

Éder Milton Schneider

**Formação de comunidades estruturais mediadas
por confinamento em redes complexas.**

Brasil

2023

Éder Mílton Schneider

**Formação de comunidades estruturais mediadas por
confinamento em redes complexas.**

Tese submetida à Universidade Federal do Rio Grande do Sul como requerimento parcial à obtenção ao título de Doutor em Ciências, sob supervisão do Prof. Doutor Sebastián Gonçalves (Instituto de Física - UFRGS, Porto Alegre, Brasil).

Universidade Federal do Rio Grande do Sul

Instituto de Física

Programa de Pós-Graduação em Física

Orientador: Prof. Doutor Sebastián Gonçalves

Brasil

2023

*Este trabalho é dedicado aos laços que nos unem,
que nos fazem comunidade, mesmo no isolamento.*

AGRADECIMENTOS

Os agradecimentos principais são direcionados a minha família, pelo suporte e pelas infinitas alegrias concedidas; ao Professor Doutor José Roberto Iglesias que, embora não seja orientador deste trabalho no papel é meu eterno orientador; ao professor Doutor Sebastián Gonçalves que me acolheu e dedicou centenas e centenas de horas orientando esta tese, constantemente provocando o debate científico e desafiando minhas pré-concepções.

Agradecimentos especiais são direcionados ao Instituto de Física da UFRGS e a todos os professores e funcionários que resistem e fazem sobreviver o papel indispensável das universidades públicas no desenvolvimento da ciência brasileira. Agradeço ao colega e pesquisador Doutor Bruno Requião da Cunha, cujas discussões e comentários contribuíram desde a gênese desta tese. Agradeço também ao indispensável suporte do CNPq na realização desta pesquisa.

*“It has been said that all models are wrong but some models are useful.
In other words, any model is at best a useful fiction—there never was,
or ever will be, an exactly normal distribution or an exact linear relationship.
Nevertheless, enormous progress has been made by entertaining such fictions
and using them as approximations.”*

(George E. P. Box)

RESUMO

Sistemas complexos são formados por múltiplas partes interconectadas, apresentando uma série de propriedades notáveis. As chamadas redes complexas são redes representativas de sistemas empíricos, através de coleta de dados dos sistemas, ou derivadas de modelos que buscam emular as características dessas redes empíricas. O campo de ciência de redes, nas últimas três décadas, demonstrou a existência de uma série de propriedades comuns entre sistemas dos mais variados campos. Uma dessas propriedades notáveis é o surgimento, existência e dissolução de comunidades estruturais, partes da rede com elementos altamente interconectados e fracamente conectados ao restante da rede.

Neste trabalho é apresentada a tese de que espaços confinados, dentro de uma rede dinâmica, são um elemento de organização da rede, capaz de explicar a emergência de topologias não triviais, típicas de redes complexas: formação de cliques, distribuições de grau de conectividade heterogêneas, com formação de hubs altamente conectados e de comunidades estruturais. Um modelo computacional simples é proposto, baseado no confinamento temporário de elementos e uma dinâmica de redirecionamento de conexões relativa ao confinamento. É demonstrado que dito modelo é capaz de produzir redes dinâmicas com coeficientes de agrupamento elevado, distribuição de graus livre de escala e formação de comunidades estruturais.

O modelo proposto demonstra estabilidade de características topológicas em uma rede dinâmica, sendo capaz de crescimento e diminuição de rede, nascimento e envelhecimento de nodos, escalável para diferentes tamanhos de rede. Este modelo dá suporte à tese de confinamento como mediador de emergência de propriedades complexas de rede, em particular das comunidades estruturais, mas também pode servir como base para representação e estudo de sistemas empíricos, onde a característica dinâmica da rede e a presença de módulos são de especial interesse.

Palavras-chaves: Sistemas complexos, redes complexas, redes dinâmicas, modularidade, confinamento e comunidades estruturais.

ABSTRACT

Complex systems are made up of multiple interconnected parts, and show a number of notable properties. The so-called complex networks are representative networks of empirical systems, either through data collection from the systems, or produced from models that seek to emulate the characteristics of these empirical networks. The field of network science has demonstrated in the last three decades the existence of a series of common properties between systems originating from the most varied fields of scientific activity. One of these notable properties is the emergence, existence and dissolution of structural communities, parts of the network with elements highly interconnected and weakly connected to the rest of the network.

This document presents the thesis that confined spaces within a dynamic network are an element of network organization, capable of explaining the emergence of non-trivial topologies typical of complex networks: formation of cliques, heterogeneous connectivity degree distributions with formation of highly connected hubs, and emergence of structural communities. A simple computational model is proposed, based on the temporary confinement of elements and a dynamics of redirection of connections relative to the confinement. It is demonstrated that this model is capable of producing dynamic networks with high clustering coefficients, scale-free degree distribution and formation structural communities.

The proposed model demonstrates stability of topological characteristics in a dynamic network, is capable of network growth and shrinkage, birth and aging of nodes, and is scalable to different network sizes. This model supports the thesis of confinement as a mediator of the emergence of complex network properties, in particular structural communities, but can also serve as a basis for representing and studying empirical systems where the dynamic characteristic of the network and the presence of modules are of special interest.

Key-words: complex systems, complex networks, dynamic networks, modularity, confinement. structural communities.

LISTA DE ILUSTRAÇÕES

- Figura 1 – Pontes de Könisberg e representação de uma topologia como rede. *Painel esquerdo:* Mapa de Könisberg (hoje Kalingrad) na época de Euler. Existem sete pontes conectando duas margens do rio e duas ilhas. *Painel direito:* representação da topologia utilizando um grafo. As pontes são as linhas, ou conexões, e os pontos, ou nodos, são as margens e ilhas. Nesta representação a visualização pode ser distorcida, sem modificar a topologia do problema. 24
- Figura 2 – Diferentes formas de representação de redes: A) Representação simples com pontos representando nodos e linhas representando conexões; B) Representação de uma rede direcionada, onde as conexões são representadas por setas e a relação entre nodos tem algum tipo de direcionalidade. Existe uma representação de dupla conexão entre os nodos B e C; C) Rede conectada com pesos, as conexões têm diferentes espessuras de acordo com seu peso, existe um nodo com autoconexão, representado por uma linha curva no nodo G; D) Rede multicamada, um sistema composto de diferentes camadas de redes, conectadas entre si. 25
- Figura 3 – Representação de rede como matriz de adjacência. *À esquerda:* representação gráfica de uma rede de tamanho $N = 7$, não direcional e sem pesos, os nodos são numerados de 1 a 7 e as linhas entre eles são as conexões. *À direita* representação da mesma rede como matriz de adjacência: nesta situação, as linhas e colunas representam os nodos. Os elementos são $A_{ij} = 1$, em caso de conexão, e $A_{ij} = 0$, quando não conectados. 26
- Figura 4 – Algoritmo W-S. *Painel superior:* processo de randomização entre os limites da rede ordenada e uma rede randômica (*fonte Watts e Strogatz (1)*). *Painel inferior:* efeitos da randomização sobre os parâmetros caminho médio $\langle d \rangle$ e coeficiente de agrupamento C em termos dos valores iniciais $\langle d_0 \rangle$ e C_0 . Para uma razão muito pequena de conexões randomizadas, a rede ordenada se torna um "mundo pequeno", mantendo boa parte de seu ordenamento. 36

- Figura 5 – Comparação entre distribuições de grau $P_{(k)}$ de uma rede randômica (Poisson) e uma rede livre de escala (lei de potência). As duas distribuições correspondem a redes geradas computacionalmente, utilizando o mesmo tamanho $N = 10^5$ e densidade, com grau médio $\langle k \rangle = 10$. *Painel esquerdo*: o grau médio das duas distribuições $\langle k \rangle = 10$ é apontado com uma linha vertical. A cauda longa da lei de potências possibilita o aparecimento de hubs na rede. A figura é, na verdade, um zoom na escala de k para melhor visualização da distribuição randômica, o grau máximo da rede livre de escala é da ordem de 10^3 . *Painel direito*: $\log(P_{(k)})$ vs $\log(k)$. É mais fácil identificar distribuições livres de escala na forma $\log(P_{(k)}) = -\gamma \log(k)$ 37
- Figura 6 – Fase de Fluxo: Em uma rede de tamanho N , uma quantidade de nodos $N_c = Np_s$ formam o subconjunto confinado. A cada passo temporal uma quantidade $N_f = Np_f$ de nodos pertencentes ao confinamento deixam o conjunto N_c , e a mesma quantidade de nodos não confinados passa a integrar o confinamento. Durante a dinâmica nodos são confinados e libertados a cada passo temporal. 51
- Figura 7 – Fase de reconexões: *Painel A*: No esquema conexões internas ℓ_{in} são representada em vermelho, externas ℓ_{out} em azul e conexões interfaciais ℓ_{io} em laranja. *Painel B*: Uma quantidade de nodos interfaciais (linhas tracejadas) é selecionada randomicamente para reconexão. *Painel C*: conexões interfaciais selecionadas são convertidas em conexões internas, ligando o nodo confinado original a outro nodo confinado, aumentando a densidade do espaço confinado. 52
- Figura 8 – Evolução temporal do coeficiente de agrupamento global C devido a dinâmica de confinamento com um único espaço confinado. Redes iniciais têm tamanho $N = 10^4$, e $\langle k \rangle = 8$, com agrupamento muito próximo a zero $C = 0.0004$. O espaço confinado \mathcal{P} tem tamanho relativo $p_s = 0.01$ para todas as curvas apresentadas, enquanto o fluxo relativo p_f é diferente para cada curva. Após um transiente de passos proporcionais a p_f , C alcança um estado estacionário. C é medido apenas na maior componente conexa. 56
- Figura 9 – Mapeamento do coeficiente de agrupamento (da MCC) em função dos parâmetros p_s e p_f . Um fluxo menor resulta em nodos, permanecendo mais tempo dentro de \mathcal{P} , aumentando a probabilidade de formar triângulos no espaço confinado. A rede inicial é aleatória com $N = 10^4$ e $\langle k \rangle = 8$, com um coeficiente de agrupamento próximo de zero, $c = 0,0004$ 57

Figura 10 – Propagação de cliques na rede. Representação gráfica de uma rede $N = 10^4$, $\langle k \rangle = 4$ gerada com parâmetros $p_s = 0.03$ e $p_f = 0.0025$. Um algoritmo de força, com condição de não sobreposição de nodos, é usado para a visualização; o tamanho dos nodos é proporcional ao seu coeficiente local de agrupamento C_i . O espaço confinado é representado pela cor vermelha. Os cliques são formados no espaço confinados, em vermelho, mas disseminados em toda a rede. 58

Figura 11 – A partir de uma rede randômica inicial, de tamanho $N = 10^4$ e $\langle k \rangle = 4$, a dinâmica de confinamento cria uma rede estável e mais complexa. A figura mostra o efeito de um único espaço confinado, onde os parâmetros dinâmicos são $p_s = 0,03$ e $p_f = 0,0025$. Nesta representação visual da rede resultante, um algoritmo de força é usado para o posicionamento dos nodos (2), resultando em nodos fortemente conectados sendo agrupados. O tamanho dos nodos é proporcional ao grau de conectividade. *Painel superior:* os nodos são codificados por cores de acordo com o particionamento resultante do algoritmo de detecção da comunidade (3). *Painel Inferior:* Com a mesma disposição de nodos e conexões, os nodos são codificados por cores, como confinados em vermelho ou livres em azul. Durante a dinâmica, os nós entram e saem do confinamento, e os links são religados, mas as características gerais da rede permanecem as mesmas. 60

Figura 12 – Modularidade associada ao particionamento das redes pelo algoritmo de detecção, dados valores diferentes de p_s e p_f . Painel esquerdo: evolução temporal para diversos parâmetros a partir da rede aleatória inicial. Painel direito: mapeamento de modularidade como uma função de p_s e p_f . A rede inicial, em todas simulações, é randômica, com $N = 10^4$ e $\langle k \rangle = 8$ 63

Figura 13 – *Painel Esquerdo:* Evolução temporal de Φ para alguns parâmetros de escolha, (os mesmos usados na figura 12). A razão inicial superior a $\Phi = 0.6$, e diminui com o tempo em todos os casos onde $p_s \neq p_f$, atingindo um estado estável. *Painel direito:* diagrama de fases mostrando a razão Φ em relação aos parâmetros p_s e p_f . A rede inicial em todas simulações é randômica, com $N = 10^4$ e $\langle k \rangle = 8$ 64

Figura 14 – A partir de uma rede inicial randômica de tamanho $N = 10^4$ e $\langle k \rangle = 8$ e distribuição de grau de Poisson, $P_{(k)}$ se transforma tornando-se mais heterogênea. Esta figura procura isolar os efeitos de diferentes parâmetros dinâmicos p_s e p_f na heterogeneidade de grau. *Painel esquerdo:* usando uma taxa de fluxo fixa $p_f = 0.003$, três curvas são mostradas, cada uma correspondendo ao $P_{(k)}$ estável final de um tamanho de confinamento diferente p_s . *Painel direito:* usando um tamanho de confinamento fixo $p_s = 0.03$, outras três distribuições são mostradas, cada uma usando um valor diferente de p_f . Os valores fixos foram escolhidos por estarem aproximadamente no centro da região de interesse de p_s e p_f apresentadas neste capítulo 65

Figura 15 – Heterogeneidade de grau produzida pela dinâmica de confinamento único: diferentes parâmetros dinâmicos geram formas distintas de distribuição de graus. Para o valor fixo do espaço confinado $p_s = 0.035$, são mostradas duas distribuições, uma com p_f maior (triângulos vermelhos) justaposta com ajuste de função log-normal (linha pontilhada preta) e a outra com p_f menor (círculos roxos) ajustada a uma lei de potência. Em geral, um tempo maior dentro de \mathcal{P} significa a formação de mais hubs e maior heterogeneidade de grau, portanto um espaço confinado maior com menos fluxo resulta em distribuições mais heterogêneas. 66

Figura 16 – Tamanho da maior componente conexa (MCC) no modelo de confinamento. *Painel esquerdo:* evolução temporal da maior componente conexa para uma rede aleatória inicial com $N = 10^4$ e $\langle k \rangle = 8$. O tamanho da componente é dado como fração do tamanho total N da rede e são mostradas curvas para diferentes conjuntos de parâmetros. Após um estado transiente, a componente se estabiliza, enquanto os nós saem e entram novamente na MCC devido a dinâmica de confinamento. *Painel direito:* diagrama de fase mostrando o tamanho da MMC em função dos parâmetros p_s e p_f . O tamanho da componente é novamente dado em relação ao tamanho N da rede. Regiões azuis escuras representando um componente maior. Maiores espaços de confinamento resultam em maior redução do grau de conectividade dos nós fora de \mathcal{P} , resultando em mais desconexões. Da mesma forma, com um fluxo menor, menos nós componentes desconexas são religadas. 67

Figura 17 – Invariância de métricas frente a mudança de tamanho. *Painel esquerdo:* evolução temporal de algumas das medidas topológicas para um tamanho absoluto de confinamento fixo (o número de nós confinados é $N_c = 50$ e o número de nós de fluxo é $N_f = 5$. O número de espaços confinados para cada tamanho de rede N é $\mathcal{P} = N/1000$. *Painel direito:* diferentes tamanhos de rede resultam em distribuição de graus diferentes, mas com o método proposto a forma da distribuição de graus é mantida, salvo por um fator multiplicativo de tamanho (ou deslocamento vertical em escala logarítmica). 70

LISTA DE TABELAS

Tabela 1 – Comparação de parâmetros topológicos de redes empíricas com os mesmos parâmetros esperados em uma rede randômica de tamanho e densidade de conexões similar. Fonte: Albert & Barabási (4).	34
Tabela 2 – Alguns exemplos de ocorrência de redes livre de escala em sistemas empíricos. Redes não direcionais são marcadas com asterisco e o coeficiente de potência γ_{in} e γ_{out} são iguais. Para redes direcionadas coeficientes γ_{in} e γ_{out} são distintos. Fonte: Albert & Barabási (4)	38

LISTA DE ABREVIATURAS E SIGLAS

B-A	Barabási-Albert.
CV	Comando vermelho.
E-R	Erdős-Rényi.
W-S	Watts-Strogatz.
MCC	Máxima componente conexa.
PCC	Primeiro comando da capital.

LISTA DE SÍMBOLOS

\mathcal{N}	Designação genérica de um grafo / rede.
N	Tamanho de rede dado em número de nodos.
ℓ	Número de conexões do grafo.
n_j	Referência a nodo individual numerado j .
$\ell_{i,j}$	Conexão.
N_α	Referência a um subgrupo de nodos α dentro da rede.
A	Matriz de conectividade.
A_{ij}	Elemento da matriz de conectividade.
k_i	Grau de conectividade de um nodo individual n_j .
$\langle k \rangle$	Grau de conectividade médio.
$P_{(k)}$	Distribuição de graus de conectividade;
\mathcal{P}	Subgrafo confinamento; espaço confinado.
$d_{i,j}$	Distância (tamanho de caminho) entre dois nodos arbitrários n_i e n_j , dado em número de conexões percorridas.
$\langle d \rangle$	Distância média da rede.
D	Diâmetro da rede (maior caminho presente entre dois nodos na rede).
M	Modularidade global de dado particionamento de rede.
m	Modularidade local de uma partição.
p_s	Tamanho relativo do espaço confinado dado em percentual de nodos totais.
p_f	Tamanho relativo do fluxo dado em percentual de nodos totais.

Observação: todos os símbolos listados para referência rápida são introduzidos e/ou contextualizados ao longo do texto.

SUMÁRIO

	INTRODUÇÃO	17
1	TEORIA DE REDES COMPLEXAS	20
1.1	Redes como metodologia de estudo de sistemas complexos	20
1.2	Redes complexas	23
1.2.1	Matriz de adjacência	26
1.2.2	Grau de conectividade	27
1.2.3	Distribuição de grau de conectividade	27
1.2.4	Densidade de rede	28
1.2.5	Coefficiente de Agrupamento	29
1.2.6	Caminho, caminho médio e diâmetro de rede	30
1.2.7	Maior componente conexa e componente gigante	30
1.3	Considerações finais do capítulo	31
2	REDES EMPÍRICAS E SUAS REPRESENTAÇÕES	32
2.1	Redes randômicas e o modelo Erdős–Rényi	32
2.2	Mundos pequenos, agrupamento e o modelo Watts–Strogatz	34
2.3	Redes livres de escala e o Modelo B-A	36
2.4	Considerações finais do capítulo	39
3	COMUNIDADES ESTRUTURAIS EM REDES COMPLEXAS	41
3.1	Hipóteses gerais e modularidade	41
3.2	Particionamento de rede e Modularidade	43
3.2.1	Algoritmo de Louvain para detecção de comunidades	44
3.3	Limitações do algoritmo de detecção	45
3.4	Algoritmo Girvan-Newman	46
3.5	Considerações finais do capítulo	46
4	REDES DINÂMICAS MEDIADAS PELO CONFINAMENTO	47
4.1	Caracterização de confinamento	47
4.1.1	Dinâmica de confinamento temporal	48
4.2	Redes dinâmicas	49
4.3	Modelo de redes dinâmicas mediadas pelo confinamento	49
4.4	Metodologia	53
5	RESULTADOS	55
5.1	Aumento do coeficiente de agrupamento	55

5.2	Formação de comunidades estruturais	58
5.3	Distribuição de graus de conectividade	64
5.4	Fragmentação da Rede	66
5.5	Propriedades adicionais do modelo	68
5.5.1	Efeitos de múltiplos confinamentos	68
5.5.2	Independência da topologia em relação a redes iniciais e modificação dinâmica de parâmetros	69
5.5.3	Crescimento, redução da rede e transformação contínua da rede	70
5.6	Considerações finais do capítulo	71
	 CONCLUSÕES	 72
	 REFERÊNCIAS	 77
	 APÊNDICES	 87
	 APÊNDICE A – IMPLEMENTAÇÃO COMPUTACIONAL	 88

INTRODUÇÃO

Sistemas complexos são formados por diversos sub-sistemas interconectados, onde a disposição dos elementos, a topologia de suas interações contém os mecanismos de seu funcionamento coletivo. Esta classe especial de sistemas é bem representada por redes complexas: representações teóricas das partes e suas iterações que permitem estudar a complexidade do ponto de vista relacional. De fato, todo o campo interdisciplinar das ciências de rede é dedicado a enfrentar esta classe de problema.

Um dos desafios centrais, no campo de ciências de rede, é o estudo de comunidades. Estas estruturas topológicas, também chamadas de clusters ou módulos, são tipicamente caracterizadas como regiões da rede mais densas e de iteração mais forte. A identificação destas subestruturas tem diversas utilidades: a caracterização de subsistemas, classificação de componentes similares, correlação de efeitos e comportamentos, planejamento de intervenções estabilizadoras ou destrutivas em sistemas, dentre muitas outras. Comunidades estruturais surgem em todas as disciplinas perpassadas pelo estudo da complexidade como rede e as motivações e aplicações são múltiplas e, em grande parte, ainda inexploradas.

O estudo da dinâmica, surgimento, funcionamento e dinâmica de comunidades estruturais no contexto de redes dinâmicas é um assunto na ponta da lança da pesquisa e desenvolvimento de ferramentas de análise computacional, coletas de dados e simulações. A caracterização destas comunidades em redes dinâmicas não é trivial e eleva os desafios de coleta e análise de dados de redes empíricas a uma barra mais elevada, e desafiam as limitações dos modelos teóricos estáticos.

Neste trabalho, trata-se do assunto de comunidades estruturais no contexto de redes dinâmicas, em particular, é levantada a hipótese que os espaços confinados agem como limitadores de iteração, mas também como organizadores de complexidade de redes. A dinâmica de confinamento que é proposta neste trabalho é ao mesmo tempo simples, intuitiva e rica em resultados, sendo capaz de reproduzir uma série de comportamentos característicos de redes empíricas complexas sem utilização de parâmetros diretos. Neste sentido, procura-se demonstrar que o confinamento é um elemento organizador de fenômenos associados à complexidade e que este mecanismo é relevante em uma série de sistemas de interesse.

Para dar suporte a esta hipótese, um sistema simples de dinâmica de rede é apresentado, baseado em confinamento temporário e reconfiguração de conexões. Através de resultados obtidos de simulações deste modelo, características típicas de sistemas complexos emergem. De particular interesse é o surgimento de comunidades estruturais, ou módulos, em redes dinâmicas estáveis. Acredita-se que o modelo apresentado pode ser utilizado

como substrato para outros modelos e estudos, onde a presença de espaços confinados ou a existência de comunidades estruturais dinâmicas representam papel importante. O modelo também lança entendimento adicional sobre as mecânicas de emergência de fenômenos complexos em redes dinâmicas, buscando um avanço no entendimento de seus mecanismos organizadores.

Uma das características notáveis do modelo proposto é produzir uma topologia complexa estável em uma rede dinâmica. Este fenômeno, em que os componentes do sistema mudam de posição na topologia, mas a forma e função geral se mantém é uma característica importante dos sistemas complexos em geral, e está associado às semelhanças topológicas que surgem entre fotografias de redes em diferentes áreas e em diferentes tempos, um indicativo de emergência de ordenamentos e dinâmicas de grupo que o sistema proposto consegue reproduzir.

O documento está estruturado da seguinte forma: nos três primeiros capítulos é feita uma revisão bibliográfica do estado da arte, estando o escopo desta revisão limitado aos conceitos associados diretamente aos resultados principais obtidos. No primeiro capítulo, “*Teoria de redes complexas*”, as redes complexas são definidas como forma de representação de sistemas complexos em geral, são apresentados exemplos de representações e aplicações destas representações em diferentes campos de estudo (engenharias, ciências sociais, sistemas biológicos, economia etc.). Na sequência, são introduzidos conceitos estatísticos e ferramentas de análise destas redes complexas, novamente, com foco nos conceitos pertinentes ao modelo e à tese apresentada.

No segundo capítulo, “*Redes empíricas e suas representações*”, alguns modelos de representação de sistemas empíricos são expostos introduzindo características notáveis deste sistemas como presença de subgrupos altamente conectados, distribuição heterogênea de conexões, e efeitos de mundo pequeno. As semelhanças e limitações destes modelos são discutidas, mas também no contexto de limitações gerais de modelos geradores de redes.

No terceiro capítulo, “*Comunidades estruturais em redes complexas*”, trata-se de comunidades estruturais. As definições de comunidades estruturais são apresentadas, apoiadas por hipóteses que as fundamentam. A caracterização de comunidades em uma rede complexa está diretamente ligada à sua possibilidade de aplicação, na detecção destas estruturas em redes empíricas grandes, requerendo uso de algoritmos de otimização computacional. O algoritmo utilizado para detecção de comunidades nesta pesquisa é apresentado, e as suas limitações e vantagens discutidas.

No quarto capítulo, “*Redes dinâmicas mediadas pelo confinamento*”, é introduzida a ideia de dinâmica de rede e de rede dinâmica. O conceito de confinamento é definido de maneira específica, e um modelo de redes dinâmicas mediadas pelo confinamento é apresentado. Este modelo, inicialmente proposto no artigo “*Dynamic Modular Networks Model Mediated by Confinement*” (5), é utilizado, junto com simulações computacionais,

como metodologia de estudo deste trabalho.

No quinto capítulo, “*Resultados*”, são apresentados os resultados das simulações. Resultados notáveis, como formação de cliques altamente conectados, surgimento de redes livres de escala, e de comunidades estruturais. Os mecanismos de organização de rede do modelo são explorados mais profundamente no contexto dos parâmetros propostos do sistema. Finalmente, propriedades adicionais do modelo referentes à fenomenologia de redes dinâmicas, como capacidade de crescimento e diminuição de rede, transformação de topologia como parte da dinâmica, e resistência a perturbações, são discutidas.

Acredita-se que o modelo apresentado é original, ao mesmo tempo simples e rico em geração de complexidade, e que lida com um assunto importante de interesse central em sua área de estudo: evolução de comunidades estruturais em redes dinâmicas. Existe também o potencial de uso deste modelo como substrato para aplicações correlacionadas e abertura de um espectro grande de possíveis pesquisas futuras.

1 TEORIA DE REDES COMPLEXAS

Neste capítulo, será feita uma revisão bibliográfica sobre o tema de ciência de redes, que é um campo de estudos interdisciplinar, onde a tese defendida neste documento está situada. Este campo de estudo será contextualizado dentro da área de estudos de sistemas complexos. A forma de representação destes sistemas como redes complexas será apresentada, e os conceitos principais da disciplina, pertinentes à discussão proposta nesta tese, serão delineados.

