

UNIVERSIDADE FEDERAL DE VIÇOSA

JULIANE TEIXEIRA DE MORAES

**ESTUDO DE SÉRIES DE PREVALÊNCIA EPIDÊMICA
UTILIZANDO O GRAFO DE VISIBILIDADE**

VIÇOSA - MINAS GERAIS

2020

JULIANE TEIXEIRA DE MORAES

ESTUDO DE SÉRIES DE PREVALÊNCIA EPIDÊMICA UTILIZANDO O
GRAFO DE VISIBILIDADE

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Física para obtenção do título de *Magister Scientiae*.

Orientador: Silvio da Costa Ferreira Junior.

VIÇOSA - MINAS GERAIS
2020

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

M827e
2020

Moraes, Juliane Teixeira de, 1996-
Estudo de séries de prevalência epidêmica utilizando o grafo de visibilidade / Juliane Teixeira de Moraes. – Viçosa, MG, 2020.
68 f. : il. (algumas color.) ; 29 cm.

Orientador: Silvio da Costa Ferreira Junior.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.64-68.

1. Epidemias - Modelos matemáticos. 2. Teoria das redes complexas (Física). 3. Análise de séries temporais.
I. Universidade Federal de Viçosa. Departamento de Física.
Programa de Pós-Graduação em Física. II. Título.

CDD 22. ed. 530.15

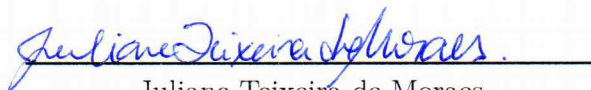
JULIANE TEIXEIRA DE MORAES

ESTUDO DE SÉRIES DE PREVALÊNCIA EPIDÊMICA UTILIZANDO O
GRAFO DE VISIBILIDADE

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Física para obtenção do título de *Magister Scientiae*.

APROVADA: 10 de novembro de 2020.

Assentimento:



Juliane Teixeira de Moraes
Autora



Silvio da Costa Ferreira Junior
Orientador

Agradecimentos

A Deus, por tudo.

Aos meus pais, por me educarem e me transmitirem valores de honestidade e humildade. Ao meu irmão, por estar sempre comigo em minhas conquistas e me deixar participar das suas também. E a todos os meus familiares, amigos e namorado, por me incentivarem e acreditarem em mim.

Ao meu orientador, Prof. Silvio Ferreira, pelo apoio, paciência e por todo conhecimento transmitido.

Aos amigos do GISC, Guilherme e Diogo, pela colaboração e dicas importantes para este trabalho.

Aos amigos do mestrado.

A todos os bons professores que passaram por minha trajetória, por me inspirarem.

À Universidade Federal de Viçosa, por toda a infraestrutura e pelo suporte computacional através do NSSC.

Ao CNPq, à CAPES e à FAPEMIG, pelo apoio financeiro.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“A imaginação é a faculdade da descoberta, predominantemente. É ela que atravessa os mundos invisíveis que nos rodeiam, os mundos da ciência.”
(Ada Lovelace)

Resumo

MORAES, Juliane Teixeira de, M.Sc., Universidade Federal de Viçosa, novembro de 2020.
Estudo de séries de prevalência epidêmica utilizando o grafo de visibilidade.
Orientador: Silvio da Costa Ferreira Junior.

Séries temporais emergem dos mais diversos tipos de sistemas reais. Podemos encontrar sequências de dados ordenados em valores de uma ação no mercado financeiro, em medidas da temperatura da superfície do mar, na atividade elétrica cerebral por meio de técnicas de imagem como a ressonância magnética, entre muitos outros exemplos. Porém a análise dessas séries temporais, em geral, não é uma tarefa fácil pois os dados disponíveis nem sempre possuem qualidade e quantidade ideal. Além disso, em geral é preciso fazer a reconstrução do espaço de fase do sistema, exigindo a determinação de parâmetros que depende da habilidade do pesquisador. Um método moderno de análise de séries temporais é mapeá-las em redes complexas e analisar essas estruturas. Assim, para este trabalho utilizamos a ferramenta chamada grafo de visibilidade para transformar séries temporais em redes complexas. Esse método pode ser usado a princípio para qualquer tipo de série, mas a aplicação que escolhemos está relacionada a processos de disseminação. Um exemplo é a propagação de epidemias, campo de estudo que ganhou ainda mais relevância em 2020 com a pandemia de COVID-19. Conhecer a estrutura de rede em que a doença se dissemina pode ser muito importante para compreender a evolução da propagação e elaborar estratégias para combater a doença. Estudamos o modelo Suscetível-Infetado-Suscetível (SIS), usado para modelar processos de disseminação. Simulamos esse modelo para diferentes estruturas: redes quadradas, redes regulares aleatórias e redes com distribuição de grau em lei de potência. Para cada uma delas estudamos a possível transição de fase para o estado absorvente ao variarmos a taxa de infecção do modelo. Utilizando métodos quase estacionários, geramos as séries de prevalência epidêmica, que são valores da fração de indivíduos infectados na população em função do tempo. Desta forma, mapeamos essas séries em grafos de visibilidade e utilizamos quantidades conhecidas em análise de redes complexas para estudá-los, como distribuição de grau, coeficiente de agregação, correlação de grau e menor caminho médio. Através dessas análises foi possível destacar fortes evidências de que o grafo de visibilidade é capaz de diferenciar comportamentos próximo e longe da transição e de captar diferenças entre os mecanismos de ativação relacionados às redes originais em que o processo de disseminação ocorre.

Palavras-chave: Prevalência epidêmica. Redes complexas. Séries Temporais.

Abstract

MORAES, Juliane Teixeira de, M.Sc., Universidade Federal de Viçosa, November, 2020. **Study of epidemic prevalence time series using the visibility graph.** Adviser: Silvio da Costa Ferreira Junior.

Time series emerge from diverse types of real systems. We can find sequences of ordered data in the values of an exchange in financial market, in measurements of sea surface temperature, in the brain electrical activity by means of imaging techniques such as magnetic resonance, and many other examples. However, time series analysis is not an easy task, because the available data are often not ideal in quality and quantity. In addition, reconstruction of the system phase space is usually necessary, requiring the determination of parameters that depend on the researcher skills. A modern method of time series analysis is to map them into complex networks and analyze these structures. Thus, for this work, we have used a tool called visibility graph to transform time series into complex networks. This method can be used, in principle, for any type of time series, but the application which we have chosen is related to spreading processes. An example is the epidemic spreading, a field that received even more attention in 2020 due to the COVID-19 pandemic. Knowing the network structure in which the disease spreads can be very important to understand the spreading evolution and to elaborate strategies for combating the disease. We studied the susceptible-infected-susceptible (SIS) model, used for modeling spreading processes. We simulate this model for different structures: square lattices, regular random networks and networks with power-law degree distributions. For each one, we studied the absorbing-state phase transition by varying the infection rate of the model. Using quasistationary methods, we generate the epidemic prevalence time series, which are the fraction of infected individuals in the population as a function of time. Therefore, we mapped these time series into visibility graphs and calculated some quantities of complex networks analysis in order to study them, such as the degree distribution, clustering coefficient, degree correlation and shortest average path length. By means of these analyzes, we could present strong evidences that the visibility graph is able to distinguish behaviors near and far from the transition, and also to discriminate between activation mechanisms related to the original networks in which the dissemination process takes place.

Keywords: Epidemic prevalence. Complex networks. Time series.

Lista de abreviaturas e siglas

RRA	Rede Regular Aleatória
SIS	Suscetível-Infetado-Suscetível
GV	Grafo de Visibilidade
RB	Ruído Branco
MB	Movimento Browniano
HMF	<i>Heterogeneous Mean Field</i>
QMF	<i>Quenched Mean Field</i>
UCM	<i>Uncorrelated Configuration Model</i>
BFS	<i>Breadth-First Search</i>
OGA	<i>Optimized Gillespie Algorithm</i>

Sumário

I	Introdução	9
II	Revisão de redes complexas	12
2.1	Conceitos básicos	12
2.2	Distribuição de grau	16
2.3	Coefficiente de agregação	19
2.4	Correlação de grau	21
III	Processos de disseminação	23
3.1	Modelo SIS e teorias de campo médio	23
3.1.1	Teoria de campo médio homogênea	26
3.1.2	Teoria de campo médio heterogênea (HMF)	27
3.1.3	Teoria de campo médio microscópica	29
3.2	Séries de prevalência epidêmica	30
3.3	Mecanismos de ativação	31
IV	Mapeando séries em redes complexas	34
4.1	O Grafo de Visibilidade (GV)	35
4.2	GV para séries de prevalência epidêmica - exemplo em rede quadrada	40
V	Metodologia	42
5.1	Gerando redes complexas	42
5.2	Algoritmo BFS	44
5.3	Algoritmo de Gillespie otimizado	45
5.4	Método quase estacionário	47
VI	Resultados e discussões	49
6.1	Escolha das redes utilizadas	49
6.2	Análise de tamanho finito para o GV na fase crítica	51
6.3	Análise do GV para diferentes fases epidêmicas	53
6.4	GV para redes com diferentes mecanismos de ativação	57
VII	Considerações finais	61
	Referências	64

I Introdução

O que um eletrocardiograma, medidas de temperatura da superfície do mar e oscilações no valor de uma ação do mercado financeiro possuem em comum? Todos esses sistemas reais podem ser descritos por séries temporais, um conjunto de dados ordenados que pode ser analisado via métodos estatísticos comuns. Dados reais expressos em séries temporais estão por toda parte. Muitos deles se tornaram um objeto interessante de estudo em sistemas dinâmicos, por apresentarem evolução temporal complexa e previsibilidade fortemente limitada devido à sensibilidade às condições iniciais que esses sistemas podem exibir [1].

Além disso, sabe-se hoje que muitos sistemas reais em vários domínios da natureza, tecnologia e sociedade podem ser descritos por estruturas de redes: conexões entre voos de aeroportos [2], estações de distribuição de energia [3], interação de pessoas em redes sociais como *Facebook* e *Twitter* [4], a atividade sincronizada de regiões no cérebro [5], metabolismo de bactérias [6], entre muitos outros exemplos. Assim, desde que seja possível definir os nós e as ligações para cada situação, que são os componentes de uma rede complexa, pode-se usar de um conjunto comum de ferramentas de análise. A verificação dessa universalidade em redes complexas foi um dos grandes motivos para que a área ganhasse tanta relevância nas duas últimas décadas - o interesse nesses sistemas é multidisciplinar.

Destaca-se também o caráter computacional [7, 8] e estatístico da área para estudo de redes com grande número de nós e ligações, que só é possível via simulação numérica e análise de propriedades globais da rede. Já foi observado que fenômenos críticos, como transições de fase [9], também ocorrem em estruturas de rede, logo, é possível utilizar o ferramental da Física Estatística.

Séries temporais e redes complexas seguiram caminhos de pesquisa paralelos em sistemas dinâmicos e ambos são de grande interesse para investigação de sistemas reais. Porém, análise de séries temporais envolve uma gama de métodos bastante trabalhosos por inúmeras razões: quantidade e qualidade nem sempre ideais dos dados disponíveis, existência de ruído nos dados, necessidade de se reconstruir o espaço de fase do sistema quando as equações dinâmicas não são conhecidas, determinação de parâmetros para análise que muitas vezes depende da experiência do pesquisador [10], entre outros. Já o aparato de análise de redes complexas tem ganhado mais robustez com grande variedade de métodos [11].

Portanto, neste trabalho foi estudado um método de mapeamento de uma série temporal em uma rede complexa, através do grafo de visibilidade [12]. E as séries em questão são de prevalência epidêmica, geradas através de um processo de disseminação em uma rede.

Diversos tipos de processos de disseminação estão presentes na natureza, tecnologia e sociedade. Suas propriedades podem ser estudadas por meio de redes complexas, já que a propagação, seja de uma informação ou doença, é intermediada por algum tipo de contato entre partes, caracterizando uma estrutura de rede. No contexto de propagação de vírus de computador, por exemplo, verificou-se que é necessário se considerar a natureza temporal das interações sociais [13]. Já para disseminação de vírus em telefones móveis, notou-se que a transição de fase, devido ao surgimento da componente gigante na rede, desempenha um papel importante no sistema [14].

