os vértices na mesma comunidade e constrói-se uma nova rede em que os nós são as comunidades da fase anterior. Qualquer conexão entre vértices da mesma comunidade agora são tratados como *loops* e arestas de vértices da mesma comunidade ligando-se a um nós de outra comunidade são representadas por uma aresta pesada entre as duas comunidades. À medida que esta etapa acaba, é possível reaplicar a primeira fase do algoritmo à essa nova rede e iterar o processo.

3. As duas etapas são iteradas até não haver mais ganhos e um máximo de modularidade é alcançado.

Como será abordado e justificado em maior profundidade no capítulo que trata de ataques modulares, este será o método de deteção de comunidades utilizado ao longo de todo o trabalho.

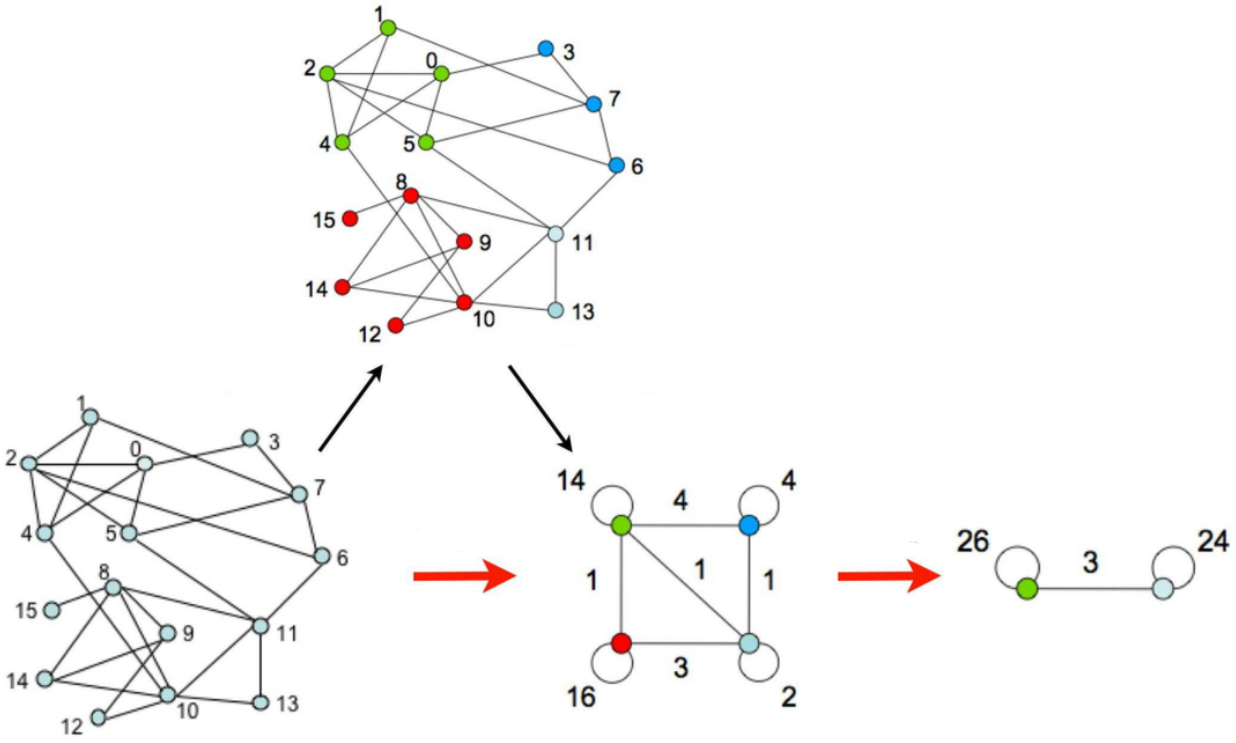

Figura 2.2: Visualização dos passos do algoritmo *Louvain*. Cada etapa consiste num momento em que a modularidade é otimizada ao se permitir apenas mudanças locais nas comunidades. Na segunda etapa as comunidades extraídas são agregadas para se construir uma nova rede de comunidades. As etapas são repetidas iterativamente até não haver mais ganho na modularidade. A figura foi adaptada do artigo original de Blondel *et al* [59].

2.2 Fragilidades estruturais em redes complexas

Geralmente, a fragmentação de uma rede pode ser representada pela remoção de nós ou arestas. Deletar nós apresenta uma vantagem sobre apagar apenas arestas já que a remoção de um vértice sempre resulta na remoção de todas as arestas ligadas a este nó. Contudo, dependendo do sistema real utilizado, um ataque por nós ou arestas pode não fazer sentido. Por exemplo, no caso de uma rede de rodovias, pode-se vislumbrar o bloqueio do tráfego entre duas cidades, contudo remover um nó significaria apagar uma cidade inteira do mapa! Por outro lado, em sistemas biológicos a remoção de nós faz todo sentido já que metabólitos individualmente são suscetíveis a ataques e falhas.

Vulnerabilidades estruturais de sistemas reais é um dos tópicos que mais atrai a atenção da comunidade científica [43, 60] tanto do ponto vista do ataque (quando se está interessado em desativar ou fragmentar uma rede com o menor esforço possível) [61], quanto do ponto de vista da segurança (quando se quer criar redes mais seguras ou se deseja defendê-las de ataques dirigidos ou maliciosos) [62]. Por exemplo, a propagação de uma epidemia [63], a operatividade do crime organizado e de células terroristas [55, 64], ou a segurança de uma rede elétrica [65, 66], são alguns exemplos de redes nas quais se está interessado ora em estratégias eficientes de fragmentação [67] ora na adoção de ações defensivas para prevenir o desmantelamento do sistema [68].

Portanto, a procura pelo conjunto mínimo de nós ou bordas estruturais cuja remoção deixaria uma determinada rede fragmentada é um dos mais importantes tópicos em ciência de redes. Embora muitos avanços tenham aparecido recentemente, este problema ainda está aberto ao debate. Um maneira pouco prática de se conseguir uma lista de alvos de uma rede com N vértices a serem removidos seria por força bruta: tentar todas as listas possíveis até encontrar aquela que reduz a rede até o tamanho desejado com um número mínimo de remoções. Todavia, isso é impraticável já que significa checar $N!$ listas possíveis, o que é computacionalmente impossível para qualquer rede com $N \geqslant 12$. Em contrapartida, a maneira mais simples, porém pouco eficiente, é pela remoção aleatória de nós. Dessa forma, a atomização do grafo acaba sendo um processo muito lento nesses casos. Uma maneira mais eficiente e factível de se degradar redes é pela remoção de nós conforme a sua importância estrutural no funcionamento da rede. Nesse sentido, ataques tradicionais costumam focar no ordenamento de vértices segundo algum índice de centralidade –que tem uma performance muito melhor que ataques randômicos [8, 69–73].

Redes podem ser estruturalmente danificadas tanto pela remoção aleatória de elementos (falhas) quanto pela exclusão dirigida de alvos específicos (ataque malicioso) [9, 74–76]. Os ataques dirigidos visam a interromper o sistema pela remoção de uma pequena fração de nós ou arestas. Tradicionalmente, esses ataques se concentram na classificação de vértices de acordo com sua importância na arquitetura da rede, isto é, de acordo com algum índice de centralidade. Basicamente, existem duas abordagens para a fragmentação: ataques não-adaptativos (ou simultâneos) e adaptativos (ou sequenciais). Na primeira abordagem, a lista de nós atacados é produzida apenas uma vez, antes do início do procedimento de remoção [8]. Na segunda, a lista de alvos é atualizada após cada supressão pelo recálculo da centralidade utilizada para classificar os vértices ou arestas [10, 77]. Consequentemente, os ataques

adaptativos (quando o nó com maior centralidade é iterativamente removido) demandam muito mais tempo de processamento, mas por outro lado o método geralmente produz mais dano por remoção se comparado com a abordagem não-adaptativa. A razão é mais ou menos óbvia: se a lista de ataque for medida apenas uma vez, o método não consegue levar em conta as mudanças na ordem de centralidade devido às remoções anteriores. Assim, a versão adaptativa de um procedimento é no mínimo tão boa quanto a abordagem não-adaptativa, no entanto é geralmente melhor.

Muitos índices de centralidade foram propostos recentemente para se medir a importância estrutural de nós e arestas em sistemas reais [8,69]. No entanto, as centralidades de intermediação, de proximidade e de grau são as mais utilizadas para se construir ataques dirigidos, sejam eles adaptativos ou não [8]. No entanto, Morone e Makse [78] mapearam recentemente o problema de encontrar o conjunto mínimo de vértices que se retirados despedaçariam redes complexas em um problema de minimização de energia de um sistema de muitos corpos, resultando na chamada centralidade de Influência Coletiva (CI). Mais tarde, Morone *et al* [79] desenvolveram ainda mais o método, desta vez com menor complexidade computacional. Além disso, Braunstein *et al* [80] também estudaram o problema do conjunto mínimo para o desmantelamento de redes propondo um ataque que consiste em um algoritmo de três etapas min-sum. Ambos grupos de autores demonstraram que seus métodos estão próximos do conjunto ideal de fragmentação, o que se mostrou ser um problema não-local.

No entanto, redes reais tendem a se organizar em estruturas comunitárias— aglomerados densamente conectados internamente, mas escassamente ligados entre si [81]. Além do mais, os nodos que unem essas comunidades são ainda mais cruciais em manter redes reais coesas do que vértices altamente conectados, impedindo que elas desmoronem [82, 83]. Nesse sentido, Faqeeh *et al* enfatizaram o papel proeminente das pontes entre comunidades na robustez de rede complexas e sua importância em quantificar surtos epidêmicos [84]. Ainda mais recentemente, Shekhtman *et al* resumiram os avanços sobre robustez de redes de redes [85].

2.3 Teoria de controle

Teoria de controle é uma disciplina bem desenvolvida em diversos ramos da engenharia como circuitos eletrônicos, processos de manufatura, comunicação, navegação, robótica entre outros [86, 87]. A noção intuitiva é a de que controle é a habilidade de levar o comportamento de um sistema dinâmico até um estado desejado qualquer por meio de uma escolha adequada de estímulos externos, como fazer um míssil acertar seu alvo na velocidade e momento certo. Nesse contexto [88–90], um sistema dinâmico é dito controlável se pela escolha correta de *inputs* é possível levá-lo desde um estado inicial arbitrário até um estado final qualquer (*output*) durante um intervalo finito de tempo. Estritamente, segundo a teoria clássica de controle o sistema

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}, \tag{2.13}$$

em que o vetor $\mathbf{x} = (x_1, ..., x_N)^T$ representa o estado dos vértices, $A \in \mathbb{R}^{N\times N}$ é a matriz de adjacência que representa o acoplamento do sistema, já que a_{ij} denota o peso de casa aresta entre os nós j e i, $\mathbf{u}$ é o vetor de controladores m $\mathbf{u} = (u_1, u_2, ..., u_m)^T$ e B é a matriz $N \times m$ de controle, é controlável se

podemos alcançar um estado final $\mathbf{x}$ do sistema pela escolha adequada da matriz B que age sobre os controladores.

Todavia, diversos sistemas cuja arquitetura se apresenta em forma de redes são enormes em tamanho e complexidade. Então não é possível se controlar cada um dos elementos da rede para se ter total controle do sistema. Nesses casos, é mais adequado procurar por um subconjunto de vértices-alvo cujo controle garanta total domínio da rede. Esse foi exatamente o propósito de trabalhos recentes de Gao *et al* quando propuseram a chamada teoria estrutural de controle [91] e de Yuan *et al* sobre a controlabilidade exata de redes complexas [92].

Assim, de acordo com a teoria da controlabilidade exata de redes complexas a fração mínima de controladores n_D de uma rede de estrutura topológica arbitrária é dada pelo posto[3] da matriz de adjacência A da seguinte forma:

$$n_D = \frac{1}{N} max\{1, N - rank(A)\} \tag{2.14}$$

O resultado vale para redes com arquiteturas quaisquer: dirigidas, não-dirigidas, com *loops*, arestas múltiplas etc.

Outra abordagem importante, mas que se resume a redes dirigidas, a teoria de controle estrutural indica que a tarefa de se identificar o número de controladores pode ser mapeado no problema de se encontrar a correspondência máxima da rede. Uma correspondência é um subconjunto de arestas que não compartilham vértices iniciais ou finais. Assim, um nó é dito correspondido se uma aresta da correspondência máxima aponta para ele. Nesse sentido, o problema de correspondência máxima reflete a dependência funcional de n_D com as correlações de grau [93]. Portanto, pode-se investigar as correlações de grau entre os graus de entrada e saída usando-se o coeficiente de correlação de Pearson r (também conhecido como assortatividade) da seguinte maneira:

$$r^{(\alpha-\beta)} = \frac{\frac{1}{E}\sum_e \left(k_e^{(\alpha)} - \overline{k^{(\alpha)}}\right)\left(j_e^{(\beta)} - \overline{j^{(\beta)}}\right)}{\sigma^{(\alpha)}\sigma^{(\beta)}}, \tag{2.15}$$

onde a soma é sobre todas as arestas E, $\alpha, \beta \in \{in, out\}$ é o tipo de grau, k^{α} é o grau do nó-fonte (in), j^{β} é o grau do nó-alvo (out), e $\overline{j}_{\alpha} = \frac{1}{E}\sum_e j_e^{\alpha}$ é o grau médio dos vértices no início de cada aresta, $\sigma_{\alpha}^2 = \frac{1}{E}\sum_e \left(k_e^{(\alpha)} - \overline{k^{(\alpha)}}\right)^2$ é a variância com $\overline{k^{(\beta)}}$ e $\sigma^{(\beta)}$ são definidos similarmente. Para correlações altas (em módulo) out-in apenas uma aresta pode estar na correspondência. As arestas restantes apontam ou para hubs (caso assortativo) ou para nós com baixo grau (caso desasortativo). Desta maneira, as arestas conectadas ao nó original não possuem correspondência, o que aumenta n_D. Por outro lado, altos valores da correlação out-out implica que se um vértice possui alto grau-out, seu vizinho deve também apresentar alto grau-out gerando um efeito em cascata que inibe futuras correspondências de arestas, também aumentando n_D. O mesmo raciocínio vale para correlações in-in. Finalmente, não há dependência de n_D com relações do tipo in-out.

[3] O posto de uma matriz é o número de linhas (colunas) não-nulas quando ela está escrita na sua forma reduzida escalonada por linhas (colunas). Equivalentemente, trata-se do número de linhas (colunas) linearmente independentes da matriz.

Um trabalho recente [94] sobre controlabilidade de redes complexas mostrou que diversas redes sociais são caracterizadas por valores pequenos de n_D se comparadas com redes biológicas ou tecnológicas. Isso indica que sistemas sociais são, *ceteris paribus*, controladas por uma fração relativamente pequena de indivíduos. A aplicação da teoria clássica de controle a redes complexas é ainda algo muito novo na literatura especializada e pretende-se explorar o tópico com mais profundidade em trabalhos futuros como se verá na seção adequada.

Capítulo 3

Resultados

3.1 Ataques dirigidos por comunidades

Conforme já debatido, comunidades ou estruturas modulares são partições de grafos cuja concentração interna de arestas é maior que a densidade de conexões entre os módulos. Tem-se, pois, grandes aglomerados fracamente conectados entre si. Esta configuração permite identificarem-se quais são os vértices e arestas que servem de pontes entre essas estruturas. Destarte, a remoção de dessas poucas pontes deve separar eficientemente esses grandes aglomerados (comunidades ou módulos), fragmentando eficientemente a rede. É exatamente esta a ideia conceitual do chamado ataque baseado em módulos (MBA), que será explorado em detalhes a seguir.

O conceito de pontes entre comunidades na topologia de redes complexas foi trazido à baila recentemente por Valente *et al* [70]. Por outro lado, Hwang *et al* [71] definiram uma centralidade de ponte para caracterizar a localização de nós centrais entre vértices muito conectados. O método é bem sucedido em identificar módulos funcionais, mas não mostra resultados significativamente melhores que ataques por centralidade de intermediação quando se procura fragmentar redes complexas. Marcus e Hilgetag [72] especularam que conexões entre aglomerados poderiam ser importantes para se prever vulnerabilidades e que as posições dessas conexões poderiam ser identificadas utilizando-se a medida de frequência de arestas, isto é, a centralidade de intermediação das bordas).

Posteriormente, Bu *et al* [73] estudaram como a remoção de pontes afeta o tamanho de epidemias, mas com foco em estratégias locais com conhecimento limitado da topologia da rede. Em geral, essas contribuições identificam os nós que conectam comunidades distintas como sendo aqueles com maior centralidade de intermediação. Além do mais, esses trabalhos foram publicados antes da disponibilidade de algoritmos de detecção de comunidades. Logo, os autores não extraíram comunidades no sentido formal usualmente utilizado hoje. Mais recentemente, Shai *et al* [82] estudaram analiticamente a vulnerabilidade de redes Erdõs-Rényi modulares frente a falhas aleatórias e a ataques dirigidos a pontes. Assim, juntando essas ideias sobre ataque às pontes entre comunidades e recentes desenvolvimentos nos algoritmos de extração de comunidades em grafos complexos [59, 95] tem-se um caminho promissor para desenvolver estratégias eficientes de ataque. Por outro lado, apesar dos muitos

avanços feitos nos últimos anos, os efeitos de ataques a redes complexas baseados em comunidades é ainda um tópico em aberto. E esse é precisamente um dos tópicos que será abordado nessa contribuição.

A importância estrutural de um nó depende tanto de medidas locais quanto de medidas não locais. Assim, no âmbito do método de ataque por módulos (MBA), centralidade e detecção de comunidades são os temas que se devem abordar a fim de desenvolver um ataque efetivo. Como foi salientado nos trabalhos de Iyer *et al* [8] e de Holme *et al* [10], nós com alta centralidade de intermediação e alto grau geralmente são fortemente correlacionadas e ambos os ataques acabam tendo eficiências semelhantes. Além disso, o trabalho de Iyer mostra que para redes reais os métodos baseados em intermediação são em geral os mais eficientes. Por isso, no método de ataque por módulos (MBA), todas as comparações são realizadas tomando-se como referência o ataque por centralidade de intermediação.

Da mesma forma, os vértices de ligação entre comunidades geralmente têm alta centralidade de intermediação uma vez que muitos caminhos curtos passam por eles. Por outro lado, já que se espera menos conexões entre comunidades, as pontes não são necessariamente aqueles nós com maior grau. Por conseguinte, a fim de separar as comunidades de uma maneira eficiente, o ataque MBA proposto vagamente se assemelha à ideia original de laços fracos introduzida por Granovetter [96] para redes sociais e posteriormente desenvolvida no âmbito de comunidades topológicas por De Meo, Ferrara *et al* [97].

Assim, o procedimento de ataques por módulos é constituído pelos seguintes passos:

1. Extraem-se comunidades usando um algoritmo heurístico de detecção.
2. Opta-se por atacar nós ou arestas.
3. Faz-se uma lista com os nós (ou arestas) que participam em ligações intercomunitárias.
4. Ordena-se a lista em ordem decrescente da centralidade de intermediação de nó ou aresta.
5. Excluem-se os alvos um a um a partir do primeiro da lista.
6. Embora centrada na remoção de nós, uma vez que um nó de uma ligação entre duas comunidades é excluído, sua contraparte é ignorada (não há necessidade de removê-lo), a menos que ele também participe em outra ligação intercomunitária.
7. O ataque é sempre restrito à maior componente conectada da rede. Em outras palavras, se em algum momento o próximo nó da lista não pertence à maior componente conectada, aquele alvo é ignorado durante aquele passo.

Para se comparar os métodos mais conhecidos de fragmentação simultânea de redes, ilustram-se diversos ataques na figura 3.1, que resumem os efeitos das estratégias de ataque por centralidade de intermediação, por grau, por maior caminho [69] e por módulos sobre a rede de distribuição de energia elétrica dos EUA. Podem-se observar na figura as representações por nós e modular (figura 3.1B) e fotos da rede quando 1%, 2%, e 3% dos nós são removidos segundo as prescrições por centralidade de intermediação (*High Betweenness* - HB) e por módulos (figura 3.1C). Notavelmente, o método

de ataque por módulos quebra a rede original de 4941 nós em diversos fragmentos menores que 210 nós ($\approx 4\%$ do tamanho original) ao removerem-se apenas 142 nós (menos de 3%) identificados pelo novo procedimento. Por comparação, os outros métodos removem apenas 18% da rede original com a retirada da mesma quantidade de estruturas, isto é, mais de 4000 nós continuam conectados após o ataque. Essa fragmentação extrema da rede é visualmente apresentada na figura 3.1C .

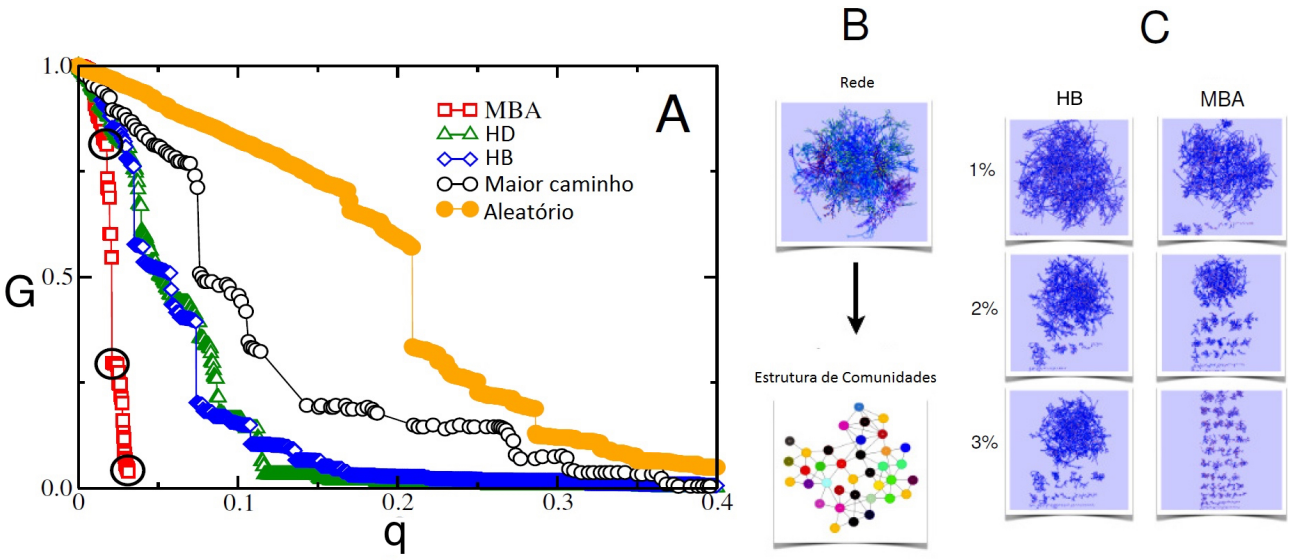

Figura 3.1: Comparação entre o efeito dos ataques à rede de distribuição de energia dos EUA: por intermediação, por grau, por maior caminho, aleatório e por módulos. (A) Tamanho da maior componente conectada em termos do tamanho original, G, como função da fração de nós retirados, q. (B) Representação modular da rede. (C) Fotos da representação gráfica da rede quando 1%, 2% e 3% dos nós são retirados utilizando os métodos de ataque por centralidade de intermediação (HB) e por módulos (MBA).