1.1 Redes como metodologia de estudo de sistemas complexos

Quando se diz que um sistema é complexo, é feita uma afirmação sobre a natureza do sistema que direciona a forma como seu estudo precisa ser abordado: de que o sistema em questão é constituído de múltiplas partes, e a forma como estas partes interagem é chave para o entendimento de seu comportamento. Mais do que isso, é proposta a impossibilidade de entendimento do sistema como soma simples de suas partes. Em geral, sistemas complexos apresentam algumas propriedades/características notáveis, como seguem abaixo:

Não linearidade: as perturbações do sistema não geram reações proporcionais facilmente previsíveis pela soma dos efeitos ou partes individuais, de forma que os estados, variáveis ou métricas resultantes não são proporcionais às perturbações de entrada.

Podem ser abertos: sistemas complexos são capazes de absorver e dissipar energia, aumentar ou diminuir de tamanho, perder ou adquirir componentes enquanto mantêm suas características principais de funcionamento. Com frequência, todos os componentes individuais de um sistema complexo podem ser substituídos ao longo do tempo, enquanto sua forma geral permanece reconhecível.

Presença de sistemas aninhados: sistemas complexos, são tipicamente compostos de elementos que, observados individualmente, também são sistemas complexos. Sistemas sociais humanos, por exemplo, podem ser divididos em indivíduos e suas interações, os indivíduos, por sua vez, são entes biológicos individualmente complexos. É possível (e normalmente preferível) estudar estes sistemas em diferentes escalas, sem levar em consideração toda a complexidade individual dos componentes, mas apenas seus comportamentos de interconexão.

Vulnerabilidade a falhas em cascata: em sistemas interconectados, modificações ou falhas de componentes individuais, podem ter consequências na totalidade do sistema. Mesmo quando estes componentes interagem com uma quantidade muito pequena da tota-

lidade, as falhas e alterações podem se propagar rapidamente, descaracterizando o sistema ou mudando suas propriedades gerais. Assim, em sistemas bancários interconectados, a falta de liquidez de uma instituição financeira, pode gerar uma sequência de inadimplência que se propaga rapidamente, com a quebra de diversas instituições financeiras de maneira quase simultânea. Em sistemas biológicos, outro exemplo, uma alteração ou dano a um único órgão pode levar à falha geral do sistema e à morte. Em cadeias produtivas e logísticas, a falta de um único insumo ou interrupção de uma via de fornecimento pode se propagar para uma crise de abastecimento generalizada em diversas áreas não diretamente conectadas.

Circuitos de realimentação: são sistemas sensíveis ao seu próprio estado, ou seja, as variáveis de saída do sistema podem ser variáveis de entrada, gerando ciclos de amortecimento e estabilidade ou de amplificação rápida de perturbações internas sem a necessidade de perturbações externas.

Presença de fenômenos emergentes: sistemas complexos apresentam características, propriedades e fenômenos que são observáveis apenas através da interação das partes. Padrões não aleatórios ou propriedades emergem como efeito das interações entre partes e com frequência podem apenas ser medidas ou observadas no contexto destas interações. O entendimento de propriedades emergentes são tanto uma propriedade como o objetivo do estudo de sistemas complexos. Exemplificando esta característica: os surtos epidêmicos de doenças, em diferentes espécies, têm um grau de impacto significativo em todos os campos da atividade humana. Mesmo que os mecanismos de ação da doença nos organismos sejam conhecidos, o entendimento da propagação do surto só podem ser observados ou compreendidos no contexto da rede de contágio. A epidemia é uma propriedade emergente da população afetada, mais do que da doença específica.

A natureza é, em geral, complexa e não linear, com o problema da complexidade perpassando todas as áreas de estudo. Representar a disposição e interação entre as partes em uma estrutura teórica é fundamental para o seu entendimento. A representação destes sistemas em forma de redes, associadas ao aumento da capacidade de coleta, armazenamento e processamento de dados, que ocorreu nas últimas três décadas, tem lançado uma nova compreensão sobre a complexidade, tornando-se uma área interdisciplinar destacada: a *ciência de redes*.

A natureza interdisciplinar da chamada *ciência de redes* está associada à gama diversa de técnicas utilizadas em sua formalização (álgebra linear, estatística, campos diversos da matemática, modelos físicos, ciências da computação, engenharia de sistemas da informação, etc) mas também pela generalidade de sua aplicação: sistemas complexos são objeto de estudo em quase todos os campos das ciências, e as redes destes sistemas apresentam similaridades notáveis. Redes empíricas com características estatísticas similares surgem na representação de conexões entre dispositivos de comunicação (antenas,

servidores, links de satélites, etc.) (6, 7), *hyperlinks* em *webpages* (8, 9), redes de transportes e locomoção (10, 11, 12), cadeias produtivas (13, 14, 15). Em ciências biológicas, redes complexas características se apresentam nas ligações entre neurônios (16, 17, 18), nas iterações entre cadeias de proteínas (19, 20, 21), interpretação de marcadores genéticos (22, 23), oncologia (24, 25, 26), entre outros.

Os diversos campos da atividade humana, e os fenômenos relacionados, ocorrem dentro de um substrato que é a rede de conexões interpessoais. Essas são as chamadas *redes sociais humanas*, que no estudo de ciências de redes recebem destaque em estudo e aplicações. Aqui se refere a categoria de redes associadas a interações humanas em geral (que, embora incluam, não se limitam a plataformas digitais). Redes sociais são redes de amizades, contatos profissionais e de parentesco, redes de contato romântico e sexual, topologias de posicionamento geográfico e movimentações diárias, etc, que possuem tipicamente indivíduos ou grupos de indivíduos como nodos, e algum tipo de relação ou iteração como conexão.

Esta é uma definição frequente, mas não rígida, com redes sociais podendo ser definidas com indivíduos ou grupos como conexões e os nodos como alguma outra classe de conceito de interesse social. De maneira mais ampla, redes empíricas associadas à atividade social ou estudo social são, às vezes, referenciadas na literatura como redes sociais, mesmo que os nodos não sejam pessoas ou que as relações entre eles não sejam classificadas tipicamente como sociais. Essa definição difusa, do que constitui uma rede social, não impede seu estudo no contexto de ciência de redes. O fator importante é que essas redes representam sistemas complexos, ou seja, que o entendimento dos sistemas que representam depende da topologia dos elementos. Veremos nas seções seguintes que estas redes apresentam semelhanças estatísticas e topológicas não triviais e mensuráveis.

A consolidação das ciências de redes ocorre a partir do final do século XX. No final da década de 90, o acesso à infraestrutura computacional atingiu um limiar em que a coleta, armazenamento de dados e a análise destes dados tornou possível o levantamento de uma série de redes empíricas, e ao estudo destas redes, dando origem ao que inicialmente foi chamado, na física, de “mecânica estatística das redes complexas”. Nas duas décadas que seguiram, ocorreu uma multiplicação de publicações, demonstrando novos resultados e técnicas de análise e aplicação.

A epidemiologia é um exemplo de aplicação bem sucedida das ciências de rede. Em torno de 2000-2010 foi possível, aplicando modelos clássicos de disseminação de doenças (27) e simulando-os, como fenômeno de contágio em redes complexas, explicar a ausência de um limiar epidêmico de doenças em alguns tipos de redes (28, 29) e propor melhores estratégias de imunização contra propagação viral em humanos ou computadores (30, 31). Em 2009 foi possível explicar e prever parcialmente a propagação da epidemia de Influenza H1N1 utilizando modelos baseados em propagação em redes complexas (32, 33) e, posteriormente,

o surto de ebola de 2014-2015 (34). Os modelos de disseminação de patógenos se tornaram um assunto de interesse estratégico, iniciando em 2010 e culminando na atualidade (35). Um resultado notável destes modelos é sua capacidade de representar também outros fenômenos de contágio em rede, como propagação de opiniões e comportamentos em organizações e na sociedade (36, 37, 38, 39) ou por meios digitais (40, 41, 42). Este é um indício da generalidade da abordagem de sistemas complexos utilizando as mesmas técnicas.

Esta generalidade de aplicação dos modelos e análises em diferentes campos de investigação, aparentemente não conectados, ocorre principalmente porque os objetos de estudo (as redes empíricas que representam a disposição e interação dos elementos destes sistemas complexos), apresentam similaridades matemáticas e estatísticas mensuráveis (mesmo quando representam objetos de estudo radicalmente diferentes). Nas próximas seções deste capítulo, serão introduzidos conceitos e metodologias de representação e análise de redes complexas presentes na literatura e no contexto da ciência de redes moderna. O tema é bastante amplo e uma revisão completa da disciplina foge ao escopo desta tese, mas os pontos mais importantes para o entendimento e situação do modelo e da tese apresentada serão apresentados.

1.2 Redes complexas

Como comentado previamente, nas últimas três décadas o campo de estudo das redes complexas passou por um rápido crescimento em escala de produção científica e aplicações, com a popularização de ferramentas computacionais de pesquisa e infraestrutura de armazenamento e processamento de dados, vista nas últimas décadas, desempenhando papel auxiliar indispensável. A representação de sistemas como redes ou estudo dessas representações, no entanto, não é novidade no campo da matemática. Tipicamente se atribui a Leonard Euler o início da “teoria de grafos” na matemática, através de seu célebre artigo sobre as pontes de Könisberg (43), embora este artigo tenha inspirado muito da produção neste campo posteriormente, Euler não se dedicou à formalização ou à estruturação desta área de estudo. Existem referências históricas variadas e não é possível dar um marco inicial para o uso de grafos na resolução de problemas. Leibniz teria já manifestado a necessidade de uma “geometria das posições” para certas classes de problemas, e o termo topologia teria sido cunhado (anedoticamente) por um discípulo de Gauss (44).

O exemplo das pontes de Könisberg, mostrado na figura 1, é instrutivo para introduzir a representação de um sistema por rede no contexto de uma topologia geográfica. No exemplo citado cada região da cidade pode ser reduzida a um ponto, e cada ponte pode ser reduzida a uma linha entre dois pontos. Generalizando esta representação, as linhas são

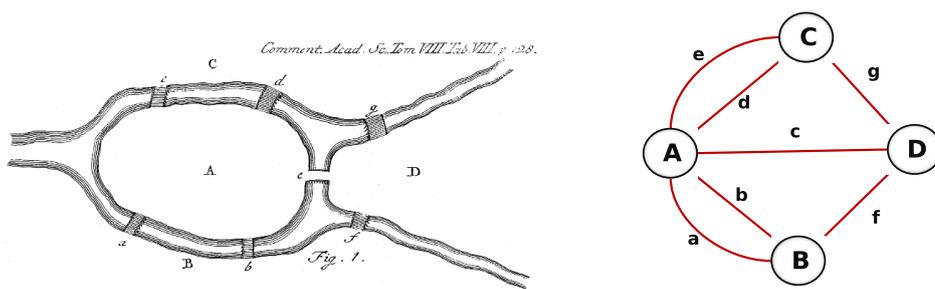


Figura 1 – Pontes de Könisberg e representação de uma topologia como rede. *Painel esquerdo*: Mapa de Könisberg (hoje Kalingrad) na época de Euler. Existem sete pontes conectando duas margens do rio e duas ilhas. *Painel direito*: representação da topologia utilizando um grafo. As pontes são as linhas, ou conexões, e os pontos, ou nodos, são as margens e ilhas. Nesta representação a visualização pode ser distorcida, sem modificar a topologia do problema.

chamadas de conexões, links, ou arestas, e os pontos são chamados de nodos ou vértices. Vértices e arestas são nomes utilizados com mais frequência na matemática, onde a forma de representação é chamada grafo. Em ciências de rede grafos são usados para representar sistemas em rede, e o nome rede é usado de maneira intercambiável frequentemente. Neste texto será utilizada a terminologia que define uma rede \mathcal{N} como um conjunto contendo um número N de nodos, conectados por um número total ℓ de conexões.

A forma de representação pode conter informação adicional, conforme demonstrado na figura 2. Conexões podem ser direcionais quando a interação representada tem uma direcionalidade, graficamente as linhas seriam, nesse caso, representadas por setas. Conexões também podem ser representadas com um peso associado, para representar a ideia que nem todas as conexões entre nodos são iguais, com diferenças de fluxo, grandeza, importância, etc. Em algumas representações, pode ser útil representar conexões duplas (duas ou mais conexões entre um par de nodos) ou autoconexões, de um nodo para si mesmo. Alguns sistemas possuem diferentes classes de nodos e conexões, cada classe é representada como uma camada com nodos e conexões, e as camadas possuem conexões entre si, formando redes multi-camadas. Para a metodologia e modelo apresentados neste documento, a representação será limitada a redes de uma camada com N nodos e ℓ conexões não direcionais, e sem pesos.

Grafos podem ser utilizados para representar as correlações internas de uma série de sistemas, bastando definir os objetos que são nodos (as partes individuais do sistema) e alguma categoria de interação, entre pares, como conexões. Na biologia podem representar co-ocorrência de doenças (45) (doenças são nodos e co-ocorrências são conexões) ou comorbidade de doenças (46) (nodos são comorbidades significativas e as doenças são as conexões), diferentes tipos de topologia cerebral (47, 48, 49) (diferentes regiões anatômicas do cérebro são nodos, e iterações são conexões), rede social de golfinhos (50) (golfinhos são

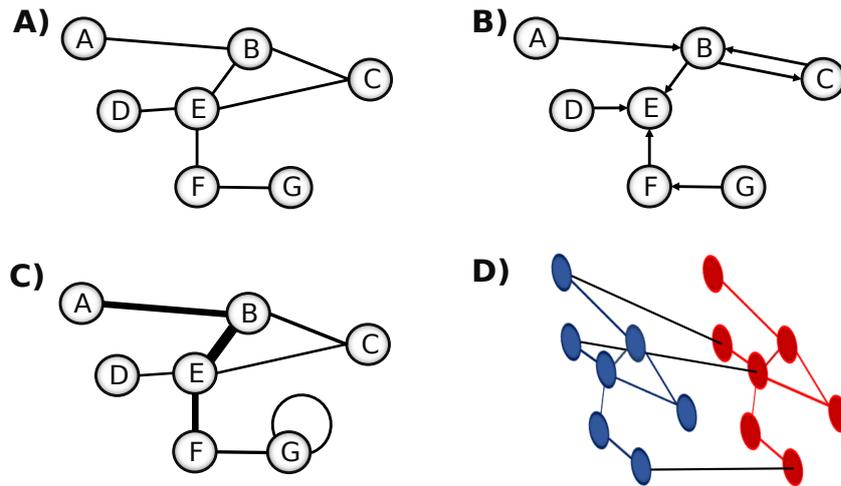


Figura 2 – Diferentes formas de representação de redes: A) Representação simples com pontos representando nodos e linhas representando conexões; B) Representação de uma rede direcionada, onde as conexões são representadas por setas e a relação entre nodos tem algum tipo de direcionalidade. Existe uma representação de dupla conexão entre os nodos B e C; C) Rede conectada com pesos, as conexões têm diferentes espessuras de acordo com seu peso, existe um nodo com autoconexão, representado por uma linha curva no nodo G; D) Rede multicamada, um sistema composto de diferentes camadas de redes, conectadas entre si.

nodos e relacionamentos são conexões). Na infra-estrutura e tecnologia podem representar redes de distribuição de eletricidade (51, 52) (nodos são hubs, centrais ou quadros de distribuição e conexões são linhas de força), redes de distribuição hídrica (nodos são reservatórios, tanques e junções, conexões são válvulas, bombas e tubos), malha rodoviária (53) (intersecções são nodos e conexões são vias). Nas áreas econômicas pode representar comércio internacional (54) (países são nodos e as relações comerciais são conexões), correlações entre diferentes ações na bolsa (55) (ações específicas são nodos, correlações são conexões), redes bancárias (56) (Bancos são nodos, fluxo de pagamentos são conexões). Em redes sociais temos representações nos mais diversos campos da atividade humana: co-autoria de artigos na área de ciência de redes (57) (nodos são autores e conexões são co-autorias), redes criminais (58, 59, 60, 61, 62) (nodos são criminosos, vítimas e pessoas de interesse, conexões são relações ou crimes). Redes de amizades (63, 64, 65, 66, 67) (nodos são pessoas, conexões são relações de amizade).

Os exemplos apresentados são pontuais e não esgotam a possibilidade de áreas de atuação ou tipos de representação, estes exemplos pontuais apenas demonstram a diversidade de aplicações em áreas não similares, e que nodos e conexões podem representar objetos ou conceitos completamente diferentes. Um dos resultados mais relevantes das últimas décadas foi a verificação de que redes representando sistemas tão diversos compar-

tilham similaridades topológicas e estatísticas não triviais. Estas redes empíricas possuem características comuns que passaram a ser consideradas propriedades da complexidade dos sistemas, resultando de mecanismos de ordenamento inerentes a sistemas interconectados de múltiplas partes. Nas próximas subseções deste capítulo, serão apresentadas as definições de alguns conceitos utilizados na análise destes sistemas.

1.2.1 Matriz de adjacência

Uma rede pode ser representada de maneira algébrica como uma matriz: a matriz de adjacência A . Nesta sub-seção será discutido o caso de matrizes de adjacência representando redes sem pesos, sem autoconexões e sem direcionalidade de conexões. Em uma rede de tamanho N , onde cada nodo individual n_i pode ou não estar conectado a outro nodo qualquer n_j , é construída uma matriz quadrada de tamanho N , onde cada linha e cada coluna correspondem a um nodo. Para uma rede onde as conexões não têm pesos, quando nodos n_i e n_j estão conectados, o elemento correspondente da matriz $A_{ij} = 1$, caso contrário $A_{ij} = 0$. A figura 3 exemplifica esta forma de representação.

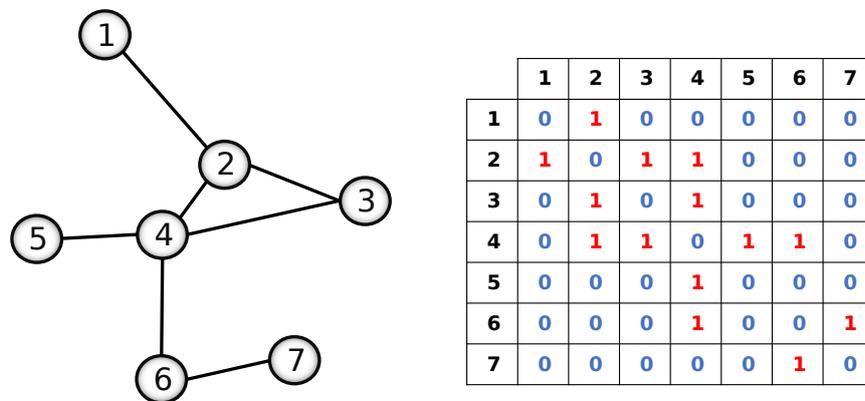


Figura 3 – Representação de rede como matriz de adjacência. À esquerda: representação gráfica de uma rede de tamanho $N = 7$, não direcional e sem pesos, os nodos são numerados de 1 a 7 e as linhas entre eles são as conexões. À direita representação da mesma rede como matriz de adjacência: nesta situação, as linhas e colunas representam os nodos. Os elementos são $A_{ij} = 1$, em caso de conexão, e $A_{ij} = 0$, quando não conectados.

Para uma matriz não direcional, sem pesos e sem autoconexões $A_{ij} = A_{ji}$, a diagonal principal $A_{i=j}$ é nula e a matriz é simétrica. Nestas condições, a quantidade total de conexões ℓ é obtida pela equação 1.1, que é válida também para obtenção de

quantidade de links em sub-partições da rede (a soma é dividida por dois pois $A_{ij} = A_{ji}$). A representação da rede, como matriz de adjacência, será utilizada, neste documento, para definir e analisar alguns conceitos e resultados, nestes casos sempre será considerado o caso de redes não direcionais, sem pesos e sem autoconexões.

$$\ell = \frac{1}{2} \sum_{ij} A_{ij} \quad (1.1)$$

A matriz A contém toda a informação topológica da rede, mas nem sempre é a melhor abordagem em ciência de redes. Redes empíricas podem ter um número de nodos muito grande e são normalmente esparsas em quantidades de conexões. Computacionalmente, o armazenamento e processamento de matrizes quadradas de tamanho N densas tende a complexidades $O(N^2)$ ou maiores. Algoritmos algébricos para lidar com matrizes esparsas (onde a maior parte dos elementos é nula) e algoritmos não algébricos são frequentemente usados para viabilizar a análise de redes complexas.

1.2.2 Grau de conectividade

O grau de conectividade k_i é uma propriedade topológica individual do nodo que representa o número de conexões do nodo n_i . Em termos de elementos da matriz A , o grau é definido pela equação 1.2.

$$k_i = \sum_j A_{ji} \quad (1.2)$$

Esta propriedade pode ser estendida para a rede, iterando sobre todos os graus individuais e obtendo o grau de conectividade médio $\langle k \rangle$, conforme equação 1.3. A segunda igualdade é obtida utilizando as equações 1.2 e 1.1.

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{\ell}{2N} \quad (1.3)$$

1.2.3 Distribuição de grau de conectividade

Para um dado tamanho de rede N e um grau de conectividade médio $\langle k \rangle$ existe uma grande possibilidade de combinações possíveis de configuração nas conexões: um único nodo conectado a todos os outros, ou uma distribuição homogênea em que todos os nodos têm o mesmo número de conexões $k_i = \langle k \rangle$. A forma desta distribuição é uma característica topológica de interesse.

Em uma rede \mathcal{N} de tamanho N qualquer, a probabilidade p_k de que um número de nodos N_k possuam uma conectividade k é dada pela equação 1.4.

$$p_k = \frac{N_k}{N} \quad (1.4)$$

Quando esta probabilidade é computada para todas as possibilidades de grau de conectividade k presentes na rede, chega-se a função distribuição de grau de conectividade, ou abreviadamente, distribuição de grau $P_{(k)}$. A normalização desta distribuição (equação 1.6) é garantida pela definição de p_k como soma de partes da rede.

$$\sum_{k=0}^{\infty} N_k = N \quad (1.5)$$

$$\sum_{k=0}^{\infty} p_k = \frac{1}{N} \sum_{k=0}^{\infty} N_k = 1 \quad (1.6)$$

A distribuição de grau $P_{(k)}$ pode assumir uma série de formas, pode-se imaginar um rede regular circular de N nodos onde cada nodo é conectado aos quatro vizinhos mais próximos: neste caso todos os nodos tem $k_i = 4$, $\langle k \rangle = 4$ e $P_{(k=4)} = 1$, e zero para todo $k \neq 4$. Uma rede com o mesmo tamanho N e grau médio $\langle k \rangle$ poderia possuir uma distribuição gaussiana de média k , ou mesmo proporcional a uma lei de potência (1.7).

$$P_{(k)} \propto \alpha^{-k} \quad (1.7)$$

Nestes exemplos, se observam dois casos (gaussiana e regular) onde a distribuição de graus é relativamente homogênea em torno do grau médio $\langle k \rangle$. No caso da lei de potência, a distribuição pode apresentar uma “cauda longa” na direção de conectividade elevada, com nodos individuais apresentando conectividades ordens de grandeza maiores do que a média ($k_i \gg \langle k \rangle$). No capítulo 2, distribuições de grau heterogêneas e leis de potência serão discutidas com mais profundidade no contexto de propriedades de redes empíricas.

1.2.4 Densidade de rede

Em uma rede de tamanho N completamente conectada, onde cada nodo está conectado a todos os demais, $\langle k \rangle = N - 1$, e o número de conexões é:

$$\ell_{max} = \frac{N(N - 1)}{2} \quad (1.8)$$

A equação 1.8 se reduz a aproximadamente $\frac{N^2}{2}$ para $N \gg 1$. Uma rede é considerada esparsa quando $\ell_{max} \gg \ell$. A densidade de rede $\rho_{\mathcal{N}}$ é uma medida do quão esparsa é uma rede \mathcal{N} . É definida como a razão entre a quantidade ℓ de conexões da rede e a quantidade máxima de conexões ℓ_{max} , e definido pela equação 1.9:

$$\rho_{\mathcal{N}} = \frac{\ell}{\ell_{max}} \quad (1.9)$$

ou, aplicando as equações 1.3 e 1.8 na expressão 1.9, a densidade é obtida em termos de N e $\langle k \rangle$ (equação 1.10).

$$\rho_{\mathcal{N}} = \frac{\langle k \rangle}{N - 1} \quad (1.10)$$

Redes empíricas de interesse, onde usualmente $N \gg \langle k \rangle$, implicando baixa densidade de rede ($\rho_{\mathcal{N}} \ll 1$), são redes esparsas em geral. Esta densidade não é necessariamente uniforme em toda rede, que pode conter algumas sub-redes altamente conectadas (cliques). Uma das maneiras de medir a presença destes cliques e flutuações de densidade é o coeficiente de agrupamento C .

1.2.5 Coeficiente de Agrupamento

Uma característica notável em diversas redes empíricas de interesse é a formação de sub-redes altamente interconectadas. Exemplos típicos, no contexto de redes sociais, são grupos familiares, ambientes de trabalho ou círculos de amizades, onde as conexões de um nodo têm grande chance de estar conectadas entre si, formando “triângulos” ou tríades de nodos. O coeficiente de agrupamento individual C_i procura quantificar esta característica para cada nodo.

Dado um nodo n_i , com conectividade k_i , define-se sua vizinhança como os nodos diretamente conectados, e quantidade de conexões entre os vizinhos como sendo e_i . Para uma vizinhança completamente conectada, portanto, tem-se $e_i = (k_i(k_i - 1))/2$. O Coeficiente de agrupamento local do nodo n_i é a razão das conexões e_i existentes e o máximo, conforme equação 1.11. O coeficiente local adota valores de 0 a 1: se os vizinhos de um nodo n_i estiverem completamente conectados entre si, $C_i = 1$, se não existirem conexões entre vizinhos, então $C_i = 0$, com valores maiores representando um coeficiente de agrupamento maior.

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \quad (1.11)$$

A partir da definição de agrupamento local é possível expandir esta ideia para um coeficiente global de agrupamento C da rede. A definição do coeficiente global não é única, podendo ser baseada em uma contagem global de triângulos (68, 69). A definição comumente adotada é a proposta por Watts e Strogatz (1), por ser computacionalmente conveniente: C é simplesmente a média dos coeficientes locais (equação 1.12). Tanto os coeficientes locais quanto a média da rede variam entre $0 \leq C \leq 1$ para uma rede qualquer.

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (1.12)$$

1.2.6 Caminho, caminho médio e diâmetro de rede

Caminho, dentro de uma rede, é definido como uma rota contínua percorrendo conexões e nodos de maneira sequencial e sem descontinuidades. Um caminho singular entre dois nodos arbitrários n_i e n_p pode ser representado por uma lista de conexões ordenada (expressão 1.13). O tamanho deste caminho é o número de links percorridos (tamanho da lista).

$$\text{Caminho} = \ell_{i,f}, \ell_{j,g}, \ell_{k,h}, \dots, \ell_{n,j} \quad (1.13)$$

A distância dos nodos $d_{i,j}$ é o mínimo tamanho entre todas as possibilidades de caminho ($d_{i,j}$ é um número inteiro). A distância $d_{i,j}$ para cada par (n_i, n_j) é única, mas a quantidade de caminhos mínimos de mesmo comprimento pode ser múltipla. A distância média $\langle d \rangle$ de uma rede é definida como a média de todas as $\frac{N(N-1)}{2}$ distâncias mínimas $d_{i,j}$ (equação 1.14).