Outro tema muito estudado recentemente é a propagação de informações em redes sociais, já que a movimentação de opiniões e ideias pode impactar diretamente o curso histórico e político de uma sociedade [4]. Nesse tipo de sistema, o papel das comunidades em redes complexas é evidenciado, sendo importante até para estratégias de *marketing* considerando uma rede de recomendação de produtos [15], por exemplo.

Além disso, a investigação de processos de disseminação de epidemias se mostrou ainda mais relevante no ano de 2020, com a pandemia de COVID-19 - uma doença infecciosa transmitida por meio do vírus SARS-CoV-2 que até o mês de outubro desse ano já alcançava a trágica marca de um milhão de mortes e mais de quarenta milhões de casos confirmados¹. O estudo da propagação de epidemias levando em conta todas as variáveis envolvidas pode ser extremamente trabalhoso ou até inviável, seja pela complexidade ou falta de dados disponíveis. Mas é possível traçar modelagens que se aproximam da realidade utilizando conjuntos de parâmetros preestabelecidos, como por exemplo a mobilidade humana entre cidades [16].

Muitos estudos a respeito de propagação de epidemias evidenciam a necessidade de se conhecer a estrutura de rede em que ele ocorre. Sabe-se hoje, que redes livres de escala facilitam o processo de disseminação [17], uma característica que pode ser importante não só para entender como a epidemia evolui mas também para erradicá-la, via estratégias de imunização [18, 19]. Assim, conhecer a arquitetura da rede que representa o sistema em estudo pode ser crucial para determinar como será o processo de disseminação em questão.

Com essas motivações, estabelecemos a aplicação deste trabalho. Utilizamos séries de prevalência epidêmica, que é a fração de indivíduos infectados da população em função do tempo, geradas a partir de um processo de disseminação. Para isso, simulamos a propagação de uma infecção através do modelo Suscetível-Infectado-Suscetível (SIS), um modelo já bastante consolidado na área [20, 21, 22, 23] que consiste em dividir os nós em compartimentos de acordo com seu estado em relação a infecção: eles podem estar suscetíveis (S) ou infectados (I) e a transição entre esses estados ocorre com taxas de infecção e cura preestabelecidas. Fizemos a implementação do modelo SIS em diferentes redes complexas e, utilizando métodos quase estacionários [24], construímos as séries de

¹ Dados retirados de: <https://covid19.who.int/> em 29/10/2020.

prevalência epidêmica.

A partir disso, utilizamos uma ferramenta chamada grafo de visibilidade (GV) [12], para mapear essas séries em redes, e assim, estudar as séries por meio das análises dos GV gerados a partir delas. Para fazer esse estudo utilizamos quantidades conhecidas em análise de redes complexas, tais como a distribuição de grau, o coeficiente de agregação, a correlação de grau e o menor caminho médio da rede [11, 25]. Além disso, utilizamos diferentes tipos de rede original para simular o modelo SIS e conseguimos fortes evidências de que a análise dos GV é capaz de captar as diferenças entre redes originais com diferentes mecanismos de ativação da epidemia.

Essa dissertação está dividida da seguinte forma: no capítulo II é feita uma breve revisão de conceitos básicos em redes complexas que serão necessários para o entendimento do trabalho. Um desses conceitos é o de correlação de grau em uma rede, relacionado a um dos resultados centrais que obtivemos. No capítulo III é feita uma discussão sucinta sobre processos de disseminação que ocorrem em diferentes redes complexas. Nesse capítulo apresentamos a aplicação deste trabalho: as séries de prevalência epidêmica geradas por meio da simulação computacional do modelo SIS. O grafo de visibilidade é discutido no capítulo IV, onde é mostrado como uma série temporal pode ser mapeada em uma rede complexa. Exemplos utilizando séries de ruído branco e de movimento Browniano são mostrados, bem como as próprias séries de prevalência epidêmica para o modelo SIS em uma rede quadrada. No capítulo V são apresentados métodos de simulação computacional, necessários para o estudo de processo de disseminação em redes complexas. Por fim, as análises feitas neste trabalho são discutidas no capítulo VI, onde são exibidas evidências encontradas utilizando diferentes tipos de redes complexas. No capítulo VII, as considerações finais são apresentadas.

II Revisão de redes complexas

2.1 Conceitos básicos

Uma rede pode ser definida como um conjunto de pontos (nós) unidos em pares por linhas (ligações) [11]. É comum também nomear esse tipo de estrutura como *grafo*, porém é importante destacar que são termos distintos: o conjunto {rede, nó, ligação} geralmente é associado ao sistema real, por exemplo redes de transporte, redes sociais ou biológicas, etc. Já quando se quer referir a representação matemática de uma rede, os termos {grafo, vértice, aresta} podem ser utilizados [25]. Toda rede é um grafo, porém nem todo grafo é uma rede, como por exemplo, o grafo de visibilidade que será utilizado nesse trabalho e não representa uma rede real. Um esquema ilustrativo a respeito dessas terminologias é mostrado na Fig. 2.1, em que temos dois tipos diferentes de redes (a) e (b) que podem ser representadas pelo mesmo grafo (c).

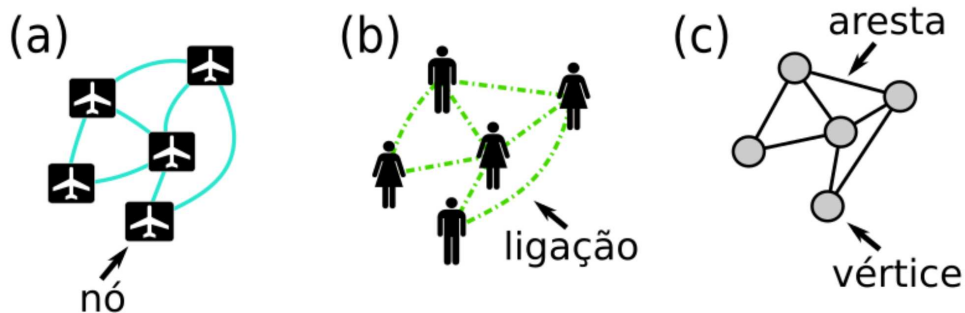


Figura 2.1 – Ilustração de redes e nomenclaturas. (a) Rede de transporte aéreo: nós são os aeroportos e ligações são as linhas de voos entre eles. (b) Rede social: nós são as pessoas e ligações são as relações de amizade entre elas. (c) Grafo referente às redes (a) e (b).

Diversos sistemas podem ser descritos por redes complexas, desde que seja possível definir os nós e as ligações, como por exemplo:

Exemplo 1: Rede de regiões cerebrais - Através de métodos de imageamento, como ressonância magnética, é possível obter informação sobre a atividade elétrica do cérebro. Os dados são coletados simultaneamente em diferentes regiões (que são os nós da rede). Se existe dinâmica sincronizada entre pares de regiões, os dois nós correspondentes estarão ligados [5]. Na Fig. 2.2 são mostradas algumas formas de se dividir o cérebro em diferentes setores. É necessário investigar como a maneira de fazer as divisões influenciará nas características da rede. Nem todos os parâmetros topológicos permanecem quantitativamente iguais, dependendo do tipo de divisão.

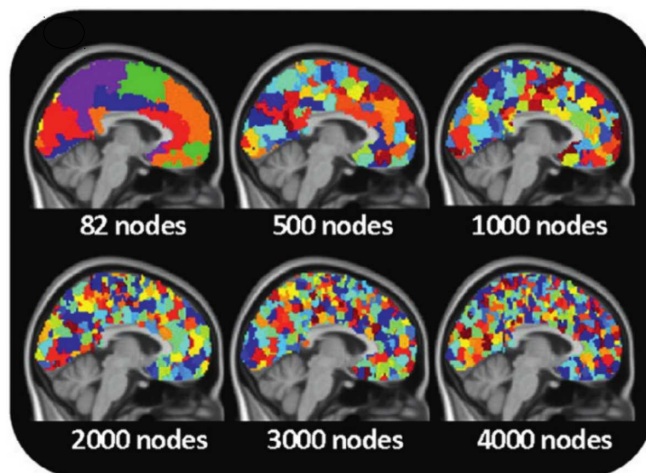


Figura 2.2 – Rede cerebral formada a partir de dados de ressonância magnética, mostrando diferentes formas de dividir o cérebro em setores. Cada área será um nó na rede gerada e as ligações são feitas quando a atividade elétrica entre elas é sincronizada. Adaptado de [5].

Exemplo 2: Rede de mobilidade humana - Pode-se também construir uma rede de unidades territoriais como mostrado na Fig. 2.3: se duas cidades possuem mobilidade de pessoas, que pode ser pendular (em que indivíduos vão e voltam em curtos períodos de tempo, para trabalhar ou estudar) ou aérea, diz-se que essas duas cidades estão conectadas. Esse tipo de rede é especialmente importante no estudo de transmissão de doenças infecciosas como a COVID-19 [16].

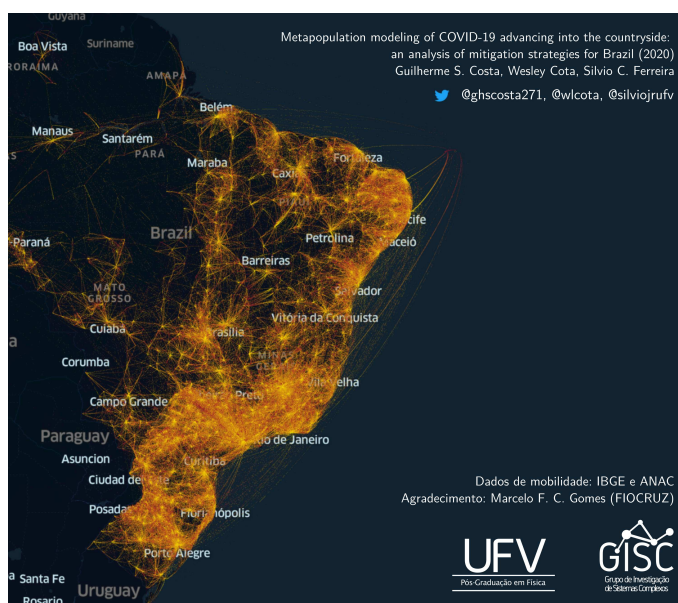


Figura 2.3 – Rede de mobilidade humana usada no estudo da COVID-19 no Brasil. A cor das ligações está relacionada à intensidade do fluxo entre cidade (nós); cores mais intensas correspondem à maior mobilidade de um lugar para outro. Adaptado de <https://covidbr.github.io/pub/1>

Esses dois exemplos mostram como são vastas as aplicações de pesquisas em redes complexas. Isso porque, independentemente do tipo de sistema em estudo, seja ele biológico, territorial, social (como *Facebook* e *Twitter* [4]) ou tecnológico, desde que sejam definidos os nós e as ligações, um conjunto de ferramentas matemáticas e estatísticas pode ser usado para sua análise. A verificação dessa universalidade em redes complexas foi um dos grandes motivos para que a área ganhasse tanta relevância nas duas últimas décadas [25].

Alguns conceitos são necessários para o estudo de redes complexas [11, 25]. Aqueles que serão usados nesta dissertação serão definidos a seguir.

- **Número de nós** (N): É o número de elementos ou tamanho da rede;
- **Número de ligações** (L): É a quantidade de conexões entre os nós.
- **Menor caminho** (d_{ij}): Em redes a distância Euclidiana entre os nós muitas vezes não é relevante, sendo substituída pela quantidade de ligações que separa dois nós em uma sequência ordenada de nós conectados. Se existe mais de um caminho possível, d_{ij} é o menor caminho disponível, como ilustrado na Fig. 2.4.