Com o objetivo de demonstrar a validade do método de ataque por módulos aplica-se o algoritmo a 10 redes reais com diferentes estruturas topológicas. Para quantificar o efeito do ataque [98], define-se $\mathcal{G}$ como sendo uma rede inicial de tamanho N, e $\mathcal{G}_q$ como a rede resultante após a remoção de uma fração q de vértices. Denota-se, então, por $\mathcal{L}_q$ a maior componente conectada $\mathcal{G}_q$, cujo tamanho é indicado por $N_{\mathcal{L}}$. Define-se, ainda, o parâmetro de ordem $G(q) = \frac{N_{\mathcal{L}}}{N}$ que permite quantificar a resposta da rede ao ataque medida pelo tamanho relativo do sistema resultante como função da fração nós (ou arestas) deletados. Nesse sentido, escolhendo um método qualquer como referência, o ganho de eficiência pode ser definido ponto-a-ponto para cada valor de q como sendo:

$$\gamma(q) = \frac{G_{null}(q)}{G(q)}. \tag{3.1}$$

Essa quantidade aumenta conforme o método de ataque se torna mais eficiente do que aquele que foi tomado como referência.

A lista de nós a serem excluídos é obtida apenas uma vez, antes do procedimento de ataque começar, no que é chamado de ataque simultâneo. Ataques sequenciais (ou ataques em cascata) [68]

são em geral mais eficazes porque as medidas topológicas são atualizados depois cada eliminação. Isto significa que a detecção de comunidades e a centralidade de intermediação têm de ser calculadas novamente após cada remoção. Isso implica um aumento do tempo de computação, tornando o ataque sequencial impraticável para grandes redes.

As redes que foram usadas para teste são de três tipos: infraestruturais (malha energética dos EUA - US power grid, estradas europeias - Euro roads, voos internacionais - Open flights, e aeroportos norteamericanos - US airports) [45, 99–105], biológicas (proteína do levedo - yeast protein, C *elegans* e H *pylori*) [76, 106, 107] e sociais (Facebook, Google+ e Twitter) [108–111]. Na rede Euro road, os nós representam cidades europeias e as arestas representam estradas. Na rede US power grid uma aresta indica uma linha de suprimento de energia e um nó é ou um gerador, ou um transformador ou uma subestação. A rede de interação da proteína do levedo é a mesma encontrada em [76]. Na rede metabólica da minhoca *Caenorhabditis elegans*, os nós são metabólitos (e.g., proteínas) e as arestas são interações entre eles. A rede *Helicobacter pylori* é a mesma interação proteica encontrada em [106]. Na subrede de relacionamentos do Facebook (NIPS) nós representam usuários e uma conexão indica amizade. Similarmente, na rede do Google+ network uma aresta significa que um usuário está no círculo de amizades do outro, enquanto que na rede do Twitter uma aresta indica que ambos usuários se seguem mutuamente.

Tabela 3.1: Dados topológicos das redes: tamanho (N), número de arestas (E), grau médio ($\langle k \rangle$), modularidade (Q), tamanho relativo da maior componente conectada (N_{mod}^{max}), fração de arestas ligando comunidades (E_{inter}), ganho global em eficiência do método de ataque por comunidades (η, ver equação (3.2) para definição). Para os quatro parâmetros relacionados à detecção de comunidades mostram-se os valores correspondentes ao caso de maior eficiência entre 10 rodadas de extração para os métodos infomap (I) e *Louvain* (L). Os dados estão representados tanto para ataques a nós quanto para ataques a arestas.

				Ataque por nó					Ataque por aresta				
Rede	N	E	$\langle k \rangle$	Q	N_{mod}^{max}	E_{inter}	η		Q	N_{mod}^{max}	E_{inter}	η	
Facebook	2888	2.981	2,06	0,81	0,262	0,012	4,19	(L)	0,81	0,262	0,012	4,19	(L)
Twitter	23.370	32.831	2,81	0,82	0,018	0,169	38,44	(I)	0,83	0,018	0,168	38,30	(I)
Google Plus	23.628	39.194	3,32	0,69	0,070	0,279	22,80	(I)	0,69	0,070	0,279	22,80	(I)
Energia Elétrica EUA	4.941	6.594	2,67	0,94	0,049	0,033	111,02	(L)	0,82	0,007	0,178	72.92	(I)
Autoestradas UE	1.174	1.417	2,41	0,79	0,016	0,203	108,40	(I)	0,79	0,014	0,198	95,16	(I)
Voos internacionais	2.939	15.677	10,67	0,65	0,184	0,142	8,30	(L)	0,65	0,182	0,139	8,14	(L)
Aeroportos EUA	1.574	17.215	21,87	0,35	0,296	0,363	4,16	(L)	0,34	0,267	0,341	4,10	(L)
Proteína levedo	1.846	2.203	2,39	0,77	0,025	0,223	36,14	(I)	0,77	0,025	0,220	35,14	(I)
H *pylori*	724	1.403	3,88	0,54	0,124	0,364	19,59	(L)	0,49	0,047	0,485	14,35	(I)
C *elegans*	453	2.025	8,94	0,43	0,163	0,423	12,04	(L)	0,43	0,163	0,423	12,04	(L)

A figura 3.3 mostra os resultado para o ataque modular por vértices para as 10 redes testadas. As simulações mostram que o ataque proposto sempre é mais eficiente que o ataque tradicional por centralidade de intermediação. Inicialmente ambos os métodos são similares, mas conforme as pontes entre comunidades são deletadas, módulos inteiros são segmentados do centro do grafo, resultando em uma rápida atomização da rede e também em uma diminuição abrupta de G (significando um aumento rápido no ganho em eficiência γ). Já para ataques por arestas, os resultados são mostrados na figura 3.4. Nesse caso, já que se apagam apenas arestas conectando módulos, a fase inicial dos ataques

é menos eficiente que o ataque tradicional para algumas redes. Nessas situações observa-se um platô em G antes das comunidades serem retiradas. Depois que se alcança este ponto, G cai abruptamente e comunidades relativamente grandes são separadas muito velozmente e a rede inteira se despedaça.

Quando se foca tanto na retirada de nós quanto na retirada de arestas, os ataques param quando a lista de ataque é exaurida, isto é, no momento em que G atinge o valor mínimo final G_e. No caso de remoção de borda $G_e = N_{mod}^{max}$, em que N_{mod}^{max} é a relação entre os tamanhos da maior comunidade e da rede. Por outro lado, a fração final de arestas removidas é $q_e = E_{inter}$, onde E_{inter} representa a relação entre o número de arestas ligando módulos e o número total de extremidades (ver tabela 3.1). No caso de remoção por nós N_{mod}^{max} representa apenas um limite superior para o valor final G, porque nós adicionais são retirados como um efeito colateral do processo, quebrando a estrutura interna das comunidades —G_e está em geral bem abaixo desse limite. Além disso, neste caso q_e está muito abaixo de E_{inter}, porque em cada eliminação dos nós todas as suas bordas são removidas. O comportamento de Q está ilustrado na figura 3.2, que mostra alta correlação entre a modularidade e a fração de arestas entre comunidades.

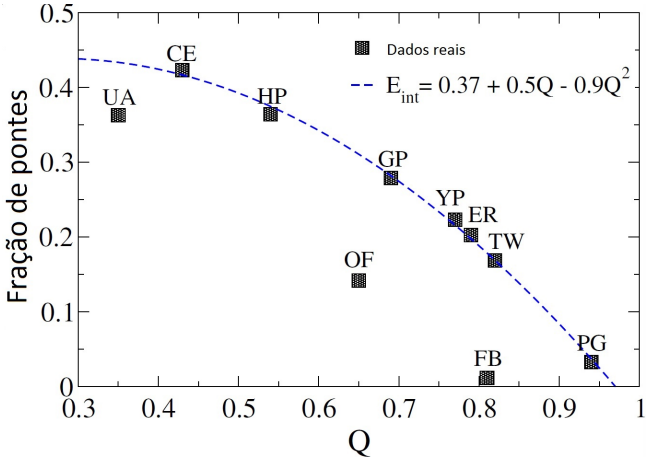

Figura 3.2: Mostra-se a relação entre a fração de nós que conectam diferentes módulos, e a modularidade, Q. Os dados correspondem a dez redes reais: Facebook (FB), Twitter (TW), Google Plus (G +), rede de energia dos Estados Unidos (PG), Estradas Europeias (ER), voos abertos (OF), aeroportos dos Estados Unidos (UA), proteína de levedura (YP), H *pylori* (HP) e C *elegans* (CE). Como esperado, observa-se uma alta correlação (negativa) entre E_{int} e Q, que é precisamente a característica que torna o método de modularidade bem colocado. A extração de comunidades foi realizada utilizando os métodos *Louvain* ou Infomap como detalhado na tabela 3.1.

Resumindo, o ponto (q_e, G_e) (a intersecção das linhas azuis tracejadas nas figuras 3.3 e 3.4) depende da estrutura modular especial de cada rede, marca onde todas as comunidades estão desconectadas sem nó ou aresta alvos dentro dos aglomerados restantes. Pode-se com segurança dizer que a rede pára de funcionar como um todo neste momento –por exemplo, as informações se manteriam empilhadas dentro das comunidades e estas estruturas não seriam capazes de se comunicar umas com as outras.

Os resultados apresentados acima podem ser resumidos na relação entre q e γ (o ganho em eficiência do ataque por módulos se comparado com o ataque tradicional por centralidade de intermediação),

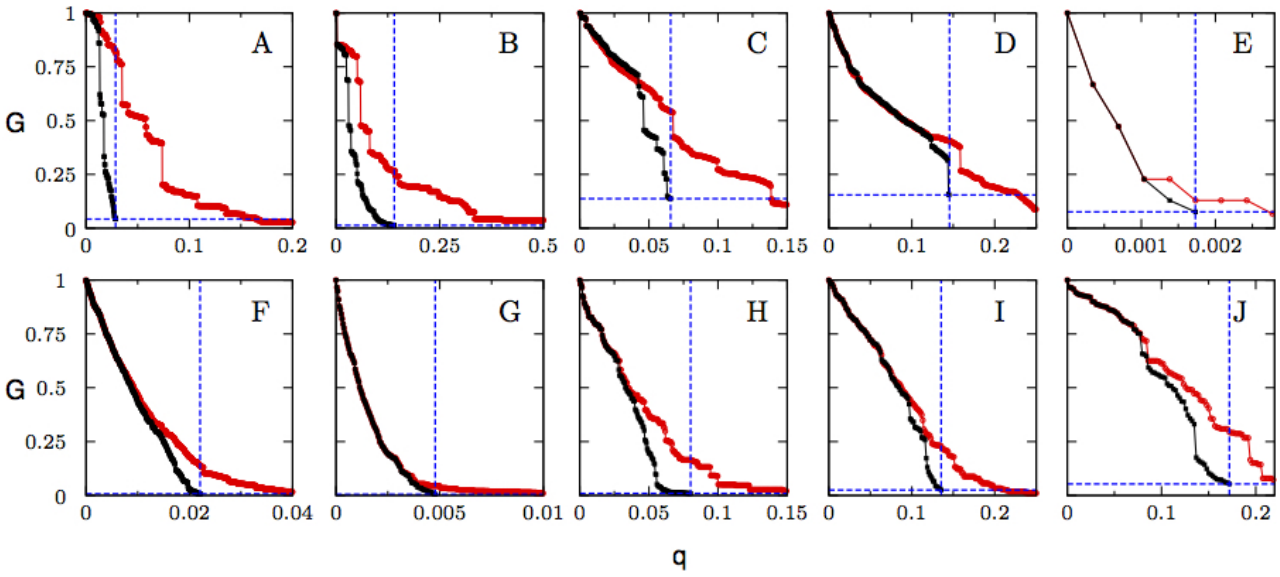

Figura 3.3: Tamanho da maior componente conectada em termos do tamanho inicial da rede, G, em função da fração de nós removidos q. Ataque a vértices por módulos (quadrados pretos), ataque a vértices por centralidade de intermediação (círculos vermelhos). (A) Rede de distribuição elétrica do oeste norteamericano. (B) Rede de estradas européias. (C) Voos abertos. (D) Aeroportos norteamericanos. (E) Facebook. (F) Twitter. (G) Google plus. (H) Proteína do levedo. (I) H *pylori*. (J) C *elegans*. As interseções das linhas azuis pontilhadas correspondem ao ponto de dano máximo da rede utilizando o ataque por módulos.

conforme demonstra a figura 3.5. Notavelmente, observa-se que a eficiência é mais que duplicada para a maioria das redes com menos de 7% dos nós retirados. O caso mais relevante é o do US power grid com quase 20 vezes de ganho com aproximadamente 3% de nós removidos. Até mesmo nos piores dos casos (H *pylori*, C *elegans* e US airports) obtêm-se ganhos de eficiência que vão de 3 a 8 vezes para 14% a 16% de vértices deletados. No caso da rede US power grid, o método proposto quebra a rede original de 4.941 vértices em vários fragmentos menores que 210 nós ($\approx 4\%$ do tamanho original) ao se remover menos de 3% da rede. Ao se comparar com outros métodos, um ataque por intermediação resulta numa componente conectada da ordem de 80% do tamanho original.

O desempenho final do procedimento de ataque por módulos relativo ao ataque por intermediação pode ser medido por quão rápido o método alcança o ponto final de ataque. Defini-se então o ganho global de eficiência como:

$$\eta = \gamma(q_e) \times \frac{q_{null}(G_e)}{q_e} \tag{3.2}$$

Na figura 3.6 mostram-se os resultados de η em função da modularidade Q para ambos ataques por nós e arestas aplicados às dez redes reais estudadas. Pelas figuras é evidente a existência de uma forte correlação entre η e Q, mostrando que redes altamente modulares são relativamente frágeis a ataques por módulos.

Em suma, apresenta-se nesta primeira etapa do projeto um método de ataque a redes complexas baseado em módulos que consiste em extrair as comunidades de uma determinada rede e apagar apenas

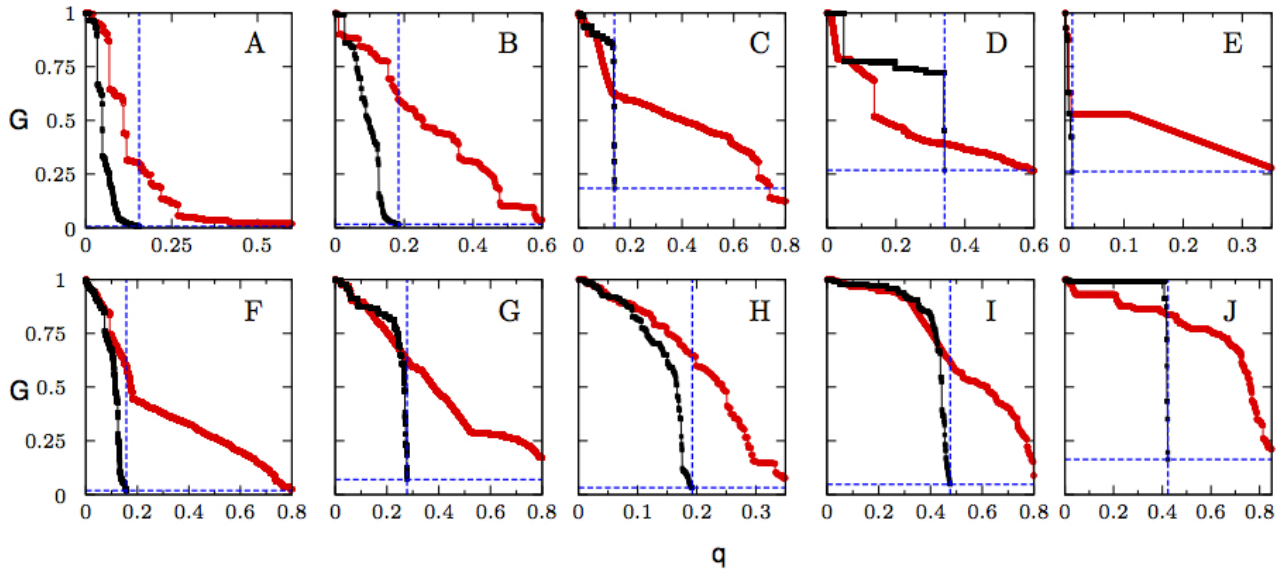

Figura 3.4: Tamanho da maior componente conectada em termos do tamanho inicial da rede, G, em função da fração de arestas removidas q. Ataque por módulos (quadrados pretos) e por centralidade de intermediação (círculos vermelhos). (A) Rede de distribuição elétrica do oeste norteamericano. (B) Rede de estradas européias. (C) Voos abertos. (D) Aeroportos norteamericanos. (E) Facebook. (F) Twitter. (G) Google plus. (H) Proteína do levedo. (I) H *pylori*. (J) C *elegans*. As interseções das linhas azuis pontilhadas correspondem ao ponto de dano máximo da rede utilizando o ataque por módulos.

os nós que ligam módulos distintos ordenados por centralidade de intermediação. As simulações computacionais em muitas redes reais mostram que o método de segmentação por módulos é mais eficiente em fragmentar redes complexas do que procedimentos tradicionais com base apenas em critérios de centralidade. Com efeito, pode-se dizer que os vértices mais conectados ou os nós que têm o valor mais alto de centralidade de intermediação não são necessariamente os mais importantes para a sobrevivência da rede. Nós que fazem pontes entre comunidades distintas são estruturalmente mais importantes e cruciais para a coesão da rede do que *hubs* ou nós altamente centrais. Se atacarem-se exatamente esses nós ou bordas, o dano produzido à rede é maior que aquele produzido por métodos tradicionais de ataque a redes complexas, eliminando-se a mesma quantidade de estruturas.

O objetivo de aplicar o presente ataque baseado em pontes entre módulos para uma dada rede é desvendar a sua vulnerabilidade estrutural, medindo o quão rápido se consegue atingir o regime em que as comunidades da rede estão todas desconectadas. Portanto, propõe-se caracterizar a vulnerabilidade modular de redes complexas precisamente pela rapidez com que o ponto crítico em que todas as comunidades estão desligadas é atingido. Excepcionalmente, o presente trabalho mostra que o ganho global de eficiência do método de ataque cresce rapidamente com a modularidade da rede, isto e, quanto maior a modularidade, mais frágil a rede se torna. Nesse sentido, a identificação de comunidades é o ingrediente essencial do nosso método. Portanto, embora estes módulos topológicos não tenham relação direta com comunidades reais, eles podem, eventualmente, divulgar algumas informações relevantes sobre a funcionalidade estrutural-modular de redes complexas.

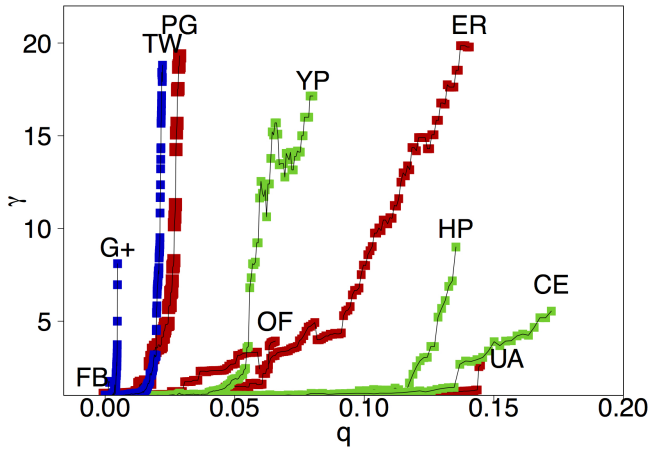

Figura 3.5: Ganho em eficiência do ataque a vértices por módulos se comparado com o ataque por centralidade de intermediação ($\gamma = G_{null}/G$) como função da fração de nós removidos, q. O código das redes é dado por Facebook (FB),Twitter (TW), Google Plus (G+), rede elétrica dos EUA (PG), estradas européias (ER), voos internacionais (OF), aeroportos dos EUA (UA), proteína de levedo (YP), H *pylori* (HP) e C *elegans* (CE). Redes de infraestrutura estão pintadas de vermelho, biológicas de verde e sociais de azul.

Algoritmos não heurísticos e bem estabelecidos para particionamento de grafos tais como aqueles que usam maximização de modularidade, inferência estatística e corte normalizado espectral [112] possuem um incoveniente computacional para redes reais. A maioria deles só é factível para algumas centenas de nós. Assim, para redes reais com milhares ou até milhões de nós devem-se usar algoritmos heurísticos que são muito menos custosos computacionalmente. Por outro lado, a escolha específica do método de detecção de comunidades deve impactar a eficiência do ataque por módulos apenas se a lista de nós e arestas pontes mudar significativamente. Nesse sentido, como apontado por Fortunato *et al* [113], o método Infomap por Rosvall e Bergstrom [95] e o método *Louvain* por Blondel e *et al* [59] são os que têm melhor performance em detectar comunidades em várias redes com topologias de referência. Assim, a figura 3.7 mostra o ataque médio no caso da rede de distribuição de energia dos EUA, representado por uma linha sólida cercada por uma sombra cinza que indica a variação correspondente a dez rodadas distintas dos métodos *Louvain* e Infomap. Como pode ser visto, a variação entre diferentes rodadas dos algoritmos de detecção de comunidades é muito pequena se comparada à diferença entre o ataque por módulos e por centralidade. Embora haja algumas diferenças na definição das comunidades, o comportamento das curvas de ataque por módulos é muito similar e as bandas de ambos os métodos se sobrepõem fortemente. Portanto, ambos os algoritmos de extração de comunidades (Infomap e *Louvain*) são igualmente adequados para a construção do ataque por módulos, que se mostra pouco sensível à escolha específica do método de extração desde que suficientemente preciso. E, independentemente do algoritmo particular utilizado para identificar as comunidades, o método de ataque baseado em módulos sempre resulta melhor que os métodos tradicionais utilizados para se fragmentar redes reais.