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{i,j} d_{i,j} \quad (1.14)$$

A maior distância da rede é chamada diâmetro da rede D . Caminho médio e diâmetro são medidas topológicas que ajudam a identificar a proximidade entre nodos na rede.

1.2.7 Maior componente conexa e componente gigante

Uma rede pode estar fragmentada em um conjunto de sub-redes não conectadas entre si, e neste caso não existe caminho dentro da rede entre, pelo menos, um par de nodos. Cada uma destas sub-redes é uma componente, e a maior componente é designada como maior componente conexa (MCC).

Quando a MCC contém uma fração significativa da rede, é comumente chamada de componente gigante (CG). Neste texto, utiliza-se o termo MCC, mas em geral a maior componente conexa representará nos resultados a componente gigante, com os nodos desconectados distribuídos em componentes menores.

1.3 Considerações finais do capítulo

No capítulo atual se introduz a representação de sistemas complexos como redes complexas, ênfase é dada à importância e à abrangência desta abordagem. As formas de representação, gráfica e como matriz de adjacência, são apresentadas, e as definições matemáticas de algumas métricas topológicas de interesse são definidas. No capítulo seguinte, estas definições são utilizadas para caracterizar redes empíricas (associadas a sistemas complexos empíricos).

2 REDES EMPÍRICAS E SUAS REPRESENTAÇÕES

Redes empíricas, também referidas na literatura como redes reais, são redes construídas baseadas em medidas, observação ou coleta de dados de sistemas empíricos. Os nodos e conexões de redes empíricas representam entidades reais observáveis ou mensuráveis. Redes empíricas se referem a uma representação de um modelo real, uma aproximação da topologia existente no sistema empírico original, obtida através de alguma metodologia.

Neste capítulo, são abordadas características de redes empíricas, utilizando os conceitos de análise topológica apresentados no capítulo anterior. Como previamente colocado, redes empíricas apresentam uma certa quantidade de características topológicas comuns entre si, que não são triviais. A análise destas redes empíricas, com o objetivo de identificar as características especiais contidas em sua estrutura é tema principal em ciência de redes.

Modelos teóricos são propostos para reproduzir característica de redes empíricas, e redes empíricas, por sua vez são utilizadas para validar estes modelos. Nas seções a seguir, características de redes empíricas serão apresentadas no contexto de alguns modelos clássicos de geração de redes, ressaltando as características topológicas que estes reproduzem, e as limitações destas representações frente às redes empíricas.

2.1 Redes randômicas e o modelo Erdős–Rényi

Pode-se questionar se as características notáveis de redes empíricas ocorrem de forma randômica. Um modelo importante a ser mencionado, antes das características destas redes empíricas, são as redes randômicas, descritas no ano de 1959 por Paul Erdős e Alfred Rényi (70), onde é proposto um algoritmo para a construção de uma rede randômica. Nas chamadas redes E-R $G_{(N,\ell)}$, parte-se de uma quantidade de nodos N , inicialmente desconectados, e ℓ conexões randômicas são sequencialmente acrescentadas entre nodos. Um segundo algoritmo de redes randômicas, $G_{(N,p)}$, foi proposto por Edgar Gilbert, também em 1959 (71). Neste algoritmo, parte-se de uma quantidade de nodos N , cada nodo tem uma probabilidade p de estar conectado a cada um dos demais nodos. Ao final do processo de conexão é formada uma rede com $\frac{N(N-1)}{2}p$ conexões aleatórias.

Estes dois modelos de geração de redes randômicas não fornecem resultados estritamente iguais, devido a natureza sequencial da forma $G_{(N,\ell)}$, onde a população de conexões possíveis diminui a cada conexão acrescentada, contra a probabilidade p homogênea do

modelo de Gilbert. No entanto, em redes esparsas estas duas formas são equivalentes e os algoritmos geram resultados com as mesmas propriedades topológicas. Devido as contribuições históricas de Erdős e Rényi (72) (73), que descobriram diversas propriedades de redes randômicas utilizando a descrição de Gilbert, o nome rede E-R é usado de maneira intercambiável entre redes randômicas e para os dois algoritmos.

Em uma rede E-R randômica $G_{(N,p)}$, a probabilidade de um nodo estar conectado a qualquer outro é p , logo o grau médio $\langle k \rangle$ é:

$$\langle k \rangle = p(N - 1) \quad (2.1)$$

Para um nodo qualquer n_i , a probabilidade de quaisquer dois de seus k_i vizinhos estarem conectados é p , então as e_i conexões entre os k_i vizinhos são $p \frac{k_i(k_i-1)}{2}$. Utilizando a definição de C_i da equação 1.11 e de C da equação 1.12, o coeficiente de agrupamento da rede randômica pode ser escrito em termos de tamanho de rede e grau médio (equação 2.2). Observa-se que para uma rede esparsa, onde $N \gg \langle k \rangle$, o coeficiente de agrupamento para uma rede randômica é muito baixo.

$$C = \frac{\langle k \rangle}{N} \quad (2.2)$$

Redes empíricas, frequentemente apresentam coeficientes de agrupamento relativamente elevados, independentemente do tamanho de rede N , da ordem de $C \propto 10^{-1}$. Em redes randômicas de conectividade equivalente, o coeficiente é muito menor, ($C \propto \frac{1}{N}$ com $N \gg \langle k \rangle$) e diminui com o crescimento da rede, uma indicação que o agrupamento, formação de cliques e grupos altamente interconectados em redes esparsas não é o efeito de conexão randômica, mas de alguma outra dinâmica de ordenamento do sistema.

A distribuição de graus de uma rede randômica é binomial (equação 2.3):

$$P_k = \binom{N}{k} p^k (1 - p)^{N-k} \quad (2.3)$$

Para um valor de tamanho de rede suficientemente grande, o limite contínuo pode ser utilizado, resultando em uma distribuição de Poisson (equação 2.4):

$$P_k(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad (2.4)$$

Essas distribuições são bastante homogêneas em torno da média, com desvio padrão convergente. Redes empíricas se diferenciam também neste quesito, apresentando frequentemente distribuições heterogêneas.

Outra característica de redes randômicas E-R é o surgimento de uma componente gigante (uma máxima componente conexa com quantidade significativa dos nodos da rede) a partir do limiar $\langle k \rangle > 1$, sendo altamente fragmentada abaixo deste limiar. Esta característica, demonstrada por Erdős e Rényi, demonstra que mesmo uma rede de densidade muito baixa é capaz de produzir, sem um ordenamento especial, redes interconectadas.

Tabela 1 – Comparação de parâmetros topológicos de redes empíricas com os mesmos parâmetros esperados em uma rede randômica de tamanho e densidade de conexões similar. Fonte: Albert & Barabási (4).

Rede	N	$\langle k \rangle$	$\langle d \rangle$	$\langle d_{rand} \rangle$	C	C_{rand}	Fonte
WWW, site level, undir.	153127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999 (8)
Internet, domain level	3015–6209	4.11	3.76	6.18	0.3	0.001	Pastor-Satorras et al., 2001 (74)
Movie actors	225226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998 (1)
LANL co-authorship	52909	9.7	5.9	4.79	0.43	0.00018	Newman, 2001 (75, 76, 77)
MEDLINE co-authorship	11520251	18.1	4.6	4.91	0.066	0.000011	Newman, 2001 (75, 76, 77)
SPIRES co-authorship	11994	173	4.0	2.12	0.726	0.003	Newman, 2001 (75, 76, 77)
NCSTRL co-authorship	70975	3.59	9.7	7.34	0.496	0.0003	Newman, 2001 (75, 76, 77)
Math, co-authorship	209293	3.9	9.5	8.2	0.59	0.000054	Barabási et al., 2002 (78)
Neurosci, co-authorship	209293	11.5	6	5.01	0.76	0.000055	Barabási et al., 2002 (78)
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2001 (79)
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2001 (79)
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2002 (80)
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2002 (80)
Words, co-occurrence	460902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001 (81)
Words, synonyms	22311	13.48	4.5	3.84	0.7	0.0006	Yook et al., 2001 (82)
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998 (1)
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998 (1)

Redes E-R possuem topologias diversas das observadas em redes empíricas e são, em geral, um modelo inadequado para representar sistemas complexos. Isso é exemplificado na tabela 1, em que se observa que redes empíricas possuem coeficientes de agrupamento global C , ordens de grandeza maiores do que os esperados em redes E-R. O modelo E-R, no entanto, ajuda a ilustrar a natureza não randômica das características topológicas que emergem nestes sistemas, indicativa de algum tipo de ordenamento ou auto-organização não randômica do sistema. Na metodologia de estudo de redes, as redes randômicas são um modelo de referência para detecção de um efeito não randômicos, através da comparação das estruturas topológicas encontradas com a estrutura equivalente, caso a distribuição fosse aleatória.

2.2 Mundos pequenos, agrupamento e o modelo Watts–Strogatz

Em uma rede formada por uma rede quadrada bidimensional de tamanho N , onde todos os nodos estão conectados apenas a seus primeiros vizinhos, para percorrer um caminho entre dois nodos é necessário trilhar um caminho na rede bidimensional, passando por alguns nodos ao longo do caminho. Neste caso, o tamanho médio $\langle d_{2D} \rangle$ e o diâmetro D_{2D} da rede são proporcionais ao tamanho da rede, ou mais especificamente, a raiz quadrada do tamanho de rede N ($\langle d_{2D} \rangle \propto N^{1/2}$ e $D_{2D} \propto N^{1/2}$).

Em redes empíricas, o caminho entre dois nodos quaisquer é tipicamente pequeno, comparativamente ao tamanho da rede. Este efeito é por vezes nomeado coloquialmente como “seis graus de separação” (83) ou “seis graus de apertos de mão”, e expressa a ideia de que, através de um “caminho de amigos de amigos”, qualquer pessoa está a poucos contatos de qualquer outro indivíduo contemporâneo. Em ciência de redes, este fenómeno de proximidade é chamado de efeito de mundo pequeno (*small world*), e é definido como uma proporção do caminho médio d_{sw} e do tamanho N proporcional ou menor a:

$$\langle d_{sw} \rangle \propto \log(N) \quad (2.5)$$

E o diâmetro da rede D_{sw} segue uma proporcionalidade similar:

$$D_{sw} \propto \log(N) \quad (2.6)$$

Efeitos tipo mundo pequeno são caracterizados também pela proporcionalidade entre caminho e tamanho de rede menores que a logarítmica. Redes complexas empíricas e esparsas apresentam (tipicamente) efeitos de mundo pequeno e coeficientes de agrupamento C elevados. Já redes randômicas E-R apresentam caminho médio e diâmetros pequenos (o que indica que o efeito de mundo pequeno pode ser obtido sem qualquer ordenamento sofisticado), mas o coeficiente de agrupamento, conforme equação 2.2, diminui com o tamanho da rede a uma proporção:

$$C_{rnd} \propto \frac{1}{N} \quad (2.7)$$

O modelo Watts-Strogatz (1), ou algoritmo W-S para geração de redes de pequeno mundo (figura 4), procura modelar estes efeitos. O algoritmo parte de uma rede ordenada circular, por exemplo, onde a geometria topológica local é uniforme e desenhada para formação de tríades, resultando em um coeficiente de agrupamento alto. O algoritmo insere um certo grau de aleatoriedade, com uma probabilidade de algumas conexões serem aleatoriamente redirecionadas, estas conexões formam “pontes” e “atalhos” na rede reduzindo o caminho médio muito rapidamente, enquanto o coeficiente de agrupamento global da rede permanece elevado.

Algoritmos geradores de redes complexas apresentam, em geral, limitações em representar a totalidade das características de redes empíricas. No caso do modelo W-S, embora o efeito de mundo pequeno e coeficiente de agrupamento alto possam ser simultaneamente reproduzidos do ponto de vista estatístico, as topologias de redes complexas empíricas em geral não são redes homogêneas ordenadas com alguns desvios de conexão. Os cliques altamente conectados, frequentemente, surgem como estruturas heterogêneas dentro de uma rede e, em geral, não partem de um ordenamento pré-

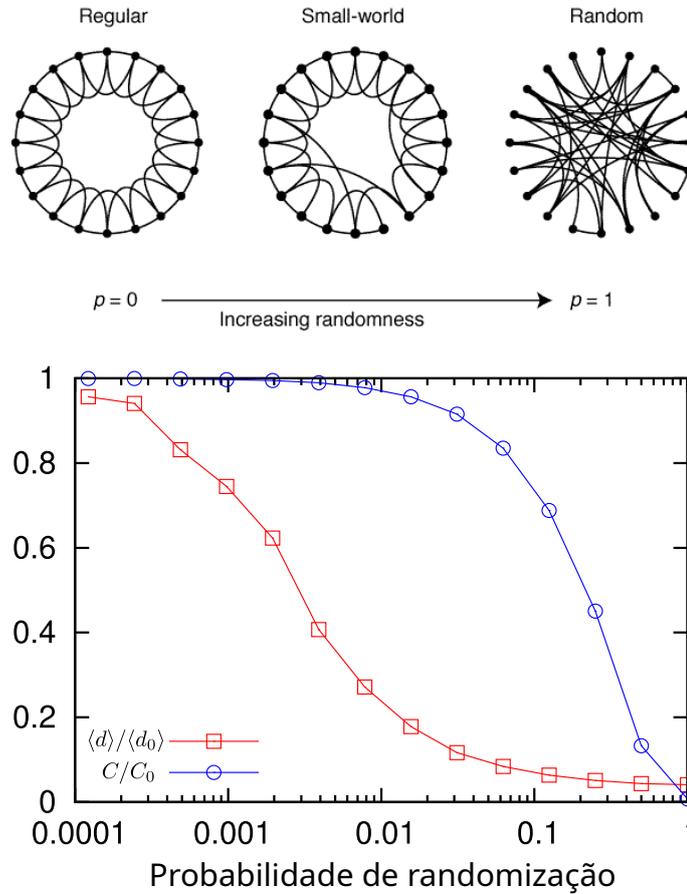


Figura 4 – Algoritmo W-S. *Painel superior*: processo de randomização entre os limites da rede ordenada e uma rede randômica (*fonte Watts e Strogatz (1)*). *Painel inferior*: efeitos da randomização sobre os parâmetros caminho médio $\langle d \rangle$ e coeficiente de agrupamento C em termos dos valores iniciais $\langle d_0 \rangle$ e C_0 . Para uma razão muito pequena de conexões randomizadas, a rede ordenada se torna um "mundo pequeno", mantendo boa parte de seu ordenamento.

estabelecido. Notoriamente, a distribuição de grau $P(k)$ do modelo W-S é homogênea, de uma maneira que frequentemente não é observada em redes empíricas, que usualmente apresentam topologias mais heterogêneas, com a existência de nodos altamente conectados (hubs). Na próxima seção, trata-se destas distribuições mais heterogêneas e das redes livres de escala.

2.3 Redes livres de escala e o Modelo B-A

Embora não seja uma generalização completa, com frequência redes empíricas de interesse, em particular redes sociais, apresentam distribuições de grau $P(k)$ de "cauda longa", ou seja, existe ocorrência de nodos altamente conectados, chamados hubs, e uma heterogeneidade dos graus de conectividades k_i em relação a conectividade média $\langle k \rangle$. Uma modalidade de distribuição de graus, especialmente importante, são as leis de potência

(equação 2.8).

$$P_{(k)} \propto k^{-\gamma} \quad (2.8)$$

Em uma lei de potências na forma da equação 2.8 e para valores de $\gamma < 3$, o primeiro momento estatístico $\langle k^2 \rangle$ é divergente, o que implica um desvio padrão da conectividade σ_k que também diverge. Na prática, isso significa que não existe uma expectativa de conectividade de um nodo escolhido aleatoriamente na rede, ou em outras palavras, não existe uma escala para os valores k_i , que podem ser muito maiores do que a média. Estas distribuições e suas respectivas redes são chamadas de livres de escala. Na figura 5 é apresentada uma comparação entre uma distribuição livre de escala e uma distribuição de Poisson para duas redes de mesmo tamanho e densidade.

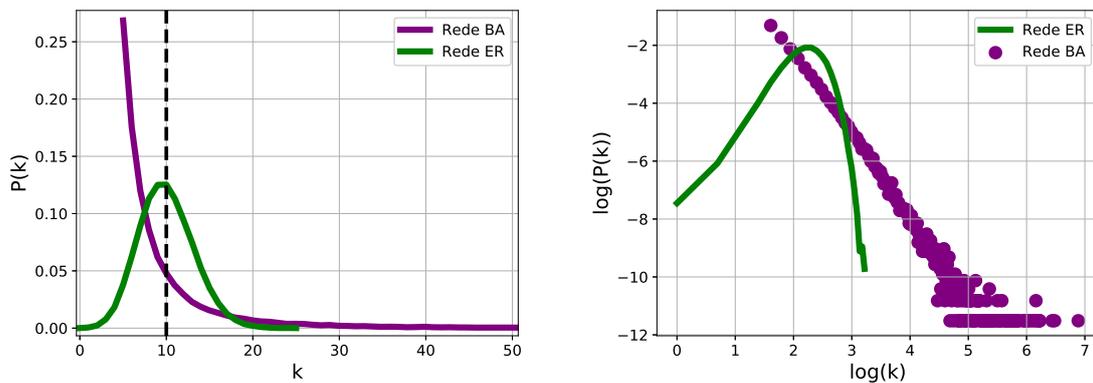


Figura 5 – Comparação entre distribuições de grau $P_{(k)}$ de uma rede randômica (Poisson) e uma rede livre de escala (lei de potência). As duas distribuições correspondem a redes geradas computacionalmente, utilizando o mesmo tamanho $N = 10^5$ e densidade, com grau médio $\langle k \rangle = 10$. *Painel esquerdo*: o grau médio das duas distribuições $\langle k \rangle = 10$ é apontado com uma linha vertical. A cauda longa da lei de potências possibilita o aparecimento de hubs na rede. A figura é, na verdade, um zoom na escala de k para melhor visualização da distribuição randômica, o grau máximo da rede livre de escala é da ordem de 10^3 . *Painel direito*: $\log(P_{(k)})$ vs $\log(k)$. É mais fácil identificar distribuições livres de escala na forma $\log(P_{(k)}) = -\gamma \log(k)$.

Redes livres de escala são encontradas (dadas algumas restrições e aproximações colocadas adiante) em uma série de sistemas de interesse. Um grande volume de publicações, a partir do final dos anos 90, apontou para uma certa prevalência de redes com distribuições livres de escala em redes empíricas originadas em diversas áreas. Estas similaridades topológicas entre redes diversas resultou em uma “corrida do ouro” para encontrar leis de potência em distribuições de grau em redes empíricas, resultando em uma grande quantidade de publicações de caracterização de redes de sistemas diversos. Na tabela 2

Tabela 2 – Alguns exemplos de ocorrência de redes livre de escala em sistemas empíricos. Redes não direcionais são marcadas com asterisco e o coeficiente de potência γ_{in} e γ_{out} são iguais. Para redes direcionadas coeficientes γ_{in} e γ_{out} são distintos. Fonte: Albert & Barabási (4)

Rede	N	$\langle k \rangle$	γ_{out}	γ_{in}	Fonte
WWW	325729	4.51	2.45	2.1	Albert, Jeong, Barabási 1999 (84)
WWW	40000000	7	2.38	2.1	Kumar et al. 1999 (85)
WWW	200000000	7.5	2.72	2.1	Broder et al. 2000 (86)
WWW, site	260000	-	-	1.94	Huberman, Adamic 2000 (87)
Internet, domain*	3015-4389	3.42 - 3.76	2.1 - 2.2	2.1 - 2.2	Faloutsos 1999 (88)
Internet, router*	3888	2.57	2.48	2.48	Faloutsos 1999 (88)
Internet, router*	150000	2.66	2.4	2.4	Govindan 2000 (89)
Movie actors*	212250	28.78	2.3	2.3	Barabási, Albert 1999 (90)
Coauthors, SPIRES*	56627	173	1.2	1.2	Newman 2001b,c (76)(75)
Coauthors, neuro.*	209293	11.54	2.1	2.1	Barabási et al. 2001 (78)
Coauthors, math*	70 975	3.9	2.5	2.5	Barabási et al. 2001 (78)
Sexual contacts*	2.810	-	3.4	3.4	Liljeros et al. 2001 (91)
Metabolic, E. coli	778	7.4	2.2	2.2	Jeong et al. 2000 (92)
Protein, S. cerev.	1870	2.39	2.4	2.4	Mason et al. 2000 (93)
Ythan estuary*	134	8.7	1.05	1.05	Montoya, Solé 2000 (80)
Silwood park*	154	4.75	1.13	1.13	Montoya, Solé 2000 (80)
Citation	783339	8.57	-	3	Redner 1998 (94)
Phone-call	53000000	3.16	2.1	-	Aiello et al. 2000 (95)
Words, cooccurrence*	460902	70.13	2.7	2.7	Cancho, Solé 2001 (81)
Words, synonyms*	22310	13.48	2.8	2.8	Yook et al. 2001 (82)

são listadas redes sociais, tecnológicas e biológicas com a propriedade livre de escala e seus respectivos coeficientes γ

Redes randômicas E-R não possuem distribuições de grau com cauda longa, apresentando tipo Poisson, com desvio padrão definido, e não formam nodos altamente conectados. Distribuições não livres de escala, tão prevalentes em redes empíricas, devem emergir de alguma forma diferente de ordenamento ou auto-organização. O modelo Barabási-Albert (90), para criação de redes livres de escala, ou modelo B-A de forma abreviada, propõe um mecanismo de crescimento e ligação preferencial capaz produzir leis de potência como distribuições de grau.

No modelo B-A, a rede começa com uma quantidade m_0 de nodos em uma pequena rede completamente conexa, sendo que esta rede cresce com a inserção de nodos e conexões sequencialmente. Cada novo nodo adicionado se conecta a m nodos pré existentes, através de um mecanismo de ligação preferencial, em que a probabilidade de conexão é proporcional à conectividade dos nodos pré existentes, ou seja, os nodos com maior número de conexões tem maior probabilidade de receber novas conexões. O sistema cresce até atingir um número total de nodos N , se N é suficientemente grande (ou no limite infinito), a rede assume a forma de uma lei de potências com $\gamma = -3$ e grau de conectividade médio $\langle k \rangle = 2m$.

O modelo B-A, originalmente proposto, exhibe efeitos de mundo pequeno e a propriedade livre de escala, mas apresenta coeficiente global de agrupamento C baixo,

diminuindo conforme a rede aumenta (equação 2.9). Existem variações diversas deste modelo capazes de reproduzir parâmetros topológicos específicos, em geral estes modelos incluem um dispositivo de ordenamento adicional, enquanto utilizam a capacidade de ligações preferenciais de gerar leis de potência com crescimento de rede.

$$C_{B-A} \propto \frac{(\ln N)^2}{N} \quad (2.9)$$

Mesmo apresentando limitações, este algoritmo contém uma característica notável: é capaz de reproduzir um fenômeno de ordenamento presente em redes empíricas através de um processo intuitivo, com analogia em vários destes mesmos sistemas empíricos. Esta associação de mecanismos de ordenamento com analogia a fenômenos do sistema estudado são capazes de indicar, por vezes, um dos mecanismos que contribuem para o ordenamento dos sistemas. No caso do algoritmo B-A, crescimento de rede e ligação preferencial.

Existem alguns cuidados a considerar quando tratamos com redes livres de escala. A ocorrência de leis de potência é comum, mas não é universal em redes empíricas, mesmo em redes de "cauda longa", e existe considerável quantidade de discussão sobre a validade do uso de leis de potência como o melhor ajuste para distribuições empíricas tradicionalmente consideradas como livres de escala.

Com a melhora da qualidade da coleta de dados, aumento de rigor estatístico de análise e levantamento de milhares de redes adicionais, a prevalência de leis de potência e das propriedades de escalabilidade e crescimento das mesmas deixou de ser unanimidade. Outros modelos como distribuições exponenciais, distribuições log-normais e distribuições de Pareto são frequentemente propostas como alternativas melhores. Por outro lado, várias destas curvas podem ser interpretadas como leis de potência com saturações na faixa de graus elevados e/ou mais baixos.

Embora não universais em redes empíricas, mesmo entre as que apresentam cauda longa, os formalismos e percepções de redes livres de escala são úteis no estudo de uma série de sistemas, e permanecem relevantes. Nesta tese, são reproduzidas distribuições heterogêneas diversas, com similaridades notáveis tanto a leis de potência quanto a outras distribuições de cauda longa, e também com as distorções e saturações observadas em redes empíricas. A classificação ou não destas distribuições como livres de escala não é o foco do trabalho, mas a produção de distribuições heterogêneas similares a redes empíricas, sim.

2.4 Considerações finais do capítulo

No presente capítulo, é apresentada uma definição mais clara de redes empíricas, e são discutidas propriedades topológicas recorrentes nas mesmas, que são de interesse

para os resultados do trabalho. São apresentados três modelos clássicos da literatura para ilustrar estas propriedades, e é dado destaque à natureza não randômica do ordenamento destas propriedades. No próximo capítulo, a revisão bibliográfica é concluída, com a introdução de uma característica mais sofisticada da topologia de redes empíricas: a existência de comunidades estruturais.

3 COMUNIDADES ESTRUTURAIS EM REDES COMPLEXAS

Como será visto no capítulo 5, onde os resultados são apresentados, a formação de comunidades em redes dinâmicas é um dos resultados principais obtidos pela aplicação do modelo proposto. Neste capítulo, procura-se apresentar a definição usada do que é uma comunidade no contexto de redes complexas, além de introduzir os métodos escolhidos de detecção.

A definição do que constitui uma comunidade em diversas áreas é bastante ampla e difusa, podendo representar grupos de mesma etnia, religião, alinhamentos políticos, etc. A denominação representa, frequentemente, uma similaridade entre os membros que a constituem. Embora muitas destas comunidades formem grupos altamente conectados, na definição de senso comum, uma comunidade é um grupo de similares, estando esses conectados ou não por alguma interação complexa.

Essa definição de comunidade não é errada, ela possui utilidade do ponto de vista do discurso e estudo em uma série de áreas. Dentro da área de ciências de redes, no entanto, o termo é utilizado para designar algo mais específico. Comunidades são, em geral, consideradas grupos de nodos altamente interconectados dentro de uma rede mais esparsa. A distinção importante é que o termo geral linguístico de comunidade é mais abrangente e menos específico, podendo ignorar completamente a topologia de relações entre os membros da comunidade, enquanto em ciência de redes, comunidade se refere a uma estrutura topologicamente definida pelas conexões: a estrutura de comunidade está contida completamente na matriz de adjacência, dispensando rótulos nos nodos. Na definição geral, uma comunidade pode ser apenas uma propriedade individual compartilhada pelos nodos, enquanto em ciência de redes, é uma propriedade topológica da configuração de conexões.