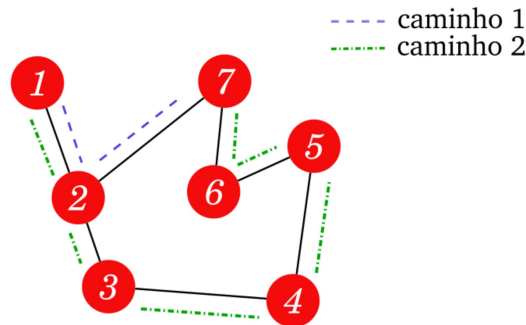


Figura 2.4 – Ilustração de possíveis caminhos em uma rede. A distância entre os nós 1 e 7 pode ser medida pelos caminhos 2 e 1, sendo esse último o menor caminho $d_{1,7} = 2$.

- **Diâmetro** (d_{\max}): O maior valor de d_{ij} possível entre quaisquer dois nós i e j determina o diâmetro da rede. Para o exemplo de rede da Fig. 2.4 $d_{\max} = d_{1,5} = 4$.
- **Menor caminho médio** ($\langle d \rangle$): É a média de d_{ij} entre todos os pares (i, j) da rede, dada pela expressão

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{\substack{i,j=1,N \\ i \neq j}} d_{ij}. \quad (2.1)$$

Para uma rede pequena como a da Fig. 2.4 ($\langle d \rangle = 2$) a aplicação da Eq.(2.1) é simples, porém torna-se inviável para tamanhos muito grandes. Então, para determinar todas as distâncias d_{ij} em redes com N qualquer, utiliza-se o algoritmo de busca em largura, ou *Breadth-First Search (BFS)*, que será discutido na seção 5.2.

- **Grau de um nó** (k_i): É o número de ligações que um nó i possui, ou seja, o número de vizinhos de i . Uma rede pode ser classificada quanto ao grau de seus nós: será *homogênea* se todos os nós possuem grau próximo do grau médio, ex. rede quadrada (Fig. 2.5(a)) e rede regular aleatória (RRA)¹ (Fig. 2.5(b)) ou *heterogênea* se existem nós de grau muito diferente do grau médio, ex. lei de potência (discutida na seção 2.2) (Fig. 2.5(c)).

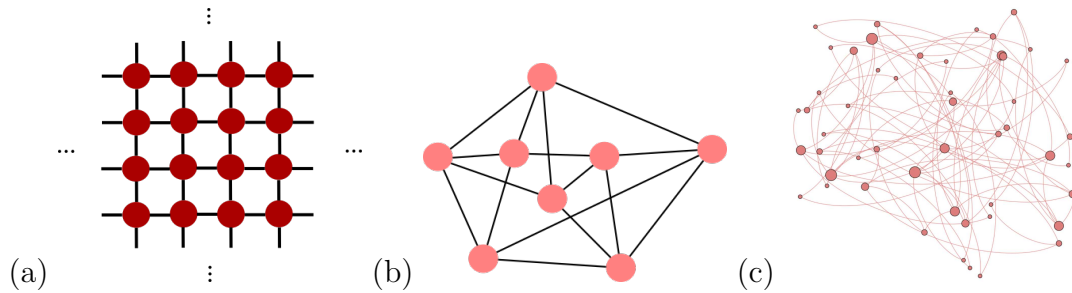


Figura 2.5 – Exemplos de redes com distribuição de grau diferentes. (a) Rede regular quadrada com $m = 4$. (b) Rede regular aleatória com $m = 4$. (c) Rede com distribuição de grau (seção 2.2) em lei de potência.

- **Grau médio** ($\langle k \rangle$): Definido por

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (2.2)$$

Na Fig. 2.5(c) temos um exemplo de rede heterogênea em que os nós estão com tamanhos diferentes. Pois estão representados de acordo com seu grau k_i . Aqueles que possuem $k_i \gg \langle k \rangle$ são chamados *hubs* da rede.

- **Matriz de adjacência** (A_{ij}): Definida por

$$A_{ij} = \begin{cases} 1, & \text{se } i \text{ conecta com } j \\ 0, & \text{caso contrário} \end{cases} \quad (2.3)$$

Portanto, cada entrada da matriz é uma possível ligação na rede. Um exemplo é mostrado na Fig. 2.6.

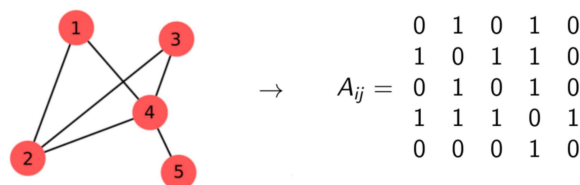


Figura 2.6 – Exemplo de rede não direcionada e a matriz de adjacência correspondente.

¹ Para RRA todos os nós possuem o mesmo grau $k = m$ e as conexões são feitas ao acaso [21].

A matriz A_{ij} carrega toda a informação a respeito das ligações na rede. Além disso, algumas características também podem ser extraídas dela: quando $A_{ij} = A_{ji}$, ou seja, a matriz é simétrica, ela representa uma rede *não direcionada*, como no exemplo acima. Nesse caso o número de ligações é dado por

$$L = \frac{1}{2} \sum_{i=1}^N k_i, \quad (2.4)$$

em que o fator $1/2$ se deve justamente ao fato de que nessa soma conta-se duas vezes a mesma ligação. Quando a matriz A_{ij} é assimétrica, ou seja $A_{ij} \neq A_{ji}$, a representação é de uma rede *direcionada*, como ilustrado na Fig. 2.7.

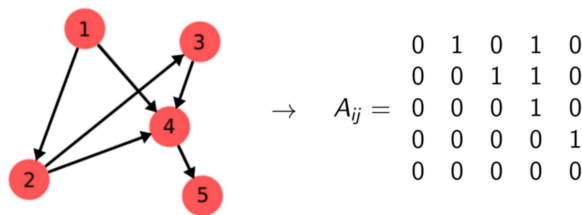


Figura 2.7 – Exemplo de rede direcionada e a matriz de adjacência correspondente. Para esse tipo de rede, a direção das ligações, ilustrada pelo sentido das setas na figura, é importante: existe a ligação (1, 2) mas não a (2, 1).

2.2 Distribuição de grau

A distribuição de grau de uma rede é a probabilidade de que um nó escolhido aleatoriamente tenha grau k [11], ou seja,

$$p_k = \frac{N_k}{N}, \quad (2.5)$$

em que N_k é o número de nós que possuem grau k . Além disso, sendo uma probabilidade, p_k é normalizada $\sum_{k=1}^{\infty} p_k = 1$. Muitas propriedades da rede são calculadas utilizando a distribuição de grau, como por exemplo, o grau médio e os outros momentos, definidos por

$$\langle k^n \rangle = \sum_{k=1}^{\infty} k^n p_k. \quad (2.6)$$

Para redes aleatórias, seguindo o modelo de Erdős-Renyi [26], os N nós da rede são conectados em pares com uma dada probabilidade p . Como exemplo, temos a Fig. 2.8(a). A distribuição de grau para redes Erdős-Renyi é do tipo binomial, porém para os casos em que $\langle k \rangle \ll N$, é possível fazer uma aproximação para distribuição de Poisson. Na figura Fig. 2.8(b) temos a comparação da distribuição de grau de redes com mesmo tamanho $N = 10^5$ e diferentes valores de p , logo diferentes $\langle k \rangle$.

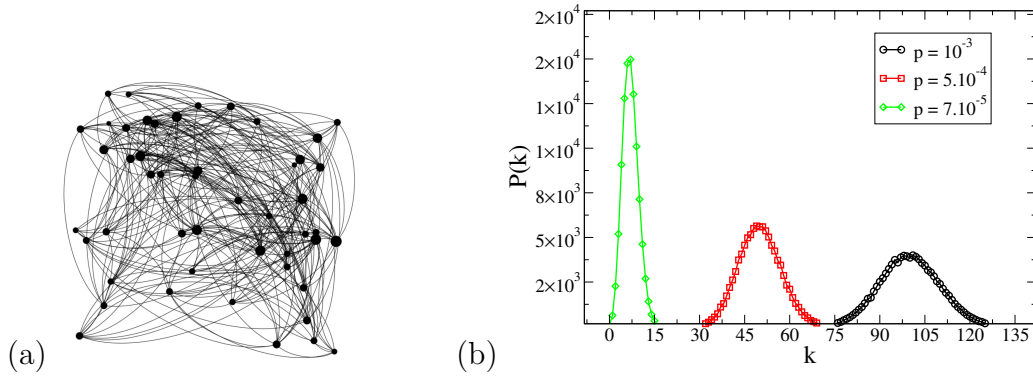


Figura 2.8 – (a) Ilustração de rede Erdős-Rényi com tamanho $N = 50$ e $p = 0.25$. (b) Distribuições de grau para redes de Erdős-Rényi com tamanho $N = 10^5$, para diferentes probabilidades de conexão.

Redes de Erdős-Rényi são um bom objeto de estudo para compreender diversos fenômenos. Dentre eles podemos citar o chamado *efeito de mundo pequeno*, o qual estabelece que dois nós em uma rede estão separados por uma distância média muito menor do que o tamanho da rede, tipicamente $\langle d \rangle \sim \ln(N)$ [25]. Além disso, nessas redes é observado que uma transição pode ocorrer conforme $\langle k \rangle$ varia. A partir de $\langle k \rangle = 1$, um conjunto grande de nós de tamanho N_G conectados aparece na rede, a chamada *componente gigante*, e é possível estudar os regimes existentes nessa transição.

Temos que para redes desse tipo, se $\langle k \rangle \ll N$, a dispersão é dada por $\sigma_k = \langle k \rangle^{1/2}$. Isso implica que para redes grandes a maioria dos nós terá grau da mesma ordem de $\langle k \rangle$, sem o aparecimento de *hubs*. O avanço da computação e a obtenção dados de redes reais possibilitaram testes da teoria prevista por Erdős-Rényi. Em 1999, A. -L. Barabási e R. Albert [27] observaram que redes de *World Wide Web* seguiam uma distribuição de grau do tipo lei de potência $p_k \sim k^{-\gamma}$, muito diferente da previsão para redes de Erdős-Rényi. Na Fig. 2.9 pode-se notar essa discordância entre a distribuição de Poisson (construídas utilizando o grau médio de cada rede) e os dados de redes reais.

A fim de se encontrar uma modelagem que descrevesse melhor as redes reais, o modelo de Barabási-Albert traz dois elementos importantes, que são o crescimento da rede com o tempo e a ligação preferencial entre nós. E assim, é possível gerar redes com distribuição de grau em lei de potência que descrevem melhor o comportamento de redes reais. Outro modelo que gera redes do tipo lei de potência (e também com outros tipos de distribuição de grau) é o *Uncorrelated Configuration Model (UCM)* [28] que será detalhado no capítulo V. Através dele é possível gerar redes como a da Fig. 2.10(a) que seguem uma distribuição de grau como mostrado na Fig. 2.10(b).

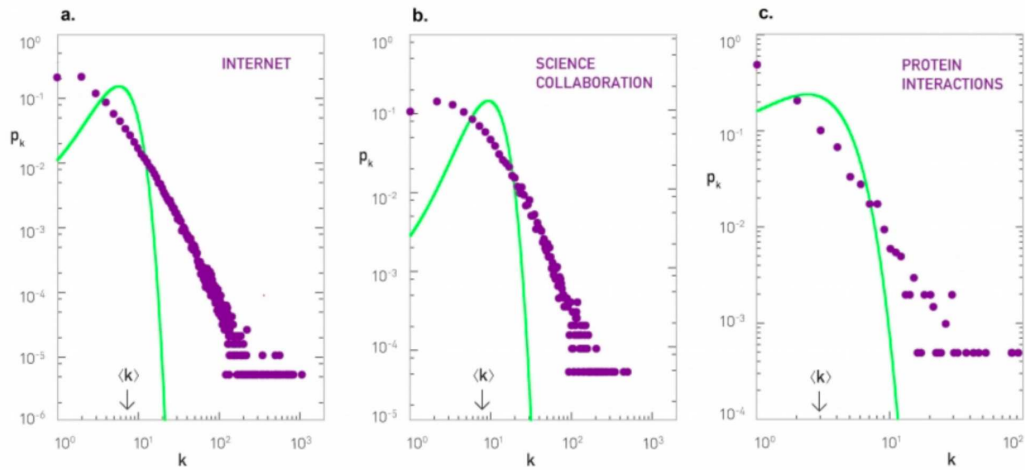


Figura 2.9 – Distribuição de grau para dados reais (símbolos roxos) e previsão pelo modelo de Erdős-Rényi (curvas verdes) para redes de (a) Internet, (b) colaboração científica e (c) interação de proteínas. Retirado de [25].