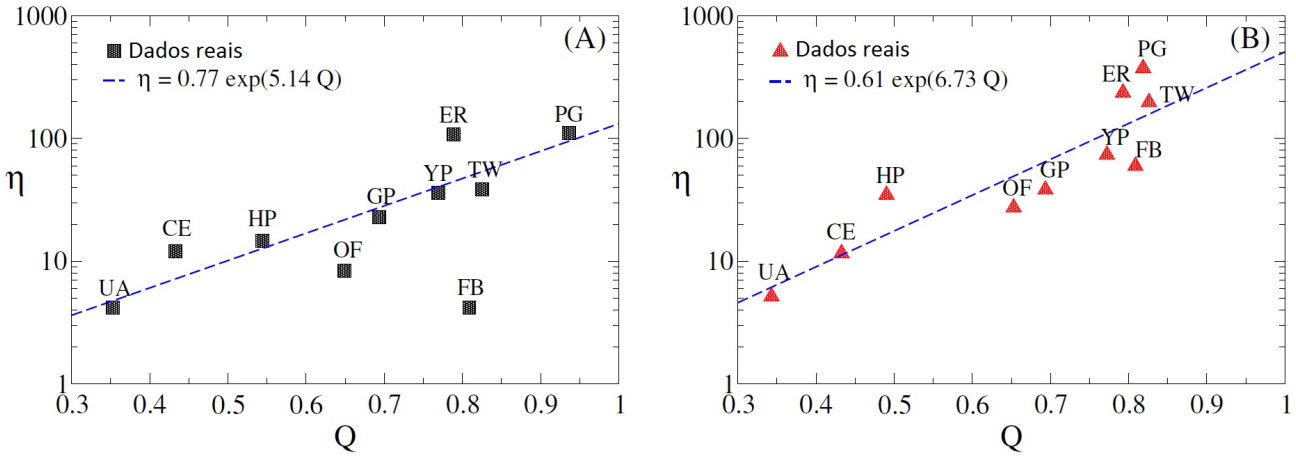

Figura 3.6: Ganho global de eficiência (η) do ataque por módulos relativo ao ataque por intermediação como função da modularidade Q para remoção de vértices (A) e de arestas (B). O eixo vertical está em escala logarítmica e o eixo horizontal em escala linear. As redes atacadas são Facebook (FB),Twitter (TW), Google Plus (G+), rede elétrica dos EUA (PG), estradas européias (ER), voos internacionais (OF), aeroportos dos EUA (UA), proteína de levedo (YP), H *pylori* (HP) e C *elegans* (CE).

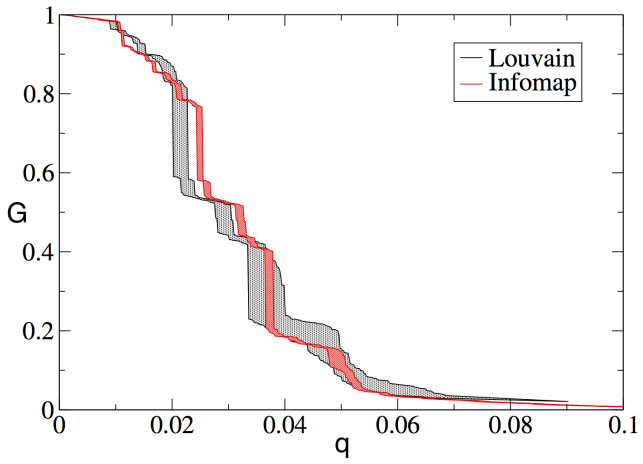

Figura 3.7: Vinte rodadas distintas do ataque por módulos segundo os algoritmos de *Louvain* e Infomap. Mostram-se os ataques aos nós no caso da rede elétrica dos EUA e a média deles para demonstrar a típica sensibilidade do método proposto à escolha particular do algoritmo de extração de comunidades.

3.2 *Performance* de métodos de ataque

A relação entre o dano causado a uma rede, o custo computacional e a complexidade temporal necessária para se atingir um nível desejado de fragmentação é uma questão importante ainda pouco abordada em contribuições na literatura. Tal conhecimento pode ser muito útil em aplicações práticas onde é necessário guia para se escolher o método de ataque mais apropriado— quando os recursos são limitados, por exemplo. Nesta seção, pretende-se preencher precisamente essa lacuna ao se estudar a relação custo/benefício de estratégias de ataque bem estabelecidas, tanto adaptativas quanto simultâneas, em redes modulares.

Como assinalado em [78], o método de Influência Coletiva (CI) supera em eficiência os ataques adaptativo e não-adaptativo por grau, PageRank, adaptativo por proximidade, adaptativo por autovetor, adaptativo por intermediação e k-core ataques em termos de atomizar grandes redes com a remoção do menor conjunto de nós. No entanto, o método CI não foi testado em redes de modularidade moderada ou alta, nem contra o método MBA que apresenta por sua vez resultados poderosos conforme a modularidade do sistema aumenta. Além disso, outros métodos adaptativos como o HBA não foram comparados com CI, talvez por causa do alto custo computacional do método anterior para sistemas muito grandes. Portanto, nesta seção comparam-se os métodos: CI, MBA, HDA e HBA. Para tanto, discutem-se as complexidades dos métodos de fragmentação estudados a seguir (ver tabela 3.2)— isto é, a relação assintótica entre o comprimento do vetor de entrada e número de passos de um algoritmo.

- **Adaptativo por alto grau— HDA**: neste método, os vértices são classificados por grau e removidos iterativamente— o grau dos nós remanescentes são recalculados após cada remoção. Esse algoritmo roda com complexidade $O(N^2 + NE)$, onde N é o número de vértices e E é a quantidade de arestas, o que o torna realizável para aplicações reais. No caso de grafos esparsos, a complexidade computacional do método é de $O(N^2)$.

- **Adaptativo por intermediação— HBA**: a centralidade de intermediação de um nó é basicamente a fração de todos os menores caminhos que passam por ele. Neste método iterativo, a centralidade de intermediação é recomputada após cada remoção. Isto acontece a um alto custo computacional. Os melhores algoritmos para cálculo da intermediação têm complexidade computacional de $O(NE)$, o que para grafos esparsos resulta em uma complexidade de $O(N^3)$ para o método HBA. Isto não é rápido o suficiente para redes reais muito grandes e o método pode se tornar impraticável nesses casos.

- **Influência Coletiva— CI**: neste método, o problema de se encontrar o conjunto mínimo de vértices que se removido resultaria no desmantelamento total da rede é mapeado pela Influência Coletiva de cada nó i pela seguinte equação:

$$CI_k(i) = (k_i - 1) \sum_{j \in \partial Ball(i,l)} (k_j - 1) \tag{3.3}$$

  onde k_i é o grau do nó e $\partial Ball(i, l)$ é o conjunto de todos os nós a uma distância l do vértice i. Esse método possui uma complexidade teórica de $O(N^2 \log N)$. Contudo, o algoritmo pode ser rodado em $O(N \log N)$ usando uma estrutura de dados max-heap que evita a classificação total dos valores de CI. [79]

- **Ataque baseado em módulos— MBA**: a abordagem baseada em módulos consistem em encontrar os vértices que ligam comunidades distintas de uma determinada rede, classificá-los por sua

centralidade de intermediação e removê-los em ordem decrescente, focando sempre apenas na maior componente conectada a cada remoção. Trata-se de um método simultâneo no qual a lista de ataque é gerada apenas uma vez. Algoritmos de extração rápida de comunidades como o de *Louvain* possui complexidade $O(N \log N)$ [59] enquanto que a centralidade de intermediação é calculada apenas para os vértices interconectados (ou seja as pontes entre as comunidades), então a dependência deste termo vai com $O(N_d E)$, onde N_d é o número de pontes entre as comunidades e E é o número total de arestas. Isto resulta em uma complexidade total de $O(N \log N + N_d E)$ e para grafos esparsos $O(N + N_d N)$. De fato, o algoritmo MBA é conhecido por ter um excelente desempenho em redes altamente modulares, casos em que $N_d \ll N$. Todavia, é possível manter um regime linear de complexidade contanto que $N_d < \log N$, o que é bastante razoável para redes com comunidades bem definidas.

Ataque	**Complexidade**
HDA	$O(N^2)$
HBA	$O(N^3)$
MBA	$O(N + N_d N)$
CI	$O(N \log N)$

Tabela 3.2: Complexidade temporal dos algoritmos de ataque estudados neste no caso de grafos esparsos: HDA, HBA, MBA e CI.

Portanto, a partir das complexidades computacionais, os métodos CI e MBA parecem os mais viáveis para grandes redes. Ainda assim, para se comparar esses métodos e escolher o algoritmo mais eficiente, a complexidade computacional deve ser balanceada com a resposta de cada rede aos diferentes ataques. Este *feedback* de um sistema a uma determinada estratégia é geralmente chamado de robustez da rede. Apesar de haver muitas pesquisas recentes sobre a robustez de redes complexas, não existe uma definição única do termo [8, 9, 74, 75]. Robustez pode ser definida como a capacidade de um sistema de manter um dado conjunto de funções ou serviços quando submetidos a perturbações ou ataques [114]. Este comportamento é fortemente acoplado ao objetivo do sistema real subjacente a sua representação matemática de tal forma que a robustez seja dependente do serviço desempenhado pela rede [115]. Portanto, a robustez de grafos abstratos complexos, nos quais a dinâmica do sistema não é abordada, deve ser descrita quer uma transição de fase estrutural quer, na ausência de tal comportamento, por uma função genérica representando a resposta geral da topologia da rede a uma estratégia de fragmentação ou ataque [7, 69, 116–119]. Dentro deste contexto, a robustez é tipicamente considerada em um quadro de percolação e quantificada pela fração crítica q_c de vértices que, uma vez removidos, levam à atomização completa da rede. Nesse ponto, pode-se dizer com segurança que a rede deixa de funcionar como um todo, porque não há nenhuma componente gigante conectando o sistema de acordo com o critério Molloy-Reed. Para se quantificar o efeito dos ataques sobre redes [98] geralmente define-se o parâmetro de ordem $G(q) = \frac{N_\mathcal{C}}{N}$ que é o tamanho do maior aglomerado conectado em relação ao tamanho da rede original como função de q, a fração de nós excluídos.

A fim de comparar diferentes métodos de fragmentação, para tamanhos de redes arbitrários, a maioria das medidas de robustez leva em conta listas de ataques do mesmo comprimento para todas as estratégias—- isto é, o número de nós da rede. Schneider *et al* propuseram uma medida única para quantificar a robustez de um sistema que enfrenta ataques [120] que se resume à área sob a curve $G(q)$. No entanto, na estratégia baseada em módulos por exemplo, a lista de ataque é muito menor que o tamanho da rede e o sistema perde funcionalidade muito antes da remoção de todos os vértices. É necessária, então, uma abordagem para explicar tais situações. Para tanto, é proposta uma generalização da medida de robustez proposta por Schneider *et al*, considera-se a área sob a curva $G(q)$ quando uma fração genérica $q_{max} \leq 1$ é removida da rede da seguinte forma (ver figura 3.8 para detalhes):

$$G_d = \frac{N_{max}}{N} \tag{3.4}$$

Onde q_{max} é o ponto em que termina o ataque e G_{min} é o tamanho relativo da maior componente conectada no ponto q_{max}. Esta quantidade mede a área abaixo da curva $G(q)$ relativa à área máxima de ataque, isto é, a área do retângulo delimitado pelos pontos $(0, G_{min}), (0, 1), (1, 1), (1, G_{min})$. Em particular, Bagrow *et al* estudaram a robustez dos sistemas modulares quando da falha de seus elementos [121]. Os autores mostraram que a organização modular desempenha um papel crucial na funcionalidade de redes, e que as comunidades podem se desacoplar muito mais cedo que a atomização completa do sistema em caso de falha ou ataque. Este ponto, onde todas as comunidades estão desconectadas do grafo original e quando a rede deixa de funcionar como um todo, chama-se o ponto de desativação modular da rede, $P_d = (G_d, q_d)$, onde $G_d = \frac{N_{max}}{N}$, é o tamanho do maior módulo relativo ao tamanho da rede original e q_d é a fração de pontes entre as comunidades. Neste caso, $q_{max} = q_d$ e $G_{min} = G_d$. Conseqüentemente, pode-se definir a fragilidade generalizada de uma rede frente um dado ataque pela função inversa da robustez como $f = 1/R$.

Com efeito, a contribuição central desta seção é o que agora chama-se de performance, definida para um dado ataque como o balanço entre a fragilidade generalizada e o número de passos do algoritmo como segue:

$$\mathcal{P} = \frac{f}{t} \tag{3.5}$$

onde $f = 1/R$ é a fragilidade da rede e t é o número de passos necessários para se completar o procedimento. Em outras palavras, $\mathcal{P}$ mede a troca entre eficiência de ataque, medida pelo inverso da robustez da rede (ou fragilidade frente ao ataque), e o tempo necessário para se completar o ataque— um algoritmo rápido que não é muito eficiente em fragmentar uma rede terá performance similar de um método lento, mas muito eficiente.

A fim de comparar o desempenho $\mathcal{P}$ do métodos MBA, HDA, HBA e CI, deve-se primeiro construir redes modulares de referência cujos resultados são independentes da escolha específica do método de extração de comunidades. Muitos métodos para detecção de comunidades foram propostos ao longo dos últimos anos, e testar a exactidão desses algoritmos é, por conseguinte, muito importante. Para verificar o desempenho dos métodos de extração (ou identificação) de comunidades, recentemente foram propostas várias redes artificiais de referência com estruturas comunitárias bem

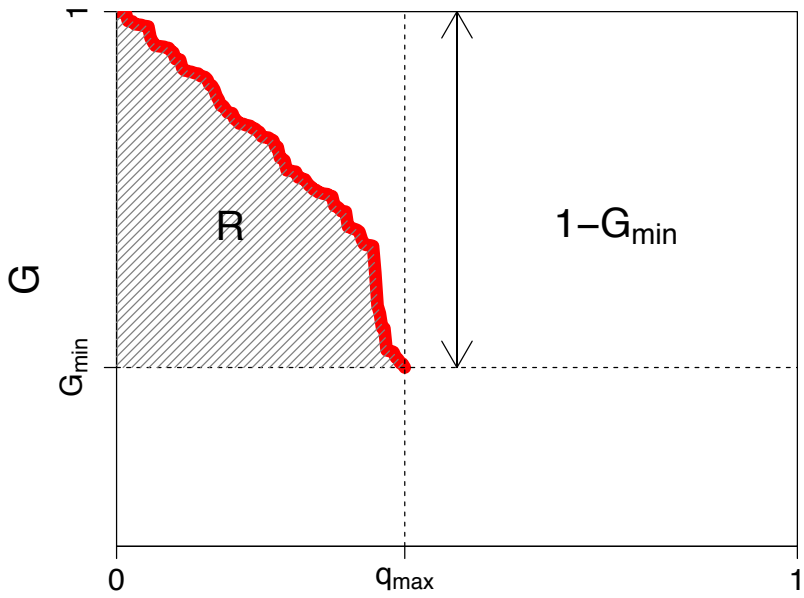

Figura 3.8: Representação geométrica da robustez generalizada como a razão entre a área sob a curva vermelha e a área máxima de ataque delimitada pelo retângulo com lados $1 - G_{min}$ e 1 conforme a definição equação 3.4.

definidas. Um dos primeiros critérios de referência com essa finalidade é uma classe de redes artificiais não dirigidas propostas por Girvan e Newman, e conhecidos como o Modelo de Bloco Estocástico (SBM) [122]. Essas referências consistem em redes com nós com aproximadamente o mesmo grau, como em um grafo aleatório, mas com vértices preferencialmente conectados aos nós do mesmo módulo. No entanto, redes reais possuem distribuições heterogêneas tanto de grau quanto de tamanho de comunidades. Esse comportamento é responsável por características de redes reais, tais como resiliência a falhas/ataques e à ausência de um limiar para percolação ou epidemia. Nesse sentido, Karrer e Newman introduziram uma distribuição heterogênea de grau no modelo SBM, introduzindo o Modelo de Bloco Estocástico corrigido [123]. Mais recentemente, Lancichinetti, Fortunato e Radicchi propuseram redes de referência não dirigidas (LFR) [124], que pressupõem que as distribuições de tamanho e grau apresentam comportamento invariante em escala. Nas redes de referência tipo LFR, a modularidade é controlada pelo parâmetro de mistura μ, que é a relação entre o número de arestas ligando um vértice a outras comunidades e o seu próprio grau. Em outros palavras, cada nó compartilha uma fração μ de suas bordas com nós da rede inteira e uma fração $1 - \mu$ de suas arestas com nós de sua própria comunidade. Daí, pequenos valores de μ indicam altos valores de modularidade, enquanto valores maiores significam que as comunidades não estão bem definidas.

Vale ressaltar que o tamanho médio dos módulos geralmente depende da precisão e do limiar de detecção do método de identificação de comunidades. No entanto, esta-se interessados em encontrar informações independentes da escolha específica do algoritmo. Nesse sentido, Lancichinetti e Fortunato propuseram uma análise comparativa de vários métodos de extração da comunidade, incluindo os

algoritmos espectrais, de inferência estatística e heurísticos [113]. Os autores concluíram que, entre outros, o algoritmo *Louvain*, também conhecido como método multinível, proposto por Blondel *et al* [59] funciona muito bem em redes tipo SBM e LFR. Além disso, tal como referido no mesmo artigo, o método *Louvain* tem baixa complexidade computacional e pode ser facilmente utilizado em grafos muito grandes. Em um artigo mais recente, Yang *et al* chegaram a resultados semelhantes, afirmando que ao levar em conta a precisão e o tempo de computação, o algoritmo de *Louvain* supera outros métodos testados em um amplo conjunto de redes tipo LFR [125]. Por conseguinte, a fim de se obter dados independentes do algoritmo de comunidades, nas simulações daqui em diante usam-se redes artificiais do tipo LFR e o método *Louvain* para detectar suas comunidades.

Após a aplicação do método às redes modulares de referência, estuda-se a performance dos principais métodos de ataque às mesmas redes reais apresentadas na tabela 3.1, com a inclusão da rede de energia elétrica da UE [126] que possui $N = 1.494$, $E = 2.322$ e $Q = 0,89$. Novamente, todas as redes são tratadas como não-dirigidas. Além disso, *loops* e bordas múltiplas foram removidos.

Para garantir a uniformidade de avaliação de desempenho, as simulações foram avaliadas dentro das mesmas condições de *software* e *hardware* (*igraph* v 1.0.1 [127] e R v 3.2.3). Para comparar todos os métodos, gera-se um conjunto de redes tipo LFR com modularidade e tamanho variáveis. Redes com modularidade próxima de $0,41$ apresentam uma partição ótima que se aproxima do caso de duas comunidades distintas enquanto que valores maiores implicam os módulos serem mais bem distinguíveis [1]. Por outro lado, valores de modularidade menores ($Q \sim 0,22$) indicam partições subótimas que não conseguem identificar corretamente comunidades [1]. Destarte, já que se está interessado apenas em redes com módulos bem definidos, a modularidade das redes artificiais utilizadas variam entre $0,48$ e $0,99$. Os sistemas apresentam tamanhos entre 10^3 e 10^4

Redes maiores são computacionalmente muito dispendiosas para o método HBA e por isso não foram estudadas. Na figura 3.9 traça-se o equilíbrio entre o tempo de execução de cada algoritmo e a robustez da rede $\mathcal{P}$ como função da modularidade Q para uma rede LFR com tamanho $N = 10^4$ (resultados semelhantes se aplicam para $10^3 < N < 10^4$) em escala semi-log. Pode-se ver que $\mathcal{P}$ não é particularmente sensível a aumentos na modularidade para os métodos HBA e HDA, para o método CI é mais sensível, mas ainda mais para o algoritmo MBA.

Para o intervalo de valores de modularidade verificados aqui, o MBA supera todos os outros métodos. Além disso, a performance do método modular é de até duas ordens de magnitude maiores do que a do método CI para grandes valores de modularidade.