Nesta tese, estas comunidades topológicas existentes em redes complexas serão chamadas *comunidades estruturais* ou, de maneira intercambiável, de *módulos* quando obtidos a partir da hipótese de modularidade, que é introduzida adiante no presente capítulo. Na próxima seção, são colocadas algumas hipóteses fundamentadoras das definições possíveis de comunidades estruturais.

3.1 Hipóteses gerais e modularidade

Não existe uma definição única formal sobre o que constitui uma comunidade estrutural dentro da topologia de uma rede; há uma ideia geral de que uma comunidade é

um grupo de nodos e conexões (uma sub-rede) com um grau de interconexão maior entre si do que com o restante da rede.

A hipótese “*de cliques*” prevê que comunidades estruturais são formadas por sub-redes completamente conectadas, formando um único clique. Esta hipótese é o limite mais restritivo, porém não depende do restante da rede nem de nenhum parâmetro de qualidade; a comunidade é classificada como um clique ou não. Embora redes empíricas apresentem cliques em quantidade (principalmente para os casos comuns de agrupamento C elevado de muitas redes empíricas), o tamanho de cada clique é pequeno.

A hipótese “*hard*” considera que todos os nodos pertencentes a uma comunidade estrutural possuem conectividade interna superior à externa, ou seja, para cada nodo n_i individual, a condição 3.1 deve ser verdadeira. Na prática, a ocorrência de comunidades duras de tamanhos significativos em redes empíricas é rara, sendo a hipótese “*hard*” de utilidade limitada.

$$k_i^{int} > k_i^{ext} \quad (3.1)$$

A hipótese “*soft*” considera que comunidades estruturais têm mais conexões entre seus nodos componentes que com nodos fora dela. Dado um conjunto de nodos N_c dentro de uma rede, com conectividade total interna k^{int} conexões internas, e conectividade total externa k^{out} , a sub-rede é caracterizada como comunidade estrutural se a conectividade interna é maior do que a conectividade externa (equação 3.2). Esta hipótese é mais inclusiva, mas partição que verifica a relação 3.2 não é única; existem muitas partições em comunidades em que uma rede pode ser bipartida satisfazendo essa relação.

$$\sum_i^c k_i^{int} > \sum_i^c k_i^{ext} \quad (3.2)$$

A Hipótese de *modelo nulo* parte do pressuposto que existe uma topologia equivalente à rede analisada que não apresenta comunidades. A modularidade, por exemplo, parte da hipótese que uma rede puramente randômica não apresenta comunidades estruturais (96). Como comentado previamente, a rede randômica pode ser uma barra de comparação para existência de efeitos não randômicos. A modularidade é um índice numérico representativo da diferença da expectativa da distribuição randômica das conexões, com a relação de conectividade existente na rede.

Neste trabalho, o critério de qualidade para detecção de comunidades é a modularidade. Esse não é o único critério existente, mas é muito usado devido a facilidade de implementação computacional. A modularidade é definida na próxima seção.

3.2 Particionamento de rede e Modularidade

Detectar comunidades estruturais em uma rede é propor uma forma de particionamento, de divisão da rede em sub-redes, onde essas correspondem a comunidades estruturais. A qualidade deste particionamento precisa ser mensurada de alguma forma e por algum critério, que é derivado de uma hipótese do que é uma comunidade.

A modularidade local (M_c) de uma partição constituída por N_c nodos (uma sub-rede dentro da rede completa) é definida como a comparação entre um modelo nulo, a referência de uma rede não modular (que neste caso é a expectativa local randômica de conectividade) e a distribuição de conexões existentes. A modularidade local M_c é definida (96) pela equação 3.3:

$$M_c = \frac{1}{2\ell} \sum_{i,j}^{N_c} \left(A_{i,j} - \frac{k_i k_j}{2\ell} \right) \quad (3.3)$$

A soma é feita entre os N_c nodos da partição, ℓ é o número total de conexões da rede, $A_{i,j}$ é o elemento da matriz de adjacência A e o elemento $\frac{k_i k_j}{2\ell}$ é a probabilidade de que o elemento os nodos n_i com conectividade k_i e n_j com conectividade k_j estejam conectados no modelo nulo escolhido, que neste caso é a distribuição aleatória de conexões. Se existirem mais conexões do que o esperado em uma distribuição randômica, o termo positivo da soma é predominante, e a modularidade é positiva. Se o número de elementos for igual à expectativa de uma distribuição aleatória, a modularidade é zero. A modularidade pode atingir valores negativos se o número de conexões internas for inferior ao esperado em uma distribuição randômica equivalente.

A expressão 3.3 pode ser simplificada com algumas considerações: a soma do termo positivo é o dobro das conexões internas da partição N_c , ou seja, $\sum_{i,j}^{N_c} A_{i,j} = 2\ell_c$. A soma no termo negativo $\sum_i^{N_c} k_i \sum_j^{N_c} k_j$ é apenas o quadrado da soma da conectividade total k_c da partição N_c . A equação 3.3 pode ser então reescrita na forma da equação 3.4.

$$M_c = \frac{\ell_c}{\ell} - \left(\frac{k_c}{2\ell} \right)^2 \quad (3.4)$$

A modularidade global da rede M é obtida somando as modularidades locais de todas as c partições da rede (equação 3.5).

$$M = \sum^c \left[\frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right] \quad (3.5)$$

M (equação 3.5) serve como um critério de qualidade de particionamento: $M = 0$ denota que o particionamento é o esperado em um modelo nulo randômico, enquanto modularidades positivas denotam que a quantidade de conexões internas são maiores do

que as esperadas em um sistema desordenado. Algoritmos de detecção de comunidades utilizam padrões de qualidade para atingir o melhor particionamento possível, e neste caso, o de maior modularidade.

3.2.1 Algoritmo de Louvain para detecção de comunidades

Como previamente comentado, detectar comunidades é propor um particionamento de rede, onde cada partição é uma comunidade. Em uma rede pré existente, a rede é subdividida em sub-redes (partições) em que a conectividade interna das sub-redes é superior à conectividade externa entre estas.

Teoricamente, poderia-se testar todos os particionamentos possíveis de uma rede e escolher a partição com máxima modularidade como a que melhor identifica suas comunidades. O problema computacional associado é que a quantidade de particionamentos possíveis em uma rede, de complexidade combinatorial, torna essa abordagem impossível mesmo para redes relativamente pequenas ($N > 200$). Para uso prático em estudo de redes grandes, algoritmos de detecção, ou seja, de particionamento, são necessários.

O chamado algoritmo de Louvain, criado por Vincent Blondel et al (3), procura detectar comunidades a partir do agrupamento local de nodos, fundindo módulos, procurando maximizar a modularidade. É até o momento o algoritmo de detecção mais rápido, com velocidade de execução proporcional ao número de conexões da rede, permitindo trabalhar com redes maiores, onde tipicamente $\ell \propto N$.

O algoritmo analisa a rede executando duas etapas de maneira intercalada: inicialmente, na primeira etapa, cada nodo individual é considerado uma comunidade. Calcula-se, para cada nodo, a modificação de modularidade que resultaria se o mesmo fosse agrupado na partição de algum de seus vizinhos imediatos na rede. Se a troca resulta em aumento da modularidade total da rede, o nodo é movido para a nova partição. O processo é repetido até atingir um platô de modularidade. Então se passa para a segunda etapa do algoritmo.

Na segunda etapa, uma nova rede é construída, onde cada comunidade do passo 1 é considerada um nodo. As conexões internas das comunidades são tratadas como autoconexões, e todas as conexões entre duas comunidades são fundidas em uma conexão única, com peso atribuído proporcional à quantidade de links. O algoritmo não modifica a rede original, o agrupamento de partições em nodos em uma nova rede é executado apenas para fins de cálculo de máxima modularidade. O particionamento retornado é referente à rede original. Uma vez concluído esse passo, repete-se alternadamente o primeiro e o segundo passos até que a modularidade atinja um máximo.

A vantagem do algoritmo de Louvain é ser um método rápido de detecção de comunidades, linearmente proporcional ao tamanho de redes esparsas, sendo capaz de lidar em tempos realistas com redes grandes. Este algoritmo e suas variações são chamados

também, em algumas literaturas, de algoritmos tipo multinível.

3.3 Limitações do algoritmo de detecção

Tanto a hipótese de maximização de modularidade, quanto o algoritmo de Louvain para detecção de módulos, são úteis e amplamente usados na detecção de comunidades estruturais, produzindo resultados positivos verificáveis em redes de referência (*benchmarks*). De todo modo, o método possui limitações importantes a serem observadas.

Em primeiro lugar, *modularidade positiva não é um indicador objetivo da existência de comunidades estruturais*, significa apenas que um particionamento favorável foi encontrado. Se existirem comunidades estruturais na rede, o particionamento encontrado com modularidade positiva tende a ser uma boa representação destas comunidades. Modularidade positiva é, no entanto, encontrada quando o algoritmo é aplicado a redes randômicas E-R. Isto ocorre porque o algoritmo procurará partições nas flutuações locais da rede randômica, detectando um grande número de comunidades menores resultantes destas flutuações de densidade.

Existe um limite de resolução no tamanho de comunidades detectadas, isto ocorre porque em determinadas condições de comunidades com k_c similares, fundir duas comunidades estruturais distintas resulta em um aumento marginal na modularidade global, forçando comunidades menores em fundirem-se em uma única comunidade maior (97). Algumas variações do algoritmo de Louvain apresentam um fator resolução para contrabalançar este efeito, mas o uso adequado desta ferramenta exige algum conhecimento sobre tamanhos esperados de comunidade, ou algum outro critério empírico de separação da estrutura. Algoritmos multinível tendem a detectar comunidades de tamanhos similares.

Comunidades estruturais em redes empíricas podem estar aninhadas, esta propriedade está diretamente associada à existência de sistemas aninhados em sistemas complexos. Quando uma representação do sistema na forma de rede é feita, a existência de subestruturas em estruturas maiores é refletida. Dificuldade em reconhecer estes sistemas aninhados é uma limitação distinta da resolução, já que comunidades aninhadas integram as estruturas maiores detectadas e estão completamente contidas nas mesmas, enquanto o problema de resolução está associado à fusão de estruturas distintas.

Os algoritmos particionam a rede e atribuem a cada uma das partições o status de comunidade estrutural. A existência de sub-redes que não formam comunidades não é contemplada (não existe diferenciação entre partes modulares e não modulares de uma rede). Esta é uma limitação importante, de que nem todas as comunidades detectadas terão estrutura, densidade e importância igual.

A maximização de modularidade não é determinística, e o máximo de modularidade

é, em geral, um platô de configurações possíveis. Em redes empíricas estas configurações são geralmente muito similares, e esta limitação torna-se mais relevante apenas quando avaliamos a colocação de nodos individuais entre comunidades vizinhas. Nestes casos, um algoritmo quasi-determinístico, como o algoritmo de Girvan-Newman seria mais apropriado.

3.4 Algoritmo Girvan-Newman

O algoritmo Girvan-Newman é uma metodologia para a detecção de comunidades em redes complexas. O princípio fundamental do algoritmo baseia-se na ideia de que as conexões entre comunidades distintas têm um papel crucial na estrutura da rede e, portanto, são mais propensas a possuírem centralidade de caminho elevada. O algoritmo remove iterativamente as conexões com a maior centralidade de caminho, recalculando a centralidade após a remoção de cada conexão. Esse processo continua até que um número significativo de comunidades seja formado, tipicamente identificado quando o aumento na modularidade (uma medida da estrutura de comunidade numa rede) não é mais observado. Esse método é eficaz para identificar comunidades naturais dentro de uma rede, revelando a estrutura modular sem a necessidade de predefinir o número de comunidades. Computacionalmente é complexo, com tempos de execução da ordem $O(\ell^2 N)$, e, não viável na presente pesquisa.

3.5 Considerações finais do capítulo

Neste capítulo, a revisão bibliográfica é concluída. A ciência de redes é, no entanto, um campo de estudo amplo, e apresentar todas as áreas e resultados obtidos foge do escopo e objetivo deste documento. Nos capítulos 1, 2 e 3, uma revisão dos conceitos principais pertinentes ao modelo proposto foi feita.

No presente capítulo, discussões adicionais com foco das hipóteses de existência e identificação de comunidades estruturais são apresentadas por serem pertinentes ao modelo. A formação de comunidades estruturais, a partir da dinâmica de confinamento, será discutida nos resultados obtidos. Por este motivo, o algoritmo de detecção destas estruturas, o critério de qualidade e hipóteses usadas em sua identificação foram apresentados, bem como suas limitações.

O estudo de comunidades em redes e sistemas complexos é um dos tópicos mais ativos da área. As aplicações de identificações de comunidades estruturais em redes empíricas são muitas, e abrange diversas áreas. De modo geral, estas estruturas representam subestruturas, anormalidades ou mecanismos de funcionamento importantes, e sua identificação é fundamental no entendimento, modelagem, predição e controle dos sistemas complexos analisados.

4 REDES DINÂMICAS MEDIADAS PELO CONFINAMENTO

Nos capítulos anteriores foram, introduzidas métricas topológicas, também modelos clássicos de redes complexas, que procuravam caracterizar e emular algumas características topológicas não triviais, gerando redes com organizações similares às empíricas. Nesses modelos estavam contidos mecanismos de ordenamento. Neste capítulo, é introduzido o conceito de confinamento e seu papel como ordenador da topologia de redes complexas.

Em sistemas empíricos, confinamentos podem representar limitações rígidas de separação física, como encarceramento, quarentenas ou barreiras geográficas, mas também podem representar formas mais sutis de separação: barreiras sociais onde as interações com outros grupos são limitadas, seja por segregação ou normas sociais. Pode-se definir bolhas ideológicas ou informacionais como uma forma de confinamento, onde o contato com elementos da mesma ideologia é facilitado, enquanto o tráfego de informação e contato, com elementos fora destas bolhas é dificultado.

Exemplos de confinamentos estão em toda parte : vizinhanças geográficas ou divisões regionais, barreiras linguísticas, ambientes de trabalho, instituições de estudo, clubes; espaços semi permeáveis que fortalecem as conexões internas em detrimento das conexões externas. Dentro de uma rede, define-se o confinamento como uma sub-rede \mathcal{P} com uma fração de nodos da rede total, e conexões internas e externas ao confinamento.

Na próxima seção, a dinâmica é proposta e exemplificada conceitualmente. A dinâmica de confinamento tem analogia direta com mecanismos observados em redes empíricas, o que torna o modelo útil do ponto de vista de entendimento de parte dos mecanismos de ordenamento de redes empíricas em geral, mas também da emulação de redes empíricas onde a dinâmica de confinamento se encontra presente de maneira explícita.

4.1 Caracterização de confinamento

Para apresentar a dinâmica de confinamento sobre a topologia de uma rede, é preciso, primeiramente, definir topologicamente o que se quer dizer por confinamento. Espaços confinados de um sistema são regiões de um sistema complexo onde três condições são atendidas:

- 1) Elementos do sistema, no espaço confinado, têm acesso dificultado aos elementos externos ao confinamento. Em termos de topologia de rede, isso significa que conexões entre nodos confinados e não confinados (ou confinados em espaços diferentes) são mais

difíceis de serem mantidas.

2) Elementos do sistema, confinados no mesmo espaço, tendem, por proximidade ou por necessidade de formar novas conexões, a se interconectarem. Topologicamente, em termos de rede, isso significa que as conexões entre nodos, dentro do mesmo espaço confinado, são mantidas com facilidade, e novas conexões tendem a surgir.

3) O espaço confinado representa uma parcela reduzida do sistema, esta condição representa a ideia de que um espaço do tamanho do sistema não pode ser considerado um confinamento dentro do sistema. Se um espaço contém mais da metade do sistema, por exemplo, pode-se dizer que, se existe confinamento, ele está na parte menor e não neste espaço maior. Do ponto de vista da topologia da rede, isso significa que a quantidade de nodos dentro de um confinamento deve representar uma parte pequena da rede.

4.1.1 Dinâmica de confinamento temporal

Topologicamente, define-se o espaço de confinamento \mathcal{P} não como um conjunto de elementos fixos da rede, mas como um espaço abstrato que os nodos ocupam temporalmente, podendo ser confinados ou libertados ao longo da dinâmica. Com isso se quer dizer que os nodos podem ser confinados e deixarem o confinamento, mas o espaço confinado \mathcal{P} continua existindo, como espaço ocupado por nodos, mas que estes podem mudar ao longo do tempo. Alguns exemplos empíricos podem tornar esta dinâmica menos abstrata.

Uma instituição de ensino pode ser considerada um espaço confinado, existe a estrutura física e a organização institucional. Ao longo do tempo, novos alunos ingressam nestes espaços confinados e passam a ocupar suas turmas e espaços. Com o tempo e progressão, eventualmente, funcionários e alunos deixam a instituição por transferência, abandono ou graduação, enquanto novos alunos e funcionários ingressam. As pessoas passam temporalmente pela instituição, mas a estrutura geral se mantém. Ao mesmo tempo, a topologia da rede social é modificada por esta passagem. Conexões novas são formadas dentro de turmas e nas relações institucionais, enquanto ligações de escolas, turmas e amizades anteriores podem enfraquecer ou desaparecer. Alguns anos depois, fora da instituição, alguns cliques sociais e conexões permanecem, mas a maioria das conexões enfraqueceu.

Prisões são espaços confinados por definição. Detentos novos são admitidos, enquanto prisioneiros que cumprem a pena ou avançam no regime são libertados. Dentro deste espaço reduzido de interação, conexões são formadas internamente, enquanto as conexões externas sofrem rápido declínio, gerando uma organização interna da rede e da vizinhança topológica imediata. No caso especial do sistema prisional brasileiro, o confinamento age como espaço de organização topológica de facções criminosas sofisticadas como o PCC e o Comando Vermelho (98, 99, 100, 101).

A dinâmica de confinamento pode ser descrita como uma mistura da mecânica alienadora do confinamento, que fortalece conexões internas em detrimento das conexões externas, com o fluxo de entrada e saída de nodos do espaço confinado. É razoável supor que estes mecanismos reconfigurem a topologia da rede, que evolui temporalmente no contexto de uma rede dinâmica.

4.2 Redes dinâmicas

Sistemas complexos empíricos raramente são estáticos. As redes empíricas que representam estes sistemas tendem a ser também dinâmicas. Em uma rede dinâmica, a matriz de adjacência $A_{(t)}$ é uma função do tempo, pois a topologia da conexão entre os nodos evolui temporalmente. Em termos de representação de rede falamos aqui de grafos variáveis no tempo (102, 103, 104). Novos nodos podem surgir na rede, enquanto outros podem desaparecer, e a rede pode ter variações de tamanho, além de conectividade. Comunidades estruturais em redes dinâmicas não são permanentes, mas caracterizadas por evolução temporal: surgimento, crescimento, divisão, fusão ou absorção, redução ou dissolução completa. Em redes sociais, por exemplo, as dinâmicas de grupo sofrem alterações ao longo do tempo, com organizações surgindo e desaparecendo constantemente, enquanto outras se dividem ou se fundem.

Nos capítulos de revisão teórica, redes complexas foram tratadas como objetos estáticos e caracterizadas com métricas aplicáveis a matrizes de adjacência fixa. Para representar redes dinâmicas, cada configuração apresentada pode ser considerada uma medida única em uma série temporal. As redes empíricas evoluem ao longo do tempo, e são “fotografada” a intervalos específicos, não dissimilares a um período de amostragem, formando uma série temporal. Suas métricas como coeficiente de agrupamento C , tamanho N , conectividade média $\langle k \rangle$, distribuição $P_{(k)}$, etc, são calculadas a cada ponto, e a evolução destas métricas pode ser analisada.

4.3 Modelo de redes dinâmicas mediadas pelo confinamento

Nesta seção, é introduzido um modelo computacional que reproduz a dinâmica do confinamento, este modelo é usado para para ilustrar a capacidade desta dinâmica em produzir resultados relevantes e no entendimento de seus mecanismos. O modelo proposto consiste em uma rede dinâmica, onde o conceito de “confinamento” apresentado exerce o papel principal na mudança topológica.

A rede dinâmica é caracterizada por uma rede inicial \mathcal{N} (não necessariamente conexa), onde cada nodo pode assumir um de dois estados: *confinado* ou *livre*. As conexões são não direcionadas e não possuem qualquer propriedade ou estado além dos

nodos que conectam. Esta rede inicial é transformada, modificando o estado desses nodos e redirecionamento das conexões, o que ocorre a cada passo temporal, até que um estado topologicamente estável é atingido do ponto de vista de certas métricas escolhidas: distribuição de grau de conectividade, coeficiente de agrupamento, modularidade e número de comunidades detectadas. A seguir, os passos da dinâmica são descritos detalhadamente.

1) O ponto de partida é qualquer rede com tamanho N nodos e grau de conectividade médio $\langle k \rangle$. Nas simulações apresentadas, salvo quando apontado o contrário, as redes iniciais são redes randômicas $\mathcal{N}(\langle k \rangle)$, construídas usando o algoritmo Erdős-Rényi $G_{(p,N)}$ (70). Configuração inicial das conexões não modifica os resultados importantes da dinâmica, e as redes E-R são usadas apenas pelo efeito colateral de partir de uma configuração que não contém os ordenamentos que resultam desta dinâmica.

2) Um subgrupo inicial de N_c nodos, randomicamente escolhidos, são designados como confinados; este subgrupo é designado como *espaço confinado* ou *confinamento* e representado pelo símbolo \mathcal{P} , os nodos restantes são designados *livres*. Durante a dinâmica, nodos entram e saem do espaço confinado modificando seu estado, mas a quantidade de nodos confinados N_c permanece constante (na realidade empírica os espaços confinados não tem número ou tamanho fixo, então esta é uma aproximação simplificada). Como meio de referenciar o tamanho do confinamento ao tamanho da rede utilizada, é definido o parâmetro p_s (equação 4.1), como sendo a razão do tamanho do subgrupo N_c e o tamanho total da rede N , sendo o tamanho relativo do confinamento p_s um dos parâmetros chave do modelo.

$$p_s = \frac{N_c}{N} \quad (4.1)$$

3) O algoritmo representa uma rede dinâmica, ou seja, com variação de configurações de nodos e conexões da rede ao longo do tempo. No modelo, a passagem do tempo é representada computacionalmente por passos discretos em que ocorrem operações de rearranjo dos status dos nodos e religamentos de links: *a fase de fluxo* e *fase de reconexões*.

4) Na fase de fluxo, representada na figura 6, uma quantidade de N_f nodos, fora do confinamento, são aleatoriamente selecionados e redefinidos como *confinados*, simultaneamente N_f nodos confinados são também randomicamente escolhidos e redefinidos como *livres*. Apenas o status dos nodos é modificado nesta etapa de fluxo e a quantidade de nodos N_f pode ser representada, relativamente ao tamanho da rede, pelo parâmetro p_f :

$$p_f = \frac{N_f}{N} . \quad (4.2)$$

5) Todas as conexões da rede podem ser classificadas em um de três tipos (ℓ_{in} , ℓ_{out} , ℓ_{io}) identificados da seguinte forma: ℓ_{in} é o número de conexões entre nodos no mesmo espaço confinado, ou seja, conexões internas do confinamento; ℓ_{out} é o número de conexões

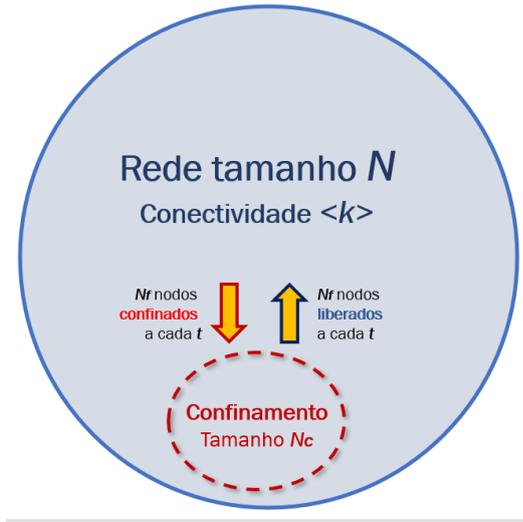


Figura 6 – Fase de Fluxo: Em uma rede de tamanho N , uma quantidade de nodos $N_c = Np_s$ formam o subconjunto confinado. A cada passo temporal uma quantidade $N_f = Np_f$ de nodos pertencentes ao confinamento deixam o conjunto N_c , e a mesma quantidade de nodos não confinados passa a integrar o confinamento. Durante a dinâmica nodos são confinados e libertados a cada passo temporal.

entre nodos livres (não confinados), ou seja, conexões externas ao confinamento; ℓ_{io} denota o número de conexões de interface (conexões entre um nodo livre e um nodo confinado). Essas três quantidades podem variar com a dinâmica (ou seja são funções de tempo), mas o número total de conexões da rede, ℓ , é conservado (equação 4.3). Esta é uma aproximação, já que em sistemas empíricos o número de conexões e nodos, tipicamente, não se conserva.

$$\ell = \ell_{in}(t) + \ell_{out}(t) + \ell_{io}(t). \quad (4.3)$$

6) A topologia muda na fase de reconexão: cada um dos N_c nodos confinados pode, potencialmente, perder uma de suas conexões para um nodo livre. Esta conexão é então redirecionada a outro nodo confinado, ou seja, uma conexão interfacial ℓ_{io} é redirecionada, tornando-se uma conexão interna ℓ_{in} . Isto representa a tendência dos nodos confinados de se alienarem do exterior do confinamento ao mesmo tempo que aumentam suas conexões internas. Assim, cada nodo individual n_j^c dos N_c nodos confinados tem uma de suas conexões modificadas, se tiver, ao menos, uma conexão de interface (com um nodo fora do confinamento) e se existir, ao menos, um outro nodo no confinamento com o qual já não esteja conectado. Nesse caso, uma das conexões interfaciais de n_j (escolha randômica, se existirem várias) é redirecionada para outro nodo no confinamento, ao qual n_j ainda não esteja conectado. A fase de reconexão é representada na figura 7.