Sabe-se que, para redes de Erdős-Rényi, a maioria dos nós possui grau que está em uma determinada variação: $k \in [\langle k \rangle - \langle k \rangle^{1/2}, \langle k \rangle + \langle k \rangle^{1/2}]$, de modo que $\langle k \rangle$ é uma espécie de escala na rede. Para redes com distribuição de grau em lei de potência, porém, temos que se $2 < \gamma < 3$ o segundo momento $\langle k^2 \rangle$, e também a variância, divergem para $N \rightarrow \infty$. Isso indica que as flutuações ao redor de $\langle k \rangle$ podem ser arbitrariamente grandes [25], de maneira que não existe uma escala característica para os graus na rede. Por isso, esse tipo de rede é chamado *livre de escala*.

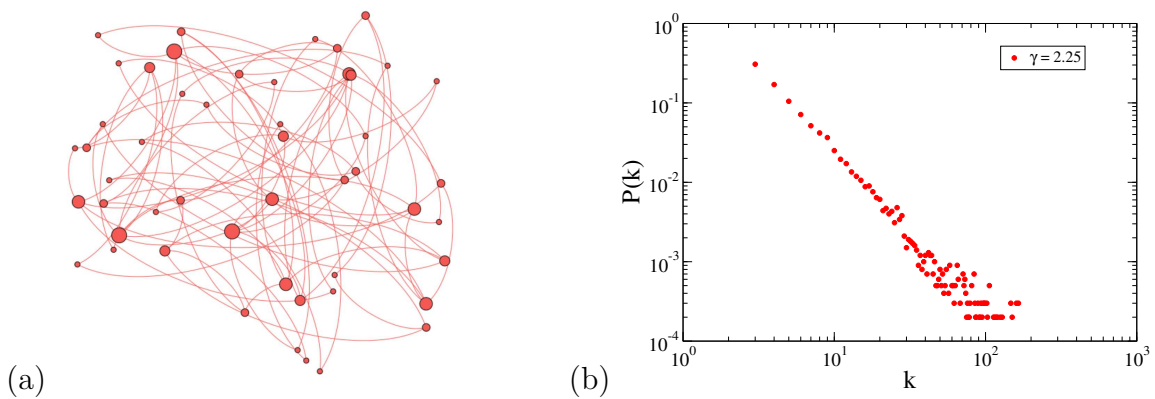


Figura 2.10 – (a) Ilustração de rede com distribuição de grau em lei de potência $\gamma = 2.25$ e $N = 50$ gerada pelo modelo UCM (seção 5.1). (b) Distribuição de grau para rede UCM $\gamma = 2.25$ e $N = 10^3$.

O estudo de redes com distribuição em lei de potência com diferentes valores de γ permite inferir uma série de propriedades sobre a rede. Uma delas é que as distâncias em uma rede mudam drasticamente: para $\gamma > 3$ temos que $\langle d \rangle \sim \ln(N)$ (um comportamento semelhante ao que acontece para redes de Erdős-Rényi); já para $2 < \gamma < 3$, $\langle d \rangle \sim \ln(\ln(N))$, pois a presença dos *hubs* gera vários atalhos que diminuem muito o valor

de $\langle d \rangle$, dando origem ao *efeito de mundo ultra pequeno*. Esse é um exemplo de como propriedades da rede dependem de γ .

Em 2001, R. Pastor-Satorras e A. Vespignani [17] mostraram como redes livre de escala possuem características que favorecem processos de disseminação (capítulo III), seja de informações ou doenças infecciosas. Desta forma, é notável a importância de se investigar a distribuição de grau de uma rede.

2.3 Coeficiente de agregação

Outra quantidade importante a ser analisada em uma rede é o coeficiente de agregação, C_i , definido pela probabilidade de que dois vizinhos de um nó escolhidos ao acaso estejam ligados [29]:

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad 0 < C_i < 1 \quad (2.7)$$

onde L_i é a quantidade de ligações que de fato existe entre os vizinhos de i e $k_i(k_i - 1)/2$ é a quantidade de ligações possíveis entre os vizinhos de i , obtida por meio uma combinação simples. Portanto, C_i mede o número de triângulos fechados, visto que L_i , da Eq.(2.7), é nada mais que o número de triângulos em que o nó i está inserido [30]. Dois exemplos são mostrados na Fig. 2.11 para duas configurações de vizinhos de i .

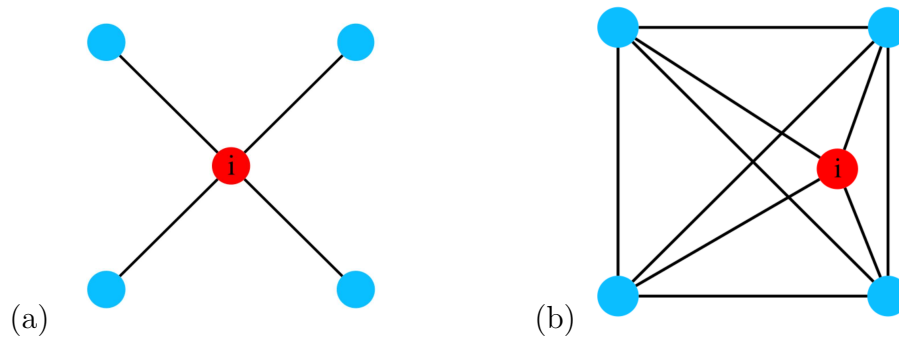


Figura 2.11 – Exemplo de coeficiente de agregação de um nó i para duas configurações de vizinhos diferentes (a) $C_i = 0$. (b) $C_i = 1$

Na Fig. 2.11(a) os vizinhos do nó i não estão ligados entre si, logo segundo a Eq.(2.7), $C_i = 0$. Exemplificando com uma rede de amigos poderia se dizer que os amigos de um determinado indivíduo (nó i da rede) não se conhecem. Já para a situação oposta, em que todos os vizinhos estão conectados como na Fig. 2.11(b), o coeficiente de agregação é máximo. O tipo de rede formada a partir dessa configuração, em que todos os nós então conectados entre si, é chamado *grafo completo*. Pode-se ter estruturas como a da Fig. 2.11(b) dentro de uma rede qualquer, para esse caso, em que temos um nó com todos seus vizinhos conectados entre si, esse conjunto é chamado *clique*.

Pode-se dizer que o coeficiente de agregação mede a densidade local de ligações [25], quanto mais densa a vizinhança de um nó, maior C_i . Porém, em muitos casos pode ser mais informativo observar como o coeficiente de agregação varia com o grau, utilizando a Eq.(2.8) :

$$C(k) = \frac{1}{N_k} \sum_{i=1}^N C_i \delta_{k,k_i} \quad (2.8)$$

com N_k sendo o número de nós com grau k .

Ao longo de diversos estudos em redes reais, foi observado que a topologia livre de escala coexistia com comportamento de grande coeficiente de agregação da rede, e que muitas delas possuíam *modularidade hierárquica* [31], que é codificada na seguinte relação

$$C(k) \sim k^{-1}. \quad (2.9)$$

Essa propriedade está relacionada à estrutura de comunidades na rede. Uma comunidade pode ser formada se um conjunto de nós tiver alta densidade de conexões entre si, ligando-se mais ao próprio conjunto do que com o restante da rede. Alta modularidade pode indicar que a rede está dividida em comunidades bem definidas [32]. O termo hierárquica pode ser entendido observando a Eq.(2.9) - nós com grau pequeno tem alto $C(k)$ pois residem em comunidades densas. Já para altos valores de k o coeficiente de agregação tende a ser menor já que nós com esses graus se conectam com comunidades diferentes. Como exemplo, temos a Fig. 2.12 que mostra o comportamento delineado pela Eq.(2.9) para uma rede de colaboração científica.

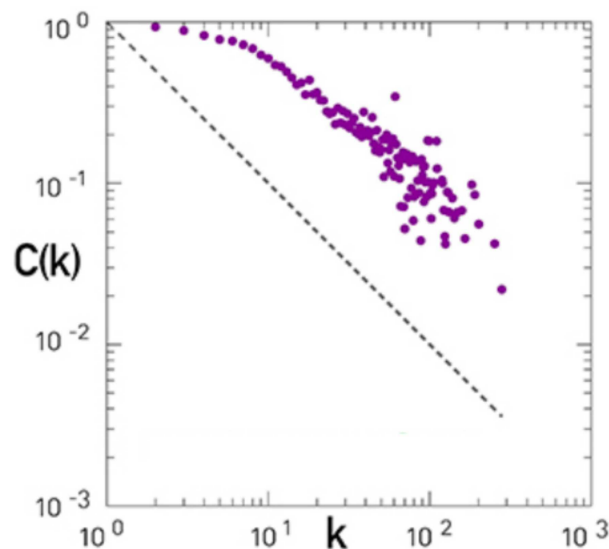


Figura 2.12 – Coeficiente de agregação em função do grau para uma rede de colaboração científica. Curva roxa são dados reais e a linha tracejada é $C(k) \sim k^{-1}$. Adaptado de [25].

Um fato a se destacar é que em modelos como Erdős-Renyi ou Barabási-Albert $C(k)$ é independente de k , logo as redes geradas a partir deles não possuem modularidade

hierárquica. Entretanto essa propriedade está constantemente presente em redes reais [31] e também nos grafos de visibilidade (GV) que foram obtidos nesse trabalho e serão apresentados no capítulo IV.

2.4 Correlação de grau

Na seção anterior, foi mostrada uma maneira de identificar como os nós estão conectados formando padrões na estrutura da rede. Por meio do coeficiente de agregação é possível medir a densidade de conexão dos nós. Podemos também analisar se existe uma relação entre os graus dos nós que ligam uns com os outros através do conceito de *correlação de grau*. No contexto de rede social existe um exemplo simples, porém interessante, a ser analisado: por que celebridades se casam? Se a conexão entre pessoas fosse governada pelo acaso, essa situação muito raramente ocorreria, visto que se temos uma amostra de aproximadamente 10^3 celebridades em um total de 10^8 pessoas elegíveis para relacionamento, a chance de ligação entre famosos seria da ordem de 10^{-5} ! Assim, pode-se notar que os relacionamentos em uma rede social possuem correlação de grau - celebridades (que seriam os *hubs* da rede) tendem a se relacionar com outras celebridades.

O contrário também já foi observado em redes biológicas, como para uma rede de interação de proteínas presentes em leveduras, os *hubs* tendem a se conectar com nós de grau pequeno e a não conectarem entre si [25]. Uma consequência disso é que se a expressão de uma proteína importante falha, outras delas não serão comprometidas.