Para avaliar este raciocínio em redes reais, estuda-se agora a relação entre a fragilidade da rede e o tempo necessário para gerar cada ataque para as redes exemplificadas na tabela 3.1 além da rede de energia elétrica da União Europeia discutida acima. Nas figura 3.10 e 3.11 são traçados o processo de fragmentação e $\mathcal{P}$ para os métodos HBA, HDA, CI e MBA. De fato, devido à alta modularidade desses sistemas, as estratégias HBA e MBA atomizam as redes mais rapidamente do que os algoritmos HDA e CI. Por outro lado, apesar de todas as redes serem muito frágeis particularmente aos ataques HBA, quando a complexidade do tempo é levada em conta, o método modular MBA mostra níveis de desempenho muito mais elevados que os outros algoritmos. Os resultados indicam que o algoritmo

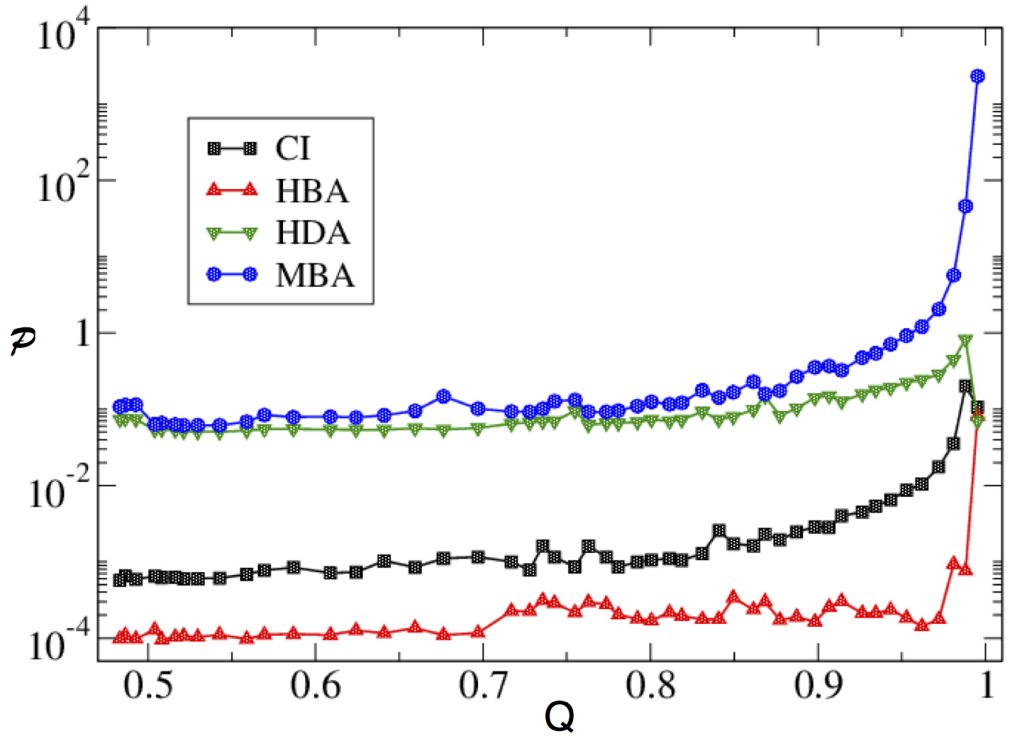

Figura 3.9: Aqui se plotam em escala semi-log as performances ($\mathcal{P}$) dos métodos HBA (triângulos vermelhos para cima), HDA (triângulos verdes para baixo), MBA (pontos azuis) e CI (quadrados pretos) em redes de referência LFR com tamanho $N = 10^4$ e modularidade variando entre $0,48 < Q < 0,99$. Uma bola com raio $l = 3$ foi utilizada nas simulações CI após um processo de otimização de raios.

HDA roda mais rápido do que o CI, o que significa que, nesta implementação o método CI está em execução mais lenta que o limite teórico $O(N \log N)$.

Nesta contribuição estudou-se a interação entre a fragilidade de redes modulares e o custo computacional das seguintes estratégias de ataque HDA, HBA, MBA e CI. Para tanto, introduziram-se uma medida de robustez generalizada e uma quantidade empírica chamada de performance que podem ser úteis como guias para escolher o método de ataque a redes complexas mais apropriado para cada caso real.

Resultados em redes artificiais de referência e sistemas reais com altos níveis de modularidade indicam que, se forem tidos conta tanto a robustez como a complexidade, o método MBA é a melhor escolha para vários casos. Esta conclusão é assegurada teoricamente quando a fração de pontes N_d entre comunidades é menor que $\log N$. Além disso, a robustez de redes reais altamente modulares frente ao ataque simultâneo MBA é muito semelhante à do método HBA, enquanto os algoritmos HDA e CI não parecem ser muito sensíveis a variações na modularidade. Em termos apenas da robustez, o método HBA supera os outros, porém à custa de uma complexidade muito maior, o que o torna o método inaplicável as grandes redes.

Em suma, esses resultados podem ter impactos importantes no planejamento de estratégias de ataque eficazes para desmantelar sistemas reais, tais como como doenças, epidemias ou redes sociais dirigidas a práticas criminosas.

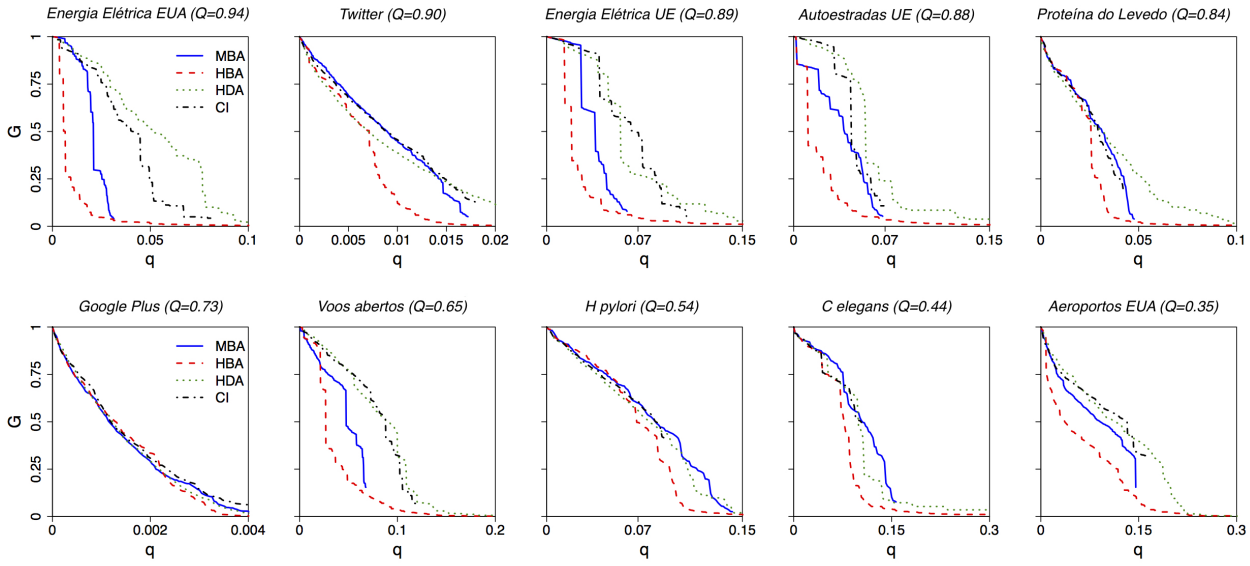

Figura 3.10: A figura mostra o processo de fragmentação, isto é, o tamanho da maior componente conectada (G) em função da fração de nós removidos (q), das redes de energia elétrica da União Europeia e dos Estados Unidos da América, o sistema de rodovias da União Europeia, o proteoma do levedo, extratos do Google Plus, voos abertos, proteoma do H *pylori* e do C *elegans* além dos aeroportos dos Estados Unidos da América sob os ataques HBA (linhas vermelhas pontilhadas), HDA (linhas verdes pontilhadas), MBA (linhas azuis) e CI (linhas tracejadas pretas). Uma bola com raio $l = 3$ foi utilizada para as simulações CI após um processo de otimização de raios.

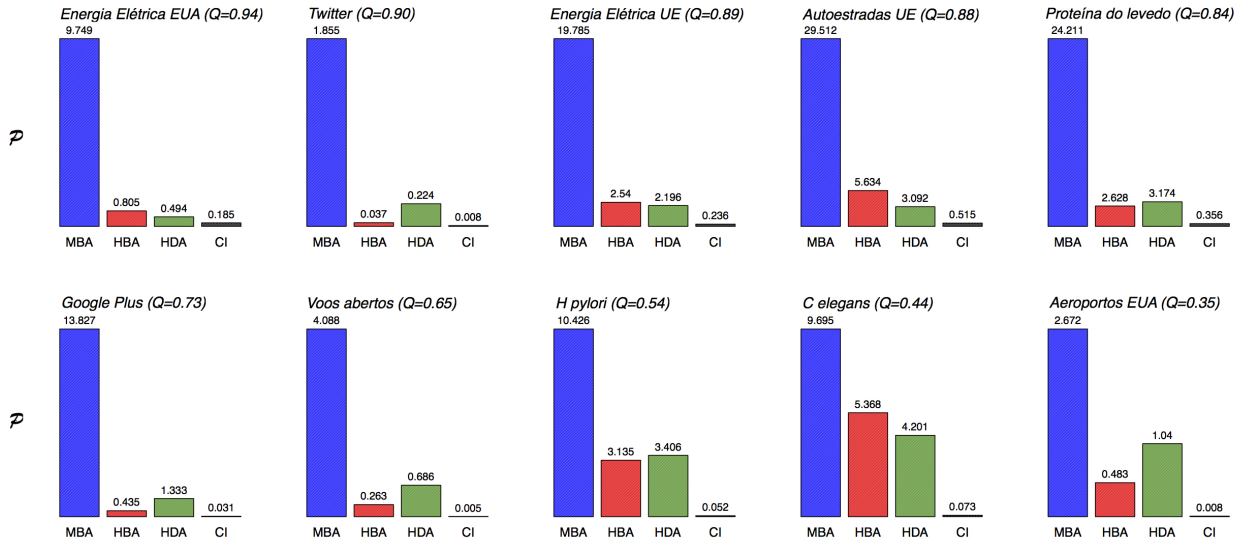

Figura 3.11: A figura retrata os histogramas da performance ($\mathcal{P}$) para cada attack (azul), HBA (vermelho), HDA (verde) e CI (preto) em cada rede real.

3.3 Redes criminais

Apesar do constante esforço das agências brasileiras de aplicação da lei em combater o crime organizado, o panorama social não vem mostrando mudanças significativas— as taxas de homicídio atingiram um pico em 2014 alcançando mais de 29 mortes para cada cem mil pessoas [128], escândalos de corrupção e lavagem de dinheiro alcançaram grandes empresas e importantes figuras políticas, além do mais o país se tornou recentemente um dos maiores consumidores mundiais de cocaína e um importante corredor para o narcotráfico [129]. O problema é multivariado, passando por questões culturais e históricas até o anacronismo do sistema de segurança pública [130]. Contudo, do ponto de vista da ciência de redes, outra razão importante é que as intervenções policiais tradicionais não costumam levar em conta a topologia das redes de relacionamentos criminais. Todavia, esse panorama vem mudando. Por exemplo, Agreste *et al.* estudaram recentemente a estrutura de rede da Mafia Siciliana (também conhecida como *Cosa Nostra*) [64]. Este trabalho resultou em uma rede bipartite (uma camada de contatos e outra camada de fato criminal) que se mostrou robusta a ataques— o subgrafo de contatos é muito mais frágil a ataques dirigidos que a camada criminal. Apesar disto, os autores não estudaram a modularidade da rede mafiosa e sua robustez a MBAs [56] ou a outros métodos mais eficientes de ataque [78]. Além disso, outros autores já estudaram corporações mafiosas e os resultados apontaram para a hierarquização da estrutura da rede com alguns poucos *capi* (chefões) comandando a atividade criminosa [131].

Uma característica comum das redes criminais é o equilíbrio entre o sigilo de suas atividades ilegais e a eficiência de suas conexões, o que está diretamente relacionado à sua topologia de rede [16]. Assim, a densidade de arestas e a eficiência topológica de uma rede geralmente estão associadas ao "brilho"do sistema no sentido de que um grande número de conexões entre criminosos significa que se um indivíduo for apanhado pela polícia seria possível, em princípio, extrair informação crítica sobre a estrutura da rede. Por outro lado, uma rede mais "escura"significa que a transferência direta de informação dentro do próprio sistema é diminuída devido ao número reduzido de caminhos entre os indivíduos. Com efeito, tanto a densidade da rede quanto sua eficiência revelam informações importantes sobre o balanço entre a segurança e a difusão efetiva de informação e dados. Essa característica difere o crime organizado de células terroristas no sentido de que OrCrims geralmente são mais eficientes devido à sua inerente motivação econômica em oposição à motivação ideológica do terrorismo. Consequentemente, OrCrims tendem a ser mais frágeis a ataques dirigidos [132]. Duijn *el al.* [55] apontaram recentemente, ao estudar uma rede relacionada ao narcotráfico, que as organizações criminosas podem se tornar mais eficientes em resposta a ataques dirigidos. O inconveniente positivo é a que a atuação policial diminui a segurança da rede, oferecendo oportunidades estratégicas para o planejamento de operações efetivas de fragmentação da rede. Esses e outros aspectos estruturais são importantes para se entender a natureza desse fenômeno. Além disso, a fragilidade característica de cada sistema real é de suma importância para se pensar maneiras concretas de aprimorar o combate ao crime. Nesse sentido, estudam-se a seguir duas importantes redes reais: a de crimes federais brasileiros e a de um fórum ilícito escondido na chamada *deep web*.

3.3.1 Crimes federais

Ao se ter acesso exclusivo e criptografado ao histórico de operações da Polícia Federal (PF)– entre abril de 2013 e agosto de 2013– foi possível se construir uma rede única de crimes federais. Conforme a Constituição Federal no seu Art. 144, $\S 1^{\circ}$, incs I e II a Polícia Federal destina-se a:

- Apurar infrações penais contra a ordem política e social.
- Apurar infrações penais praticadas em detrimentos de bens, serviços e interesses da União ou de suas entidades autárquicas e empresas públicas.
- Apurar outras infrações penais cuja prática tenha repercussão interestadual ou internacional e exija repressão uniforme, segundo se dispuser em lei [1].
  - Sequestro, cárcere privado e extorsão mediante sequestro.
  - Formação de cartel.
  - Relativas à violação de direitos humanos, que a República Federativa do Brasil se comprometeu a reprimir em decorrência de tratados internacionais de que seja parte.
  - Furto, roubo ou receptação de cargas, inclusive bens e valores, transportadas em operação interestadual ou internacional, quando houver indícios da atuação de quadrilha ou bando em mais de um Estado da Federação.
  - Falsificação, corrupção, adulteração ou alteração de produto destinado a fins terapêuticos ou medicinais e venda, inclusive pela internet, depósito ou distribuição do produto falsificado, corrompido, adulterado ou alterado.
  - Furto, roubo ou dano contra instituições financeiras, incluindo agências bancárias ou caixas eletrônicos, quando houver indícios da atuação de associação criminosa em mais de um Estado da Federação.
- Prevenir e reprimir o tráfico ilícito de entorpecentes e drogas afins.
- Prevenir e reprimir o contrabando e o descaminho, sem prejuízo da ação fazendária e de outros órgãos públicos nas respectivas áreas de competência.

Trata-se, na verdade, de atribuições em investigações que vão desde o crime organizado transnacional (Leis 9.034/05 e 12.850/13), a lavagem de dinheiro (Lei 9.613/98), o terrorismo (Lei 13.260/16), até aos crimes cibernéticos (Lei 8.069/90) entre outros. Para mais detalhes ver por exemplo a obra Crimes Federais de Fábio Araújo e Rogério Sanches [133]. Na rede construída, cada vértice corresponde a uma pessoa investigada e as arestas indicam que em algum momento da investigação ambos indivíduos se relacionaram de alguma maneira relevante para o respectivo Inquérito Policial. O sistema então consiste em $N = 23.666$ vértices e $E = 35.913$ arestas. Uma representação gráfica do sistema é apresentada na figura 3.12.

[1] Lei 10.446/02

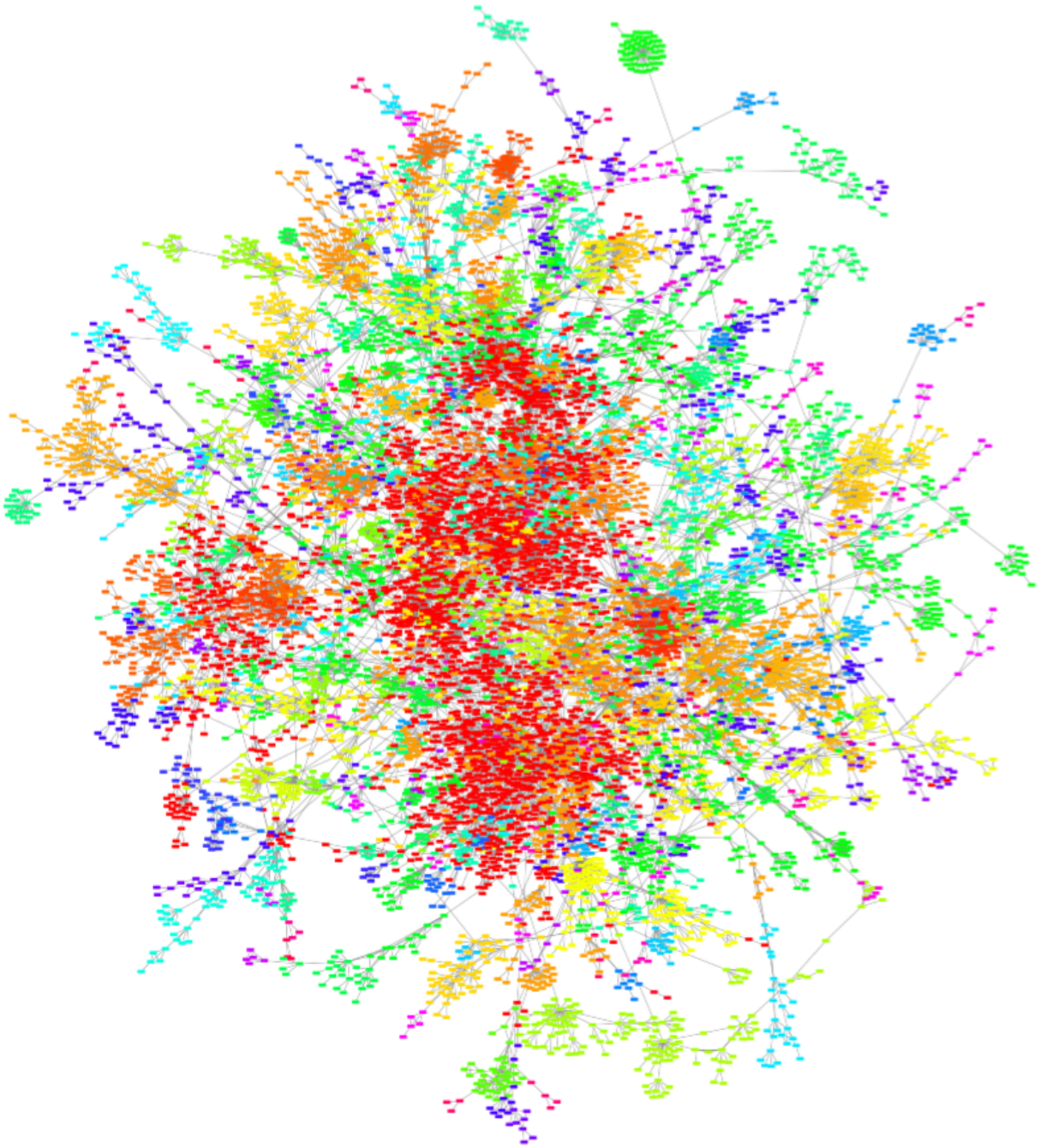

Figura 3.12: Representação da maior componente conectada da rede de crimes federais brasileiros consistindo em 9.887 indivíduos e 91 módulos. As cores representam vértices da mesma comunidade conforme extraídos pelo método de *Louvain*.

	δ	E_{ff}	N	E
Facebook (NIPS)	0,0071	$29,3\%$	2.888	2.981
Crime	0,0043	$21,5\%$	829	1.473
Hamsterster	0,0056	$20,8\%$	2.426	16.631
PF	0,0004	$8,4\%$	9.887	19.744

Tabela 3.3: Dados comparativos (densidade, eficiência, número de vértices e número de arestas) entre a rede de crimes federais e outros sistemas sociais: um subgrafo do Facebook [108, 111], um conjunto de dados criminais da Polícia de Saint Louis nos Estados Unidos da América por volta de 1990 [134], relacionamentos entre usuários do sítio hamsterster.com [135] e a própria rede de crimes federais brasileiros.

Estrutura

Neste caso, a rede resultante é simples (não-dirigida, não-ponderada e sem *loops* ou arestas múltiplas) e possui 3.425 componentes desconexas com um tamanho médio de apenas 7 indivíduos. Contudo, o coeficiente de complexidade da rede $\langle k^2\rangle/\langle k\rangle = 7,42$ é muito maior que o critério de Molloy-Reed. Isso significa que a rede está na verdade em um regime no qual apenas uma componente gigante permeia o sistema inteiro [6]. Este é o primeiro resultado importante desta seção, já que não se esperaria a existência de uma componente única reunindo estatisticamente crimes tão diversos quanto tráfico de drogas e crimes ambientais, por exemplo. Por conseguinte, a partir de agora foca-se apenas nesta maior componente conectada do sistema, uma vez que as componentes fragmentárias podem ser consideradas apenas como fenômenos criminais residuais, presentes em qualquer sociedade, enquanto que uma componente gigante representa uma fase de crime generalizado e auto-organizado que é, de fato, mais perigosa de um ponto de vista da segurança nacional. Assim, esta componente relevante aponta para 9.887 vértices e 19.744 arestas (40% do total de nós e 54% de todos os relacionamentos).

Como visto nos capítulos anteriores, a densidade e a eficiência da rede estão relacionados à segurança e à eficiência de comunicação dentro da rede. E este é precisamente o caso desta rede de crimes federais que é bem mais escura que redes sociais tradicionais, isto é, ela possui níveis de densidade de arestas pequenos e ao mesmo tempo apresenta pouca eficiência (ver tabela 3.3). O mapa tipo radar apresentado na figura 3.14 mostra as diferenças topológicas entre a rede federal e sua versão randomizada na qual todas as arestas são aleatoriamente reconectadas mantendo-se o grau médio e a densidade de arestas constantes. Os dados destacam o comportamento típico de redes de pequeno-mundo [6] já que o sistema possui um menor caminho médio reduzido se comparado com o diâmetro da rede, porém com valores significantes do coeficiente de agrupamento.