Existem duas condições limitadoras entre os parâmetros p_s e p_f . O número de nodos em fluxo não pode ser maior do que o confinamento, pois, no caso limite em que

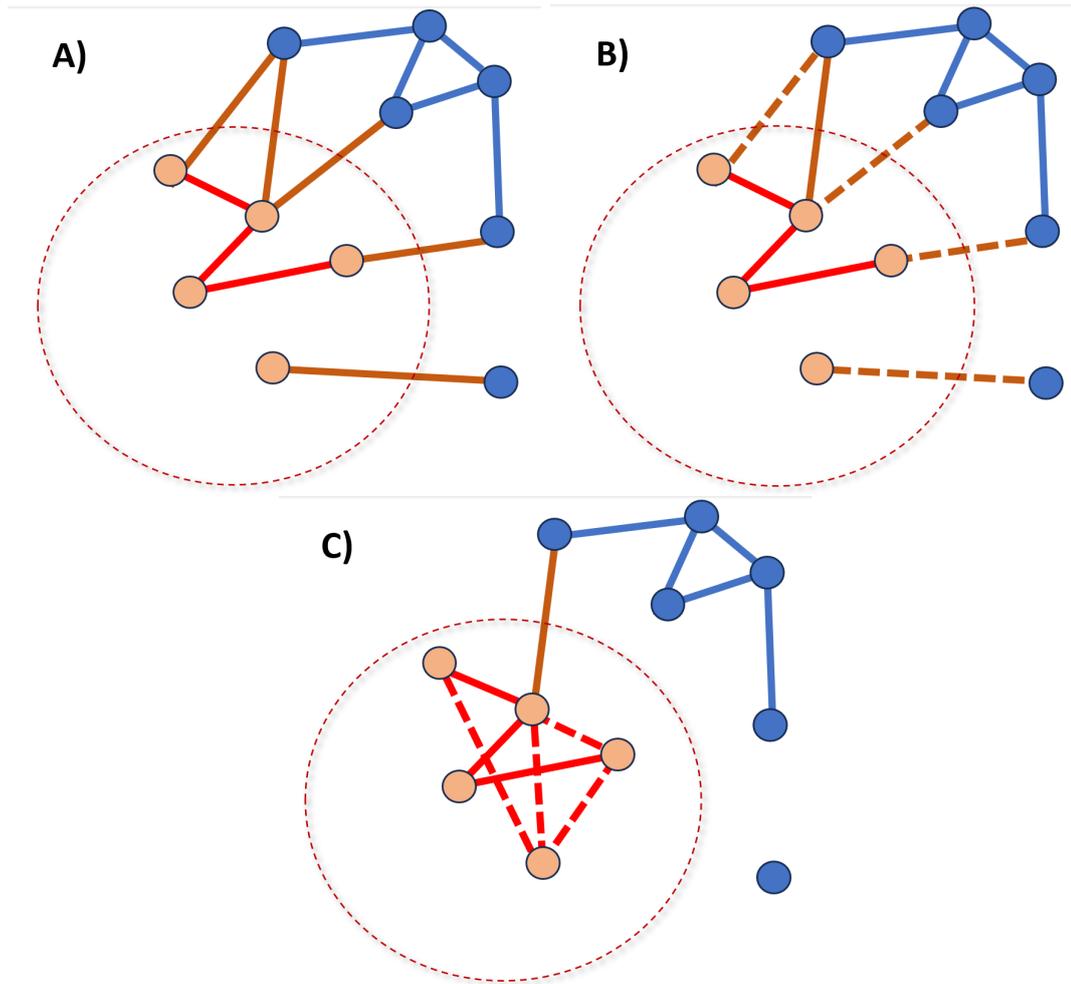


Figura 7 – Fase de reconexões: *Painel A:* No esquema conexões internas ℓ_{in} são representada em vermelho, externas ℓ_{out} em azul e conexões interfaciais ℓ_{io} em laranja. *Painel B:* Uma quantidade de nodos interfaciais (linhas tracejadas) é selecionada randomicamente para reconexão. *Painel C:* conexões interfaciais selecionadas são convertidas em conexões internas, ligando o nodo confinado original a outro nodo confinado, aumentando a densidade do espaço confinado.

$N_c = N_f$, todos os nodos confinados deixam o confinamento a cada passo temporal e são substituídos. Assim, $N_c \geq N_f$, em termos dos parâmetros do sistema:

$$p_s \geq p_f. \quad (4.4)$$

Por outro lado, $N_f \leq N - N_c$ precisa ser satisfeito para que existam nodos livres (N_f) suficientes para substituir os nodos confinados em fluxo, o que em termos dos parâmetros do modelo, traduz-se em:

$$p_s + p_f \leq 1, \quad (4.5)$$

No entanto para qualquer cenário em que $N_c \leq \frac{N}{2}$, ou seja, em que o confinamento é menor que metade da rede, $p_f \leq 0.5$, e com a condição imposta pela Eq. 4.4, torna a Eq. 4.5 sempre satisfeita. Como comentado na definição conceitual, o confinamento é definido como uma pequena delimitação da rede total com $N_c \ll N$.

Conforme a dinâmica progride, a rede é transformada pelos redirecionamentos de conexões e, com exceção no número de nodos e conexões, suas características topológicas mudam até que um estado de estabilidade seja alcançado. Neste estágio, os papéis dos nodos individuais e suas conexões estão em transformação contínua, mas a forma geral da rede permanece estável. No capítulo 5, o modelo descrito é utilizado para demonstrar, através de simulações computacionais, o comportamento complexo que emerge da dinâmica mediada por confinamento e explorar a influência no tamanho relativo do confinamento p_s e do fluxo relativo p_f na rede dinâmica resultante.

4.4 Metodologia

A implementação da dinâmica é feita através de simulações computacionais e registro da série temporal dos parâmetros. Simulações múltiplas são utilizadas para obter tanto evoluções temporais diversas quanto os estados de estabilidade de cada configuração de parâmetros apresentada. Para fins de reprodução de resultados ou livre adaptação do modelo, uma implementação em Python foi colocada no Apêndice A, com instruções de utilização e comentários.

A menos que seja explicitamente mencionado, todas as simulações partem de redes randômicas (geradas com o algoritmo E-R) com $N = 10^4$ nodos e grau de conectividade médio $\langle k \rangle = 8$ como condições iniciais. Tamanhos de rede e densidade de conexões diferentes modificam os resultados quantitativos, mas, inicialmente, neste texto estas influências são isoladas, fixando tamanho e densidade para que os efeitos dos parâmetros de confinamento: p_s (tamanho relativo do confinamento) e p_f (fluxo relativo) em uma rede com um único espaço confinado, possam ser apresentadas e melhor compreendidas. O tópico do tamanho, crescimento e redução de rede é discutido, no capítulo 5, no contexto de confinamentos múltiplos.

Cada simulação é executada até que um estado topologicamente estável seja alcançado. Cada passo temporal executado é definido como uma fase de reconexão, que varre todos os nodos confinados uma vez, seguido por uma fase de fluxo. O número de passos individuais m , necessários para ocorrer estabilização, depende dos parâmetros p_s e p_f . Para todos os resultados apresentados tipicamente $10^3 < m < 2 \times 10^4$ são suficientes. Em geral, um parâmetro de espaço confinado relativo p_s menor (que representa um espaço confinado \mathcal{P} menor) e um parâmetro de fluxo p_f menor (com menos nodos ingressando e deixando o confinamento por passo) requerem tempos maiores para atingir estabilidade.

É importante esclarecer o que está sendo definido como *estado estável* de redes dinâmicas. Durante a evolução transiente de m passos temporais, são computadas as seguintes variáveis topológicas: o coeficiente de agrupamento médio C (1), a distribuição de graus de conectividade dos nodos $P_k(k)$, e a fragmentação da rede (tamanho e variação da maior componente conexa e modularidade). Como a dinâmica pode fazer emergir comunidades estruturais na rede, o algoritmo de Louvain (3), baseado na maximização de *modularidade* (96, 105), é utilizado para encontrar partições da rede que podem definir esses agrupamentos. Modularidade da rede, número de comunidades detectadas e tamanho médio das mesmas são três parâmetros adicionais para determinar estabilidade.

Uma vez alcançada a estabilidade, a dinâmica do confinamento pode prosseguir indefinidamente. Comunidades estruturais se formam, crescem e são diluídas ao longo do tempo, hubs altamente conectados emergem e eventualmente perdem suas conexões, tríades são formadas e desfeitas, mas a forma geral da rede, frente aos parâmetros escolhidos, mantém-se. Os resultados da aplicação da metodologia são apresentados no capítulo a seguir.

5 RESULTADOS

No presente capítulo, é aplicada a metodologia proposta no capítulo anterior. As simulações são feitas, salvo quando apontado, usando redes de tamanho $N = 10^4$ e conectividade média $\langle k \rangle = 8$. Fixar N e $\langle k \rangle$ foi uma escolha, feita para que o foco da discussão permanecesse nos parâmetros de tamanho p_s e fluxo p_f .

Os resultados são qualitativamente muito similares aos apresentados no artigo “*Dynamic Modular Networks Model Mediated by Confinement*” (5), que propõe esta dinâmica originalmente, mas um valor diferente de $\langle k \rangle$ foi utilizado (no artigo $\langle k \rangle = 4$). Esta mudança é deliberada, para demonstrar que, qualitativamente, os resultados são similares e os mecanismos de ordenamento são os mesmos.

O sistema é capaz de reproduzir uma série de propriedades topológicas complexas empíricas, utilizando processos com analogias em processos empíricos correspondentes. Nas próximas seções, estas características são exploradas individualmente, acompanhadas de discussões dos resultados e dos mecanismos de ordenamento. Destaque especial será dado à modularização da rede, e emergência de comunidades estruturais.

Finalmente, na seção “Propriedades adicionais do modelo”, deste capítulo, é discutida a generalização natural do modelo para redes abertas (onde número de nodos e conexões não são conservados) e variação dos parâmetros durante a dinâmica.

5.1 Aumento do coeficiente de agrupamento

Como apresentado anteriormente, no capítulo 2, a maior parte das redes empíricas de interesse (particularmente redes sociais) apresentam sub-redes altamente interconectadas (cliques). O conceito de coeficiente de agrupamento global C ou transitividade (69) é uma das maneiras de mensurar este efeito. É um fato reconhecido em estudo de redes complexas que redes puramente randômicas apresentam um coeficiente global de agrupamento C muito baixo se comparado com redes empíricas. Em grafos de redes sociais, são tipicamente encontrados coeficientes $C \propto 10^{-1}$ (4, 91), e o agrupamento relativamente elevado está relacionado a algum comportamento não randômico típico a iterações sociais, por exemplo.

A figura 8 mostra a evolução temporal do coeficiente de agrupamento C , começando da condição inicial de rede randômica (com agrupamento muito próximo a zero), para um valor fixo de confinamento $p_s = 0.01$ e diversos parâmetros p_f ; pode-se observar, que após um transiente temporal, o coeficiente de agrupamento global C aumenta e atinge um estado de estabilidade. A dinâmica de confinamento transforma a rede inicial E-R randômica, elevando o coeficiente C , dependendo dos parâmetros p_s e p_f , atingindo, após

um transiente, uma valor estável.

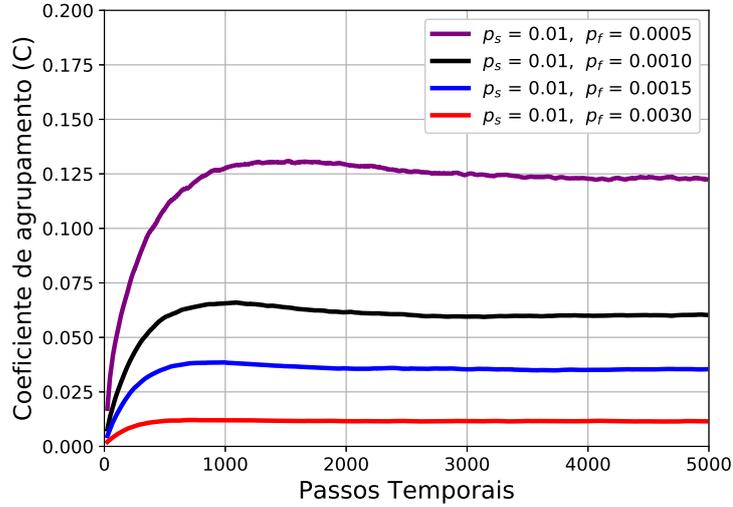


Figura 8 – Evolução temporal do coeficiente de agrupamento global C devido a dinâmica de confinamento com um único espaço confinado. Redes iniciais têm tamanho $N = 10^4$, e $\langle k \rangle = 8$, com agrupamento muito próximo a zero $C = 0.0004$. O espaço confinado \mathcal{P} tem tamanho relativo $p_s = 0.01$ para todas as curvas apresentadas, enquanto o fluxo relativo p_f é diferente para cada curva. Após um transiente de passos proporcionais a p_f , C alcança um estado estacionário. C é medido apenas na maior componente conexa.

O aumento do agrupamento é gerado a partir do confinamento \mathcal{P} , onde os nodos tendem a se conectar internamente, graças ao mecanismo da fase de reconexão, elevando a densidade de conexões dentro do confinamento. Isto reflete o comportamento de uma série de processos observados em redes complexas de natureza social, aumentando a densidade de conexões em espaços confinados/próximos às custas de perda de conexões com o exterior/regiões distantes. Estes nodos e tríades formadas são disseminados em toda a rede através do mecanismo de fluxo, e estas relações / conexões tendem a sobreviver por um tempo, especialmente para tríades formadas por nodos livres.

A Figura 8 mostra um aumento de C para todos os valores do parâmetro de fluxo p_f em relação à condição inicial randômica. Simulações mostram que para tamanhos relativos de confinamento e fluxo pequenos, o coeficiente de agrupamento C fracamente aumenta com o tamanho relativo do confinamento p_s e também com a diminuição do fluxo p_f . Este comportamento pode ser observado no diagrama de fase de C no plano $p_s - p_f$ (Figura 9).

O aumento do agrupamento C é causado pelo aumento da densidade dentro do espaço confinado. Na fase de reconexões, observa-se um aumento da conectividade do espaço confinado, correspondendo a um fator de $p_s N$ vezes o intervalo de tempo considerado. Isso parte do pressuposto de que a dinâmica parte de uma rede inicialmente aleatória

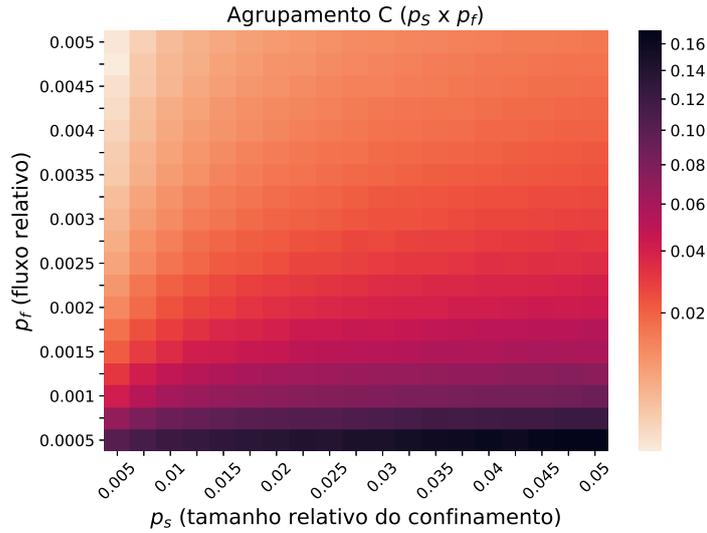


Figura 9 – Mapeamento do coeficiente de agrupamento (da MCC) em função dos parâmetros p_s e p_f . Um fluxo menor resulta em nodos, permanecendo mais tempo dentro de \mathcal{P} , aumentando a probabilidade de formar triângulos no espaço confinado. A rede inicial é aleatória com $N = 10^4$ e $\langle k \rangle = 8$, com um coeficiente de agrupamento próximo de zero, $c = 0,0004$.

e que o fluxo p_f é extremamente baixo ou inexistente. Sob estas condições, o grau médio da rede cresce significativamente em poucos passos temporais — potencialmente dobrando — especialmente porque a densidade de conexões em uma rede aleatória e esparsa é tipicamente baixa. Além disso, os links são redistribuídos aleatoriamente dentro de uma sub-rede consideravelmente menor, de tamanho $p_s N$. Considerando a relação de Agrupamento da rede randômica com grau médio e tamanho de rede da equação 2.2, o agrupamento máximo C_{max} neste espaço de baixo fluxo tende à equação 5.1.

$$C_{max} \propto \frac{2}{p_s} C_{rand} \quad (5.1)$$

Essa aproximação é válida para valores de fluxo muito baixos, fluxos mais elevados evitam a saturação de conexões dentro do espaço confinado. O tamanho do confinamento exerce uma função dualística na dinâmica de formação de cliques, ao mesmo tempo que a quantidade de novas conexões internas ℓ_{in} por passo temporal é proporcional a p_s , um número de nodos maior reduz a probabilidade de uma religação formar uma tríade. As tríades recém formadas são eventualmente transferidas para fora do confinamento, contribuindo para o aumento geral de C , fora do confinamento, conforme demonstrado na Figura 10. A sobrevivência de uma grande quantidade de cliques externos é facilmente explicada pela presença de tríades formadas puramente de conexões externas, que não são desfeitas até que algum dos nodos deste clique seja aprisionado.

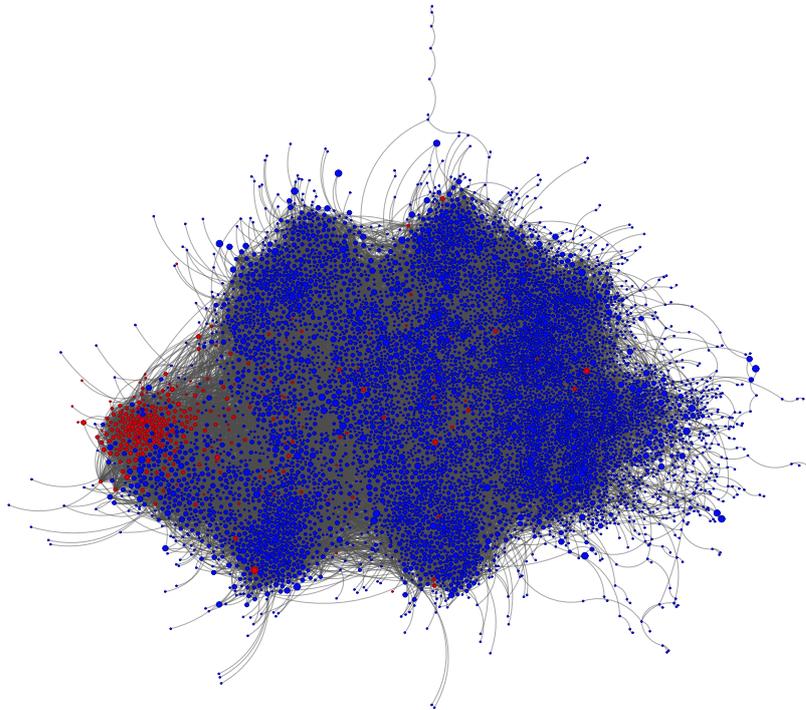


Figura 10 – Propagação de cliques na rede. Representação gráfica de uma rede $N = 10^4$, $\langle k \rangle = 4$ gerada com parâmetros $p_s = 0.03$ e $p_f = 0.0025$. Um algoritmo de força, com condição de não sobreposição de nodos, é usado para a visualização; o tamanho dos nodos é proporcional ao seu coeficiente local de agrupamento C_i . O espaço confinado é representado pela cor vermelha. Os cliques são formados no espaço confinados, em vermelho, mas disseminados em toda a rede.

5.2 Formação de comunidades estruturais

Além de um aumento no agrupamento global, a dinâmica de confinamento proposta gera subgrupos de nodos altamente conectados e fracamente conectados com outras partes da rede, o tipo de estrutura de rede normalmente associada com a ideia de comunidades estruturais. A detecção e caracterização de comunidades (ou módulos) em grandes redes

é um assunto de grande interesse na ciência de redes moderna (106) e a emergência de estruturas de comunidade é um resultado marcante da dinâmica proposta.

Os resultados apresentados usam como base a modularidade M (96), e um algoritmo de maximização de modularidade (o “método Louvain” (3)) é usado para particionar a rede, medir a modularidade M e detectar comunidades estruturais. Adicionalmente, os resultados são observados usando representações gráficas de rede e usando algoritmos de força (107, 2, 108) (Figura 11). É importante destacar que não existe valor pré-fixado de modularidade na dinâmica proposta, nem de números de módulos ou de tamanho médio desses. Comunidades emergem durante as simulações sem uso de qualquer metadado de agrupamento, *a priori*. Não se quer implicar com isso que a geração de comunidades não está implícita no modelo, de fato não é inesperado que comunidades estruturais sejam criadas dentro dos espaços confinados, já que a dinâmica de reconexões é justamente a criação de conexões internas e desconexão externa que define uma comunidade estrutural, mas a mesma dinâmica é responsável por formação e permanência de comunidades *externas* ao espaço confinado, o que não é trivial.

De forma a melhor compreender como comunidades estruturais emergem durante a dinâmica, considera-se o espaço confinado \mathcal{P} como sendo uma partição da rede denominada c , e sua contribuição para a modularidade global M_c . Primeiramente, retomam-se a equação da contribuição da modularidade de uma única partição c da rede apresentada no último capítulo (equação 3.3) (96): sendo ℓ o número total de conexões da rede (ou seja, $\sum_{i=1}^N k_i = 2\ell$), A_{ij} a matriz de adjacência, e $\frac{k_i k_j}{2\ell}$ o número esperado de links i e j quaisquer para uma rede randômica, M_c para a partição c dos nodos contidos em \mathcal{P} é dada pela equação:

$$M_c = \frac{1}{2\ell} \sum_{i,j}^c \left(A_{ij} - \frac{k_i k_j}{2\ell} \right) \quad (5.2)$$

Onde a soma $\sum_{i,j}^c$ é feita sobre os N_c nodos dentro da partição c . A expressão para M_c , como visto na equação 3.5, pode ser escrita de forma mais simples:

$$M_c = \frac{\ell_c}{\ell} - \left(\frac{k_c}{2\ell} \right)^2 \quad (5.3)$$

onde $k_c = \sum_i^c k_i$ é a soma de todos os graus da partição c .

Assumindo a condição inicial de uma rede randômica, analisa-se os efeitos de uma única reconexão de um dos nodos confinados: nodos dentro de \mathcal{P} recebem novas conexões, então a matriz de conectividade recebe mais termos com $A_{ij} = 1$. Para cada nodo redirecionado, ℓ_c e k_c aumentam em 1, e como, em geral, quando $p_s \ll 1$, então $k_c \ll \ell$, pode-se desprezar o fator quadrático da Eq. 5.3 e reescrever M_c como:

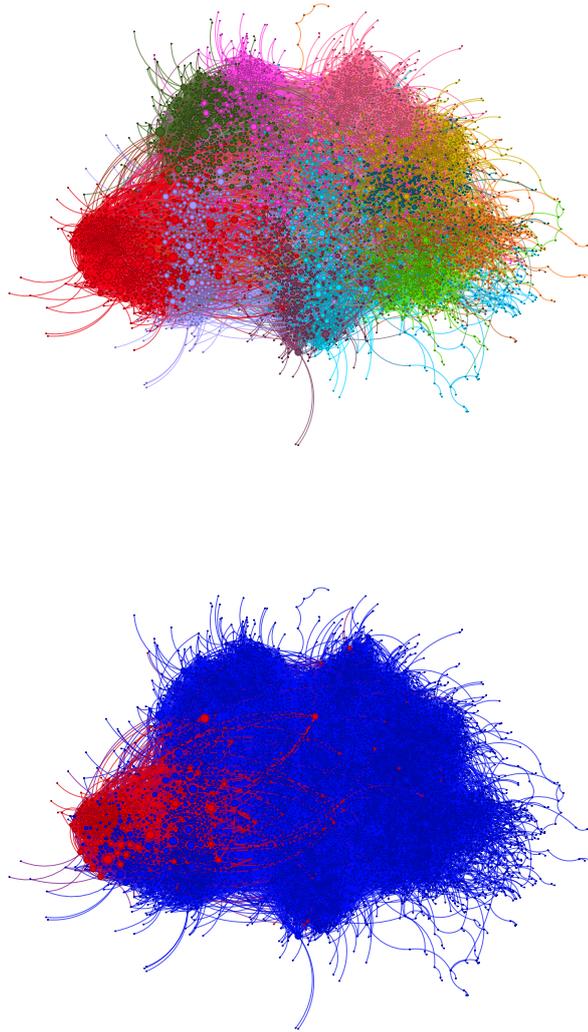


Figura 11 – A partir de uma rede randômica inicial, de tamanho $N = 10^4$ e $\langle k \rangle = 4$, a dinâmica de confinamento cria uma rede estável e mais complexa. A figura mostra o efeito de um único espaço confinado, onde os parâmetros dinâmicos são $p_s = 0,03$ e $p_f = 0,0025$. Nesta representação visual da rede resultante, um algoritmo de força é usado para o posicionamento dos nodos (2), resultando em nodos fortemente conectados sendo agrupados. O tamanho dos nodos é proporcional ao grau de conectividade. *Painel superior:* os nodos são codificados por cores de acordo com o particionamento resultante do algoritmo de detecção da comunidade (3). *Painel Inferior:* Com a mesma disposição de nodos e conexões, os nodos são codificados por cores, como confinados em vermelho ou livres em azul. Durante a dinâmica, os nós entram e saem do confinamento, e os links são religados, mas as características gerais da rede permanecem as mesmas.

$$\Delta M_c \approx \frac{1}{\ell}; \quad (5.4)$$

Esta contribuição para a modularidade local do espaço confinado é referente a um único redirecionamento. Inicialmente, uma quantidade $p_s N$ de conexões (uma para cada nodo confinado) é redirecionada, multiplicando este fator a contribuição unitária da equação 5.4 obtemos:

$$\Delta M_c \approx p_s N \frac{1}{\ell} \quad (5.5)$$

Como $\ell = \frac{\langle k \rangle N}{2}$ chegamos à quantidade positiva:

$$\Delta M_t \approx \frac{2p_s}{\langle k \rangle} \quad (5.6)$$

Esse mecanismo é responsável pelo aumento da modularidade dentro do espaço confinado e sua vizinhança. É necessário ressaltar que esta aproximação na contribuição local, para a modularidade (positiva e proporcional a p_s), só é válida no início da dinâmica, assumindo rede inicial randômica. Com o aumento das conexões internas, k_c aumenta, e algumas reconexões podem se tornar impossíveis devido a impossibilidade de conexões duplas e autoconexões, e a modularidade do espaço confinado deve atingir um máximo. Além disso, nodos deixam o confinamento de maneira contínua devido ao fluxo, e portanto não ficam conectados aos que estão confinados de maneira permanente. Porém, justamente por esta última característica, de que os nodos deixam o confinamento após algum tempo devido ao fluxo, observa-se uma formação contínua e dinâmica de comunidades estruturais fora do espaço confinado.