Uma forma de se mensurar a correlação de grau é por meio da média dos graus dos vizinhos de um nó, que pode ser escrita como

$$K_{nn}(i) = \frac{1}{k_i} \sum_{j=1}^N A_{ij} k_j, \quad (2.10)$$

em que o subscrito *nn* significa *nearest neighbors*, ou seja $K_{nn}(i)$ é o grau médio dos primeiros vizinhos de i . Para analisar como essa quantidade varia com o grau, utiliza-se a o valor médio para os nós de mesmo grau k , definido como

$$K_{nn}(k) = \frac{1}{N_k} \sum_{i=1}^N K_{nn}(i) \delta_{k,k_i}. \quad (2.11)$$

Em 2001, Pastor-Satorras *et al* [33] identificaram propriedades de correlação não triviais para a Internet e propuseram que, para essa e outras redes reais, K_{nn} poderia assumir a seguinte lei de escala

$$K_{nn}(k) \sim k^\mu, \quad (2.12)$$

onde μ é o *expoente de correlação* e pode assumir três comportamentos:

- se $\mu > 0$ a rede é associativa, ou seja, os nós tendem a se conectar com outros de grau semelhante. A Fig. 2.13(a) mostra um exemplo de associatividade em uma rede de colaboração científica.
- se $\mu < 0$ a rede é desassociativa, isto é, os nós tendem a ter mais conexões com outros nós de grau diferente, como mostrado no exemplo da Fig. 2.13(b) para uma rede de metabolismo da bactéria *E. Coli*.
- se $\mu = 0$ a rede é neutra, não exibe correlação de grau. Pois como $K_{nn}(k) = ak^\mu$ com a sendo uma constante, se $\mu = 0$ $K_{nn}(k) = a$. Isso pode ser observado na Fig. 2.13(c), onde é mostrado o comportamento da correlação de grau em função do grau para uma rede de distribuição de energia.

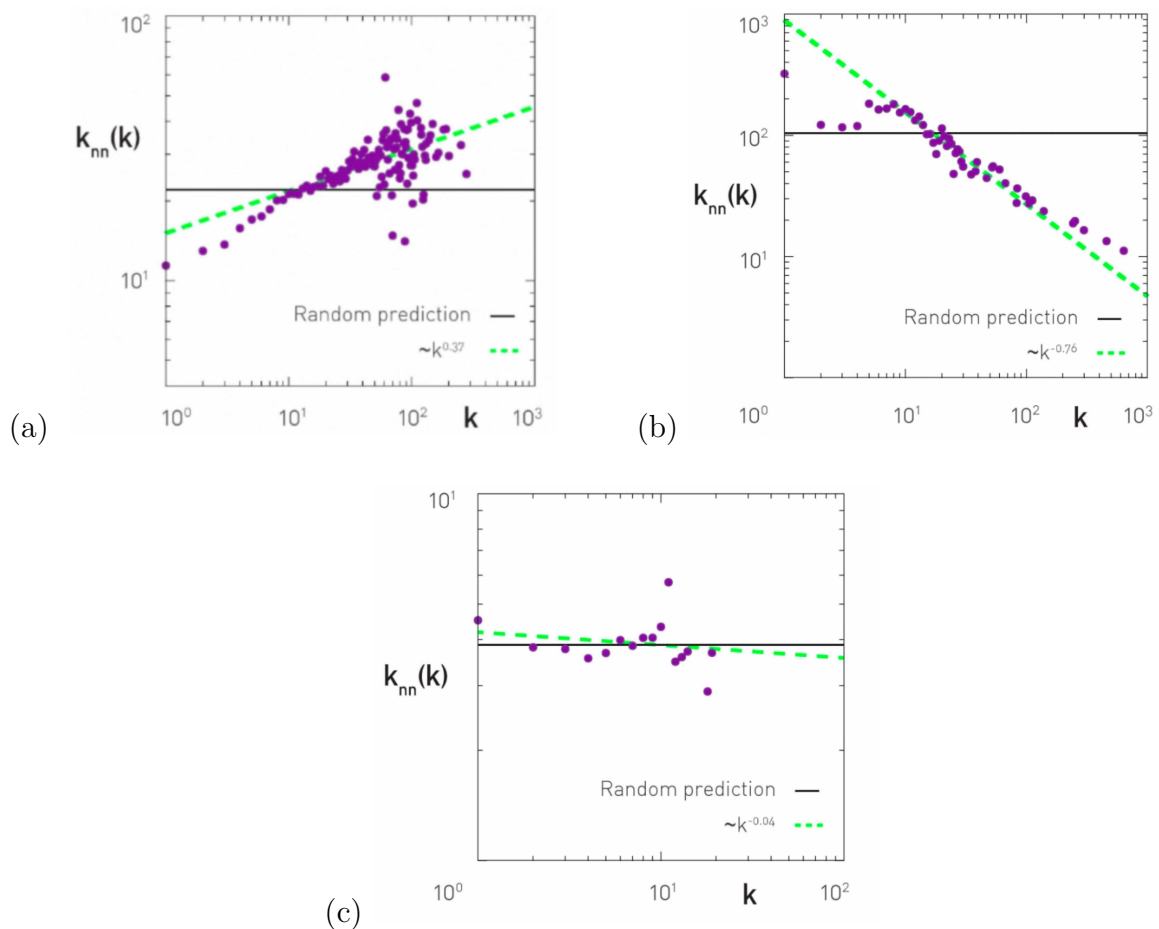


Figura 2.13 – Correlação de grau $K_{nn}(k)$ em função do grau para diferentes redes. Os símbolos roxos são dados reais, a linha verde tracejada é a relação $K_{nn}(k) \sim k^\mu$ e a linha preta é uma previsão teórica do K_{nn} para uma rede cujas conexões são completamente aleatórias, permitindo também auto e múltiplas conexões. (a) Rede de colaboração científica entre pesquisadores com $\mu \approx 0.37$. (b) Rede de metabolismo da bactéria *E. Coli*, em que os nós são os metabólitos e as ligações são reações químicas. Para essa rede temos $\mu \approx -0.76$. (c) Rede de distribuição de energia elétrica - os nós são transformadores e as ligações são as conexões cabeadas entre eles, nesse caso $\mu \approx 0.04$. Adaptado de [25].

III Processos de disseminação

Como já discutido no capítulo I, existem vários processos de disseminação presentes na sociedade e tecnologia. Apesar de possuírem natureza distinta, esses fenômenos podem ser estudados pela Física Estatística. Uma das formas para isso é através dos modelos compartimentais, em que os nós da rede são divididos em compartimentos, de acordo com seu estado em relação ao processo de disseminação. Exemplos de compartimentos podem ser suscetível (S) - o nó está saudável mas com possibilidade de contrair a infecção; infectado (I) - o nó encontra-se infectado e podendo transmitir a doença; removido ou recuperado (R) - o nó não pode ser mais infectado nem transmite a doença ou não faz mais parte da rede. Esses três compartimentos constituem os modelos mais simples para propagação de epidemias: suscetível-infectado (SI), suscetível-infectado-suscetível (SIS) e suscetível-infectado-recuperado (SIR). Para tais são estabelecidas taxas de infecção e cura a fim de se escrever equações diferenciais que descrevem a dinâmica da disseminação. Além de métodos analíticos também é possível simular computacionalmente esses processos, como foi feito para este trabalho utilizando o modelo SIS (seções 3.1 e 5.3).

3.1 Modelo SIS e teorias de campo médio

Um dos modelos compartimentais mais estudados é o modelo SIS, em que os indivíduos são divididos nos compartimentos **suscetível** e **infectado**. Um nó no estado I pode transmitir a doença para cada um dos seus contatos suscetíveis com uma *taxa de infecção* λ ou se curar espontaneamente com uma *taxa de cura* μ . O processo é esquematizado na Fig. 3.1.

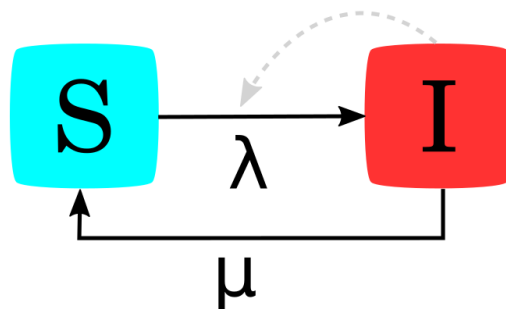


Figura 3.1 – Diagrama ilustrando o modelo SIS. As caixas representam os compartimentos (suscetível e infectado) e as setas sólidas são as possíveis transições entre esses dois estados. A seta pontilhada significa que a transição $S \rightarrow I$ depende do contato com um indivíduo infectado.

Uma quantidade relevante para esse modelo é a chamada prevalência epidêmica, ou densidade de infectados, definida por $\rho = N_i/N$, com N_i sendo o número de nós infectados.

Observe na Fig. 3.2 que o modelo SIS pode apresentar o chamado *estado absorvente* com todos os nós suscetíveis. Nessa configuração, a dinâmica permanece estagnada e não há evolução do sistema, pois como não há aparecimento espontâneo de nós infectados a disseminação termina.

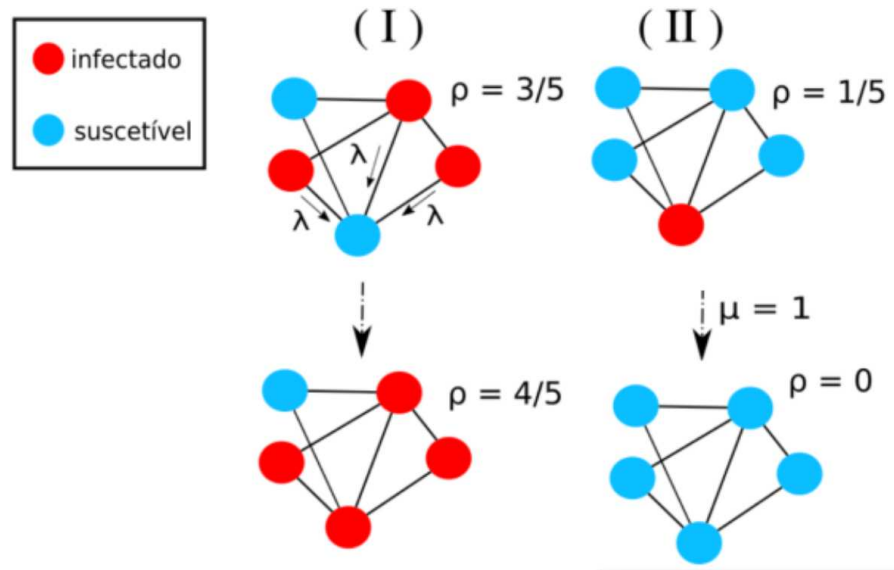


Figura 3.2 – Exemplos de infecção (I) e cura (II) no modelo SIS.

Esse modelo consegue explicar os casos em que há possibilidade de reinfeção pela doença, ou seja, quando a imunidade não é conferida ou é insuficiente [34]. Para essas situações pode ocorrer a chamada *fase endêmica*, na qual a doença permanece em circulação por longos períodos de tempo [35]. A fase endêmica também ocorre mesmo com imunidade, por meio da renovação da população.

Um dos motivos para que o modelo SIS seja um tema de grande interesse na área de Física Estatística é o fato de possuir características análogas as de sistemas de não equilíbrio que exibem transição de fase. Uma transição de fase é uma mudança abrupta no estado de um sistema para outro e esses são caracterizados por propriedades diferentes. Isso é realizado variando um determinado parâmetro de controle [36].

Para um sistema de tamanho infinito, a transição é caracterizada por um parâmetro de ordem que tem valor zero em uma das fases e diferente de zero em outra. Temos que, para o modelo SIS, a densidade de infectados (ρ) é um parâmetro de ordem e λ um parâmetro de controle. Para $N \rightarrow \infty$, ao variarmos λ até um valor crítico λ_c (limiar epidêmico) uma transição de fase ocorrerá, como ilustrado na Fig. 3.3, para $\langle \rho \rangle$.

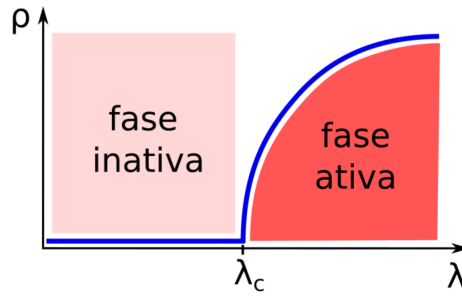


Figura 3.3 – Transição de fase para o estado absorvente caracterizada pela densidade de infectados (ρ) estacionária ao variarmos λ .

Para valores de $\lambda < \lambda_c$ o sistema decai para a configuração absorvente com um tempo característico, sendo necessária a utilização de estados quase estacionários (seção 5.4), que são perturbações na dinâmica do sistema que evitam o estado absorvente. Assim, é possível construir séries temporais de ρ , que serão tratadas na seção 3.2, e fazer sua média no estado estacionário ($t \rightarrow \infty$).