A distribuição de grau do grafo é de extrema importância para se revelar a natureza do sistema subjacente. Por exemplo, dentre outras implicações práticas, redes com distribuições homogêneas de grau, na qual a probabilidade $P(k)$ de que um vértice arbitrário tenha grau k cai exponencialmente com o grau k, apresentam uma transição de uma fase totalmente conectada se uma fração q_c é retirada aleatoriamente da rede [1]. Por outro lado, grafos heterogêneos são robustos a falhas aleatórias dos nós, mas fracos a ataques dirigidos a seus vértices mais centrais ou *hubs* [1]. Exemplos desse tipo de sistema incluem a Internet, a World Wide Web, e a grande maioria de redes sociais [6]. Nesses casos,

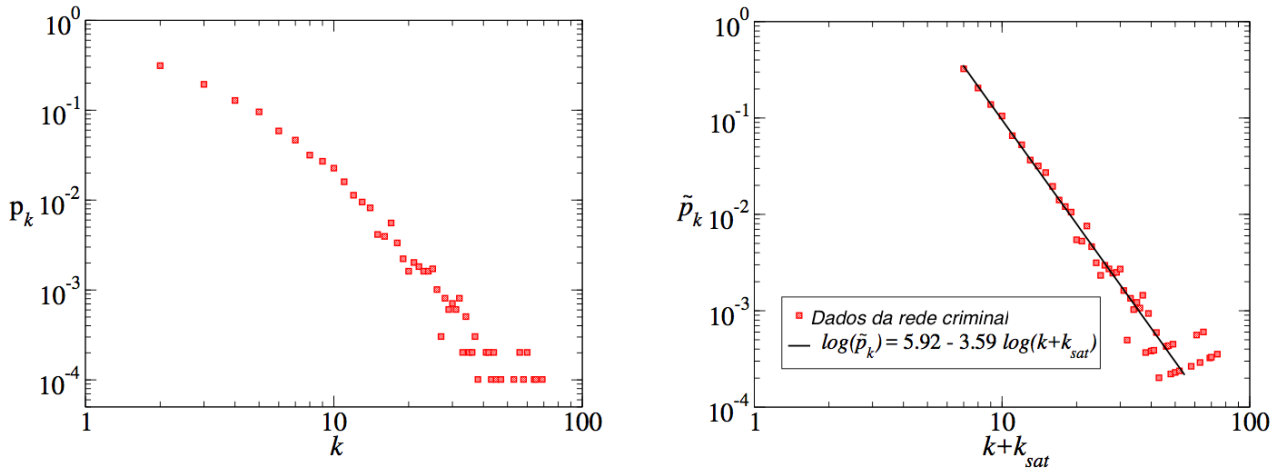

Figura 3.13: Distribuição de grau, p_k, para a rede de crimes federais em escala log-log. A figura mostra a característica típica de redes sociais com saturação de baixa conectividade e corte para alto grau (painel à esquerda). Já no painel à direita tem-se a distribuição de grau reescalonada p_k como função de $k + k_{sat}$ também em escala log-log e respectivo ajuste para lei de potência.

a distribuição de grau geralmente segue uma lei de potência em um regime típico de invariância em escala ($p(k) \propto k^{-\gamma}$, com $2 < \gamma < 3$) e revela um model generativo associado a ligação preferencial linear, ou seja, vértices mais novos tendem a se conectar aos mais populares em um mecanismo de rico-fica-mais-rico. Entretanto, sistemas reais raramente apresentam distribuições do tipo leis de potência puras [1]. Em geral, dois fenômenos competem: saturação de baixa conectividade e corte para alta conectividade [1]. Assim, o número de vértices com baixa conectividade é costumeiramente menor que o esperado para uma lei de potência pura devido a uma atração inicial de todos os vértices. Já o segundo comportamento indica uma queda abrupta em $p(k)$ para $k > k_{cut}$ devido à limitação inerente ao número de arestas de cada nó. Para redes sociais típicas, essa limitação está fortemente relacionada à limitação humana em manter mais que 150 laços fortes (uma característica conhecida como o número de Dunbar [136]). Contudo, no caso criminal, além dessa restrição cognitiva, o corte para alta conectividade também é devido à falta de confiança entre os criminosos que é necessária para esconder as atividades ilegais da rede, diminuindo o brilho dela como discutido anteriormente. Leis de potência generalizadas com saturação e corte são descritas pela seguinte equação (ver figura 3.13):

$$p(k) \propto (k + k_{sat})^{-\gamma} \exp(\frac{-k}{k_{cut}}). \tag{3.6}$$

A assortatividade (r) é outro aspecto importante a se estudar. Em redes assortativas ($r > 0$), os vértices apresentam a tendência de se ligar a outros com um grau similar, enquanto que em redes desassortativas ($r < 0$), nós com alto grau tem um viés de se conectar a outros com baixo grau. Para o sistema de crimes federais tem-se $r = 0,02$. Essa é um fenômeno bem conhecido em redes sociais [47, 48], isto é, pessoas muito prósperas preferem se relacionar com outras do mesmo estrato social. Em relacionamentos empresariais empreendedores preferem colaborar com outros nomes consolidados em busca por sucesso, reputação, influência e *status* social. Aparentemente, a mesma

regra vale para redes de crime organizado, que de certa forma são casos particulares de redes de negócios. Todavia, o valor de A é próximo de zero, o que revela que uma natureza neutra. Nessas situações, nós com grau maior que um corte estrutural apresentam desassortatividade estrutural [137]. Em outras palavras, não há arestas o suficiente entre os *hubs* para manter a natureza neutra do sistema, então para alto k, há uma forte queda em $p(k)$. Contudo, esse não é o caso da rede de crimes federais já que o grau máximo ($k_{max} = 68$) é menor que o corte estrutural para grafos simples ($k_s \sim (\langle k \rangle N)^{1/2} = 185,47$). De modo que o comportamento próximo da neutralidade está intimamente ligado ao problema da confiança entre criminosos como abordado nas linhas anteriores.

Vulnerabilidades e controle

Do ponto de vista da ciência em rede, um grafo pode ser impedido de funcionar como um todo removendo-se seus nós ou removendo-se apenas suas arestas (mantendo os nós). Nesse sentido, as operações policiais geralmente visam a identificar e prender criminosos. Portanto, a prisão de indivíduos está diretamente relacionada à remoção de arestas, uma vez que os nós não são de fato excluídos da rede. Por outro lado, a supressão de nós significa a remoção completa desses indivíduos da rede criminosa– um cenário que só ocorreria por morte ou por re-socialização e não diretamente por ações policiais. Numa perspectiva topológica, a remoção de nós é mais eficaz na atomização de redes complexas causando mais danos por eliminação do que a remoção de bordas uma vez que a retirada de um único nó da rede resulta na eliminação de todos os links ligados a ele [8, 75].

Este é um segundo resultado importante com relevantes implicações sociológicas, isto é, a partir de uma perspectiva de ciência em redes, a re-socialização (por exemplo pela educação ou pelo trabalho) é em geral uma estratégia mais eficaz para se reduzirem os níveis de criminalidade do que a prisão. Ainda assim, deve-se notar que, de acordo com este raciocínio, mesmo que repreensível ética e legalmente, a morte dos indivíduos-chave iria alcançar resultados semelhantes, *ceteris paribus*.

Simulam-se, a seguir, as rupturas de borda e nó na componente gigante da rede de crimes federais considerando dois procedimentos diferentes: ataques de alta centralidade, quando uma fração de nós ou bordas é excluída simultaneamente de acordo com uma lista previamente ordenada por um índice de centralidade escolhido (o que não é exclusivo e mede a importância estrutural de nós e bordas para manter a rede coesa) e ataques de alta centralidade adaptativos quando atacamos componentes individuais da rede de acordo com uma lista ordenada iterativamente por um índice de centralidade e atualizada após cada remoção [1].

Para testar a fragilidade estrutural da rede, interrompemo-la por meio dos seguintes ataques baseados em vértices: High Degree Adaptive (HDA), High Betweenness Adaptive (HBA), High Degree (HD), High Betweenness (HB) e Módulo (MBA) (HBA), High Betweenness (HB), ataques baseados em Módulos (MBA) e Influência Coletiva (IC) (ver figura 3.15). A centralidade de grau é apenas o número de conexões que um nó tem e a centralidade de intermediação basicamente mede a fração de caminhos mais curtos que passam por um determinado vértice [8]. A influência coletiva de

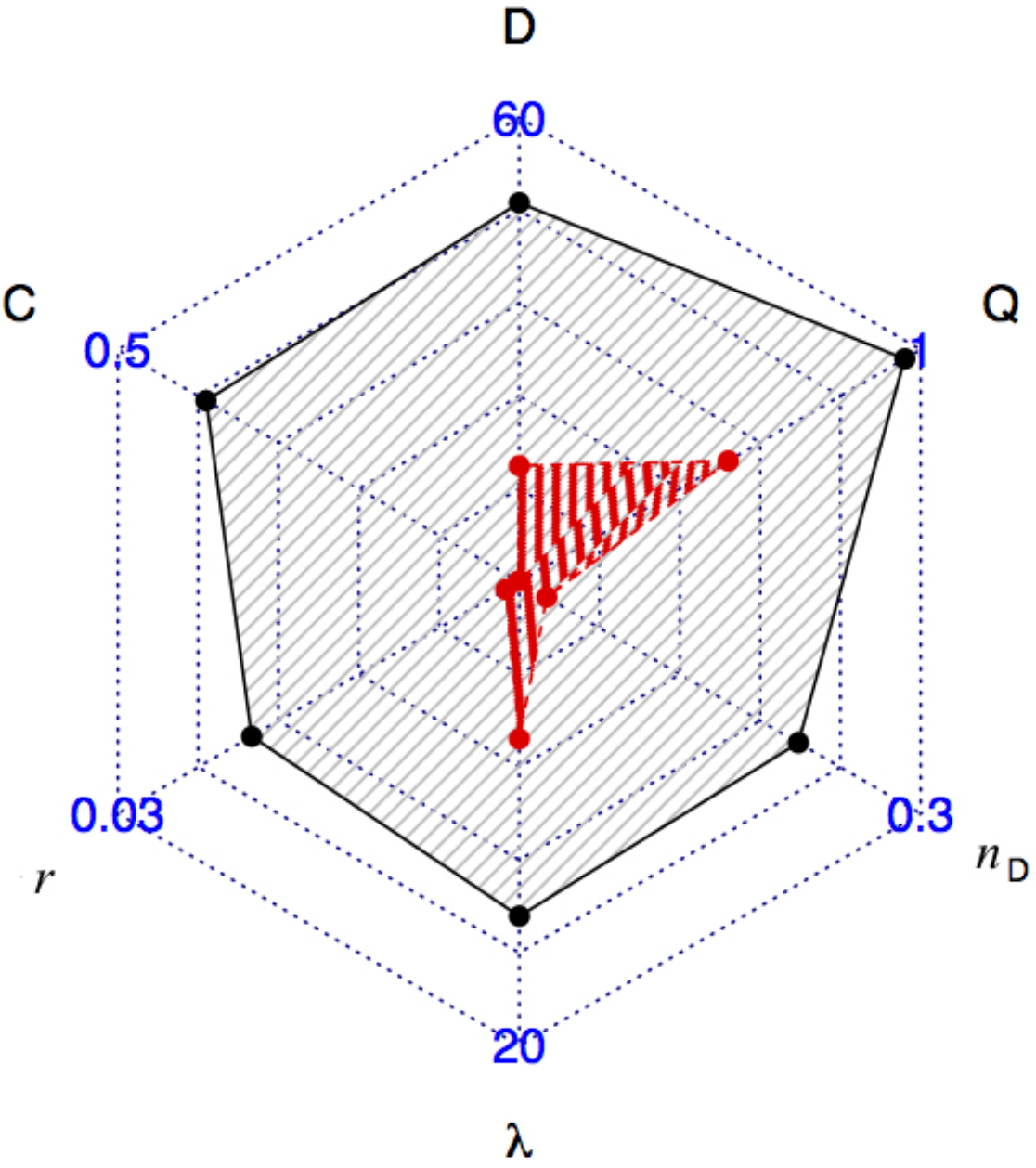

Figura 3.14: O mapa tipo radar apresenta os seguintes parâmetros para a rede da PF (padrão em cinza) e sua contraparte randomizada mantendo-se o grau médio e a densidade de arestas constantes (padrão em vermelho): diâmetro ($D = 49$ e 15), modularidade ($Q = 0,96$ e $0,52$), fração de controladores ($n_d = 0,21$ e $0,02$), comprimento médio de caminho mais curto ($\lambda = 14,43$ e $6,78$), assortatividade ($r = 0,017$ e $0,001$), coeficiente de agrupamento ($C = 0,391$ e $0,001$).

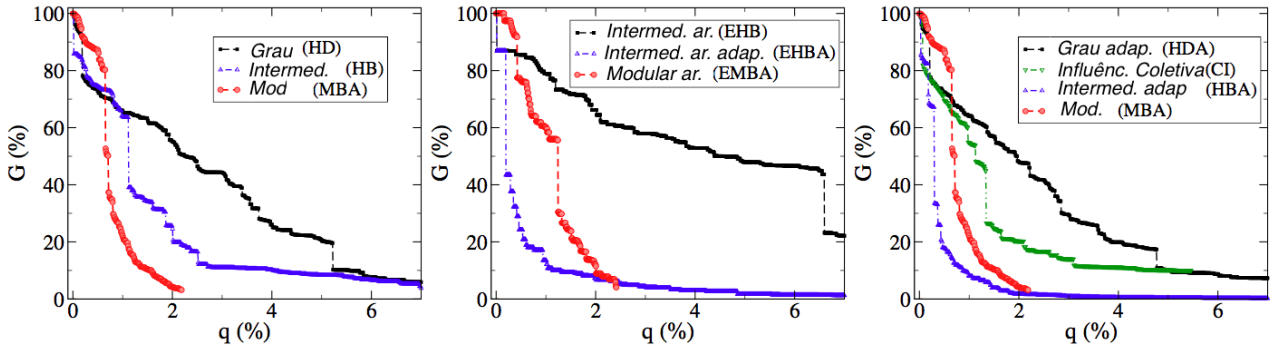

Figura 3.15: As figuras mostram as curvas de fragmentação da rede de crimes federais a partir do tamanho relativo da maior componente conectada G em função da fração de nós/arestas removidos q de acordo com os seguintes procedimentos: (esquerda), remoção de vértices por grau (HD - quadrados pretos), por intermediação (HB - triângulos azuis) e por módulos (MBA - círculos vermelhos); (centro), remoção de arestas por intermediação (HB - quadrados pretos), por intermediação adaptativa (HBA - triângulos azuis) e por módulos (MBA - círculos vermelhos); (direita), remoção de vértices por grau adaptativo (HDA - quadrados pretos), por influência coletiva (CI - triângulos verdes para baixo), por intermediação adaptativa (HBA - triângulos azuis) e por módulos (MBA - círculos vermelhos).

um nó leva em conta o grau de seus vizinhos a uma determinada distância l da seguinte maneira:

$$CI_k(i) = (k_i - 1) \sum_{j \in \partial Ball(i,l)} (k_j - 1) \tag{3.7}$$

onde k_i é o grau do vértice e $\partial Ball(i,l)$ é o conjunto de todos os nós a uma distância l do vértice i. Conforme estudos anteriores este método é muito próximo do conjunto mínimo de desmantelamento de redes [78]. O ataque baseado em módulos está relacionado com a natureza modular das redes reais, o fenômeno de que redes complexas tendem a se agrupar em aglomerados densamente conectados internamente, mas apenas fracamente conectados entre eles. A densidade de ligações que conectam comunidades diferentes quando comparadas com a densidade interna de bordas é geralmente medida pela modularidade da rede, Q que varia de -1 a 1, e depende ligeiramente do algoritmo de extracção de comunidade utilizado [122]. Foi mostrado recentemente [56] que as redes altamente modulares são frágeis a MBA. Nesse sentido, seria de se esperar que o crime organizado mostrasse características altamente modulares, uma vez que a fraca conexão entre as comunidades favoreceria a obscuridade da rede, enquanto a alta densidade interna das comunidades seria um cenário adequado para administrar negócios internamente de maneira eficiente. De fato, a rede PF tem uma modularidade muito alta usando os métodos *Louvain* [59] ($Q = 0,96$) e usando o Infomap [95] ($Q = 0,88$).

Para quantificar os efeitos de cada estratégia de interrupção na rede, medimos o tamanho da maior componente conectada em relação ao tamanho original da rede, $G(q)$, em função da fração de objetos removidos, q. Como apontado em [138] a robustez generalizada de uma rede para uma determinada estratégia de ataque é dada pela métrica:

$$R = \frac{1}{N(1 - G_{min})} \sum_{q=0}^{q_{max}} G(q) \tag{3.8}$$

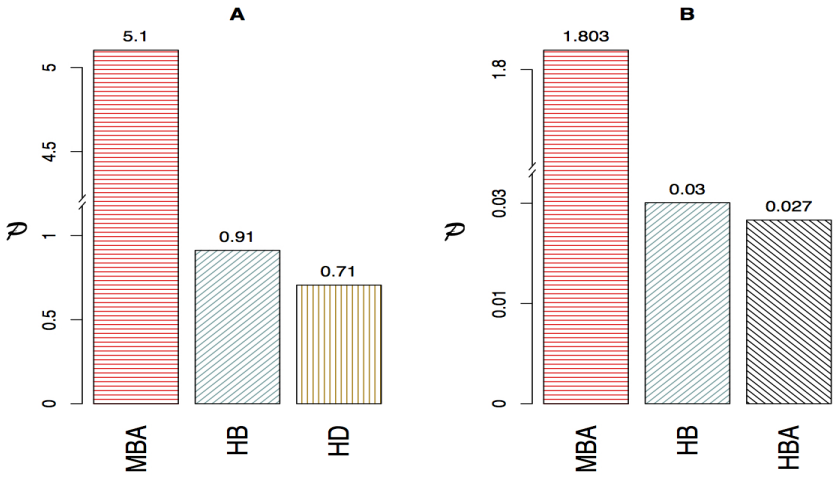

Figura 3.16: Os histogramas mostram a performance dos três melhores ataques sobre a rede de crimes federais. O painel (A) mostra os ataques por vértices MBA (sombreado horizontal vermelho), HB (sombreado azul inclinado) e HD (sombreado dourado), enquanto o painel (B) mostra MBA, HB e HBA (sombreado preto inclinado) para ataque a arestas.

onde N é o número de nós na rede, q_{max} é o ponto em que termina o ataque e G_{min} é o valor do tamanho relativo do maior componente conectada em q_{max}. No entanto, para avaliar a eficácia de cada estratégia, é importante medir o trade-off entre a robustez (R) e o tempo (t) necessário para calcular a lista de ataques. Nesse sentido, o desempenho de um ataque é medido pela relação $\mathcal{P} = t^{-1} \times R^{-1}$ onde t é o tempo necessário para completar o procedimento em segundos e R é a robustez.

De acordo com estas considerações, a estratégia de ataque com maior desempenho (ver figura 3.16) é o MBA tanto para ataques por nós quanto por bordas, como esperado devido à alta modularidade da rede. No entanto, a rede é um pouco menos robusta para HBA, que por sua vez leva muito mais tempo de cálculo. Em outras palavras, a rede seria completamente atomizada depois da remoção de aproximadamente 2% de seus vértices e quase 5% de suas bordas por HBA. Além disso, o ponto de desativação em que todas as comunidades são destacadas do núcleo do grafo original é atingido pela prescrição MBA quando quase 2% de suas arestas ou nós são removidos. Esses resultados significam que, embora os ataques por nós sejam em geral mais eficientes que os ataques por bordas, particularmente nessa rede, ambas as estratégias são muito semelhantes– por exemplo, o MBA de borda tem maior desempenho e robustez semelhante ao HBA de nó. Esse é outro resultado importante, já que a rede se fragmentaria completamente por ações de aplicação da lei tradicionais (ataques aleatórios) após a falha aleatória de 80% dos nós e 86% das arestas. Outra característica importante é que o sistema é muito mais fraco frente aos ataques HBA e MBA do que frente à estratégia CI, conforme ilustrado na figura 3.15. No seu trabalho seminal Liu *et al* mostraram que redes sociais geralmente apresentam valores relativamente baixos para a fração de controladores. Os resultados da rede federal de crimes vão ao encontro desse trabalho pretérito, mostrando que com acesso a aproximadamente 20% ($n_D = 0.21$) dos vértices poderíamos controlar toda a rede criminal.

3.3.2 Darknet

Em 2014 e 2016, a Polícia Federal no Rio Grande do Sul deflagrou as fases I e II da Operação Darknet [139, 140]. Tratou-se da maior operação da história do Brasil de combate à exploração sexual de crianças e adolescentes. Conforme a Comunicação Social da Polícia Federal, policiais monitoraram a atividade de usuários brasileiros em um fórum de pedofilia escondido em uma camada obscura da internet conhecida como deep web– trata-se na verdade de todo conteúdo não-indexado da World Wide Web inacessível por motores de busca comuns, é dita ser muitas ordens de magnitude maior que a internet tradicional (também conhecida como surface). Mais precisamente, a operação teve enfoque em usuários do Projeto Tor, que é composto de uma rede de túneis dinâmicos virtuais com comunicação criptografada que permite organizações e indivíduos compartilharem informações construídas sobre redes públicas sem comprometer sua privacidade [141]. A investigação durou aproximadamente dois anos e resultou em mais de 100GB de informações destinadas a pesquisas científicas dirigidas a novas soluções para o combate a crimes cibernéticos. Neste período foram identificados 182 indivíduos que responderam pelos crimes previstos na Lei 8.069/1990 (Estatuto da Criança e do Adolescente), *in verbis*:

- Art. 241-A. Oferecer, trocar, disponibilizar, transmitir, distribuir, publicar ou divulgar por qualquer meio, inclusive por meio de sistema de informática ou telemático, fotografia, vídeo ou outro registro que contenha cena de sexo explícito ou pornográfica envolvendo criança ou adolescente.

- Art. 241-B. Adquirir, possuir ou armazenar, por qualquer meio, fotografia, vídeo ou outra forma de registro que contenha cena de sexo explícito ou pornográfica envolvendo criança ou adolescente.

Também houve o salvamento de 6 crianças que estavam em situação de fragilidade. Nesses casos os responsáveis também responderam pelo crime previsto no Art. 217-A do Decreto-Lei 2.848/40 (Código Penal), qual seja:

- Art. 217-A. Ter conjunção carnal ou praticar outro ato libidinoso com menor de 14 (catorze) anos. Nesses casos, as penas somadas podem chegar a 25 anos de reclusão.

A rede foi construída diretamente a partir do fórum online investigado [2].A interação mais importante era aquela por interesse, ou seja, pela visualização dos tópicos. Assim, havia os usuários que apenas observavam as postagens (leachers) e aqueles que de fato produziam o material ilícito (putters). Obteve-se assim uma rede dirigida com $N = 10.407$ vértices e $E = 842.247$ arestas. Uma representação gráfica da rede criminal estudada é apresentada na figura 3.17.