Um raciocínio similar ao usado, para explicar o aumento da modularidade no espaço confinado, pode ser usado para explicar porque comunidades persistem fora do espaço confinado. Assume-se um agrupamento de nodos altamente conectado (usando novamente a equação 5.2) para um agrupamento interconectado que acaba de sair do espaço confinado \mathcal{P} nos passos temporais prévios devido a dinâmica de fluxo. Este módulo livre recém formado, que aqui é considerado como a partição hipotética c , é rico em conexões internas e fortemente conectado ao espaço confinado (sendo, inicialmente, detectado como uma única partição em conjunto com o espaço confinado). A longo dos próximos passos da evolução temporal, a soma dos links internos $\sum_{ij}^c A_{ij}$ permanece a mesma, se nenhum nodo livre for re-confinado, mas a segunda parte da soma, a parte negativa, se torna menos significativa na medida que $\langle k_c \rangle$ diminui devido a dinâmica de reconexão, em que as conexões ao espaço confinado são perdidas, resultando no aumento da modularidade M_c desta partição hipotética, e na diminuição da modularidade desta partição, se fundida ao confinamento.

Com a quantidade de links internos relativamente preservada e a quantidade de conexões ao espaço confinado diminuídas, a partição livre é eventualmente detectada como uma comunidade estrutural distinta. Em resumo, existe um processo constante de formação

de comunidades estruturais dentro e na vizinhança topológica do espaço confinado. Com o tempo estas estruturas deixam o confinamento e se separam deste, populando a rede. Estas estruturas, eventualmente, desaparecem e são substituídas por novas comunidades criadas pela dinâmica permanente de confinamento e fluxo, mas o número médio de partições detectadas, seu tamanho médio e a modularidade do particionamento permanecem as mesmas, salvo flutuações temporais causadas pelo tamanho finito da rede (figura 11).

A detecção e evolução de comunidades em redes dinâmicas é um tema bastante desafiador, e algoritmos de detecção não são o escopo deste trabalho, então a investigação sobre a modularidade da rede é feita através de análise da modularidade do particionamento da rede a cada passo discreto de tempo e observando a estabilidade da quantidade, tamanho e modularidade do particionamento. Esta distinção é importante, pois algoritmos de detecção de comunidades em redes dinâmicas empíricas se preocupam em caracterizar comunidades específicas e monitorar sua divisão, fusão e crescimento ou dissolução, logo um componente de memória ao longo do tempo precisa estar presente, e uma série de critérios muito particulares para cada tipo de rede precisam ser pré-definidos. Neste trabalho, a preocupação é com o comportamento geral de redes dinâmicas modulares e a presença e estabilidade do particionamento entre passos de tempo é considerada suficiente.

A Figura 12 mostra a evolução temporal da modularidade para alguns valores de tamanho relativo do confinamento p_s , e alguns valores de fluxo relativo p_f . Particionamento da rede é realizado a cada passo temporal, utilizando o "algoritmo de Louvain". É demonstrado no painel esquerdo que a modularidade atinge um estado estacionário após um transiente (o estado transiente depende da rede inicial, a modularidade final dos parâmetros da dinâmica). O painel direito mostra um diagrama de fases parcial na área de interesse dos parâmetros p_s e p_f , mostrando uma maior modularidade na área vermelho escura, associada com pequeno fluxo e um tamanho de confinamento intermediário. Esta região de confinamento suficientemente grande e fluxo baixo é consistente com a formação de agrupamentos altamente conectados dentro do confinamento e sua lenta separação, formando uma comunidade estrutural independente.

A modularidade é útil para avaliar a qualidade de um determinado particionamento de rede, no entanto os algoritmos de detecção de comunidades têm limitações, sendo uma delas a não garantia de existência de comunidades estruturais, mesmo com uma modularidade positiva. Outra peculiaridade é, por exemplo, a descoberta de múltiplas comunidades com modularidade superior a zero, quando o algoritmo de Louvain é aplicado a uma rede completamente aleatória. Isso se deve a flutuações locais em grandes redes, nas quais possíveis partições de comunidade podem resultar do algoritmo de maximização da modularidade. Contudo que estas limitações não sejam esquecidas (109, 110), algoritmos de detecção de comunidades permanecem bastante úteis.

Uma vez que no modelo apresentado comunidades estruturais não são predetermi-

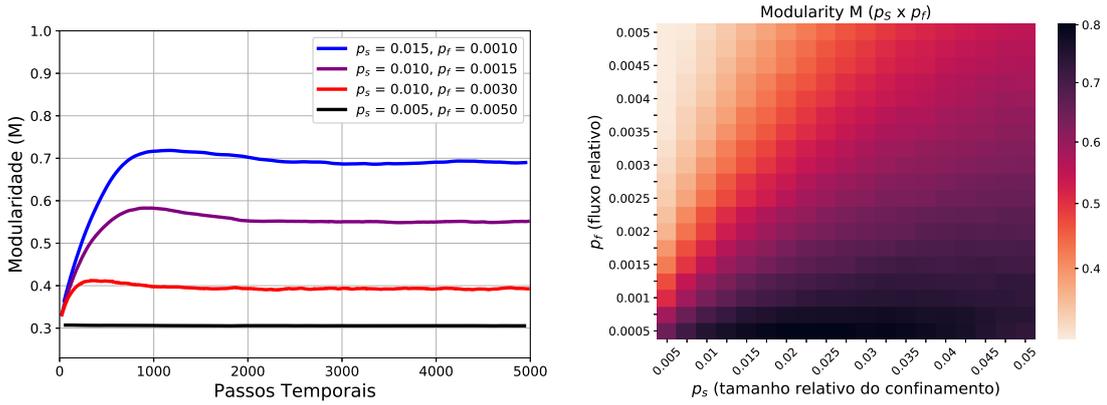


Figura 12 – Modularidade associada ao particionamento das redes pelo algoritmo de detecção, dados valores diferentes de p_s e p_f . Painel esquerdo: evolução temporal para diversos parâmetros a partir da rede aleatória inicial. Painel direito: mapeamento de modularidade como uma função de p_s e p_f . A rede inicial, em todas simulações, é randômica, com $N = 10^4$ e $\langle k \rangle = 8$.

nadas, mas emergem como resultado indireto da dinâmica de confinamento, um segundo parâmetro numérico é utilizado para medir a existência e qualidade das estruturas encontradas. Define-se o parâmetro $\phi(c)$ de uma dada partição c , de tamanho N_c , como a razão entre os números de links externos e totais dessa partição. De modo que para um nó n_j da partição c , k_j^{ext} é o número de links externos e k_j o grau do nodo, e $\phi(c)$ é escrito na forma da equação 5.7.

$$\phi(c) = \frac{\sum_{j=1}^{N_c} k_j^{ext}}{\sum_{j=1}^{N_c} k_j}. \quad (5.7)$$

Portanto, para uma partição com apenas conexões internas (uma comunidade completamente fechada) $\phi(c) = 0$, enquanto em uma partição sem nenhuma conexão interna $\phi = 1$. $\phi(c)$ está relacionado a condição de verificação da hipótese de uma comunidade “soft” (equação 3.2), já que a hipótese torna-se válida para $\phi < 0.5$. A razão média de todas as partições pode ser usada como forma complementar de acompanhar a emergência e estabilidade das comunidades estruturais, para isso é definido um $\Phi(\mathcal{N})$ global, com as somas feitas ao longo de todas as partições da rede:

$$\Phi(\mathcal{N}) = \frac{\sum_{j=1}^N k_j^{ext, N_c}}{\sum_{j=1}^N k_j}. \quad (5.8)$$

Aqui k_j^{ext, N_c} é o número de links externos do nodo j em sua partição respectiva N_c , mas a soma é feita ao longo de todas as partições e nodos da maior componente conexa. A Figura 13 mostra uma clara correspondência com os dados de modularidade da Figura 12; usando os mesmos conjuntos de parâmetros p_s e p_f , vê-se que a curva com a menor

modularidade assintótica tem a maior razão $\Phi(\mathcal{N})$ assintótica, representando concordância entre os critérios de validação.

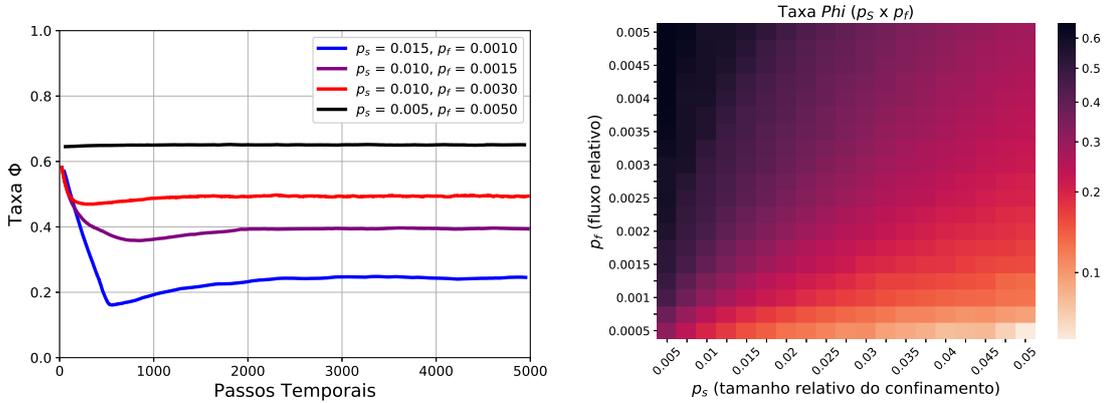


Figura 13 – *Painel Esquerdo*: Evolução temporal de Φ para alguns parâmetros de escolha, (os mesmos usados na figura 12). A razão inicial superior a $\Phi = 0.6$, e diminui com o tempo em todos os casos onde $p_s \neq p_f$, atingindo um estado estável. *Painel direito*: diagrama de fases mostrando a razão Φ em relação aos parâmetros p_s e p_f . A rede inicial em todas simulações é randômica, com $N = 10^4$ e $\langle k \rangle = 8$.

5.3 Distribuição de graus de conectividade

Partindo da rede inicial no tempo $t = 0$, a dinâmica de reconexões muda o grau de conectividade de nodos individuais e, eventualmente, um estado estacionário é alcançado em um determinado tempo $t = t_s$ (dependente da rede inicial transiente), no qual a distribuição de graus $P_{(k)}$ torna-se estável. Neste estado, o grau de nodos individuais muda ao longo do tempo, mas a distribuição preserva sua forma geral. A forma assintótica estável de $P_{(k)}$ depende do tamanho da rede N , do grau médio $\langle k \rangle$ e dos parâmetros p_s e p_f da dinâmica. A Figura 14 mostra as distribuições finais estáveis obtidas de simulações, usando alguns conjuntos dos parâmetros p_s e p_f , e são obtidos resultados de maior heterogeneidade de grau para p_f menores e maiores p_s (fluxo menor e tamanhos de confinamento maiores).

As distribuições $P_{(k)}$ apresentadas, na Figura 14, são alguns exemplos das muitas formas diferentes que podem ser geradas, cada uma correspondendo a um conjunto distinto de parâmetros. No caso especial em que $p_s = p_f$, a rede exibe distribuição de graus de Poisson como uma rede aleatória de mesmo tamanho. A medida que o tamanho do confinamento aumenta e/ou o fluxo diminui, as distribuições de graus resultantes mudam para o tipo de caudas mais longas com maior ocorrência de “hubs” altamente conectados, aumentando a heterogeneidade. Entre as distribuições $P_{(k)}$, encontradas nas simulações, estão curvas que podem ser melhor ajustadas com distribuições Poisson ou Gaussianas,

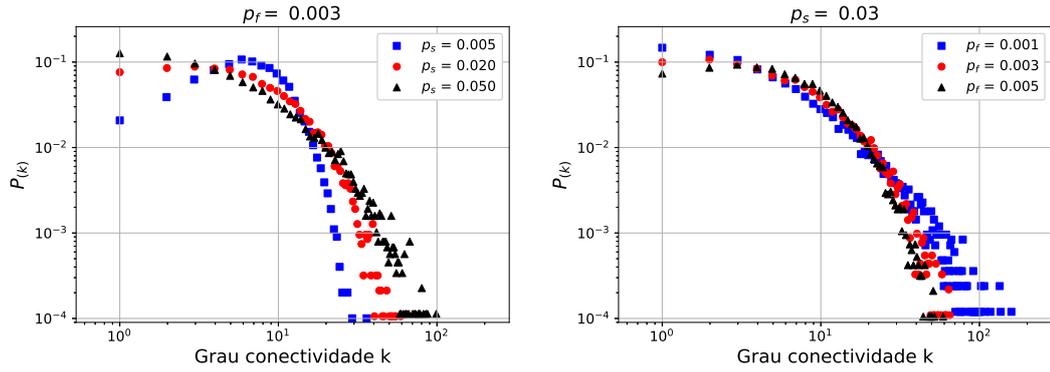


Figura 14 – A partir de uma rede inicial randômica de tamanho $N = 10^4$ e $\langle k \rangle = 8$ e distribuição de grau de Poisson, $P_{(k)}$ se transforma tornando-se mais heterogênea. Esta figura procura isolar os efeitos de diferentes parâmetros dinâmicos p_s e p_f na heterogeneidade de grau. *Painel esquerdo*: usando uma taxa de fluxo fixa $p_f = 0.003$, três curvas são mostradas, cada uma correspondendo ao $P_{(k)}$ estável final de um tamanho de confinamento diferente p_s . *Painel direito*: usando um tamanho de confinamento fixo $p_s = 0.03$, outras três distribuições são mostradas, cada uma usando um valor diferente de p_f . Os valores fixos foram escolhidos por estarem aproximadamente no centro da região de interesse de p_s e p_f apresentadas neste capítulo

mas a maioria dos parâmetros resulta em ajustes log-normais e de potência (como visto na Fig. 15). As distribuições log-normal e leis de potência são obtidas na região inferior de p_f combinados a tamanhos de confinamento p_f maiores.

Através do mecanismo de religação, os nodos confinados tendem a aumentar o grau de conectividade, enquanto os nós livres tendem a perder conexões. Pelo mecanismo de fluxo, os nodos são confinados aleatoriamente e também saem do confinamento aleatoriamente. Isso significa que o tempo gasto em confinamento para $t \rightarrow \infty$ é semelhante para todos os nodos. Qualquer heterogeneidade de grau dos nodos deve ter origem no passado recente da série temporal, em aprisionamentos recentes. Assumindo que um nodo n_j está confinado no tempo $t = t'$, a probabilidade $P_f^j(t = t' + 1)$ de que o mesmo nodo continua confinado na próxima fase de fluxo em $t = t' + 1$ é igual a:

$$P_f^j(t = t' + 1) = 1 - \frac{p_f}{p_s}. \quad (5.9)$$

Como $p_s \geq p_f$, essa probabilidade está bem definida no intervalo $[0 : 1]$, com uma razão menor resultando em chances menores de deixar o confinamento. Como todos os nodos são escolhidos aleatoriamente, alguns passam rapidamente pelo confinamento, enquanto outros podem demorar mais tempo (existem flutuações). A probabilidade de que o nodo n_j permaneça confinado continuamente no instante $t'' > t'$, após ser continuamente

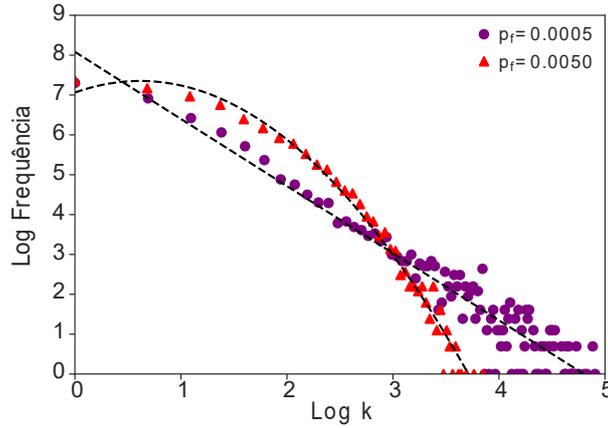


Figura 15 – Heterogeneidade de grau produzida pela dinâmica de confinamento único: diferentes parâmetros dinâmicos geram formas distintas de distribuição de graus. Para o valor fixo do espaço confinado $p_s = 0.035$, são mostradas duas distribuições, uma com p_f maior (triângulos vermelhos) justaposta com ajuste de função log-normal (linha pontilhada preta) e a outra com p_f menor (círculos roxos) ajustada a uma lei de potência. Em geral, um tempo maior dentro de \mathcal{P} significa a formação de mais hubs e maior heterogeneidade de grau, portanto um espaço confinado maior com menos fluxo resulta em distribuições mais heterogêneas.

confinado $\Delta_t = t'' - t'$ é:

$$P_f^j(\Delta_t) = \left(1 - \frac{p_f}{p_s}\right)^{\Delta_t} ; \quad \Delta_t \rightarrow 1, 2, 3, \dots, \infty. \quad (5.10)$$

$P_f^j(\Delta_t)$ é exponencial, diminuindo e se aproximando de zero conforme $\Delta_t \rightarrow \infty$. A velocidade dessa diminuição depende da razão p_f/p_s , com razões mais próximas a $p_f/p_s = 1$ resultando probabilidades maiores para pequenos Δ_t . Os nodos tendem a ganhar conexões enquanto confinados devido à dinâmica de religação, e aqueles confinados por muito tempo provavelmente ficarão mais conectados. Esta permanência é mais provável quanto menor for a razão p_f/p_s , isso explica porque a heterogeneidade de grau aumenta à medida que p_f diminui com p_s constante (painel direito da Fig. 14) e com aumento de p_s e p_f constante (painel esquerdo da figura 14). Adicionalmente, o aumento nos links internos ℓ_{in} é proporcional a p_s , possibilitando uma aquisição maior de links mais rapidamente, e de um grau máximo maior.

5.4 Fragmentação da Rede

A dinâmica de reconexão, especialmente quando associada a um fluxo baixo, eleva a densidade de conexões dentro do espaço confinado e sua vizinhança conexa. Como o número

de conexões é conservado em toda a rede, os nodos livres fora de \mathcal{P} ($N_i \notin \mathcal{P}$) tendem a ter seu grau de conectividade reduzido devido à religação, chegando eventualmente a $k = 0$. Uma única religação da dinâmica redireciona uma conexão interfacial, transferindo a mesma para o interior de \mathcal{P} , e reduzindo o grau de conectividade médio fora do confinamento. A depender de parâmetros, pode ocorrer fragmentação de nodos (que ficam desconectados da rede) ou separação de algumas sub-redes, diminuindo o tamanho da maior componente conexa.

Essa fragmentação é contrabalançada pelo fluxo constante de entrada e saída de nós aleatórios, portanto, nos casos em que $p_f > 0$, um nó órfão ou um agrupamento separado pode ganhar novas conexões ao ser confinado em \mathcal{P} , reconectando-se a outros nós ou agrupamentos em \mathcal{P} . Se, em vez disso, assumirmos o fluxo máximo, com $p_f = p_s$ (o que significa que o fluxo é igual ao tamanho de \mathcal{P}), teremos apenas uma religação aleatória de links na rede e o tamanho do maior componente conexa será o mesmo de uma rede aleatória de tamanho N e grau médio de conectividade $\langle k \rangle$. É de interesse conhecer o nível de fragmentação da rede no estado estacionário e sua dependência dos parâmetros da dinâmica. Na Figura 16, são apresentados resultados de evolução temporal, e mapeamento do tamanho da MCC, conforme associados aos parâmetros da dinâmica.

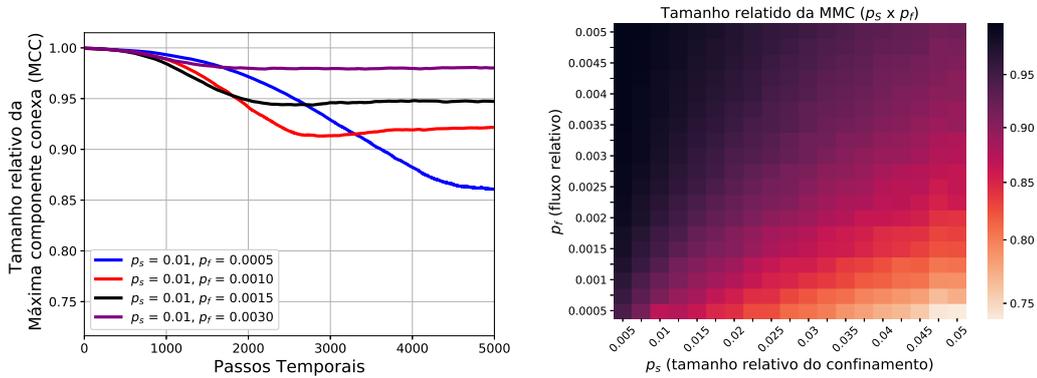


Figura 16 – Tamanho da maior componente conexa (MCC) no modelo de confinamento. *Painel esquerdo:* evolução temporal da maior componente conexa para uma rede aleatória inicial com $N = 10^4$ e $\langle k \rangle = 8$. O tamanho da componente é dado como fração do tamanho total N da rede e são mostradas curvas para diferentes conjuntos de parâmetros. Após um estado transiente, a componente se estabiliza, enquanto os nós saem e entram novamente na MCC devido a dinâmica de confinamento. *Painel direito:* diagrama de fase mostrando o tamanho da MMC em função dos parâmetros p_s e p_f . O tamanho da componente é novamente dado em relação ao tamanho N da rede. Regiões azuis escuras representando um componente maior. Maiores espaços de confinamento resultam em maior redução do grau de conectividade dos nós fora de \mathcal{P} , resultando em mais desconexões. Da mesma forma, com um fluxo menor, menos nós componentes desconexas são religadas.

O painel esquerdo da Figura 16 mostra que, após um transiente, esses dois efeitos (fragmentação por religação e reconexão por fluxo) se equilibram e um estado estacionário de tamanho da MCC é alcançado. Seu tamanho depende dos quatro parâmetros da dinâmica: tamanho e densidade da rede e tamanho relativo do espaço confinado p_s e fluxo relativo p_f . O diagrama de fase de p_s e p_f , o painel direito da Figura 16, mostra o aumento da fragmentação para o tamanho relativo maior do confinamento p_s , que é um resultado esperado uma vez que o número de conexões redirecionadas ao subgrafo \mathcal{P} é proporcional a $p_s N$, resultando em diminuição do grau de conectividade fora da componente que inclui \mathcal{P} . Da mesma forma, com fluxos p_f menores, nodos fora de \mathcal{P} são menos propensos a serem confinados e reconectados a MCC. Adicionalmente, o grau médio fora de \mathcal{P} diminui com menor fluxo, pois a conectividade dos nodos fora do confinamento depende apenas dos conjuntos de links interfaciais ℓ_{io} e externos ℓ_{oo} , que são criados apenas através da saída de nós do confinamento.

5.5 Propriedades adicionais do modelo

Nesta seção, a discussão dos resultados é concluída com a apresentação de propriedades do sistema que viabilizam sua escalabilidade, crescimento de rede, redução e a mudança dinâmica de parâmetros. Essas propriedades estão presentes nas dinâmicas de redes empíricas e, portanto, são consideradas importantes para que o modelo seja representativo das mesmas, para que o modelo tenha aplicabilidade.

5.5.1 Efeitos de múltiplos confinamentos

A dinâmica com um único espaço confinado é instrutiva para introduzir o assunto e ilustrar os mecanismos ordenadores que resultam no aumento do agrupamento, na formação de comunidades estruturais e em distribuições de grau heterogêneas. No entanto, para tamanhos relativos de confinamento fixos p_s e fluxos p_f fixos, conforme aumenta-se o tamanho da rede, o tamanho em nodos do espaço confinado aumenta e a densidade do espaço confinado tende a cair. Em geral diferentes tamanhos e densidades de rede produzem redes com diferentes métricas e uma distribuição de graus que não escala com o aumento de tamanho. No entanto, no mundo real, esperamos que o número de espaços confinados associados a redes empíricas acompanhem o tamanho da rede. O caso de espaço confinado único é apenas um caso especial do modelo, aplicável a redes pequenas. Conforme as redes aumentam de tamanho, é esperado o surgimento de quantidade maior de módulos e de espaços confinados.

Pode-se mostrar que a introdução de vários espaços confinados, dentro da mesma rede, faz com que o modelo seja estável de acordo com o tamanho. Começa-se com uma rede pequena $N = 1000$, com $p_s = 0.05$ e $p_f = 0.005$. Neste caso, tem-se um espaço

confinado de tamanho 50, com 5 nodos sendo confinados e libertados na fase de fluxo. Quando o tamanho da rede duplica ($N=2000$), em vez de aumentar o confinamento para 100 e o fluxo para 10 nodos, em um único confinamento, adiciona-se um segundo espaço confinado do mesmo tamanho que o primeiro. Repete-se o processo à medida que a rede cresce, mantendo a quantidade de nodos de tamanho do confinamento e o número de nodos de fluxo, mas aumentando o número de espaços confinados para que a rede ainda tenha um número relativo fixo de nós confinados p_s e fluxo relativo p_f (divididos entre espaços confinados independentes). Cada espaço confinado opera apenas com os nós livres (apreensão, liberação e religação conforme descrito anteriormente). Uma vez libertado, um nodo pode ser reapreendido por outro espaço confinado ou até mesmo pelo mesmo.

Neste cenário de confinamento múltiplo, o agrupamento e o tamanho relativo do componente gigante possuem invariância de tamanho, como pode ser visto na Figura 17, à esquerda. Enquanto isso, a modularidade exibe uma fraca dependência com o tamanho da rede N , aumentando lentamente com N (Figura 17, à esquerda). Os algoritmos de detecção de comunidades estruturais, como previamente dito, têm limitações, principalmente no que diz respeito à resolução. A “verdadeira” estrutura comunitária que emerge do modelo proposto não é conhecida a priori. É uma consequência emergente, dinâmica e indireta dos parâmetros-chave do modelo. Portanto, os tamanhos e a modularidade das partições são bons indicadores da estrutura observada, mas não são exatos. Em todo caso, o número de comunidades detectadas tende a ser proporcionais ao número de confinamentos, o que é um resultado razoável.

A forma da distribuição de graus também é mantida utilizando múltiplos confinamentos, embora se espere que em tamanhos realmente grandes, distorções características de redes empíricas surjam limitando o tamanho máximo dos hubs. No modelo isto ocorre devido a limitação de tamanho de cada um dos múltiplos confinamentos, nas redes empíricas este efeito é atribuído a fatores diversos, sendo a incapacidade dos nodos de suportar um limiar elevado de conexões a razão mais comum.