Para exemplificar a transição da fase sem doença para uma fase endêmica, consideremos o modelo SIS numa rede quadrada de tamanho $N = L^2$. Conforme aumentamos L o comportamento da curva de $\langle \rho \rangle \times \lambda$ (Fig. 3.4) se aproxima mais da Fig. 3.3.

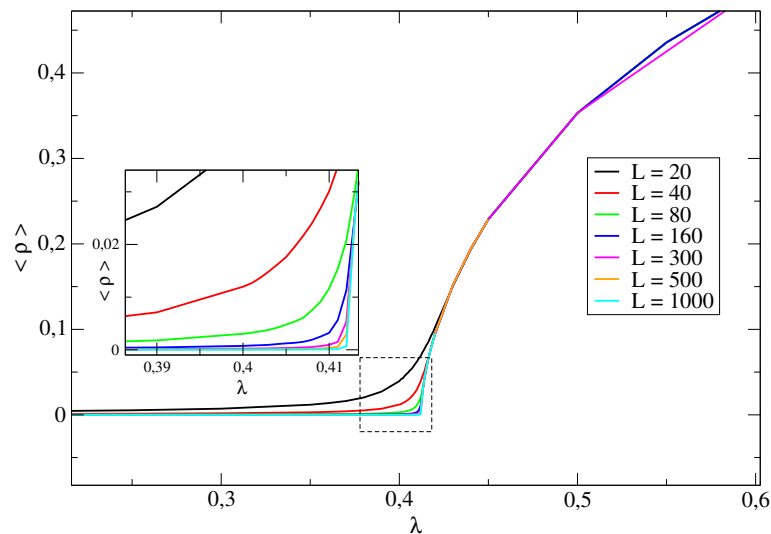


Figura 3.4 – Curvas de densidade quase estacionária de infectados em função da taxa de infecção, para diferentes tamanhos da rede quadrada. O destaque mostra uma ampliação na região destacada no gráfico principal.

Considerando os estados estacionários próximos ao ponto crítico λ_c , a prevalência segue um comportamento dado pela seguinte lei de escala [37]

$$\langle \rho \rangle \sim (\lambda - \lambda_c)^\beta \quad (3.1)$$

quando $N \rightarrow \infty$, na qual β é um expoente crítico associado ao parâmetro de ordem $\langle \rho \rangle$.

Além disso, ainda analisando o estado estacionário, pode-se medir a variância de ρ através do cálculo da suscetibilidade. Essa medida fornece uma estimativa das flutuações de ρ . Para redes complexas utiliza-se a suscetibilidade dinâmica (χ) [21] dada por

$$\chi \equiv N \frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle}. \quad (3.2)$$

Devido ao aumento nas flutuações de ρ à medida que se aproxima do ponto crítico, a suscetibilidade dinâmica apresenta um pico ao redor de λ_c , para uma rede quadrada esse valor é aproximadamente $\lambda_c = 0.41225$. Observe na Fig. 3.5 que, conforme se aumenta o tamanho da rede, o pico da curva é cada vez mais evidenciado no valor do ponto crítico de forma que no limite termodinâmico ocorrerá uma divergência em $\lambda = \lambda_c$. Assim, a suscetibilidade dinâmica é uma maneira de se determinar uma estimativa do ponto crítico através de simulações computacionais, como será discutido no capítulo V. Mas também é possível estimar λ_c via métodos analíticos, utilizando as teorias de campo médio. Nessas abordagens, são feitas aproximações que desprezam a dependência entre estados de nós distintos, ou seja, a correlação dinâmica entre eles. De acordo com cada tipo de aproximação temos uma teoria de campo médio diferente, como será tratado a seguir.

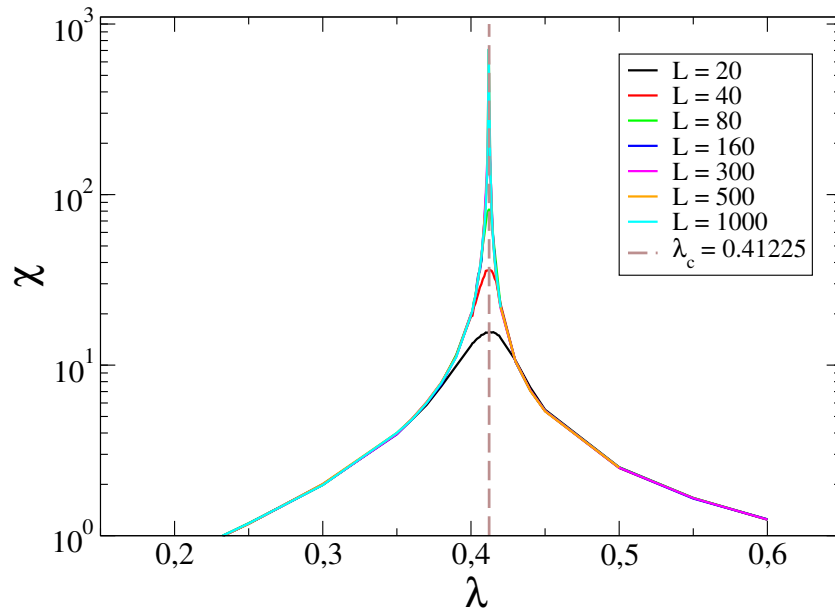


Figura 3.5 – Curvas de suscetibilidade dinâmica em função da taxa de infecção, para diferentes tamanhos da rede quadrada.

3.1.1 Teoria de campo médio homogênea

Considera-se que todos os nós da rede possuem o mesmo grau $\langle k \rangle$ e que a densidade de infectados independe de características da rede. Temos então que ρ é a probabilidade de um nó estar infectado e ω a probabilidade de estar suscetível, tal que $\rho + \omega = 1$.

Desta forma é possível construir uma equação para a evolução de ρ , dada por

$$\frac{d\rho}{dt} = -\mu\rho + \lambda \langle k \rangle \omega\rho. \quad (3.3)$$

Vamos usar $\mu = 1$ para cura espontânea, definindo assim a unidade de tempo. Note que o primeiro termo do lado direito da Eq.(3.3) se refere aos indivíduos que saem do compartimento de infectados (por isso o sinal negativo), ou seja, que são curados. Enquanto o segundo termo diz respeito a novas infecções.

Para $t \rightarrow \infty$ ($d\rho/dt = 0$) temos dois estados estacionários possíveis: $\rho = 0$ ou $\rho = 1 - \frac{1}{\lambda\langle k \rangle}$. Fazendo uma análise de estabilidade linear temos que $\rho = 0$ é estável se $\lambda < 1/\langle k \rangle$ e $\rho = 1 - 1/\lambda\langle k \rangle$ é estável se $\lambda > 1/\langle k \rangle$. Logo, $\lambda = \lambda_c = 1/\langle k \rangle$ é o limiar epidêmico para a aproximação de campo médio homogêneo. Assim, para $\lambda < \lambda_c$ a epidemia está numa fase absorvente, ou livre da doença e para a solução $\lambda > \lambda_c$ temos a chamada fase endêmica em que o número de novos infectados é aproximadamente igual ao de recuperados da doença, logo a prevalência média não muda com o tempo.

3.1.2 Teoria de campo médio heterogênea (HMF)

Como proceder quando a rede em questão tem distribuição de grau do tipo lei de potência? Para esse caso podemos considerar o sistema como sendo uma *mistura heterogênea* e por isso o nome *heterogeneous mean field* (HMF). Para tanto iremos modificar a Eq.(3.3) para que o grau dos nós infectados seja levado em conta. Porém ainda sim, é necessário fazer aproximações: assume-se que todos os nós de mesmo grau são equivalentes e as correlações dinâmicas entre nós são desprezadas, de modo que a fração de nós ρ_k com grau k infectados possua dependência somente em k . Assim, a equação dinâmica correspondente é [33, 36]

$$\frac{d\rho_k}{dt} = -\rho_k + \lambda k(1 - \rho_k) \sum_{k'} P(k'|k) \rho_{k'}. \quad (3.4)$$

A prevalência epidêmica é dada por

$$\rho = \sum_k \rho_k P(k). \quad (3.5)$$

Note que o segundo termo da Eq.(3.4) foi modificado com relação a Eq.(3.3). Agora considera-se o contato entre um nó de grau k com um nó de grau k' e $P(k'|k)$ está relacionado a correlação de grau da rede (seção 2.4). A solução da Eq.(3.4) irá depender da forma de $P(k'|k)$, não sendo viável encontrar uma solução analítica geral. Porém é possível fazer uma análise de estabilidade da Eq.(3.4) já que o estado livre da doença ($\rho_k = 0$) é mais uma vez uma possível solução. Assim para valores pequenos de ρ_k ,

podemos linearizar a Eq.(3.4), obtendo

$$\frac{d\rho_k}{dt} = -\rho_k + \lambda k \sum_{k'} P(k'|k) \rho_{k'}. \quad (3.6)$$

Podemos definir a matriz Jacobiana como sendo $J_{kk'} = -\delta_{kk'} + \lambda k P(k'|k)$. A solução ρ_k será instável se existir pelo menos um autovalor da matriz $J_{kk'}$ que seja positivo [38]. Mas $J_{kk'}$ pode ser reescrita utilizando a matriz de correlação $C_{kk'} = kP(k|k')$ como $J_{kk'} = -\delta_{kk'} + \lambda C_{kk'}$. Assim, como $\delta_{kk'}$ representa a matriz identidade, encontrar os autovalores de $J_{kk'}$ é nada mais que encontrá-los para $C_{kk'}$. Para redes não correlacionadas, temos que

$$C_{kk'} = \frac{kk'P(k')}{\langle k \rangle}. \quad (3.7)$$

Utilizando a equação de autovalores $\sum_{k'} C_{kk'} v_{k'} = \Lambda v_{k'}$ e verificando que um dos autovetores de $C_{kk'}$ é $v_{k'} = k$ com autovalor Λ , é possível escrever, pela Eq.(3.7), a seguinte relação

$$\sum_{k'} C_{kk'} k = \sum_{k'} \frac{k'^2 P(k')}{\langle k \rangle} k = \frac{\langle k^2 \rangle}{\langle k \rangle} k = \Lambda k, \quad (3.8)$$

de modo que $\Lambda = \langle k^2 \rangle / \langle k \rangle$. Assim, pelo teorema de Perron-Frobenius [11], que estabelece: "o maior autovalor de uma matriz que tem todos os elementos não negativos (como é o caso de $C_{kk'}$) é único e o autovetor correspondente tem todos os elementos não negativos e é único", temos que $\Lambda = \langle k^2 \rangle / \langle k \rangle$ é o maior autovalor de $C_{kk'}$.

Assim o maior autovalor de $J_{kk'}$ será $-1 + \lambda \Lambda$ e a solução $\rho_k = 0$ será instável se $-1 + \lambda \Lambda > 0$ e o limiar epidêmico para a teoria HMF é

$$\lambda_c = \frac{1}{\Lambda} = \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (3.9)$$

Note que para redes livres de escala [25]

$$\langle k^2 \rangle = \int_{k_{\min}}^{k_{\max}} k^n P(k) dk \approx \frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{3-\gamma}, \quad (3.10)$$

logo, $\langle k^2 \rangle \sim k_{\max}^{3-\gamma}$. Além disso, temos que o número de nós com grau maior que k_c é da ordem de 1

$$N \int_{k_c}^{\infty} P(k) dk \sim 1. \quad (3.11)$$

Assim, para $P(k) \sim k^{-\gamma}$

$$k_c \sim N^{\frac{1}{\gamma-1}}. \quad (3.12)$$

Logo, como $k_{\max} \sim k_c$, se $N \rightarrow \infty$, temos $k_{\max} \rightarrow \infty$ e o mesmo ocorre com $\langle k^2 \rangle$, conseqüentemente, $\lambda_c \rightarrow 0$. Mas para distribuições do tipo $P(k) \sim k^{-\gamma}$ com $\gamma > 3$, $\langle k^2 \rangle$ é finito, contradizendo resultados matemáticos [39] e de simulação numérica [21] que mostram que $\lambda_c \rightarrow 0$ no limite termodinâmico. Portanto, é necessário utilizar outra

aproximação, que considere a estrutura real da rede. Essa análise será explicada em linhas gerais a seguir.