[2]Um site online de discussões onde os usuários podem interagir na forma de mensagens postadas e da troca de arquivos em conversas conhecidas como tópicos.

Figura 3.17: Representação da rede criminal da deep web. Vértices vermelhos e maiores possuem valores mais altos de grau interno. Para facilitar a visualização as arestas foram omitidas e dos vértices mostram-se apenas os 25% mais conectados.

Estrutura

A rede de usuários da deep web foi construída seguindo-se as seguintes regras:

- Se um usuário, $E1$, visualizar um tópico postado por outro usuário, $E2$, cria-se uma aresta dirigida $E1 \rightarrow E2$.
- Se o mesmo usuário visualizar múltiplas vezes o mesmo tópico, a aresta é ponderada pelo número de visualizações.

O sistema resultante consiste em um grafo conectado com 10.407 vértices e 842.247 arestas dirigidas, com um grau médio de $\langle k \rangle = 161,86$ e uma dispersão $d = \sigma^2/\langle k \rangle = 1.369,12$, indicando que a distribuição de grau da rede apresenta um comportamento típico de distribuições negativas com maiores intervalos de contagens altas e baixas que uma distribuição de Poisson. Além disso, redes dirigidas apresentam grau interno (k_{in}) e grau externo (k_{out}), com cada aresta dirigida apontando de uma fonte para um alvo. Nesse sentido, do número total de vértices, apenas 769 apresentam grau-interno diferente de zero, enquanto que 10.404 nós apresentam grau-externo não-nulo, resultando em $\langle k_{in} \rangle = 1.095,25$ e $\langle k_{out} \rangle = 80,95$.

Nas figuras 3.19 (a), (b) e (c) mostram-se as distribuições cumulativas para o grau total, para o grau-interno e para o grau-externo. A distribuição de grau não é ajustável perfeitamente por uma lei de

potência, contudo a última parte da distribuição apresenta um comportamento típico de invariância em escala com expoente $\alpha = 2,1$. Nesse sentido, há uma forte presença de *hubs* em ambos os grupos de *leachers* e *putters*. Todavia, o grau-externo médio de *leachers* que se conectam a poucos *putters* ($7,39\%$ do total de vértices) resulta em maiores valores médios de grau-interno e uma distribuição menos heterogênea. Há, então, dois fenômenos típicos de redes sociais competindo neste sistema. Primeiramente, a maioria dos *leachers* visualiza as postagens de apenas alguns *putters*, enquanto que um pequeno número de *leachers* tende a visualizar todo o conteúdo do fórum. Correspondentemente, dos já poucos *putters*, apenas alguns *hubs* costumam publicar tópicos frequentemente visualizados por uma grande quantidade de *leachers*.

Conforme mencionado anteriormente, a densidade de um grafo é definida como o número relativo de arestas se comparado com a quantidade possível de conexões da seguinte maneira:

$$\delta = \frac{2E}{N(N-1)}, \tag{3.9}$$

onde E é o número total de arestas e N é o tamanho da rede. A relação da densidade de uma rede criminal com o seu brilho, ou a segurança dos dados que trafegam dentro dela, já foi tratada no início deste capítulo. Nesse sentido, espera-se que as redes clandestinas, que operam escondidas no tecido social, como é o caso de redes ilegais criminais, tenham níveis de densidade muito baixos. Este é precisamente o caso da rede estudada aqui, que apresenta uma densidade de apenas $7,7 \times 10^{-3}$.

Além disso, um dos resultados mais importantes é que a rede estudada apresenta mistura desassortativa de conexões- *hubs* buscam conexões com indivíduos pouco conectados e vice-versa. Do ponto de vista sociológico e psicológico, na vida real o homem tem uma tendência natural a buscar relacionamentos com pessoas destacadas dentro de seu espectro relacional (seja ele econômico, social, profissional etc). Por outro lado, essas “elites"preferem se relacionar com pessoas do mesmo nível social, o que leva comumente a padrões assortativos em redes sociais do mundo real. Nas colaborações negociais, por exemplo, nomes já consolidados no mercado preferem colaborar com outros grandes nomes em busca de sucesso, reputação, influência e status. Conforme indicado por Holme *et al* [142], misturas assortativas parecem ser mais significantes apenas em ambientes dominados por estruturas competitivas. Outra origem da assortatividade de grau em colaborações profissionais é que os colaboradores são insubstituíveis, o que é geralmente decidido por interesses similares e afiliação por afinidade de grupos comuns. Ora, nestes aspectos redes sociais *online* diferem de redes sociais do mundo real. No mundo virtual, quando se cria um avatar que esconde a real identidade do usuário, os *hubs* não se recusam a se conectar mesmo com usuários de pouco relevo, até porque entendem que quanto mais conexões, maior será sua popularidade ou visibilidade– daí a desassortatividade. Relacionamentos na vida real precisam ser mantido e cultivados e isso requer esforço contínuo e custos sociais, o que não ocorre necessariamente em redes *online* onde os indivíduos não se conhecem de fato [48].

Outra característica marcante das redes reais é que elas tendem a se organizar em estruturas modulares ou comunidades, isto é, aglomerados densamente conectados internamente, mas escassamente ligados entre eles. Por conseguinte, seria de se esperar que o crime organizado se construísse em redes altamente modulares de forma a gerir eficientemente a empreitada criminosa– a comunicação seria

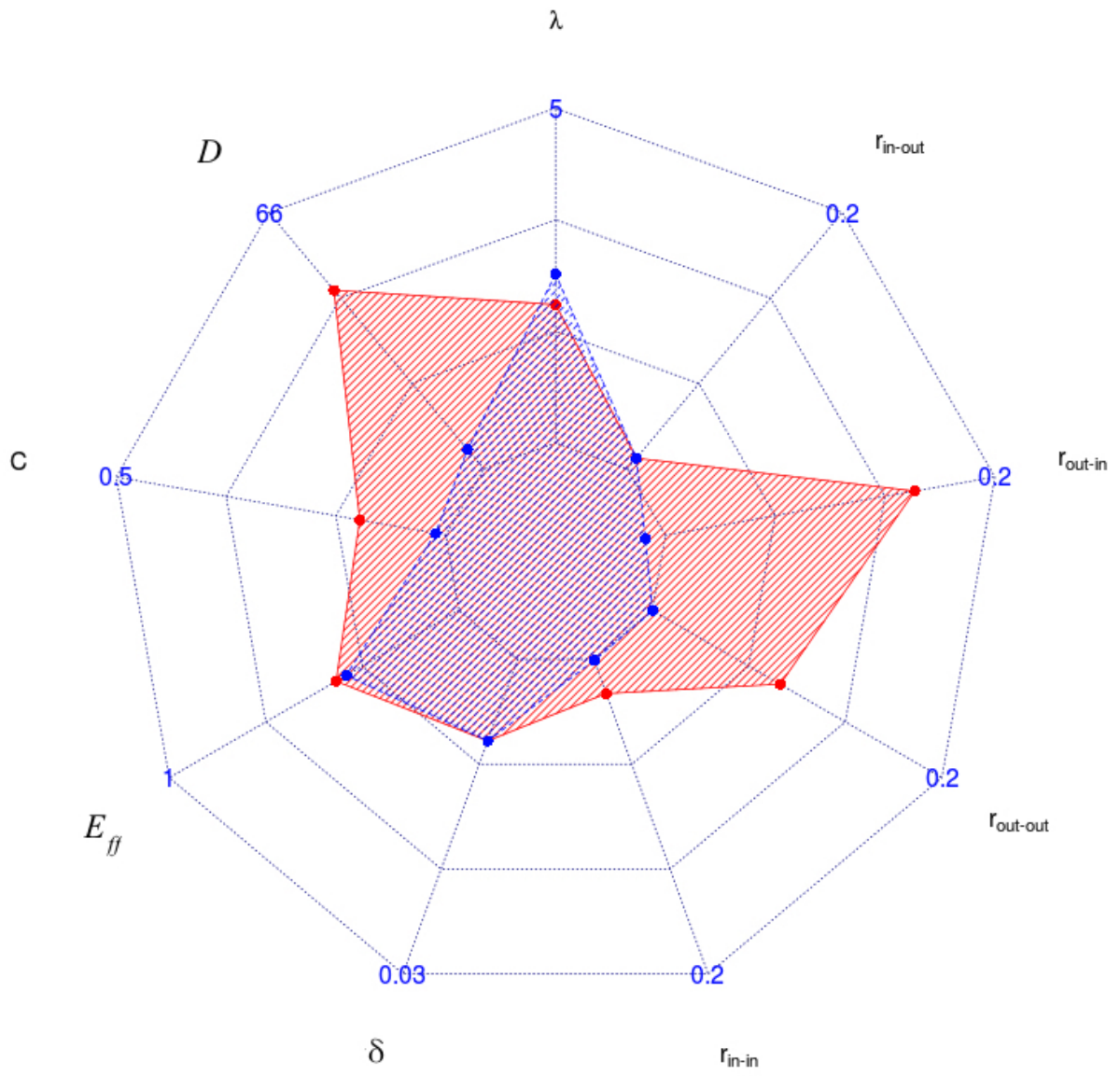

Figura 3.18: Neste mapa são mostradas características topológicas da rede da deep web (em vermelho) e sua versão aleatorizada (em azul). A figura contém o diâmetro da rede ($D = 46$ e 5), o comprimento médio de caminho mais curto ($\lambda = 2,07$ e $2,52$), densidade da rede ($\delta = 0,0078$ e $0,0078$), eficiência da rede ($E_{ff} = 0,42$ e $0,39$), coeficiente de agrupamento ($C = 0,130$ e $0,016$) e o valor em módulo das correlações de grau ($r_{in-out} = -0,08$ e $-0,02$, $r_{out-in} = -0,15$ e $0,012$, $r_{out-out} = -0,0884$ e $-0,0004$, $r_{in-in} = -0,0212$ e $-0,0007$). O valor máximo para cada quantidade é mostrado na borda da figura e cada raio equivale a um terço desse valor.

eficiente dentro de uma mesma corporação criminosa ao mesmo tempo que a ligação fraca entre grupos distintos privilegiaria a obscuridade dos negócios, como é o caso da rede de crimes federais. Todavia, esse não é o caso desta rede de pedofilia, que não mostra motivação econômica nem atividades hierárquica ou segmentação. Este aspecto é demonstrado pela modularidade da rede [122], $Q \approx 0$ quando as comunidades são extraídas pelo método Infomap [95], que foi escolhido por sua precisão e por levar em conta bordas dirigidas e ponderadas [113]. Assim, esta rede pode ser encarada como composta de apenas uma comunidade à qual pertencem todos os vértices, não havendo compartimentação, o que torna a rede mais exposta à repressão policial. Com efeito, a apreensão de um indivíduo poderia resultar numa compreensão completa do sistema. Em parte, também esse comportamento se deva à motivação da rede, que não envolve, em princípio, interesses econômicos ou competitivos como é o caso do crime organizado comum. A motivação esta mais relacionada a uma parafilia, uma obsessão ou uma compulsão, sendo classificada como uma distorção comportamental na CID-10, classe F65 [143].

Na figura 3.18 encontram-se vários aspectos topológicos da rede e de sua contraparte randomizada [3]. A figura inclui o diâmetro da rede (D), o comprimento médio de percurso mais curto (λ), as correlações de grau ($r_{in-out}, r_{in-in}, r_{out-in}, r_{out-out}$), a eficiência da rede (E_{ff}) e o coeficiente de agrupamento (C). Os dados mostram que o diâmetro da rede da deep web é muito maior que o da sua versão randomizada, apesar de o comprimento médio de percurso mais curto se manter pequeno, o que significa que a maioria dos usuários pode ser alcançada pela interação com apenas outros 2 usuários. Além disso, para a rede randomizada tem-se $\lambda/D = 0,50$ enquanto que para a rede original tem-se $\lambda/D = 0,045$, o que deixa mais clara a natureza heterogênea do sistema estudado. O coeficiente de agrupamento é também mais alto na rede original, contudo ainda é pequeno se comparado com redes sociais típicas. Esses traços, junto com o fato de que $\lambda \sim \log(N)$ mostram que a rede estudada apresenta claramente um comportamento de pequeno-mundo. As correlações de grau dizem respeito à controlabilidade da rede e serão discutidas mais adiante. Por outro lado, a eficiência da rede é definida por [55]:

$$E_{ff} = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} \frac{1}{d_{ij}}, \tag{3.10}$$

onde N é o número de vértices e d_{ij} é a distância entre os nós i e j. A eficiência da rede é de 42%, refletindo o compromisso entre compartimentação dos dados e a segurança interna da rede.

Vulnerabilidades e controle

As operações policiais destinadas a interromper atividades ilegais visam a identificar e prender criminosos. Na linguagem da ciência de redes isto é equivalente a dizer que as arestas correspondentes ao conjunto de criminosos identificados são removidas da rede. No entanto, os recursos de aplicação da lei são limitados e as investigações geralmente enfrentam múltiplas restrições legais. Portanto, é crucial conhecer o conjunto mínimo de nós que se removido resultaria no maior impacto na rede.

Basicamente, podemos romper uma rede removendo nós ou bordas. A remoção de bordas geralmente causa menos dano por objeto removido, uma vez que a eliminação de um único nó resulta no

[3] Novamente mantém-se o grau médio e a densidade de arestas constantes

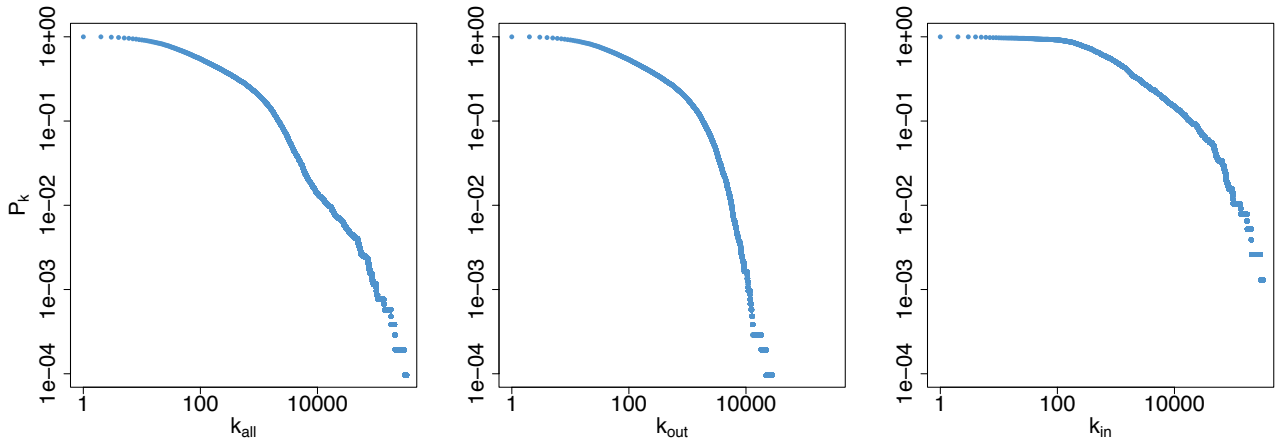

Figura 3.19: Distribuições cumulativas de grau total (a), In (b) e out (c). O gráfico é apresentado em escala log-log. O ajuste para uma lei de potência no caso da distribuição de grau total resulta em um expoente de $\gamma_{all} = 2,10$, para grau-out em $\gamma_{out} = 2.08$, e para grau-in em $\gamma_{in} = 12.74$ com um valor-p do teste de Kolmogorov-Smirnov de $0,14$, $0,16$ e $0,12$ respectivamente.

desprendimento de todas as arestas que o ligam [8, 10]. No entanto, dependendo do sistema real sob investigação ambas as abordagens podem ser aceitáveis, mas para outras situações, uma das abordagens pode não fazer sentido. Para redes criminosas, a maneira mais precisa de se simular a prisão de um criminoso é pela remoção das bordas correspondentes do vértice alvo, enquanto a remoção do nó propriamente dito é melhor interpretada como a morte ou a ressocialização completa do indivíduo. No entanto, o sistema real *in casu* consiste em um fórum *online* clandestino e, consequentemente, a prisão de uma determinada pessoa no mundo real, de fato, resultaria na remoção do usuário virtual correspondente do fórum. Portanto, apenas considera-se a remoção de nós de agora em diante.

As estratégias de rompimento de redes são, geralmente, divididas em duas abordagens: ataques simultâneos (ou não-adaptativos) e ataques seqüenciais (ou adaptativos). Na abordagem simultânea a lista de alvos é computada apenas uma vez, antes que o processo de interrupção comece. Na abordagem seqüencial, a lista de alvos é atualizada após cada exclusão pelo recálculo do índice de centralidade usado para classificar os nós. Consequentemente, os ataques sequenciais demandam mais tempo de processamento, mas por outro lado o método geralmente produz mais danos do que os ataques simultâneos, já que o método simultâneo não leva em conta as mudanças na centralidade da rede devido à remoção de elementos.

Neste sentido, como apontado em trabalhos anteriores, a heterogeneidade da distribuição de grau é uma de diversas características fundamentais ao se estudar a robustez da rede a estratégias de desestabilização ou de ataque [74, 76]. Por exemplo, as redes com distribuições de graus aleatórios desmoronam após a falha de um número crítico de nós. Por outro lado, as redes invariantes em escala são muito frágeis a ataques dirigidos aos nós ou aos núcleos centrais, e a atomização completa é atingida depois que uma pequena fracção dos vértices é removida do sistema [43]. No entanto, as redes modulares são particularmente sensíveis à extração de nós e arestas interligados (ou pontes) e é

possível ir de um estado totalmente conectado a uma fase desativada onde grandes peças densas da rede (módulos ou comunidades) estão desconectadas [56, 85, 144].

Estuda-se agora a fragilidade topológica do anel criminal frente aos ataques HDA_{all}, HDA_{in}, HDA_{out}, HD_{all}, HD_{in}, HD_{out}, HBA, HB, MBA. Como conceitualmente esperado, o ataque a *hubs* (*putters*) é altamente eficaz, pois eles são os poucos vértices mais responsáveis em manter a rede funcionando como um todo, produzindo a materialidade delitiva. Como pode ser visto, a rede mostra uma transição para uma fase desconectada quando da remoção de apenas $7,2\%$ dos nós e 40% das arestas de acordo com, respectivamente, os ataques HD_{in}, HDA_{in}, HDA_{out}, HDA_{all}. Devido à natureza dirigida do sistema, os ataques baseados na centralidade de intermediação não são tão eficazes, uma vez que a maioria dos caminhos são dirigidos de *leachers* para *putters*, e alguns vértices da rede não são acessíveis. Além disso, como a rede não é modular, os ataques baseados em módulos são ineficazes— ou seja, a rede não tem comunidades bem definidas e separá-la em partições modulares não fornece informações relevantes sobre o sistema.

A robustez generalizada de uma rede frente a uma dada estratégia de interrupção geralmente considera o tamanho relativo da maior componente conectada da rede durante o procedimento de ataque. Utiliza-se, então, a seguinte métrica de robustez:

$$R = \frac{1}{N(1 - G_{min})} \sum_{q=0}^{q_{max}} G(q) \tag{3.11}$$

onde G é o tamanho da maior componente conectada em relação ao tamanho original da rede, N é o número de nós, q é a fração dos nós removidos, q_{max} é o ponto em que o ataque termina e G_{min} é o valor do tamanho relativo da maior componente conectada em q_{max}. No entanto, também é importante medir o *trade-off* entre a robustez da rede e o tempo de computação necessário para se completar o ataque. Portanto, mede-se o desempenho de um ataque pela relação:

$$\mathcal{P} = \frac{1}{t \times R} \tag{3.12}$$

onde t é o tempo necessário para se completar o procedimento e R é a robustez. De acordo com estas definições, as melhores estratégias para se atomizar a rede seria classificar os alvos pelo método por grau simultâneo (HD), por intermediação (HB) e por grau sequencial (HDA)— esses resultados são apresentados no histograma da figura 3.20. De qualquer maneira, seria possível, em princípio, fragmentar completamente a rede removendo-se apenas 6% dos nós de acordo com a estratégia HDA (ver figura 3.21).

É possível agora comparar a fragmentação teórica baseada em critérios topológicos com o resultado efetivo alcançado pela Polícia Federal durante a Operação Darknet (ver figura 3.21). As consequências da investigação foram 182 alvos identificados responsáveis pela partilha de mídia ilícita utilizando fóruns do projeto Tor. Essa quantidade de criminosos corresponde a $q = 1,75\%$ da rede completa. A remoção de exatamente os mesmos atores identificados por Policiais Federais resultou em $3,33\%$ de fragmentação da rede original, enquanto que a remoção da mesma quantidade de vértices de acordo com um ataque de grau simultâneo resultaria em $12,73\%$, aproximadamente 4 $(3,82)$ vezes mais eficiente que os resultados reais obtidos pelo órgão repressivo federal.

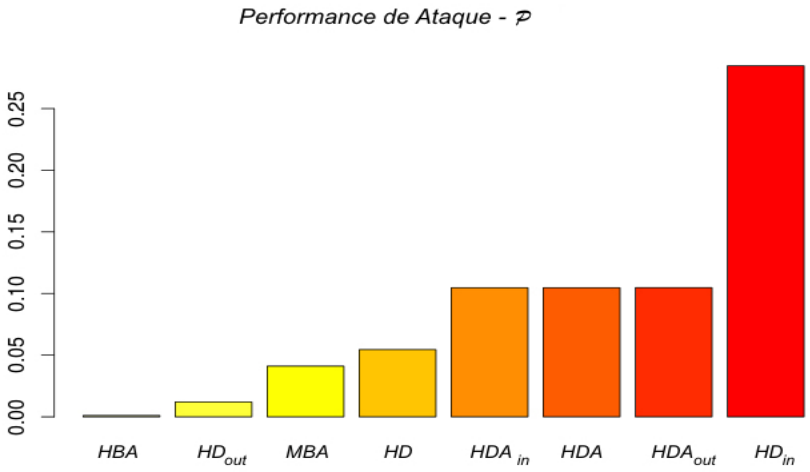

Figura 3.20: O histograma mostra a performance $\mathcal{P}$ de cada estratégia de ataque: HDA $(0,1047)$, HDA_{in} $(0,1047)$, HDA_{out} $(0,1047)$, HBA $(0,0011)$, MBA $(0,0411)$, HD $(0,0545)$, HD_{in} $(0,2847)$, HD_{out} $(0,0121)$ e HB $(0,1384)$.