5.5.2 Independência da topologia em relação a redes iniciais e modificação dinâmica de parâmetros

Os valores assintóticos de agrupamento C , distribuição de grau $P_{(k)}$, modularidade M , Φ e tamanho da MCC dependem apenas dos parâmetros p_f e p_s e do tamanho e densidade de links da rede inicial, a disposição inicial destes afeta apenas um transiente. Se rede inicial for uma rede mundo pequeno W-S ou uma rede livre de escala B-A com tamanho N e conectividade média $\langle k \rangle$, a mesma topologia assintótica é obtida, apenas com um transiente diferente. De fato, conforme a escolha da rede tipo W-S (que é normalmente utilizada para criar coeficientes de agrupamento elevados) o coeficiente de agrupamento pode, inclusive, diminuir para o valor assintótico determinado pelos parâmetros p_s e p_f .

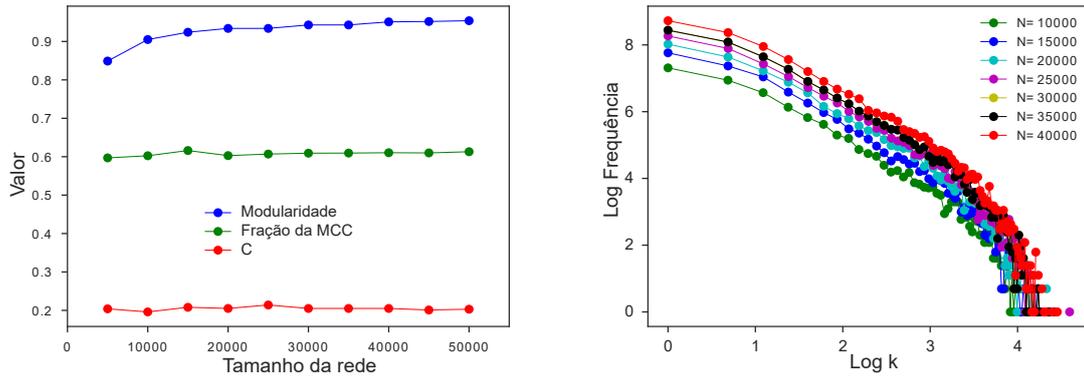


Figura 17 – Invariância de métricas frente a mudança de tamanho. *Painel esquerdo:* evolução temporal de algumas das medidas topológicas para um tamanho absoluto de confinamento fixo (o número de nós confinados é $N_c = 50$ e o número de nós de fluxo é $N_f = 5$). O número de espaços confinados para cada tamanho de rede N é $\mathcal{P} = N/1000$. *Painel direito:* diferentes tamanhos de rede resultam em distribuição de graus diferentes, mas com o método proposto a forma da distribuição de graus é mantida, salvo por um fator multiplicativo de tamanho (ou deslocamento vertical em escala logarítmica).

O sistema não aumenta ou diminui ordenamentos e métricas da rede inicial, (que não é relevante nem o objetivo da dinâmica), mas converge para sua topologia dinâmica, caracterizada inteiramente por seus parâmetros.

A razão pela qual redes iniciais diversas resultam em resultados estacionários iguais é a semi-randômica natureza dos processos. Observando os passos da dinâmica do modelo, percebe-se que, nem a fase de fluxo, nem a fase de reconexões dependem de distribuição de grau, modularidade, coeficiente de agrupamento, etc. A escolha dos nodos confinados e libertados a cada passo, bem como a escolha de links para redirecionamento dos nodos não é função de nenhum parâmetro topológico externo à dinâmica do confinamento. Assim, a reconfiguração da rede inicial escolhida torna-se menos relevante ao longo do transiente até se tornar irrelevante no estado estacionário. Como o objetivo do algoritmo é produzir uma rede dinâmica com características complexas baseadas apenas nos parâmetros da dinâmica, a descaracterização da rede inicial (que é estática) é o resultado desejável, e possibilita estabilidade frente a uma série de processos dinâmicos empíricos.

5.5.3 Crescimento, redução da rede e transformação contínua da rede

Redes complexas empíricas herdam as propriedades dos sistemas complexos, notavelmente, a de serem sistemas abertos. Isto significa que uma representação de uma rede dinâmica empírica deve ser capaz de crescer ou encolher, recebendo novos nodos e conexões ou os perdendo. No modelo mediado pelo confinamento, perturbações como renovação

de nodos por nascimentos e envelhecimentos e mudanças de tamanho são perturbações similares à colocação de uma nova rede inicial. Se a perturbação é pequena, a rede inicial é muito similar à rede não perturbada, e o transiente é praticamente nulo.

Existe um limite para esta resistência a perturbações, mas para crescimento, redução ou renovação de nodos e conexões em pequenas taxas relativas ao tamanho da rede, a dinâmica pode ser considerada semi estável, sem flutuações notáveis das métricas, utilizando uma combinação das propriedades de múltiplos confinamentos e a resistência a perturbações do tipo redes iniciais.

Por outro lado, se existe a intenção de provocar mudanças topológicas graduais como parte de uma dinâmica (em redes empíricas as métricas podem mudar ao longo do tempo), o modelo proposto é capaz de reagir rapidamente a perturbações nos parâmetros da dinâmica. A duração do transiente temporal necessário para atingir o estado assintótico é proporcional à diferença da topologia assintótica e da rede inicial, uma rede de parâmetros muito similares reage em poucos passos temporais a uma mudança de parâmetros pequena.

5.6 Considerações finais do capítulo

Neste capítulo, apresentam-se os resultados numéricos da aplicação de um modelo baseado na dinâmica de confinamento. Características topológicas empíricas são reproduzidas com uma quantidade pequena de parâmetros e emergem com analogias empíricas razoáveis. Durante o crescimento de rede, os parâmetros permanecem estáveis se a variação de tamanho de rede é proporcional ao número de confinamentos independentes, o que é um resultado também razoável empiricamente, já que diversas estruturas se multiplicam com o tamanho da população, seja para atender à infraestrutura necessária, seja pelo aumento das distâncias geográficas associadas. Finalmente, é apontada uma resiliência da dinâmica frente a perturbações que possibilita a representação de um sistema aberto, capaz de suportar de maneira estável a criação e a remoção de nodos e conexões. A rede dinâmica reage a modificações graduais nos parâmetros modificando gradualmente sua topologia, resultando em um modelo de dinâmica bastante completo.

CONCLUSÕES

Espaços confinados estão em toda parte, são caracterizados por uma mistura de isolamento com fluxo. Podem ser observados de maneira explícita, como em prisões, organizações, regiões geográficas em geral, etc. Ou em formas mais sutis, como divisões étnicas, de classe, de convenção social ou barreiras criadas por diferenças de ideologia ou preconceitos. Diante disso, é importante investigar a influência que esses confinamentos exercem sobre a organização e evolução dos sistemas complexos.

No presente trabalho, é introduzido um modelo simplificado de redes dinâmicas, centralizado na tese de que o confinamento pode funcionar como mediador da organização de redes, e catalisador da emergência de fenômenos complexos. Este modelo, introduzido no capítulo 4, é definido por parâmetros que ecoam processos observados em sistemas empíricos, particularmente de redes sociais, e reproduzem, a partir destas premissas intuitivas, estruturas complexas observadas nestes sistemas.

Os parâmetros propostos são o tamanho dos espaços confinados p_s , e o fluxo p_f para dentro e fora desses confinamentos. O tamanho e o número de espaços confinados podem ser associados a organizações do mundo real: indivíduos trabalhando em empresas, estudantes em turmas de instituições de ensino, detentos em prisões, cultos religiosos, movimentos ideológicos fechados, ou grupos de redes sociais digitais. O fluxo está relacionado à escala de tempo da dinâmica e às durações dos confinamentos: como a rotatividade de funcionários em uma organização, tempo da pena a ser cumprida em instituições disciplinares, duração de cursos, etc.

No capítulo 5, é demonstrado que o modelo dá origem à emergência e à estabilidade de fenômenos complexos observados em redes empíricas como: alto coeficiente de agrupamento, distribuições de graus heterogêneos escaláveis com o tamanho da rede e formação de comunidades estruturais. Cada um destes tópicos é explorado na discussão dos resultados.

O agrupamento C elevado resulta da formação de tríades e cliques no espaço confinado, que sobrevivem fora do confinamento na ausência de um mecanismo alienador, de forma similar a grupos de contatos comuns que mantêm contato após se encontrarem em algum espaço confinado, em algum intervalo de tempo.

De forma similar, porém mais sofisticada, os espaços confinados agem como organizadores de comunidades, primeiramente formando grupos interconectados, que se alienam do espaço confinado ao abandoná-lo, criando comunidades independentes. Embora essa não seja a única forma pela qual comunidades se mantêm, o efeito do confinamento inicial por alguma circunstância é frequentemente observado.

Tanto a formação de cliques altamente conectados, quanto de comunidades estruturais, derivam de um princípio com analogia direta a efeitos presentes em redes sociais: grupos altamente interconectados se formam em espaços confinados devido a dificuldade com interação com o exterior ou por conveniência de proximidade. Quando estes grupos saem do confinamento, muitas destas ligações desaparecem, mas frequentemente grupos altamente conectados permanecem conexos na ausência do confinamento. O processo de dissolução lenta das estruturas também tem analogia em redes empíricas e nas trajetórias de vida observadas em comunidades reais.

A dinâmica de confinamento também é capaz de reproduzir distribuições de grau de cauda longa, inclusive leis de potência. Na discussão sobre as distribuições de grau do capítulo 5, foi apontado que os hubs são formados dentro do confinamento: embora a religação seja aleatória, apenas os nodos dentro do confinamento recebem novos links. Nas redes sociais reais, o mecanismo de ligação preferencial nem sempre é explícito, e a razão para um nodo formar conexões pode ser tão sutil quanto alguém permanecer em um local confinado enquanto outros nodos transitam por este local. O modelo proposto também produz nodos com graus de tempo de vida variável, ou seja, os hubs podem emergir e desaparecer na rede dinâmica, característica que a maioria das redes sociais apresenta, e não facilmente reproduzida pelos sistemas de ligação preferencial, onde os hubs envelhecem se tornando cada vez mais conectados.

A topologia geral da rede permanece estável, enquanto os componentes estão em fluxo contínuo. Esta é uma característica importante, frequentemente manifestada em sistemas complexos reais, onde as partes do sistema são substituídas, mas a função e forma se mantêm.

O modelo não requer sistemas sofisticados de reconexão ou métricas forçadas para exibir seus resultados. De maneira notável, todos esses fenômenos emergem de apenas dois parâmetros (p_s e p_f), a partir de qualquer rede de entrada. A rede de entrada desempenha apenas um papel algorítmico na construção da rede dinâmica, e medidas topológicas iguais são obtidas a partir de redes de entrada variadas: modelos W-S e B-A, redes circulares, quadradas, etc. Aproveitando esta independência das condições iniciais, redes randômicas são usadas como redes de entrada, por motivos puramente instrutivos: a auto organização das redes resultantes torna-se evidente devido à escolha de redes aleatórias (que carecem de todas as estruturas complexas desenvolvidas durante a dinâmica) como redes iniciais e este contraste e transformação fica bastante claro no transiente. O elemento chave para tal independência das condições iniciais são os critérios aleatórios para confinamento, liberação e religação de nodos individuais, sem memória do estado inicial.

Esta independência de redes iniciais também pode ser interpretada como resistência a perturbações típicas às que são encontradas em sistemas complexos empíricos: adição e remoção de nodos e conexões, independente do critério, podem ser interpretadas como o

uso de uma nova rede de entrada. Se as perturbações forem razoavelmente pequenas, o transiente é muito rápido, ou não perceptível. A escalabilidade de tamanho é verificada usando múltiplos espaços confinados, uma escolha natural quando pensamos na ampliação de rede e sistemas empíricos, onde o número de organizações, comunidades, estruturas e espaços ocupados cresce proporcionalmente ao aumento do escopo e tamanho da rede analisada.

Aplicações

O modelo tem aplicações práticas possíveis imediatas. Pode ser usado como substrato para simulação de fenômenos em redes dinâmicas, particularmente naquelas onde o confinamento tem papel principal, ou quando a natureza modular e dinâmica é parte importante do modelo. Modelos geradores de rede são, por vezes, usados como *proxies* para redes empíricas, e o modelo mediado pelo confinamento é capaz de fornecer, simultaneamente, agrupamento elevado, efeito de mundo pequeno, distribuições de grau livres de escala, em uma rede com formação de comunidades. Além disso, a rede é dinâmica. Acredita-se que muitos modelos existentes poderiam se beneficiar desta estrutura como base.

Caracterizar a evolução temporal das comunidades, à medida que nascem ou desaparecem, fundem-se ou se dividem, é uma tarefa difícil que envolve lidar com enormes conjuntos de dados temporais, e rastrear módulos individuais da rede. Embora o modelo apresentado represente uma visão simplificada de um fenômeno complexo, acredita-se que ele pode ajudar a compreender a característica modular subjacente a redes reais complexas e, mais importante, como as comunidades podem emergir e evoluir. Em geral, obter dados empíricos de redes dinâmicas com parâmetros específicos é uma tarefa difícil e um limitador da pesquisa em redes dinâmicas em geral. O modelo apresentado pode, possivelmente, servir como substituto nestas investigações.

Oportunidades de desenvolvimentos futuros

Embora apresente resultados notáveis, o modelo mediado por confinamento foi apresentado dentro de um escopo que, não pode (como nenhum modelo pode) abranger toda a riqueza de variações existentes em fenômenos empíricos. Algumas aproximações foram feitas por simplificação e no interesse de demonstrar que uma quantidade mínima de condições era suficiente para emergência de complexidade: os espaços confinados empíricos apresentam complexidades que não foram representadas, como distribuição de tamanhos, confinamentos aninhados e também intersecção de confinamentos. Com efeito, um nodo poderia, em um modelo modificado, fazer parte de múltiplos espaços confinados não excludentes.

O modelo pode ser modificado para o uso de múltiplas comunidades de tamanhos heterogêneos e/ou variáveis. Esta é uma situação mais realista, que pode dar conta do surgimento de uma variabilidade maior de tamanho de comunidades e possibilitar graus máximos, na rede, maiores ou mais reduzidos. Confinamentos podem existir dentro de confinamentos maiores, e a implementação de uma modificação deste tipo é, provavelmente, uma condição necessária para a reprodução de comunidades estruturais, que é um assunto de grande interesse no estudo de comunidades em ciência de redes.

Espaços confinados não precisam, necessariamente, ser independentes, podem ocorrer intersecções de espaços confinados que ocorrem em níveis diferentes. Um indivíduo em uma rede social, por exemplo, pode frequentar simultaneamente vários dos confinamentos propostos: pode ao mesmo tempo estar matriculado em uma instituição de ensino, pertencer a uma ou mais organizações profissionais, morar em uma região específica e ainda pertencer a uma classe social que limita sua livre movimentação na rede. Neste sentido, o modelo de confinamentos independentes aqui apresentado seria apenas uma fotografia parcial de uma complexidade maior. Isso é esperado: em geral, modelos de sistemas complexos representam apenas uma *fatia* dos fenômenos. A inclusão de confinamentos que se intersectam é, provavelmente, necessária para o estudo de outro tópico importante no estudo de redes dinâmicas: intersecção e fusão de comunidades.

O encarceramento penal é de particular interesse como trabalho futuro: os modelos de detecção e desmantelamento de redes criminosas são de grande utilidade na geração de bem-estar público e na redução de danos pessoais e patrimoniais. A maioria dos modelos tem como objetivo desenvolver uma estratégia de ataque efetiva à rede e à infraestrutura de organizações criminosas, e supõe intervenções que removam os agentes criminosos da rede. Esta suposição pode superestimar a capacidade do sistema judicial e de encarceramento em afastar o agente da dinâmica criminosa e, ao mesmo tempo, ignorar a capacidade dos sistemas de confinamento em gerar novos contatos e reorganizar organizações criminosas. O caso brasileiro, onde estas organizações têm alcance global, a partir das prisões, é um exemplo deste tipo de organização por confinamento.

Uma linha natural de estudo seria utilizar os modelos de dinâmica criminosa e que prevejam algum mecanismo de remoção de nodos criminosos da rede, e adaptá-lo como um modelo de confinamento, onde o agente criminoso não é eliminado da rede e, eventualmente, deixa o confinamento. Qual afinal é o efeito de aprisionar uma organização criminosa? Remover a organização da sociedade ou organizar criminalmente os espaços de encarceramento? Essas são questões importantes e dignas de estudo futuro.

Considerações finais

Considera-se que o modelo apresentado é elucidativo sobre dinâmicas de sistemas complexos, demonstrando que o confinamento é um efeito forte e resiliente de organização das redes complexas. O modelo apresentado tem uma gama grande de aplicações imediatas, quando combinado a outros modelos teóricos que utilizam a topologia como substrato. Finalmente, o modelo apresenta riqueza em possibilidades futuras de modificação, de forma a reproduzir características adicionais ou melhor representar e compreender os sistemas empíricos. Assim, os resultados apresentados nesta tese não esgotam o assunto, mas servem como uma pedra fundamental para linhas adicionais de investigação.

REFERÊNCIAS

- 1 WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *Nature*, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/30918>>. Citado 6 vezes nas páginas 7, 29, 34, 35, 36 e 54.
- 2 JACOMY, M. et al. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, Public Library of Science, v. 9, n. 6, p. e98679, jun. 2014. ISSN 1932-6203. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679>>. Citado 3 vezes nas páginas 9, 59 e 60.
- 3 BLONDEL, V. D. et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2008, n. 10, p. P10008, out. 2008. ISSN 1742-5468. Disponível em: <<https://dx.doi.org/10.1088/1742-5468/2008/10/P10008>>. Citado 5 vezes nas páginas 9, 44, 54, 59 e 60.
- 4 ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, American Physical Society, v. 74, n. 1, p. 47–97, jan. 2002. Disponível em: <<https://link.aps.org/doi/10.1103/RevModPhys.74.47>>. Citado 4 vezes nas páginas 12, 34, 38 e 55.
- 5 SCHNEIDER, É. M. et al. Dynamic Modular Networks Model Mediated by Confinement. *Applied Network Science*, SpringerOpen, v. 4, n. 1, p. 1–17, dez. 2019. ISSN 2364-8228. Disponível em: <<https://appliednetsci.springeropen.com/articles/10.1007/s41109-019-0143-2>>. Citado 2 vezes nas páginas 18 e 55.
- 6 TRAVIESO, G. et al. *A complex network approach to cloud computing*. [S.l.], 2015. ArXiv:1504.02656 [physics] type: article. Disponível em: <<http://arxiv.org/abs/1504.02656>>. Citado na página 22.
- 7 LU, Y. et al. Structural Performance of Satellite Networks: A Complex Network Perspective. *IEEE Systems Journal*, v. 15, n. 3, p. 3848–3859, set. 2021. ISSN 1937-9234. Citado na página 22.
- 8 ADAMIC, L. A. The Small World Web. In: ABITEBOUL, S.; VERCOUSTRE, A.-M. (Ed.). *Research and Advanced Technology for Digital Libraries*. Berlin, Heidelberg: Springer, 1999. (Lecture Notes in Computer Science), p. 443–452. ISBN 9783540481553. Citado 2 vezes nas páginas 22 e 34.
- 9 ROY, S.; MUSAWI, A. F. A.; GHOSH, P. Inferring links in directed complex networks through feed forward loop motifs. *Humanities and Social Sciences Communications*, Palgrave, v. 10, n. 1, p. 1–13, jun. 2023. ISSN 2662-9992. Disponível em: <<https://www.nature.com/articles/s41599-023-01863-z>>. Citado na página 22.
- 10 REYNA, O. S. S.; MOTA, I. F. de la. Complex Networks of the Air Passenger Traffic in Monterrey’s Airport. *Transportation Research Procedia*, v. 48, p. 23–31, jan. 2020. ISSN 2352-1465. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352146520304154>>. Citado na página 22.

- 11 HÁZNAGY, A. et al. Complex network analysis of public transportation networks: A comprehensive study. In: *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. [S.l.: s.n.], 2015. p. 371–378. Citado na página 22.
- 12 TIAN, Z. et al. Analysis of Urban Road Traffic Network Based on Complex Network. *Procedia Engineering*, v. 137, p. 537–546, jan. 2016. ISSN 1877-7058. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877705816003179>>. Citado na página 22.
- 13 BOCCALETTI, S. et al. Complex networks: Structure and dynamics. *Physics Reports*, v. 424, n. 4, p. 175–308, fev. 2006. ISSN 0370-1573. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S037015730500462X>>. Citado na página 22.
- 14 KITO, T.; UEDA, K. The implications of automobile parts supply network structures: A complex network approach. *CIRP Annals*, v. 63, n. 1, p. 393–396, jan. 2014. ISSN 0007-8506. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S000785061400122X>>. Citado na página 22.
- 15 LI, Y. et al. Complex networks in advanced manufacturing systems. *Journal of Manufacturing Systems*, v. 43, p. 409–421, abr. 2017. ISSN 0278-6125. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0278612516300887>>. Citado na página 22.
- 16 KIM, B. J. Performance of networks of artificial neurons: The role of clustering. *Physical Review E*, American Physical Society, v. 69, n. 4, p. 045101, abr. 2004. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevE.69.045101>>. Citado na página 22.
- 17 STAM, C. J.; REIJNEVELD, J. C. Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics*, v. 1, n. 1, p. 3, jul. 2007. ISSN 1753-4631. Disponível em: <<https://doi.org/10.1186/1753-4631-1-3>>. Citado na página 22.
- 18 DOWNES, J. H. et al. Emergence of a Small-World Functional Network in Cultured Neurons. *PLOS Computational Biology*, Public Library of Science, v. 8, n. 5, p. e1002522, maio 2012. ISSN 1553-7358. Disponível em: <<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002522>>. Citado na página 22.
- 19 JEONG, H. et al. Lethality and centrality in protein networks. *Nature*, v. 411, n. 6833, p. 41–42, maio 2001. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/35075138>>. Citado na página 22.
- 20 RUAL, J.-F. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, v. 437, n. 7062, p. 1173–1178, out. 2005. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/nature04209>>. Citado na página 22.
- 21 VILLAVECES, J. M. et al. Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study. *Database*, v. 2015, p. bau131, jan. 2015. ISSN 1758-0463. Disponível em: <<https://europepmc.org/articles/PMC4316181>>. Citado na página 22.

- 22 CHASMAN, D.; SIAHPIRANI, A. F.; ROY, S. Network-based approaches for analysis of complex biological systems. *Current Opinion in Biotechnology*, v. 39, p. 157–166, jun. 2016. ISSN 0958-1669. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095816691630115X>>. Citado na página 22.
- 23 COSTA, L. d. F.; RODRIGUES, F. A.; CRISTINO, A. S. Complex networks: the key to systems biology. *Genetics and Molecular Biology*, Sociedade Brasileira de Genética, v. 31, p. 591–601, 2008. ISSN 1415-4757, 1678-4685. Disponível em: <<https://www.scielo.br/j/gmb/a/3J9wKvNkBFKS3LHFRdDmhFy/?lang=en>>. Citado na página 22.
- 24 DOMINIETTO, M. et al. Role of Complex Networks for Integrating Medical Images and Radiomic Features of Intracranial Ependymoma Patients in Response to Proton Radiotherapy. *Frontiers in Medicine*, v. 6, 2020. ISSN 2296-858X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fmed.2019.00333>>. Citado na página 22.
- 25 UTHAMACUMARAN, A. A Review of Complex Systems Approaches to Cancer Networks. *Complex Systems*, v. 29, n. 4, p. 779–835, dez. 2020. ISSN 08912513. ArXiv:2009.12693 [nlin, q-bio]. Disponível em: <<http://arxiv.org/abs/2009.12693>>. Citado na página 22.
- 26 SEO, H. et al. Network-based approaches for anticancer therapy (Review). *International Journal of Oncology*, Spandidos Publications, v. 43, n. 6, p. 1737–1744, dez. 2013. ISSN 1019-6439. Disponível em: <<https://www.spandidos-publications.com/10.3892/ijo.2013.2114>>. Citado na página 22.
- 27 KERMACK, W. O.; MCKENDRICK, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, The Royal Society, v. 115, n. 772, p. 700–721, 1927. ISSN 0950-1207. Disponível em: <<https://www.jstor.org/stable/94815>>. Citado na página 22.
- 28 PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, American Physical Society, v. 86, n. 14, p. 3200–3203, abr. 2001. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevLett.86.3200>>. Citado na página 22.
- 29 CHAKRABARTI, D. et al. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, ACM New York, NY, USA, v. 10, n. 4, p. 1–26, 2008. Citado na página 22.
- 30 COHEN, R.; HAVLIN, S.; AVRAHAM, D. ben. Efficient Immunization Strategies for Computer Networks and Populations. *Physical Review Letters*, American Physical Society, v. 91, n. 24, p. 247901, dez. 2003. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevLett.91.247901>>. Citado na página 22.
- 31 BRIESEMEISTER, L.; LINCOLN, P.; PORRAS, P. Epidemic profiles and defense of scale-free networks. In: *Proceedings of the 2003 ACM workshop on Rapid malware*. New York, NY, USA: Association for Computing Machinery, 2003. (WORM '03), p. 67–75. ISBN 9781581137859. Disponível em: <<https://doi.org/10.1145/948187.948200>>. Citado na página 22.

- 32 BALCAN, D. et al. Seasonal transmission potential and activity peaks of the new influenza a(h1n1): a monte carlo likelihood analysis based on human mobility. *BMC Medicine*, v. 7, n. 1, p. 45, 2009. ISSN 1741-7015. Disponível em: <<https://doi.org/10.1186/1741-7015-7-45>>. Citado na página 22.
- 33 BIGGERSTAFF, M. et al. Results from the centers for disease control and prevention's predict the 2013-2014 influenza season challenge. *BMC Infectious Diseases*, v. 16, n. 1, p. 357, 2016. ISSN 1471-2334. Disponível em: <<https://doi.org/10.1186/s12879-016-1669-x>>. Citado na página 22.
- 34 CHOWELL, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, v. 2, n. 3, p. 379–398, ago. 2017. ISSN 2468-0427. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2468042717300234>>. Citado na página 23.
- 35 PASTOR-SATORRAS, R. et al. Epidemic processes in complex networks. *RMP*, American Physical Society, v. 87, n. 3, p. 925–979, ago. 2015. Disponível em: <<https://link.aps.org/doi/10.1103/RevModPhys.87.925>>. Citado na página 23.
- 36 CHEN, S. S. a. Y. *The Dynamics of Public Opinion in Complex Networks*. JASSS, 2008. Disponível em: <<https://www.jasss.org/11/4/2.html>>. Citado na página 23.
- 37 WU, L. et al. *Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task*. Rochester, NY, 2008. Disponível em: <<https://papers.ssrn.com/abstract=1130251>>. Citado na página 23.
- 38 MASUDA, N. Opinion control in complex networks. *New Journal of Physics*, IOP Publishing, v. 17, n. 3, p. 033031, mar. 2015. ISSN 1367-2630. Disponível em: <<https://dx.doi.org/10.1088/1367-2630/17/3/033031>>. Citado na página 23.
- 39 GENG, L. et al. Modeling public opinion dissemination in a multilayer network with SEIR model based on real social networks. *Engineering Applications of Artificial Intelligence*, v. 125, p. 106719, out. 2023. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095219762300903X>>. Citado na página 23.
- 40 COSTA, J. B. D. da et al. Efficient data dissemination protocol based on complex networks' metrics for urban vehicular networks. *Journal of Internet Services and Applications*, v. 10, n. 1, p. 15, ago. 2019. ISSN 1869-0238. Disponível em: <<https://doi.org/10.1186/s13174-019-0114-y>>. Citado na página 23.
- 41 MALIK, H. A. M. Complex Network Formation and Analysis of Online Social Media Systems. *Computer Modeling in Engineering & Sciences*, v. 130, n. 3, p. 1737, 2022. ISSN 1526-1506. Disponível em: <https://www.academia.edu/100818819/Complex_Network_Formation_and_Analysis_of_Online_Social_Media_Systems>. Citado na página 23.
- 42 SADRI, A. M. et al. Analysis of social interaction network properties and growth on Twitter. *Social Network Analysis and Mining*, v. 8, n. 1, 2018. ISSN 1869-5450. Disponível em: <https://www.academia.edu/101245225/Analysis_of_social_interaction_network_properties_and_growth_on_Twitter>. Citado na página 23.