3.1.3 Teoria de campo médio microscópica

Ainda considerando que o estado de cada nó é estatisticamente independente do estado dos vizinhos próximos e para redes estáticas (*quenched networks*), ou seja, nas quais o conjunto de nós e de ligações não varia, podemos utilizar a chamada aproximação QMF, do inglês *quenched mean field* [36, 40]. Essa aproximação também pode ser tratada como de campo médio microscópica porque cada nó será analisado individualmente, considerando sua estrutura real de ligações na rede por meio da matriz de adjacência.

Assim, analisando a probabilidade de infecção para cada sítio i , a equação dinâmica que descreve o modelo SIS fica

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda(1 - \rho_i) \sum_j A_{ij}\rho_j. \quad (3.13)$$

Mais uma vez o primeiro termo do lado direito da equação refere-se a transição de cura (com $\mu = 1$) e o segundo termo é relativo ao aumento de nós infectados, considerando agora a estrutura real de conexões. Por meio da matriz de adjacência, as ligações de i são levadas em conta, não apenas pelo grau como nas aproximações anteriores, mas considerando seus vizinhos reais na rede e as respectivas probabilidades de estarem infectados ρ_j .

O procedimento para obter o limiar epidêmico através dessa equação é semelhante ao utilizado na HMF. Teremos novamente uma matriz Jacobiana, dessa vez dada por $J_{ij} = -\delta_{ij} + \lambda A_{ij}$. Como as redes estudadas nesse trabalho são conectadas e não direcionadas, o teorema de Perron-Frobenius também se aplica à matriz A . Logo, por meio de análise de estabilidade linear é possível mostrar que $\lambda_c = 1/\Lambda_{\max}$, em que Λ_{\max} é o maior autovalor da matriz de adjacência. Chung *et al* calcularam Λ_{\max} para redes aleatórias com distribuição em lei de potência [41], que leva à seguinte expressão para o limiar epidêmico [40]

$$\lambda_c \simeq \begin{cases} 1/\sqrt{k_{\max}}, & \gamma > 5/2 \\ \langle k \rangle / \langle k^2 \rangle, & 2 < \gamma < 5/2 \end{cases}. \quad (3.14)$$

Assim, temos que para $2 < \gamma < 5/2$ o limiar epidêmico coincide com o previsto pela teoria HMF. Para $\gamma > 5/2$, temos também $\lambda_c \rightarrow 0$ no limite termodinâmico já que k_{\max} diverge com N para qualquer valor de γ [21].

3.2 Séries de prevalência epidêmica

Na seção 3.1 vimos que a prevalência epidêmica ρ é o parâmetro de ordem para determinar a transição de fase que ocorre no modelo SIS. Utilizando o algoritmo de Gillespie otimizado (seção 5.3) e métodos quase estacionários (seção 5.4) pode-se construir séries temporais de prevalência epidêmica $\rho(t)$.

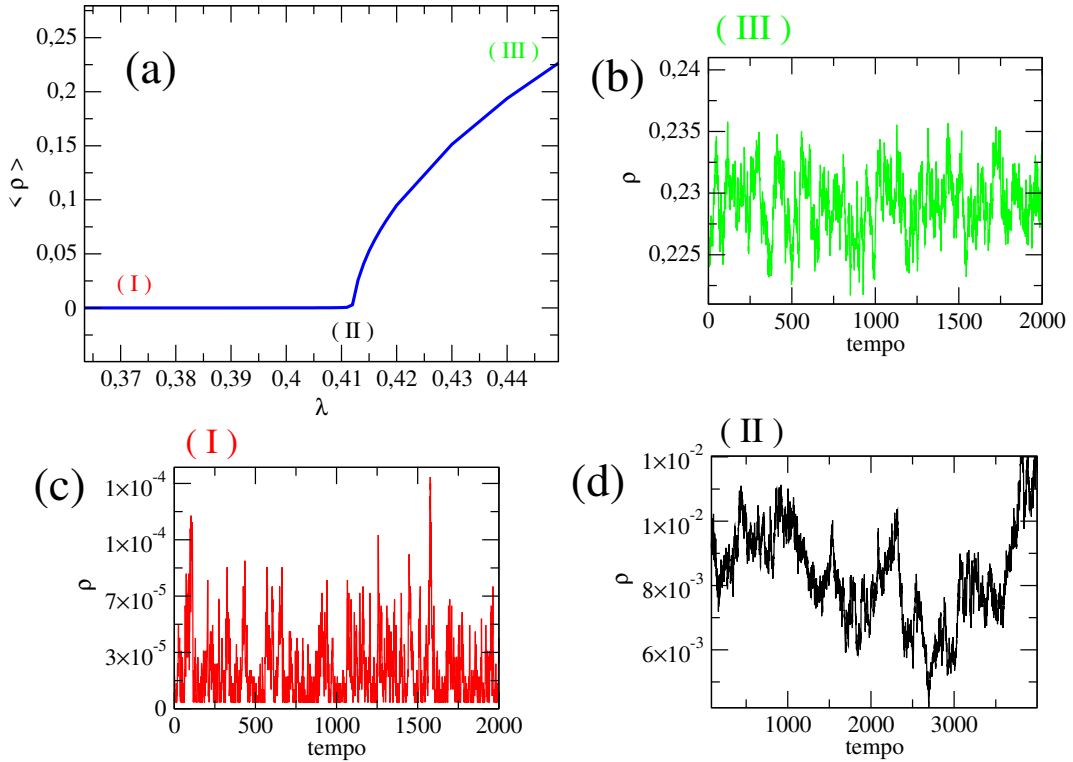


Figura 3.6 – (a) Densidade de infectados quase estacionária em função da taxa de infecção para uma rede quadrada de tamanho $N = L^2$ com $L = 500$. Séries de prevalência epidêmica para as fases: (b) supercrítica ($\lambda = 0.45$), (c) subcrítica ($\lambda = 0.37$) e (d) crítica ($\lambda = 0.412$), com valores de taxa de infecção escolhidos de acordo com a região enumerada em (a).

Na Fig. 3.6 temos exemplos de séries construídas simulando o modelo SIS em uma rede quadrada. Na Fig. 3.6(a) temos uma mudança de fase semelhante a que ocorre na Fig. 3.3. Tomando três regiões distintas, que correspondem a diferentes valores de λ , podemos construir séries temporais quase estacionárias mostradas na Fig. 3.3(b),(c) e (d), sendo $\langle \rho \rangle$ a média desses valores durante um tempo t_{med} . O método quase estacionário, em linhas gerais, consiste em evitar que ocorra o estado absorvente (aquele em que a dinâmica do sistema permanece congelada). Nessa simulação, foi utilizado o método da condição de contorno refletora, no qual toda vez que o sistema visita o estado absorvente, o nó que estava suscetível anteriormente é reativado e a simulação continua. E também o método de reativação de *hub* [42], para o qual um dos *hubs* é reinfectado toda vez que ocorre o estado absorvente. Essa discussão será feita na seção 5.4.

A partir de agora, serão usados termos emprestados do jargão de fenômenos críticos para se referir aos diferentes regimes epidêmicos: as fases subcrítica, crítica e supercrítica como sendo abaixo, na e acima da transição para o estado absorvente. É preciso fazer essa observação pois não há garantia de criticalidade dinâmica (correlação temporal de longo alcance) para o modelo SIS em qualquer tipo de rede. Para cada um dos regimes epidêmicos temos uma situação:

- (I) Regime subcrítico ($\lambda < \lambda_c$): os valores ρ são da ordem de $1/N$, ou seja, a disseminação está presente apenas uma fração muito pequena de nós na rede. Portanto, se não utilizamos métodos para evitar o estado absorvente para valores próximos do ponto crítico, a epidemia não se sustenta.
- (II) Regime crítico ($\lambda = \lambda_c$): nesse ponto há grande flutuação no parâmetro de ordem ρ , e ocorre uma transição da fase inativa da doença para a fase ativa.
- (III) Regime supercrítico ($\lambda > \lambda_c$): para esses valores da taxa de infecção a epidemia está de fato na fase ativa e pode se sustentar para sempre, levando a uma fase endêmica da doença.

Também é possível notar uma diferença nas três séries: a frequência nas fases sub e supercrítica é maior do que na fase crítica, isso porque a dinâmica das duas primeiras fases é tal que o processo de disseminação retorna rapidamente ao ponto de equilíbrio. Para a fase subcrítica um processo de infecção ocorre e logo termina. No regime supercrítico, ρ flutua em torno de um valor médio. E na fase crítica o processo é dominado por flutuações muito grandes, o que pode ser percebido observando as distâncias entre os picos/vales da série. Essas características irão refletir no mapeamento das séries em redes complexas, que é o objeto de estudo desta dissertação.

3.3 Mecanismos de ativação

Nas seções anteriores foram expostos alguns conceitos básicos em modelagem epidêmica. Sabe-se então que ao variarmos o parâmetro de controle λ até um dado valor crítico λ_c , uma transição de fase ocorre no limite termodinâmico. Assim, a epidemia passa de uma fase inativa - em que o número de infectados não é suficiente para manter a doença - para uma fase ativa na qual $\langle \rho \rangle$ é uma quantidade extensiva, fazendo com que a epidemia se sustente indefinidamente. Mas quais são as estruturas responsáveis para que isso ocorra? Qual nó deve ser infectado para a propagação seja máxima? Essas questões são cruciais no contexto de disseminação. Seja no sentido de conter a propagação (para o caso de doenças) ou de aprimorá-la para uma informação que se deseja ser passada adiante [36].

Inicialmente o papel fundamental para processos de disseminação era dado aos *hubs*. Esses nós de grau muito grande ao serem infectados irão transmitir a doença para seus vizinhos e depois de um tempo ficarão suscetíveis novamente. Mas como são muitos nós em sua vizinhança, o *hub* logo será reinfestado. Em [20] foi proposto o seguinte argumento: para um valor de $\lambda > \lambda_c$ os *hubs* e seus vizinhos se tornam ativos na rede, e essa atividade possui um tempo característico $\tau(k, \lambda)$ (ou τ_k^{rec}), dado por

$$\tau_k^{\text{rec}} \sim e^{ak\lambda^2}. \quad (3.15)$$

Logo, temos que o tempo de atividade desse conjunto (*hub* e vizinhos) cresce exponencialmente com o grau do *hub*. Se esse tempo for suficientemente grande a atividade pode ser transferida entre *hubs*, mesmo que estejam separados por distâncias $d > 1$ como explicado a seguir.

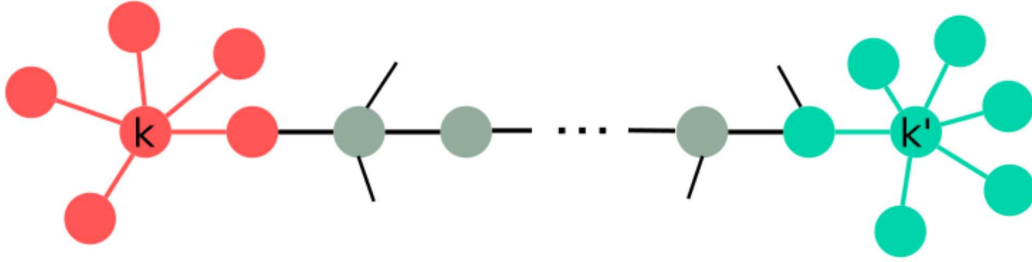


Figura 3.7 – Ativação por *hubs* separados a uma distância $d_{kk'}$.