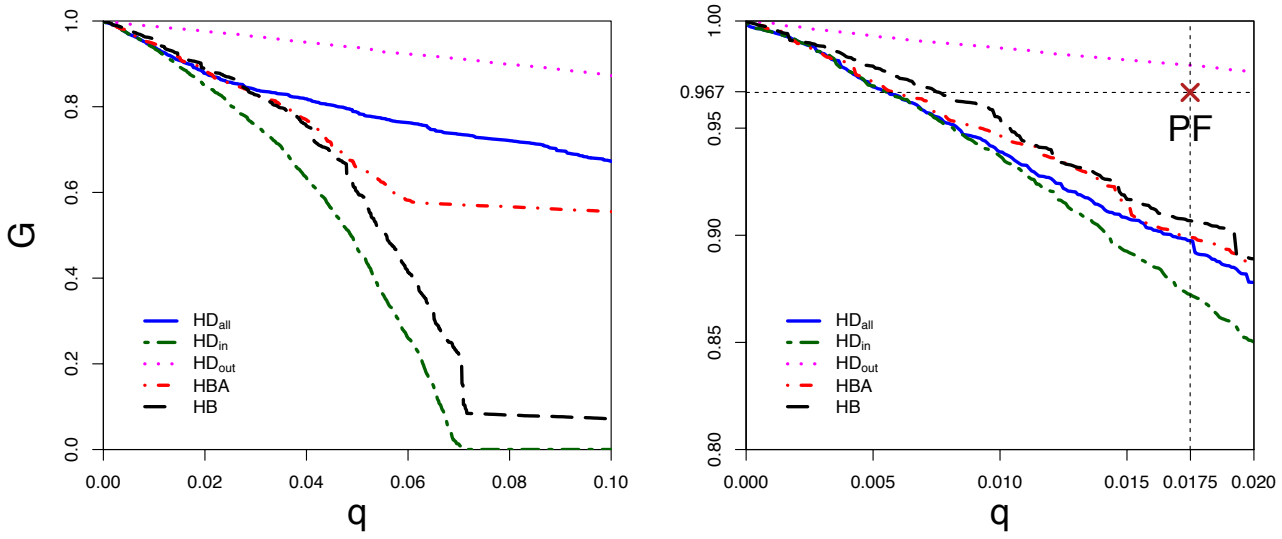

Figura 3.21: O tamanho relativo da maior componente conectada G (painel à esquerda) da rede da deep web como função da fração de vértices removidos q conforme as principais estratégias HD (linha sólida azul), HD_{in} (linha duplamente tracejada verde), HD_{out} (linha pontilhada magenta), HBA (linha ponto-tracejada vermelha) e HB (linha tracejada longa preta). Os valores em no painel à direita são uma ampliação na região onde ocorreu a ação policial. Este ponto está marcado pelo “x"vermelho e pelo acrônimo “PF"no cruzamento entre as duas linhas pontilhadas. Os policiais obtiveram mandados para 182 alvos, isto é, $1,75\%$ do total de usuários, resultando em um grafo com 1.060 usuários, ou seja, $96,7\%$ da rede original.

Conforme já debatido nos capítulos iniciais, uma rede é dita controlável matematicamente se para a equação dinâmica 2.13 é possível obter $n_D < 1$ (ver Equação 2.14). No caso trazido à baila, de fato tem-se garantia da rede ser controlável *lato sensu*. Por outro lado, a rede escondida na deep web resulta em uma fração de praticamente 93% ($n_D = 0,928$). De fato, seria necessário controlar praticamente todos os usuários do fórum. Assim, apesar da rede ser controlável no sentido matemático, isso pode não ser factível *in casu*. Como mostrado no gráfico de radar, a rede é altamente desassortativa e possui valores de correlação out-out e out-in muito altos negativamente. Isso faz com que a rede seja esparsa no número de arestas não correspondidas. Isso explica o comportamento praticamente não-controlável do sistema.

Capítulo 4

Conclusão

Nesta tese buscou-se explorar aspectos topológicos relacionados a um tipo particular de sistema social, qual seja, redes voltadas para práticas criminosas. Partindo do conceito já conhecido entre os órgãos de segurança pública de que grupos delinquentes costumam obter sigilo de suas atividades por meio de um processo de compartimentação no qual os dados relevantes são isolados em diferentes compartimentos ou células da organização, foi desenvolvido um método que relaciona a compartimentação com a modularidade de redes complexas. A partir daí identificaram-se as possíveis fragilidades destes sistemas e apresentou-se uma estratégia de ataque baseada em módulos que consiste em extrair comunidades de uma determinada rede e, em seguida, apagar apenas os nós que conectam módulos distintos ordenados pela centralidade de intermediação. Simulações computacionais em muitas redes reais mostram que o método MBA é mais eficiente em atomizar redes do que os procedimentos tradicionais baseados em critérios de centralidade. Com efeito, pode-se dizer que os vértices mais conectados ou os nós que têm o maior valor de centralidade de intermediação não são necessariamente os mais importantes para a sobrevivência da rede. Os nós que fazem as pontes entre comunidades distintas são estruturalmente mais importantes e cruciais para a coesão da rede do que *hubs* ou nós altamente centrais. Se atacarmos estas estruturas, o dano causado à rede é maior do que o causado por métodos tradicionais, eliminando-se a mesma quantidade de elementos. O objetivo de se aplicar o presente ataque baseado em módulos a uma determinada rede é revelar sua vulnerabilidade estrutural, medindo a rapidez com que se pode atingir o regime onde as comunidades da rede estão todas desconectadas. Assim, propõe-se caracterizar a vulnerabilidade modular de redes complexas precisamente pela velocidade com que o ponto final (onde todos os módulos estão desconectados) é atingido. Desta forma, o trabalho mostra que o ganho global de eficiência aumenta rapidamente com a modularidade da rede, isto é, quanto maior a modularidade, mais frágil é o sistema.

Em relação à detecção de comunidades, o limite de resolução de algoritmos baseados em modularidade é um tema de debate. No entanto, em conexão com o método de ataque proposto aqui, não é altamente relevante. O escopo do dano que se pode infringir a uma rede está relacionado ao número e tamanho dos módulos que podem ser extraídos. Por exemplo, quando grandes módulos são detectados, isso significa que a rede é decomposta em poucos módulos, o que é bom porque uma

grande parte da rede é desconectada quando um módulo é separado dos outros. A desvantagem é que o último módulo pode ser ainda grande em comparação com a rede original, como no caso da rede de aeroportos dos EUA em que o último componente conectado ainda é de 10% a 25% da rede original, para o ataque por vértices e arestas, respectivamente. Por outro lado, uma decomposição em muitas pequenas comunidades tem a vantagem de terminar com uma rede altamente fragmentada, mas à custa de maior esforço computacional e retirando-se um número muito maior de vértices e arestas. Portanto, a situação ótima está de alguma forma no meio, uma solução de compromisso em termos do tamanho médio do módulo e do tamanho da rede. A identificação de comunidades utilizando os algoritmos de detecção de módulos é o ingrediente essencial deste método. E, independentemente, do algoritmo particular utilizado para identificar as comunidades, o método de ataque baseado em módulos é sempre mais eficiente que os métodos tradicionais de fragmentação de redes reais. Portanto, embora esses módulos não tenham relação direta com comunidades reais, eles podem eventualmente revelar informações importantes sobre a funcionalidade estrutural de redes complexas.

A extensão natural deste estudo sobre a fragilidade de redes modulares frente a ataques é incluir o custo computacional de se produzir cada uma das listas de ataque. Para tanto, introduziu-se uma medida de robustez generalizada e uma quantidade empírica que podem ser úteis como guias para se escolher o método de ataque mais apropriado para cada caso real. Os resultados em redes artificiais de referência e sistemas reais com altos níveis de modularidade indicam que, se levadas em conta tanto a robustez quanto a complexidade temporal, o método MBA é a melhor escolha para a maioria dos casos. Esta conclusão é assegurada quando a fração de pontes N_d entre comunidades é menor que $\log N$. Além disso, a robustez de redes reais altamente modulares frente ao ataque simultâneo MBA é muito semelhante à do método HBA, enquanto HDA e CI não parecem ser muito sensíveis à modularidade.

Com efeito, fica evidente a aplicação direta do método modular a sistemas reais como células terroristas e estruturas de crime organizado. Assim, com o estabelecimento da eficiência de ataques a redes modulares abstratas, pode-se estudar *a posteriori* duas redes criminas reais obtidas em cooperação com a Polícia Federal. Trata-se de uma rede de crimes federais com mais 23.000 vértices e um sistema de relacionamentos virtuais em um fórum ilícito escondido na chamada deep web. A rede de crimes federais consiste em 23.666 indivíduos em 35.913 relacionamentos. Surpreendentemente, o sistema tem uma componente gigante contendo mais de 40% dos nós e 54% das arestas. Mostrou-se que a rede apresenta comportamentos de pequeno-mundo e invariância em escala, sendo mais escura que as redes sociais tradicionais, isto é, combinando valores de baixa densidade de borda e baixa eficiência da rede. A rede é particularmente fraca a ataques sequenciais por intermediação e a ataques por módulos devido principalmente à sua natureza modular significativa. Os ataques MBA mostram um desempenho mais alto, significando que o ponto de desativação onde todas as comunidades estão desarticuladas é atingido com menos esforço computacional do que a fase totalmente atomizada atingida por ataques baseados em centralidade, com ambos os pontos críticos sendo próximos um do outro.

Contra-intuitivamente, a rede é altamente controlável no sentido de que é possível, em princípio, levar qualquer variável dinâmica (como opinião, riqueza ou tendência criminal) de seu estado inicial para estados finais arbitrários, controlando-se aproximadamente apenas 20% dos nós, no que parece

ser um comportamento típico de redes sociais. No entanto, embora a controlabilidade matemática desse sistema criminal seja garantida pelo controle de menos de um quarto de seus nós, não é claro o que isso significa em termos práticos para os sistemas sociais criminais. Por exemplo, geralmente se está interessado em encontrar um desejado estado final estável ou então o sistema irá facilmente se afastar deste ponto, portanto controlabilidade matemática *per se* não fornece resultados totalmente úteis. Além disso, nas redes sociais os *drivers* são pessoas, logo até mesmo a tarefa de engenharia social de um único *input* torna-se pouco clara e discutível tanto do ponto de vista ético quanto jurídico. Portanto, uma compreensão profunda do controle social ainda é um assunto em aberto.

Argumenta-se, também, que a prisão tradicional é equivalente a ataques baseados em bordas, enquanto os ataques baseados em nós estão mais relacionados à morte ou à ressocialização completa dos criminosos. No entanto, embora em geral seja mais eficiente remover nós que bordas, particularmente nessa rede ambas as estratégias têm resultados semelhantes. Além disso, o banco de dados de crimes federais está crescendo continuamente (no momento deste trabalho, o tamanho do conjunto de dados atingiu mais de 100.000 nós), tornando virtualmente impossível gerar iterativamente N listas de ataques baseadas na centralidade de intermediação, que cresce na melhor das hipóteses como $(N \times E)^2$. Portanto, a melhor estratégia para se reduzir os níveis de criminalidade federal de acordo com a topologia apresentada seria por meio de políticas educacionais ou de prisão com os alvos escolhidos por uma abordagem modular dependendo da viabilidade política e prática de cada estratégia - por exemplo, o sistema carcerário teria que realmente trabalhar para cortar os laços sociais dos prisioneiros e para ressocializá-los ou educá-los, o que por si só são tarefas muito difíceis.

Apresentou-se, ainda, pela primeira vez uma rede criminosa clandestina operando dentro da web profunda pelo uso do navegador Tor como resultado dos dados adquiridos por uma investigação da PF. A rede consiste em um ambiente parecido a um fórum onde criminosos interagiam por mensagens escritas e transferências de arquivos. Os dados foram coletados durante a chamada operação Darknet pela Polícia Federal em 2013 e 2016. A topologia de rede revelou algumas características interessantes do sistema. A distribuição de graus é mais complexa do que uma simples lei de potência, mesmo que uma porção da cauda possa ser ajustada em uma distribuição invariante em escala com um expoente próximo a 2. Além disso, a rede mostra a forte presença de alguns *hubs* responsáveis pelo envio da materialidade criminal (menos de 8% do número total de usuários investigados). Nesse sentido, a arquitetura de rede da web profunda exibe um comportamento no qual a maioria dos *leachers* aproveita o material ilícito produzido por um pequeno número de *putters* que são verdadeiramente responsáveis por manter a rede coesa. Além disso, a densidade de borda da rede é muito pequena e pode ser interpretada como o brilho da rede, ou seja, a facilidade com a qual pode-se obter informações sobre todo o sistema, prendendo-se apenas alguns indivíduos. Este comportamento é compatível com a clandestinidade extrema da rede, na qual os usuários optaram por participar exatamente devido a suas características de anonimato. Por outro lado, a rede não é modular, não há clusters densamente conectados internamente, mas escassamente ligados a outros clusters e não há comunidades na rede, no sentido de que os usuários tendem a visualizar conteúdos gerais e não subtópicos especializados. A rede é construída de tal forma que é extremamente frágil a ataques baseados em graus, o que atomiza

totalmente o sistema depois que aproximadamente 7% de nós são removidos da rede usando tanto a estratégia em grau simultânea quanto os ataques de alto grau adaptativo. No entanto, considerando o tempo de CPU para gerar a estratégia, a melhor opção para se atacar a rede é um ataque de grau simultâneo que fragmenta completamente a rede, mas à custa de menos requisitos computacionais. A escolha correta da estratégia, daí, depende fortemente da capacidade da polícia em adquirir provas forenses e do tempo necessário para obter mandados de busca e apreensão ou de prisão. Portanto, como no Brasil costumeiramente levam-se vários meses para se atender aos requisitos jurídicos, é possível planejar a estratégia de aplicação da lei mais adequada de acordo com a topologia de rede e com o desenvolvimento da investigação. Como estudo de caso, durante a chamada Operação Darknet pela Polícia Federal $1,75\%$ do número total de nós foram atacados e removidos do fórum. Esta ação resultou em apenas $3,3\%$ de dano à maior componente conectada. Por outro lado, se a Polícia Federal tivesse utilizado critérios de topologia para planejar a investigação, quase 13% da rede seria separada do núcleo original.

Outro aspecto importante desta rede clandestina é que é matematicamente possível conduzir qualquer variável dinâmica (propagação de ideias, imitação de comportamento ou mesmo propagação intencional de um vírus informático etc), delimitada pelo acoplamento de rede, de uma condição inicial a uma condição final por uma escolha adequada de *inputs*. No entanto, ao contrário da maioria das redes sociais, o sistema estudado aqui é praticamente impossível de ser controlado, porque mais de 90% dos nós teriam de ser controlados individualmente para se ter domínio de todo o sistema. Resumindo, a rede criminosa da web profunda apresentada aqui mostra por um lado algumas características típicas de rede social, tais como pequeno-mundo, densidade de borda reduzida e distribuição de graus heterogêneos (presença de *hubs*), e por outro lado características normalmente não vistas em redes sociais tradicionais, como mistura desassortativa (alta correlação r_{out-in}), a demanda de um número extremamente elevado de nós de controle, a falta de estruturas modulares ou comunidades e pequeno coeficiente de agrupamento médio.

Assim, mostrou-se que a compartimentação típica de redes criminais corresponde matematicamente ao conceito de modularidade. Este fenômeno abre a oportunidade de se explorar uma fragilidade estrutural de redes reais pela aplicação de métodos de ataques a redes que focam nos indivíduos que fazem a ponte entre comunidades topológicas— em termos práticos pode se tratar do advogado que trabalha para duas OrCrims independentes ou o doleiro que presta serviço para vários grupos criminosos que se sobrepõem. De fato, a rede de crimes federais é propícia para a aplicação deste método, além de apresentar outras características importantes abordadas anteriormente. Por outro lado, a rede de pedofilia escondida na camada profunda da web não se mostra modular, apresentando outro tipo de comportamento particular de redes virtuais, sendo frágil, contudo, a ataques por grau. Com efeito, fica evidente a importância de aplicação de critérios topológicos para se avaliar a relevância de cada ator de uma rede criminal, isto é, pode-se obter ganhos enormes de eficiência no combate ao crime ao se optar por esta abordagem. Por termo, espera-se que esses resultados lancem alguma luz sobre a natureza dos comportamentos coletivos que formam redes criminais. Além disso, deseja-se enfatizar por este estudo a importância da aplicação de ferramentas de ciências naturais e de redes em

operações policiais e os possíveis ganhos na eficiência das estratégias de combate ao crime.

4.1 Trabalhos futuros

Conforme já explorado ao longo deste trabalho, em termos apenas da robustez, o método HBA supera outros, porém à custa de uma complexidade de tempo muito maior o que o torna inaplicável a redes reais. No entanto, pela aproximação da centralidade de intermediação por médias de Rademacher, Riondato e Upfal [145] mostraram recentemente que pode ser possível executar a centralidade de intermediação para todos os nós de grandes grafos muito mais rápido do que se pensava (o chamado algoritmo ABRA - Approximating Betweenness with Rademacher Averages), embora ainda haja muito debate sobre a complexidade do método. HBA e estratégias MBA usando o algoritmo ABRA podem se revelar mais próximos do desmantelamento ideal no futuro.

Outrossim, quando esta pesquisa foi concluída, Kobayashi *et al* [146] propuseram uma Influência Coletiva que mostrou ser mais eficiente que o método original proposto por Morone e Makse para redes com estruturas comunitárias bem definidas. Este artigo recente destaca ainda mais importância das ligações fracas, pontes e influências coletivas na manutenção de redes complexas. Esses são tópicos que merecem investigação no futuro.

Recentemente, Duijn *et al* [55] analisaram a resiliência de uma rede de narcotráfico como sendo a resposta do sistema a ataques. Para tanto, propôs-se três algoritmos de recuperação: no primeiro os nós órfãos se reconectam aleatoriamente a outros vértices, no segundo a recuperação dá preferencia estatística àqueles a uma menor distância e o último enfatiza a preferência por grau. No trabalho original, mostrou-se que a eficiência da rede é pouco afetada, indicando que tais sistemas são flexíveis e se adaptam à ação policial, mostrando alta resiliência e fácil recuperação. Contudo, o resultado é muito particular à estrutura da rede de narcotráfico apresentada. Além do mais, não foram estudados métodos de recuperação quando a rede é atacada por estratégias modulares. Assim, outra possibilidade futura é estudar como se dá a resiliência das redes de crimes federais e da deep web, além de extender a análise para a resposta a ataques modulares.

Por outro lado, muitas redes criminais são na verdade casos de sistemas multidimensionais [147, 148], nos quais cada camada representa um tipo de relacionamento distinto, *eg* conexões relacionadas ao narcotráfico, a roubos, a lavagem de dinheiro etc. Tem-se então o que se chama de redes multicamadas, com aplicações úteis na física [149], medicina [150], transporte [151] entre outros. Matematicamente, estas redes multidimensionais nada mais são que o tripleto $G = (V, E, D)$ onde V é o conjunto de vértices, E é o conjunto de arestas e D é o conjunto de dimensões (ou camadas). Cada aresta consiste, então, dos tripletos (u, v, d) com $u, v \in V$ e $d \in D$. Assim, dentre as muitas aplicações possíveis, vislumbra-se construir redes criminais multidimensionais, estudando-se suas propriedades topológicas, fragilidades e também processos difusivos [152, 153]— *trend topic* que vem chamando muita atenção da comunidade científica. De fato, propriedades de redes que dependam de algum processo difusivo são afetadas pelo acoplamento entre camadas, o que pode mostrar comportamentos inéditos no que tange à fragilidade de sistemas criminais.

Do ponto de vista de processos dinâmicos, modelos padrão de propagação de rumores como o proposto por Daley e Kendall [154] podem ser aplicados às topologias de redes criminais estudadas nesta tese. Assim, trabalhos futuros podem focar no estudo da teoria de controle de rumores e informações operando sob o tecido social do crime organizado. Também sobre processos dinâmicos, Christakis e Fowler [155] mostraram que fenômenos de rede são relevantes aos aspectos biológicos e comportamentais da obesidade, que se dissemina nos relacionamentos sociais de maneira análoga a movimentos epidêmicos. Destarte, uma pergunta relevante que surge é se a dinâmica criminal no Brasil se comporta similarmente à uma epidemia do tipo biológica.

Além disso, apesar da fragmentação de redes ser costumeiramente medida pelo tamanho relativo da maior componente conectada, a dispersão de tamanhos de todos os *clusters* também é relevante para avaliação de atomização completa de sistemas complexos, tais como o tamanho médio das geodésicas e do segundo maior aglomerado.

Por termo, a análise de fragilidades estruturais pode também ser fundamental para se fortalecer o organograma interno de relacionamentos e procedimentos da própria Polícia Federal. Abre-se, pois, a possibilidade de se utilizar ferramentas de rede para a otimização de processos burocráticos e organizacionais.

Enfim, as possíveis extensões e inovações do presente trabalho são inúmeras e podem resultar gradualmente numa compreensão mais profunda da natureza das redes criminais. Destarte, espera-se que trabalhos futuros contribuam para uma visão mais completa e objetiva do problema, resultando em efeitos práticos no dia-a-dia do combate ao crime.

Referências Bibliográficas

[1] Barabási AL. Network Science. Cambridge University Press; 2016.

[2] Strogatz SH. Exploring complex networks. nature. 2001;410(6825):268.

[3] Newman ME. The structure and function of complex networks. SIAM review. 2003;45(2):167–256.

[4] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. Physics reports. 2006;424(4):175–308.

[5] Albert R, Barabási AL. Statistical mechanics of complex networks. Reviews of modern physics. 2002;74(1):47.

[6] Dorogovtsev SN, Mendes J. Evolution of networks: From biological nets to the Internet and WWW. OUP Oxford; 2013.