- 43 EULER, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, p. 128–140, 1741. Citado na página 23.
- 44 ALEXANDERSON, G. About the Cover: Euler and the Königsberg Bridges: A Historical View. *Bulletin (New Series) of the American Mathematical Society*, v. 43, p. 567–573, out. 2006. Citado na página 23.
- 45 ZHOU, X. et al. Human symptoms–disease network. *Nature Communications*, Nature Publishing Group, v. 5, n. 1, p. 4212, jun. 2014. ISSN 2041-1723. Disponível em: <<https://www.nature.com/articles/ncomms5212>>. Citado na página 24.
- 46 ALEXANDER-BLOCH, A. F. et al. The architecture of co-morbidity networks of physical and mental health conditions in military veterans. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Royal Society, v. 476, n. 2239, p. 20190790, jul. 2020. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rspa.2019.0790>>. Citado na página 24.
- 47 REUS, M. A. d.; HEUVEL, M. P. v. d. Rich Club Organization and Intermodule Communication in the Cat Connectome. *Journal of Neuroscience*, Society for Neuroscience, v. 33, n. 32, p. 12929–12939, ago. 2013. ISSN 0270-6474, 1529-2401. Disponível em: <<https://www.jneurosci.org/content/33/32/12929>>. Citado na página 24.
- 48 BOTA, M.; SWANSON, L. W. Online workbenches for neural network connections. *Journal of Comparative Neurology*, v. 500, n. 5, p. 807–814, 2007. ISSN 1096-9861. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.21209>>. Citado na página 24.
- 49 KÖTTER, R. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac Database. *Neuroinformatics*, v. 2, n. 2, p. 127–144, jun. 2004. ISSN 1559-0089. Disponível em: <<https://doi.org/10.1385/NI:2:2:127>>. Citado na página 24.
- 50 LUSSEAU, D. et al. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, v. 54, n. 4, p. 396–405, set. 2003. ISSN 1432-0762. Disponível em: <<https://doi.org/10.1007/s00265-003-0651-y>>. Citado na página 24.
- 51 MUREDDU, M. *Representation of the German transmission grid for Renewable Energy Sources impact analysis*. [S.l.], 2016. ArXiv:1612.05532 [physics] type: article. Disponível em: <<http://arxiv.org/abs/1612.05532>>. Citado na página 25.
- 52 ZIMMERMAN, R. D.; MURILLO-SÁNCHEZ, C. E.; THOMAS, R. J. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, IEEE, v. 26, n. 1, p. 12–19, 2010. Citado na página 25.
- 53 LESKOVEC, J. et al. *Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters*. [S.l.], 2008. ArXiv:0810.1355 [physics] type: article. Disponível em: <<http://arxiv.org/abs/0810.1355>>. Citado na página 25.
- 54 SMITH, D. A.; WHITE, D. R. Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965–1980*. *Social Forces*, v. 70, n. 4, p. 857–893,

- jun. 1992. ISSN 0037-7732. Disponível em: <<https://doi.org/10.1093/sf/70.4.857>>. Citado na página 25.
- 55 ONNELA, J.-P. et al. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, v. 68, n. 5, p. 056110, nov. 2003. ISSN 1063-651X, 1095-3787. ArXiv:cond-mat/0302546. Disponível em: <<http://arxiv.org/abs/cond-mat/0302546>>. Citado na página 25.
- 56 GONZÁLEZ-AVELLA, J. C.; QUADROS, V. H. de; IGLESIAS, J. R. Network topology and interbank credit risk. *Chaos, Solitons & Fractals*, v. 88, p. 235–243, jul. 2016. ISSN 0960-0779. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960077915004099>>. Citado na página 25.
- 57 MOLONTAY, R.; NAGY, M. Two Decades of Network Science as seen through the co-authorship network of network scientists. In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. [s.n.], 2019. p. 578–583. ArXiv:1908.08478 [physics]. Disponível em: <<http://arxiv.org/abs/1908.08478>>. Citado na página 25.
- 58 LOUIS, U. of M.-S. et al. *The St. Louis Homicide Project: Local Responses to a National Problem*. University, 1991. Disponível em: <<https://books.google.com.br/books?id=umVAPQAACAAJ>>. Citado na página 25.
- 59 CUNHA, B. Requião da; GONZÁLEZ-AVELLA, J. C.; GONÇALVES, S. Fast fragmentation of networks using module-based attacks. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 11, p. e0142824, 2015. Citado na página 25.
- 60 CUNHA, B. R. da; GONÇALVES, S. Performance of attack strategies on modular networks. *Journal of Complex Networks*, Oxford University Press, v. 5, n. 6, p. 913–923, 2017. Citado na página 25.
- 61 CUNHA, B. R. da; GONÇALVES, S. Topology, robustness, and structural controllability of the Brazilian Federal Police criminal intelligence network. *Applied Network Science*, SpringerOpen, v. 3, n. 1, p. 1–20, dez. 2018. ISSN 2364-8228. Disponível em: <<https://appliednetsci.springeropen.com/articles/10.1007/s41109-018-0092-1>>. Citado na página 25.
- 62 CUNHA, B. R. da et al. Assessing police topological efficiency in a major sting operation on the dark web. *Scientific reports*, Nature Publishing Group UK London, v. 10, n. 1, p. 73, 2020. Citado na página 25.
- 63 MISLOVE, A. et al. Growth of the flickr social network. In: *Proceedings of the first workshop on Online social networks*. New York, NY, USA: Association for Computing Machinery, 2008. (WOSN '08), p. 25–30. ISBN 9781605581828. Disponível em: <<https://doi.org/10.1145/1397735.1397742>>. Citado na página 25.
- 64 BLAGUS, N.; ŠUBELJ, L.; BAJEC, M. Self-similar scaling of density in complex real-world networks. *Physica A: Statistical Mechanics and its Applications*, v. 391, n. 8, p. 2794–2802, abr. 2012. ISSN 03784371. ArXiv:1110.5609 [physics]. Disponível em: <<http://arxiv.org/abs/1110.5609>>. Citado na página 25.

- 65 MICHELL, L.; AMOS, A. Girls, pecking order and smoking. *Social Science & Medicine*, v. 44, n. 12, p. 1861–1869, jun. 1997. ISSN 0277-9536. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S027795369600295X>>. Citado na página 25.
- 66 SNIJDERS, T. A. B.; BUNT, G. G. van de; STEGLICH, C. E. G. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, v. 32, n. 1, p. 44–60, jan. 2010. ISSN 0378-8733. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378873309000069>>. Citado na página 25.
- 67 FIRE, M.; PUZIS, R.; ELOVICI, Y. *Organization Mining Using Online Social Networks*. [S.l.], 2013. ArXiv:1303.3741 [physics] type: article. Disponível em: <<http://arxiv.org/abs/1303.3741>>. Citado na página 25.
- 68 LUCE, R. D.; PERRY, A. D. A method of matrix analysis of group structure. *Psychometrika*, v. 14, n. 2, p. 95–116, jun. 1949. ISSN 1860-0980. Disponível em: <<https://doi.org/10.1007/BF02289146>>. Citado na página 29.
- 69 HOLLAND, P. W.; LEINHARDT, S. Transitivity in Structural Models of Small Groups. *Comparative Group Studies*, SAGE Publications, v. 2, n. 2, p. 107–124, maio 1971. ISSN 0010-4108. Disponível em: <<https://doi.org/10.1177/104649647100200201>>. Citado 2 vezes nas páginas 29 e 55.
- 70 ERDÖS, P.; RÉNYI, A. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, v. 6, p. 290–297, 1959. Citado 2 vezes nas páginas 32 e 50.
- 71 GILBERT, E. N. Random Graphs. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 30, n. 4, p. 1141–1144, 1959. ISSN 0003-4851. Disponível em: <<https://www.jstor.org/stable/2237458>>. Citado na página 32.
- 72 ERDÖS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. math. inst. hung. acad. sci.*, v. 5, n. 1, p. 17–60, 1960. Citado na página 33.
- 73 ERDÖS, P. On the evolution of random graphs. *Publ Math Inst Hungarian Acad Sci*, v. 5, p. 17, 1960. Citado na página 33.
- 74 PASTOR-SATORRAS, R.; VAZQUEZ, A.; VESPIGNANI, A. Dynamical and correlation properties of the Internet. *Physical Review Letters*, v. 87, n. 25, p. 258701, nov. 2001. ISSN 0031-9007, 1079-7114. ArXiv:cond-mat/0105161. Disponível em: <<http://arxiv.org/abs/cond-mat/0105161>>. Citado na página 34.
- 75 NEWMAN, M. E. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 98, n. 2, p. 404–409, 2001. Citado 2 vezes nas páginas 34 e 38.
- 76 NEWMAN, M. E. Scientific collaboration networks. I. Network construction and fundamental results. *Physical review E*, APS, v. 64, n. 1, p. 016131, 2001. Citado 2 vezes nas páginas 34 e 38.
- 77 NEWMAN, M. E. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, APS, v. 64, n. 1, p. 016132, 2001. Citado na página 34.

- 78 BARABÁSI, A. L. et al. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, v. 311, n. 3, p. 590–614, ago. 2002. ISSN 0378-4371. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378437102007367>>. Citado 2 vezes nas páginas 34 e 38.
- 79 WAGNER, A.; FELL, D. A. The small world inside large metabolic networks. *Proceedings. Biological Sciences*, v. 268, n. 1478, p. 1803–1810, set. 2001. ISSN 0962-8452. Citado na página 34.
- 80 MONTOYA, J. M.; SOLÉ, R. V. Small World Patterns in Food Webs. *Journal of Theoretical Biology*, v. 214, n. 3, p. 405–412, fev. 2002. ISSN 0022-5193. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022519301924609>>. Citado 2 vezes nas páginas 34 e 38.
- 81 CANCHO, R. Ferrer-i; SOLE, R. The small world of human language. *Proceedings. Biological sciences / The Royal Society*, v. 268, p. 2261–5, dez. 2001. Citado 2 vezes nas páginas 34 e 38.
- 82 YOON, S.-H. et al. Weighted evolving networks. *Physical review letters*, APS, v. 86, n. 25, p. 5835, 2001. Citado 2 vezes nas páginas 34 e 38.
- 83 KARINTHY, F. Chain-links. *Everything is different*, p. 21–26, 1929. Citado na página 35.
- 84 ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Diameter of the World-Wide Web. *Nature*, Nature Publishing Group, v. 401, n. 6749, p. 130–131, set. 1999. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/43601>>. Citado na página 38.
- 85 KUMAR, S. et al. The MASC/BGMP architecture for inter-domain multicast routing. In: *Proceedings of the ACM SIGCOMM '98 conference on Applications, technologies, architectures, and protocols for computer communication*. New York, NY, USA: Association for Computing Machinery, 1998. (SIGCOMM '98), p. 93–104. ISBN 9781581130034. Disponível em: <<https://dl.acm.org/doi/10.1145/285237.285264>>. Citado na página 38.
- 86 BRODER, A. et al. Graph structure in the Web. *Computer Networks*, v. 33, n. 1, p. 309–320, jun. 2000. ISSN 1389-1286. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1389128600000839>>. Citado na página 38.
- 87 ADAMIC, L. A.; HUBERMAN, B. A. Power-Law Distribution of the World Wide Web. *Science*, American Association for the Advancement of Science, v. 287, n. 5461, p. 2115–2115, mar. 2000. Disponível em: <<https://www.science.org/doi/abs/10.1126/science.287.5461.2115a>>. Citado na página 38.
- 88 FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the Internet topology. *ACM SIGCOMM Computer Communication Review*, v. 29, n. 4, p. 251–262, ago. 1999. ISSN 0146-4833. Disponível em: <<https://dl.acm.org/doi/10.1145/316194.316229>>. Citado na página 38.
- 89 GOVINDAN, R.; TANGMUNARUNKIT, H. Heuristics for Internet map discovery. *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, IEEE, Tel Aviv, Israel, v. 3, p. 1371–1380, 2000. Disponível em: <<http://ieeexplore.ieee.org/document/832534/>>. Citado na página 38.

- 90 BARABASI, A. L.; ALBERT, R. Emergence of scaling in random networks. *Science (New York, N.Y.)*, v. 286, n. 5439, p. 509–512, out. 1999. ISSN 1095-9203. Citado na página 38.
- 91 LILJEROS, F. et al. The web of human sexual contacts. *Nature*, v. 411, n. 6840, p. 907–908, 2001. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/35082140>>. Citado 2 vezes nas páginas 38 e 55.
- 92 JEONG, H. et al. The large-scale organization of metabolic networks. *Nature*, Nature Publishing Group, v. 407, n. 6804, p. 651–654, out. 2000. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/35036627>>. Citado na página 38.
- 93 MASON, R. J.; GREENE, K.; VOELKER, D. R. Surfactant protein A and surfactant protein D in health and disease. *The American Journal of Physiology*, v. 275, n. 1, p. L1–13, jul. 1998. ISSN 0002-9513. Citado na página 38.
- 94 REDNER, S. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, v. 4, n. 2, p. 131–134, jul. 1998. ISSN 1434-6036. Disponível em: <<https://doi.org/10.1007/s100510050359>>. Citado na página 38.
- 95 AIELLO, W.; CHUNG, F.; LU, L. A random graph model for massive graphs. In: *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. New York, NY, USA: Association for Computing Machinery, 2000. (STOC '00), p. 171–180. ISBN 9781581131840. Disponível em: <<https://dl.acm.org/doi/10.1145/335305.335326>>. Citado na página 38.
- 96 NEWMAN, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 103, n. 23, p. 8577–8582, jun. 2006. Disponível em: <<https://www.pnas.org/doi/10.1073/pnas.0601602103>>. Citado 4 vezes nas páginas 42, 43, 54 e 59.
- 97 FORTUNATO, S.; BARTHÉLEMY, M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 104, n. 1, p. 36–41, jan 2007. Citado na página 45.
- 98 ALVAREZ, M. C.; SALLA, F.; DIAS, C. N. Das Comissões de Solidariedade ao Primeiro Comando da Capital em São Paulo. *Tempo Social*, Departamento de Sociologia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, v. 25, p. 61–82, jun. 2013. ISSN 0103-2070, 1809-4554. Disponível em: <<https://www.scielo.br/j/ts/a/rdJQw9GPH7MwjDQBjHx8cGx/>>. Citado na página 48.
- 99 RIBEIRO, L.; OLIVEIRA, V. N.; BASTOS, L. Pavilhões do Primeiro Comando da Capital: tensões e conflitos em uma unidade prisional de segurança máxima em Minas Gerais. *O Público e o Privado*, v. 17, n. 33 jan.jun, p. 213–241, dez. 2019. ISSN 2238-5169. Disponível em: <<https://revistas.uece.br/index.php/opublicoeoprivado/article/view/2264>>. Citado na página 48.
- 100 ESPÍNDULA, F. S. Considerações sobre as principais facções criminosas brasileiras: Comando Vermelho (CV) e Primeiro Comando da Capital (PCC) e os mecanismos

- do estado no combate e prevenção ao crime organizado. 2018. Disponível em: <<https://repositorio.animaeducacao.com.br/handle/ANIMA/7533>>. Citado na página 48.
- 101 ÁVALOS, H. S. *First Capital Command - PCC*. 2017. Disponível em: <<https://insightcrime.org/brazil-organized-crime-news/first-capital-command-pcc-profile/>>. Citado na página 48.
- 102 MICHAIL, O. An introduction to temporal graphs: An algorithmic perspective. In: _____. *Algorithms, Probability, Networks, and Games: Scientific Papers and Essays Dedicated to Paul G. Spirakis on the Occasion of His 60th Birthday*. Cham: Springer International Publishing, 2015. p. 308–343. ISBN 978-3-319-24024-4. Disponível em: <https://doi.org/10.1007/978-3-319-24024-4_18>. Citado na página 49.
- 103 NICOSIA, V. et al. Components in time-varying graphs. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, AIP Publishing, v. 22, n. 2, abr. 2012. ISSN 1089-7682. Disponível em: <<http://dx.doi.org/10.1063/1.3697996>>. Citado na página 49.
- 104 CASTEIGTS PAOLA FLOCCHINI, W. Q. A.; SANTORO, N. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, Taylor & Francis, v. 27, n. 5, p. 387–408, 2012. Disponível em: <<https://doi.org/10.1080/17445760.2012.668546>>. Citado na página 49.
- 105 NEWMAN, M.; GIRVAN, M. Finding and Evaluating Community Structure in Networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, v. 69, p. 026113, mar. 2004. Citado na página 54.
- 106 BARABÁSI, A. L. et al. *Network science*. Cambridge, UK: Cambridge university press, 2016. Citado na página 59.
- 107 BASTIAN, M. et al. Gephi: an open source software for exploring and manipulating networks. *Icwm*, San Jose, California, v. 8, n. 2009, p. 361–362, 2009. Citado na página 59.
- 108 MARTIN, S. et al. OpenOrd: an open-source toolbox for large graph layout. In: *Visualization and Data Analysis 2011*. SPIE, 2011. v. 7868, p. 45–55. Disponível em: <<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7868/786806/OpenOrd-an-open-source-toolbox-for-large-graph-layout/10.1117/12.871402.full>>. Citado na página 59.
- 109 FORTUNATO, S. Community detection in graphs. *Physics Reports*, v. 486, n. 3, p. 75–174, fev. 2010. ISSN 0370-1573. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0370157309002841>>. Citado na página 62.
- 110 LANCICHINETTI, A.; FORTUNATO, S.; RADICCHI, F. Benchmark graphs for testing community detection algorithms. *Physical Review E*, American Physical Society, v. 78, n. 4, p. 046110, out. 2008. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevE.78.046110>>. Citado na página 62.

Apêndices

APÊNDICE A – IMPLEMENTAÇÃO COMPUTACIONAL

Abaixo é apresentada uma implementação em Python da dinâmica de confinamento.

Pré-requisitos: Python 3.7+ com as bibliotecas pandas, random e igraph.

Os parâmetros da rede inicial e da dinâmica são setados logo após a declaração inicial das bibliotecas, junto com o número de passos temporais a serem executados. A rede inicial é E-R, mas qualquer gerador ou arquivo de rede .gml pode ser usado com mínima alteração do código.

O programa cria como saídas: um arquivo .txt com a evolução temporal das métricas topológicas, um arquivo .gml com a componente gigante final e um arquivo .gml com a rede completa (todas as componentes finais). Arquivos .gml podem ser abertos e manipulados facilmente em uma série de programas de análise de grafos (Gephi, por exemplo).

```
# -*- coding: utf-8 -*-

#bibliotecas:
import igraph as ig
import random
import pandas as pd

#%%
##### PARAMETROS REDE INICIAL #####
N= int(10000) #Tamanho da rede N
k_med = int(8) #Conectividade Média
p_size = float(0.05) #Tamanho relativo do confinamento ps
p_flux = float(0.005) #Tamanho relativo do fluxo pf
#Numero de passos temporais a serem simulados
max_time = int(5000)

##### Definições de funções auxiliares #####

#Retorna a maior componente conexa de uma rede input net1
```

```
def giant_comp(net1):
    ggcc = net1.decompose(minelements=1)
    maxsize = 0
    max_lox= 0
    for i in range(len(ggcc)):
        if (ggcc[i].vcount()) > maxsize:
            maxsize = ggcc[i].vcount()
            max_lox = i
    return ggcc[max_lox]

# Retorna a Taxa Phi(float) na rede Network, de acordo com o
# particionamento clusters (que é gerado pelo algoritmo detector)
def conductance(Network,clusters):
    k_in = 0
    k_out = 0
    k_C = sum(Network.degree())
    for i in range(len(clusters)):
        current = Network.subgraph(clusters[i],
                                   implementation="auto")
        k_in = k_in + sum(current.degree())
    k_out = k_C - k_in
    return float(k_out)/float(k_C)

#Retorna a quantidade de links por categoria de uma rede
# l_ii links internos ao confinamento
# l_io links interfaciais
# l_oo links externos ao confinamento
def links_distribution(Network):
    confined_nodes = []
    free_nodes = []
    l_sum = sum(Network.degree())
    for i in range(N):
        if (Network.vs[i]["prison"] == 1):
            confined_nodes.append(i)
        else:
            free_nodes.append(i)
    current = Network.subgraph(confined_nodes,
                               implementation="auto")
    l_ii = sum(current.degree())
```

```

current = Network.subgraph(free_nodes,
                           implementation="auto")

l_oo = sum(current.degree())
l_io = l_sum - l_ii - l_oo
return l_ii, l_io, l_oo

# Abaixo é gerado um identificador unico e
# um preâmbulo para arquivos salvos
s_number = random.randint(0,1000000) #
Preamble = "N"+str(N)+"_K"+str(k_med)+"_"+str(p_size).replace(".", "_",1)+"_"
+"F"+str(p_flux).replace(".", "_",1)+"_"
+"Sim"+str(s_number).zfill(7)+"_"

##### Constrói rede E-R Original #####
g = ig.Graph.Erdos_Renyi(N, float(k_med)/(float(N)))

#opcionalmente geradores de redes B-A ou W-S podem ser usados:
#g = ig.Graph.Barabasi(N,int(k_med/2))
#g = ig.Graph.Watts_Strogatz(1, N, k_med/2, 0.08, loops=False,
# multiple=False)

##### Inicia o confinamento inicial #####
# o atrib. do objeto rede "prison" é 1 (confinado) ou 0 (livre)
# o atrib. do objeto rede "jailtime" é o tempo confinado
prison = sorted(random.sample(range(N),int(p_size*N)))
free = []
for i in range(N):
    if (i in prison):
        g.vs[i]["prison"] = 1
        g.vs[i]["jailtime"] = 1
    else:
        g.vs[i]["prison"] = 0
        g.vs[i]["jailtime"] = 0
        free.append(i)

##### Realiza medidas iniciais Measures #####
tevol_dataframe = pd.DataFrame()

```

```

g_giant = giant_comp(g)
gcs = g_giant.vcount()/float(N)
part = g_giant.community_multilevel(weights=None,
                                     return_levels=False,
                                     resolution=1)
clu = g_giant.transitivity_undirected(mode="zero")
Mdlarity = part.modularity
M_number = len(part)
Conductance = conductance(g_giant,part)
ii, io, oo = links_distribution(N,g)

# As metricas são salvas em um pandas Dataframe
# os nomes das colunas são as strings abaixo
data_insert = {"time":0,
               "clustering":clu,
               "gc_size":gcs,
               "modularity":Mdlarity,
               "module_number":M_number,
               "phi":Conductance,
               "l_ii":ii,"l_io":io,"l_oo":oo}
# O nome do dataframe de sries temporais da metricas
# é tevol_dataframe
tevol_dataframe = tevol_dataframe.append(data_insert,
                                         ignore_index=True)

#####
##### Laço de evolução temporal #####
for it in range(1,max_time+1):
    if it%100 == 0:
        print( "t=", it) #imprime o ciclo atual a cada 100

#####
##### Varredura de reconexões #####
for i in range(len(prison)):
    a=prison[i]
    g.vs[a]["jailtime"] = g.vs[a]["jailtime"] +1
    neigh = g.neighbors(a)
    outsiders = []

```

```

prisoners = []
for j in range(len(neigh)):
    if (g.vs[neigh[j]]["prison"] == 1):
        prisoners.append(neigh[j])
    else:
        outsiders.append(neigh[j])

if (len(outsiders) > 0):
    candidates = list(set(prison)-set(prisoners)-set([a]))
    if (len(candidates) > 0):
        b = random.sample(outsiders,1)
        g.delete_edges([(b[0],a)])
        b = random.sample(candidates,1)
        g.add_edges([(b[0],a)])

#####
#### Varredura de fluxo
#####
apreedended = random.sample(free, int(N*p_flux))
freed = random.sample(prison, int(N*p_flux))
for i in range(len(apreedended)):
    a = apreedended[i]
    g.vs[a]["prison"] = 1
    prison.append(a)
    free.remove(a)
    a = freed[i]
    g.vs[a]["prison"] = 0
    free.append(a)
    prison.remove(a)

##### Atualização de medidas no Dataframe tevol #####
g_giant = giant_comp(g)
gcs = g_giant.vcount()/float(N)
part = g_giant.community_multilevel(weights=None,
                                     return_levels=False,
                                     resolution=1)

clu = g_giant.transitivity_undirected(mode="zero")
Mdlarity = part.modularity
M_number = len(part)

```

```
Conductance = conductance(g_giant,part)
ii, io, oo = links_distribution(N,g)

data_insert = {"time":it,
               "clustering":clu,
               "gc_size":gcs,
               "modularity":Mdlarity,
               "module_number":M_number,
               "phi":Conductance,
               "l_ii":ii,"l_io":io,"l_oo":oo}

tevol_dataframe = tevol_dataframe.append(data_insert,
                                         ignore_index=True)

#Detecta o estado final das comunidades e
#salva com atributo dos nodos "louvain_id"
gcs = float(g_giant.vcount())/float(N)
clu = g_giant.transitivity_undirected(mode="zero")
part = g_giant.community_multilevel(weights=None,
                                    return_levels=False,
                                    resolution=1)

for i in range(len(part)):
    for j in range(len(part[i])):
        g_giant.vs[part[i][j]]["louvain_id"] = i

#salva um arquivo de texto com as séries temporais
tevol_dataframe.to_csv(Preamble+"Time_evolution.txt", sep="\t")
#Salva a componente gigante
ig.write(g_giant, Preamble+"final_network_giant_component.gml",
         format="gml")
#Salva a rede completa
ig.write(g, Preamble+"final_network_all_components.gml",
         format="gml")
```