[7] Callaway DS, Newman ME, Strogatz SH, Watts DJ. Network robustness and fragility: Percolation on random graphs. Physical review letters. 2000;85(25):5468.

[8] Iyer S, Killingback T, Sundaram B, Wang Z. Attack robustness and centrality of complex networks. PloS one. 2013;8(4):e59613.

[9] Crucitti P, Latora V, Marchiori M, Rapisarda A. Efficiency of scale-free networks: error and attack tolerance. Physica A: Statistical Mechanics and its Applications. 2003;320:622–642.

[10] Holme P, Kim BJ, Yoon CN, Han SK. Attack vulnerability of complex networks. Physical Review E. 2002;65(5):056109.

[11] Cohen R, Erez K, ben Avraham D, Havlin S. Breakdown of the Internet under Intentional Attack. Physical Review Letters. 2001 Apr;86(16):3682–3685.

[12] Wilson C. Searching for saddam: a five-part series on how the us military used social networking to capture the iraqi dictator. Slate. 2010;.

[13] Arquilla J, Ronfeldt D. Networks and netwars: The future of terror, crime, and militancy. Rand Corporation; 2001.

[14] Cane P, Conaghan J. The new Oxford companion to law. Oxford University Press, UK; 2008.

[15] Quinney R. Structural characteristics, population areas, and crime rates in the United States. The Journal of Criminal Law, Criminology, and Police Science. 1966;p. 45–52.

[16] Toth N, Gulyás L, Legendi RO, Duijn P, Sloot PM, Kampis G. The importance of centralities in dark network value chains. The European Physical Journal Special Topics. 2013;222(6):1413–1439.

[17] Brantingham PBP. Patterns in Crime. Macmillan New Yotk; 1984.

[18] Alves LGA, Ribeiro HV, Mendes RS. Scaling laws in the dynamics of crime growth rate. Physica A: Statistical Mechanics and its Applications. 2013;392(11):2672–2679. Available from: http://www.sciencedirect.com/science/article/pii/S0378437113001416.

[19] Picoli S, Castillo-Mussot Md, Ribeiro HV, Lenzi EK, Mendes RS. Universal bursty behaviour in human violent conflicts. Scientific Reports. 2014 04;4:4773 EP –. Available from: http://dx.doi.org/10.1038/srep04773.

[20] Ball P. Why society is a complex matter: Meeting twenty-first century challenges with a new kind of science. Springer Science & Business Media; 2012.

[21] Morselli C. Inside criminal networks. Springer; 2009.

[22] Spapens T. Interaction between criminal groups and law enforcement: the case of ecstasy in the Netherlands. Global crime. 2011;12(1):19–40.

[23] Morselli C, Petit K. Law-enforcement disruption of a drug importation network. Global Crime. 2007;8(2):109–130.

[24] Natarajan M. Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data. Journal of Quantitative Criminology. 2006;22(2):171–192.

[25] Sarnecki J. Delinquent networks: Youth co-offending in Stockholm. Cambridge University Press; 2001.

[26] Chen H, Chung W, Xu JJ, Wang G, Qin Y, Chau M. Crime data mining: a general framework and some examples. Computer. 2004;37(4):50–56.

[27] Dreżewski R, Sepielak J, Filipkowski W. The application of social network analysis algorithms in a system supporting money laundering detection. Information Sciences. 2015 2;295:18–32. Available from: http://www.sciencedirect.com/science/article/pii/S0020025514009979.

[28] McGloin J. Policy and intervention considerations of a network analysis of street gangs. Criminology & Public Policy. 2005;4(3):607–635.

[29] Sah RK. Social osmosis and patterns of crime: A dynamic economic analysis. Journal of political Economy. 1991;99(6).

[30] Glaeser EL, Sacerdote B, Scheinkman JA. Crime and Social Interactions. The Quarterly Journal of Economics. 1996 05;111(2):507–548.

[31] Morselli C. Career opportunities and network-based privileges in the Cosa Nostra. Crime, Law and Social Change. 2003;39(4):383–418.

[32] Mastrobuoni G, Patacchini E. Organized crime networks: An application of network analysis techniques to the American mafia. Review of Network Economics. 2012;11(3):1–43.

[33] Thornberry TP, Krohn MD, Lizotte AJ, Chard-Wierschem D. The role of juvenile gangs in facilitating delinquent behavior. Journal of research in Crime and Delinquency. 1993;30(1):55–87.

[34] Asch SE. Studies of independence and conformity: I. A minority of one against a unanimous majority. Psychological monographs: General and applied. 1956;70(9):1.

[35] Milgram S. Goodman S, editor. Behavioral Study of obedience. American Psychological Association; 1973.

[36] Ballester C, Calvo-Armengol A, Zenou Y. Who's who in networks. wanted: the key player. Econometrica. 2006;74(5):1403–1417.

[37] Borgatti SP. Identifying sets of key players in a social network. Computational & Mathematical Organization Theory. 2006;12(1):21–34.

[38] D'Orsogna MR, Perc M. Statistical physics of crime: A review. Physics of Life Reviews. 2015 3;12:1–21. Available from: http://www.sciencedirect.com/science/article/pii/S1571064514001730.

[39] Krebs VE. Mapping networks of terrorist cells. Connections. 2002;24(3):43–52.

[40] Baker WE, Faulkner RR. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. American sociological review. 1993;p. 837–860.

[41] Bornholdt S, Schuster HG. Handbook of graphs and networks: from the genome to the internet. John Wiley & Sons; 2006.

[42] Bollobás B. Modern graph theory. vol. 184. Springer Science & Business Media; 2013.

[43] Newman M. Networks: an introduction. Oxford university press; 2010.

[44] Holland PW, Leinhardt S. Transitivity in structural models of small groups. Comparative Group Studies. 1971;2(2):107–124.

[45] Watts DJ, Strogatz SH. Collective dynamics of small-world networks. Nature. 1998 06;393(6684):440–442. Available from: http://dx.doi.org/10.1038/30918.

[46] Godsil C, Royle GF. Algebraic graph theory. vol. 207. Springer Science & Business Media; 2013.

[47] Newman ME. Assortative mixing in networks. Physical review letters. 2002;89(20):208701.

[48] Newman ME. Mixing patterns in networks. Physical Review E. 2003;67(2):026126.

[49] Bavelas A. Communication patterns in task-oriented groups. The Journal of the Acoustical Society of America. 1950;22(6):725–730.

[50] Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977;p. 35–41.

[51] Erdos P, Rényi A. On the evolution of random graphs. Bull Inst Internat Statist. 1961;38.4:343–347.

[52] Barabási AL, Albert R. Emergence of scaling in random networks. science. 1999;286(5439):509–512.

[53] Molloy M, Reed B. A critical point for random graphs with a given degree sequence. Random structure and algorithms. 1995;6(2-3):161–180.

[54] Latora V, Marchiori M. Efficient behavior of small-world networks. Physical review letters. 2001;87(19):198701.

[55] Duijn PA, Kashirin V, Sloot PM. The relative ineffectiveness of criminal network disruption. Scientific reports. 2014;4:4238.

[56] da Cunha BR, González-Avella JC, Gonçalves S. Fast Fragmentation of networks using module-based attacks. PloS one. 2015;10(11):e0142824.

[57] Newman MEJ. Modularity and community structure in networks. Proceedings of the National Academy of Sciences. 2006;103(23):8577–8582. Available from: http://www.pnas.org/content/103/23/8577.abstract.

[58] Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E. 2006 Sep;74:036104. Available from: http://link.aps.org/doi/10.1103/PhysRevE.74.036104.

[59] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008;2008(10):P10008.

[60] Havlin S, Cohen R. Complex networks: structure, robustness and function. Cambridge University Press; 2010.

[61] Holme P. Efficient local strategies for vaccination and network attack. EPL (Europhysics Letters). 2004;68(6):908.

[62] Wang L, Singhal A, Jajodia S. Toward measuring network security using attack graphs. In: Toward measuring network security using attack graphs. ACM; 2007. p. 49–54.

[63] Hébert-Dufresne L, Allard A, Young JG, Dubé LJ. Global efficiency of local immunization on complex networks. Scientific reports. 2013;3.

[64] Agreste S, Catanese S, De Meo P, Ferrara E, Fiumara G. Network structure and resilience of Mafia syndicates. Information Sciences. 2016;351:30–47.

[65] Pagani GA, Aiello M. The power grid as a complex network: a survey. Physica A: Statistical Mechanics and its Applications. 2013;392(11):2688–2700.

[66] Habib MF, Tornatore M, Mukherjee B. Cascading-failure-resilient interconnection for interdependent power grid-optical networks. In: Cascading-failure-resilient interconnection for interdependent power grid-optical networks. Optical Society of America; 2015. p. M3I–3.

[67] Raab J, Milward HB. Dark networks as problems. Journal of public administration research and theory. 2003;13(4):413–439.

[68] Albert R, Albert I, Nakarado GL. Structural vulnerability of the North American power grid. Physical review E. 2004;69(2):025103.

[69] Pu CL, Cui W. Vulnerability of complex networks under path-based attacks. Physica A: Statistical Mechanics and its Applications. 2015;419:622–629.

[70] Valente TW, Fujimoto K. Bridging: Locating critical connectors in a network. Social Networks. 2010;32(3):212 – 220. Available from: http://www.sciencedirect.com/science/article/pii/S0378873310000146.

[71] Hwang W, Cho Yr, Zhang A, Ramanathan M. Bridging centrality: identifying bridging nodes in scale-free networks. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; 2006. p. 20–23.

[72] Kaiser M, Hilgetag CC. Edge vulnerability in neural and metabolic networks. Biological Cybernetics. 2004 May;90(5). Available from: http://dx.doi.org/10.1007/s00422-004-0479-1.

[73] Bu Y, Gregory S, Mills HL. Efficient local behavioral-change strategies to reduce the spread of epidemics in networks. Phys Rev E. 2013 Oct;88(4). Available from: http://dx.doi.org/10.1103/PhysRevE.88.042801.

[74] Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. nature. 2000;406(6794):378–382.

[75] Crucitti P, Latora V, Marchiori M, Rapisarda A. Error and attack tolerance of complex networks. Physica A: Statistical Mechanics and its Applications. 2004;340(1):388–394.

[76] Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001;411(6833):41–42.

[77] Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. Nature. 2010;464(7291):1025–1028.

[78] Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. Nature. 2015;.

[79] Morone F, Min B, Bo L, Mari R, Makse HA. Collective Influence Algorithm to find influencers via optimal percolation in massively large social media. Scientific reports. 2016;6.

[80] Braunstein A, Dall'Asta L, Semerjian G, Zdeborová L. Network dismantling. Proceedings of the National Academy of Sciences. 2016;p. 201605083.

[81] Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical review E. 2004;69(2):026113.

[82] Shai S, Kenett DY, Kenett YN, Faust M, Dobson S, Havlin S. Critical tipping point distinguishing two types of transitions in modular network structures. Physical Review E. 2015;92(6):062805.

[83] Shekhtman LM, Shai S, Havlin S. Resilience of networks formed of interdependent modular networks. New Journal of Physics. 2015;17(12):123007.

[84] Faqeeh A, Melnik S, Colomer-de Simón P, Gleeson JP. Emergence of coexisting percolating clusters in networks. Physical Review E. 2016;93(6):062308.

[85] Shekhtman LM, Danziger MM, Havlin S. Recent advances on failure and recovery in networks of networks. Chaos, Solitons & Fractals. 2016;90:28–36.

[86] Luenberger DG. Introduction to Dynamical Systems: Theory, Models and Applications. Wiley; 1979.

[87] Slotine JJ, Li W. Applied Nonlinear Control. Pearson; 1991.

[88] Kalman RE. Mathematical description of linear dynamical systems. J Soc Indus Appl Math Ser A. 1963;1:152–192. Available from: http://dx.doi.org/10.1137/0301010.

[89] Shields R, Pearson J. Structural controllability of multiinput linear systems. IEEE Trans Automat Contr. 1976;21:203–212. Available from: http://dx.doi.org/10.1109/TAC.1976.1101198.

[90] Sontag ED. Mathematical Control Theory: Deterministic Finite Dimensional Systems. Springer-Verlag; 1998.

[91] Gao J, Liu YY, D'Souza RM, Barabási AL. Target control of complex networks. Nat Commun. 2014 11;5. Available from: http://dx.doi.org/10.1038/ncomms6415.

[92] Yuan Z, Zhao C, Di Z, Wang WX, Lai YC. Exact controllability of complex networks. Nat Commun. 2013;4:2447.

[93] Pósfai M, Liu YY, Slotine JJ, Barabási AL. Effect of correlations on network controllability. Scientific Reports. 2013 01;3:1067 EP –. Available from: http://dx.doi.org/10.1038/srep01067.

[94] Liu YY, Slotine JJ, Barabási AL. Controllability of complex networks. Nature. 2011;473:167–173. Available from: http://dx.doi.org/10.1038/nature10011.

[95] Rosvall M, Axelsson D, Bergstrom CT. The map equation. The European Physical Journal Special Topics. 2009;178(1):13–23. Available from: http://dx.doi.org/10.1140/epjst/e2010-01179-1.

[96] Granovetter MS. The strength of weak ties. American journal of sociology. 1973;p. 1360–1380.

[97] De Meo P, Ferrara E, Fiumara G, Provetti A. On Facebook, Most Ties Are Weak. Commun ACM. 2014 Oct;57(11):78–84. Available from: http://doi.acm.org/10.1145/2629438.

[98] Barthélemy M. Spatial networks. Physics Reports. 2011;499(1):1–101.

[99] Collection TKN. KONECT, editor. US power grid network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/opsahl-powergrid.

[100] Collection TKN. KONECT, editor. Euroroad network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/subelj_euroroad.

[101] Šubelj L, Bajec M. Robust network community detection using balanced propagation. The European Physical Journal B. 2011;81(3):353–362. Available from: http://dx.doi.org/10.1140/epjb/e2011-10979-2.

[102] Collection TKN. KONECT, editor. OpenFlights network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/opsahl-openflights.

[103] Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks. 2010;32(3):245 – 251. Available from: http://www.sciencedirect.com/science/article/pii/S0378873310000183.

[104] Collection TKN. KONECT, editor. US airports network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/opsahl-usairport.

[105] Opsahl T. KONECT, editor. Why Anchorage is not (that) important: Binary ties and Sample selection. KONECT; 2011. Available from: http://wp.me/poFcY-Vw.

[106] Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, et al. The protein-protein interaction map of Helicobacter pylori. Nature. 2001 01;409(6817):211–215. Available from: http://dx.doi.org/10.1038/35051615.

[107] Collection TKN. KONECT, editor. Caenorhabditis elegans network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/arenas-meta.

[108] Collection TKN. KONECT, editor. Facebook (NIPS) network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/ego-facebook.

[109] Collection TKN. KONECT, editor. Google+ network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/ego-gplus.

[110] Collection TKN. KONECT, editor. Twitter lists network dataset – KONECT. KONECT; 2014. Available from: http://konect.uni-koblenz.de/networks/ego-twitter.

[111] McAuley J, Leskovec J. Learning to Discover Social Circles in Ego Networks. In: Advances in Neural Information Processing Systems. The MIT press; 2012. p. 548–556.

[112] Newman MEJ. Spectral methods for community detection and graph partitioning. Phys Rev E. 2013 Oct;88:042822. Available from: http://link.aps.org/doi/10.1103/PhysRevE.88.042822.

[113] Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis. Physical review E. 2009;80(5):056117.

[114] Jen E. Robust design: a repertoire of biological, ecological, and engineering case studies. Oxford University Press; 2005.

[115] Van Mieghem P, Doerr C, Wang H, Hernandez JM, Hutchison D, Karaliopoulos M, et al. A framework for computing topological network robustness. Delft University of Technology, Report20101218. 2010;.

[116] Schieber TA, Carpi L, Frery AC, Rosso OA, Pardalos PM, Ravetti MG. Information theory perspective on network robustness. Physics Letters A. 2016;380(3):359–364.

[117] Cohen R, Erez K, Ben-Avraham D, Havlin S. Resilience of the Internet to random breakdowns. Physical review letters. 2000;85(21):4626.

[118] Pu CL, Pei WJ, Michaelson A. Robustness analysis of network controllability. Physica A: Statistical Mechanics and its Applications. 2012;391(18):4420–4425.

[119] Boginski VL, Commander CW, Turko T. Polynomial-time identification of robust network flows under uncertain arc failures. Optimization Letters. 2009;3(3):461–473.

[120] Schneider CM, Moreira A, Andrade J, Havlin S, Herrmann HJ. Mitigation of malicious attacks on networks. Proceedings of the National Academy of Sciences. 2011;108(10):3838–3841.

[121] Bagrow JP, Lehmann S, Ahn YY. Robustness and modular structure in networks. Network Science. 2015;3(04):509–525.

[122] Girvan M, Newman ME. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002;99(12):7821–7826.

[123] Karrer B, Newman ME. Stochastic blockmodels and community structure in networks. Physical Review E. 2011;83(1):016107.

[124] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Physical review E. 2008;78(4):046110.

[125] Yang Z, Algesheimer R, Tessone CJ. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. Scientific Reports. 2016;6.

[126] Shavitt Y, Zilberman N. A Structural Approach for PoP Geo-Location. INFOCOM IEEE Conference on Computer Communications Workshops , 2010. 2010;p. 1–6.

[127] Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695(5):1–9.

[128] Cerqueira D, Ferreira H, Lima RSd, Bueno S, Hanashiro O, Batista F, et al. Atlas da Violência 2016. Instituto de Pesquisa Econômica Aplicada (Ipea); 2016.

[129] UNODC. State of crime and criminal justice worldwide. United Nations; 2015.

[130] Misse M, Costa AT, Vargas JD, Ratton JL, Azevedo RG. O inquérito policial no Brasil: uma pesquisa empírica. FENAPEF NECVU BOOKLINK; 2010.

[131] Cayli B. Italian civil society against the Mafia: From perceptions to expectations. International Journal of Law, Crime and Justice. 2013;41(1):81–99.

[132] Morselli C, Giguère C, Petit K. The efficiency/security trade-off in criminal networks. Social Networks. 2007;29(1):143–153.

[133] Araújo FR, Cunha RS. Crimes Federais. jusPODIVM; 2016.

[134] KONECT. KONECT, editor. Crime network dataset – KONECT. KONECT; 2016. Available from: http://konect.uni-koblenz.de/networks/moreno_crime.

[135] KONECT. KONECT, editor. Hamsterster full network dataset – KONECT. KONECT; 2016. Available from: http://konect.uni-koblenz.de/networks/petster-hamster.

[136] Gonçalves B, Perra N, Vespignani A. Modeling users' activity on twitter networks: Validation of dunbar's number. PloS one. 2011;6(8):e22656.

[137] Boguná M, Pastor-Satorras R, Vespignani A. Cut-offs and finite size effects in scale-free networks. The European Physical Journal B-Condensed Matter and Complex Systems. 2004;38(2):205–209.

[138] Requião da Cunha B, Gonçalves S. Performance of attack strategies on modular networks. Journal of Complex Networks. 2017;.

[139] bbc com. bbc com, editor. Brazil police crack 'darknet' in child pornography crackdown. bbc.com; 2014. [Online; posted 16-October-2014]. Available from: http://www.bbc.com/news/world-latin-america-29639241.

[140] foxnews com. foxnews com, editor. Brazilian police crack hidden 'darknet' child porn ring. foxnews.com; 2014. [Online; posted 16-October-2014]. Available from: http://www.foxnews.com/world/2014/10/16/suspects-from-all-backgrounds-caught-in-brazilian-child-porn-crackdown/.

[141] torproject org. torproject org, editor. Tor Anonymity Online. Tor hidden services; 2016. Available from: http://www.torproject.org.

[142] Holme P, Edling CR, Liljeros F. Structure and time evolution of an Internet dating community. Social Networks. 2004;26(2):155–174

[143] Organization WH. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. WHO. 1993;.

[144] Shai S, Kenett DY, Kenett YN, Faust M, Dobson S, Havlin S. Resilience of modular complex networks. arXiv preprint arXiv:14044748. 2014;.

[145] Riondato M, Upfal E. ABRA: Approximating betweenness centrality in static and dynamic graphs with Rademacher averages. arXiv preprint arXiv:160205866. 2016;.

[146] Kobayashi T, Masuda N. Fragmenting networks by targeting collective influencers at a mesoscopic level. Scientific Reports. 2016;6.

[147] Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. Journal of Complex Networks. 2014;2(3):203. Available from: +http://dx.doi.org/10.1093/comnet/cnu016.

[148] De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y, Porter MA, et al. Mathematical Formulation of Multilayer Networks. Phys Rev X. 2013 Dec;3:041022. Available from: http://link.aps.org/doi/10.1103/PhysRevX.3.041022.

[149] De Domenico M, Granell C, Porter MA, Arenas A. The physics of spreading processes in multilayer networks. Nat Phys. 2016 10;12(10):901–906. Available from: http://dx.doi.org/10.1038/nphys3865.

[150] Fiori KL, Smith J, Antonucci TC. Social Network Types Among Older Adults: A Multidimensional Approach. The Journals of Gerontology: Series B. 2007;62(6):P322. Available from: +http://dx.doi.org/10.1093/geronb/62.6.P322.

[151] Cardillo A, Gómez-Gardeñes J, Zanin M, Romance M, Papo D, Pozo Fd, et al. Emergence of network features from multiplexity. Scientific Reports. 2013 02;3:1344 EP –. Available from: http://dx.doi.org/10.1038/srep01344.

[152] Gómez S, Díaz-Guilera A, Gómez-Gardeñes J, Pérez-Vicente CJ, Moreno Y, Arenas A. Diffusion Dynamics on Multiplex Networks. Phys Rev Lett. 2013 Jan;110:028701. Available from: http://link.aps.org/doi/10.1103/PhysRevLett.110.028701.

[153] De Domenico M, Solé-Ribalta A, Gómez S, Arenas A. Navigability of interconnected networks under random failures. Proceedings of the National Academy of Sciences. 2014;111(23):8351–8356.

[154] Nekovee M, Moreno Y, Bianconi G, Marsili M. Theory of rumour spreading in complex social networks. Physica A: Statistical Mechanics and its Applications. 2007;374(1):457–470.

[155] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. n engl j med. 2007;2007(357):370–379.