Seguindo os passos das referências [43, 44], pode-se se ter uma configuração como a mostrada na Fig. 3.7, em que dois *hubs* estão conectados a uma distância $d_{kk'}$ e o tempo para que o nó de grau k (fonte) infecte o nó de grau k' (alvo) é $\tau_{kk'}^{\text{inf}}$

$$\tau_{kk'}^{\text{inf}} = \frac{1}{\lambda} \left(\frac{N \langle k \rangle}{kk'} \right)^b, \quad (3.16)$$

em que $b = \ln(1 + 1/\lambda)/\kappa$ e $\kappa = \langle k^2 \rangle / \langle k \rangle - 1$, ou seja, $\tau_{kk'}^{\text{inf}} \sim 1/\lambda$. Assim, se a condição $\tau_k^{\text{rec}} \gg \tau_{kk'}^{\text{inf}}$ é satisfeita, a epidemia será sustentada pela ativação dos *hubs* [43].

Como já foi discutido na seção 3.1, considerando uma aproximação de campo médio microscópica, o limiar epidêmico para o modelo SIS em um grafo estrela pode ser dado por $\lambda_c = 1/\sqrt{k_{\text{max}}}$ [45], sugerindo que os *hubs* são responsáveis pela sustentação da epidemia de acordo com o argumento apresentado acima, pois temos $\tau_k^{\text{rec}} \gg 1$ para $\lambda \gg 1/\sqrt{k_{\text{max}}}$.

Porém olhando ainda para a Eq.(3.14), nota-se que para $2 < \gamma < 5/2$ temos $\lambda_c \sim \langle k \rangle / \langle k^2 \rangle \sim k_{\text{max}}^{-3+\gamma}$, ou seja a dependência de λ_c com k_{max} é diferente daquela para $\gamma > 5/2$. Isso pode levantar a seguinte questão: existem outras estruturas que poderiam ser responsáveis pela manutenção da epidemia na rede?

Em 2010, Kitsak *et al* [46] verificaram que para algumas redes os nós com maior poder de disseminação se encontram em núcleos densamente conectados de determinado grau. A identificação dessa estrutura é possível através do método da decomposição de k -core, ilustrada na Fig. 3.8 e explicado a seguir:

- (i) Identificando o 1 -core: primeiro são retirados todos os nós que possuem grau 1 da rede, até não sobraem nós com esse grau (representados em azul na Fig. 3.8);
- (ii) Identificando o 2 -core: retira-se todos os nós que possuem grau 2 (em verde);
- (iii) Identificando o 3 -core: retira-se todos os nós com grau 3, note que nesse passo a rede deixa de existir após serem retirados esses nós. Assim dizemos que o 3 -core é o máximo k -core da rede.

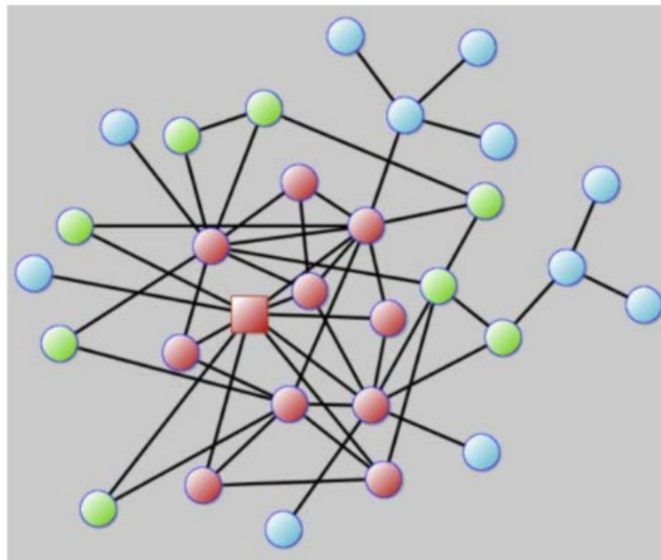


Figura 3.8 – Decomposição de k -core . Retirado de [45]

Pode-se mostrar que o limiar epidêmico para o máximo k -core em uma rede livre de escala é dado por $\lambda_c \sim k_{\max}^{\gamma-3}$ [45]. Assim, para $\gamma < 5/2$, o limiar epidêmico vai para zero mais rapidamente devido a presença do máximo k -core do que pela influencia dos *hubs* ($\lambda_c \sim 1/\sqrt{k_{\max}}$), indicando que para esse caso o máximo k -core sustenta a epidemia por longos períodos de tempo e espalha a infecção para o restante da rede.

Temos ainda o mecanismo de ativação coletivo que ocorre em redes homogêneas, para o qual todas as estruturas da rede contribuem para a sustentação da epidemia.

IV Mapeando séries em redes complexas

Uma série temporal pode ser definida como uma coleção de variáveis indexadas de acordo com a ordem em que são obtidas no tempo [47]. A análise de séries temporais complexas é um campo de investigação amplamente difundido, já que sistemas de diversas áreas podem ser estudados através dessas estruturas [48], tais como economia, medicina, meteorologia, engenharia, robótica, etc. O número de ferramentas disponíveis para investigar séries temporais é vasto [49] e a validade e robustez dos métodos têm sido aprimoradas ao longo dos anos. Uma das formas de se analisar séries temporais é transformando-as em redes complexas. Existem três principais classificações de redes geradas a partir de séries, sendo elas [1]:

- a) Redes de proximidade - geradas a partir de semelhança estatística mútua ou proximidade métrica entre diferentes segmentos de uma série;
- b) Redes de transição - geradas por meio da frequência de transição entre estados discretos da série;
- c) Grafos de visibilidade - obtidos através de um critério geométrico entre as observações sucessivas.

Os tipos de rede a) e b) são construídas utilizando o conceito de recorrência [50]. Para isso, faz-se necessária a reconstrução do atrator imerso no espaço de fase da série temporal, um procedimento introduzido por Packard *et al* em 1980 [51] e aprimorado matematicamente por Takens [52]. Esse método consiste em utilizar vetores m -dimensionais $\vec{\xi}_l$ que são construídos a partir da série temporal $\{x_i\}$ (onde $\{x_i\} = x(t_i), i = 1, \dots, N$)

$$\vec{\xi}_l = \{x(t_i), x(t_i + p) \dots x(t_i + (m - 1)p)\}, \quad (4.1)$$

em que m é a dimensão de imersão do atrator e p é o passo de reconstrução (por isso o método é conhecido também por método de atrasos temporais). Esse procedimento foi amplamente utilizado, pois o atrator reconstruído é topologicamente equivalente ao original, permitindo assim o cálculo de quantidades importantes a respeito da série temporal como expoentes de Lyapunov [53] e dimensão de correlação [54]. Porém, encontrar quais os valores adequados de m e p não é uma tarefa fácil. Muitas vezes a dimensão de imersão em que o atrator pode estar imerso é maior que três, já não sendo possível uma inspeção visual, necessitando então da utilização de outras ferramentas de análise do atrator. Além disso para séries pequenas, se p for grande, os dados não serão suficientes para grandes dimensões de imersão e não será viável observar a convergência da dimensão de correlação. Desta forma, resta o tipo de rede c) que somente faz uso da série temporal para gerar o grafo de visibilidade, que será tratado na seção a seguir.

4.1 O Grafo de Visibilidade (GV)

O conceito de visibilidade há muito já vinha sendo aplicado no contexto da geometria computacional para estudos em robótica ou em descrições topográficas do espaço e planejamento arquitetônico [55, 56]. Mas foi em 2008 que Lacasa *et al* [12] adaptaram esse conceito para séries temporais a fim de mapeá-las em redes complexas, mostrando que a rede gerada herda características relevantes da série. A definição de visibilidade para séries temporais é dada por:

Dois pontos em uma série temporal (t_a, y_a) e (t_b, y_b) terão visibilidade, ou seja, se tornarão dois nós conectados na rede associada, se qualquer valor entre eles (t_c, y_c) satisfizer:

$$y_c < y_b + (y_a - y_b) \frac{(t_b - t_c)}{(t_b - t_a)} = A. \quad (4.2)$$

A Eq.(4.2) é o critério de visibilidade que será utilizado para construir o algoritmo do GV. Esse critério vem da seguinte condição geométrica: três pontos $(t_a, y_a), (t_b, y_b)$ e (t_c, y_c) em um plano são colineares se

$$\begin{vmatrix} t_a & y_a & 1 \\ t_b & y_b & 1 \\ t_c & y_c & 1 \end{vmatrix} = 0, \quad \text{com } t_a < t_c < t_b. \quad (4.3)$$

Assim, resolvendo o determinante da Eq.(4.3) temos a condição de que os três pontos estejam alinhados em um plano dada por $y_c = A$, com A dado na Eq.(4.2). Desta forma, para que os pontos y_a e y_b possuam visibilidade, deve-se ter $y_c < A$. Para ilustrar a Eq.(4.2) temos a Fig. 4.1, que mostra um exemplo de série temporal com 16 valores e o GV gerado.

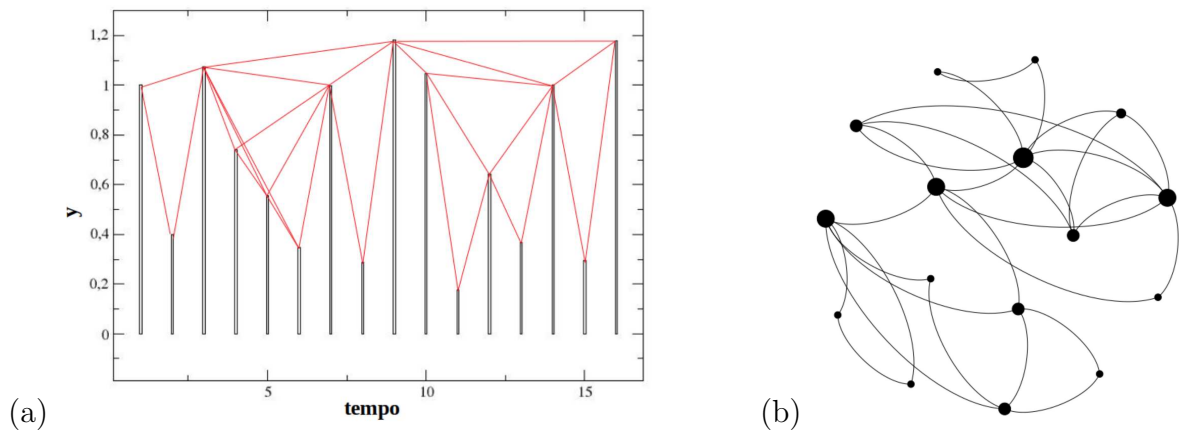


Figura 4.1 – (a) As linhas vermelhas ilustram o critério de visibilidade definido pela Eq.(4.2), mostrando como os valores da série enxergam uns aos outros. (b) GV gerado a partir da série em (a).

Note que os nós da Fig. 4.1(b) possuem tamanhos diferentes, pois estão escalonados de acordo com o grau. Os quatro maiores nós correspondem aos valores maiores da coordenada y , porém os *hubs* no GV não necessariamente serão os maiores valores da série, já que a visibilidade depende muito mais do tamanho de y em relação aos seus vizinhos. Em outras palavras, se tivéssemos um valor alto rodeado por dois pontos semelhantes a visibilidade desse seria bloqueada pelos vizinhos.

Algumas características do GV podem ser explicitadas a partir do critério de visibilidade dado pela Eq.(4.2). Pode-se dizer que o GV sempre será:

- *conectado* - já que cada nó (cada valor na série) enxerga pelo menos seus vizinhos mais próximos;
- *não-direcionado* - pela própria construção do algoritmo, os nós se enxergam mutuamente;
- *invariante sob transformações afins na série temporal* - é possível fazer reescalonamentos verticais e horizontais nos dados ou até mesmo transladar e adicionar tendência linear na série sem que o GV seja alterado.

Outra característica que deve ser destacada é o fato de que a estrutura da série temporal é conservada na topologia do GV. Assim, uma série periódica irá mapear em um grafo regular, já que foi construído através da repetição de um padrão. Além disso a distribuição de grau será uma sequência de picos como seria o espectro de potências de Fourier da série; Fig. 4.2(c). Na Fig. 4.2(a) temos uma série de 100 valores de uma função α de período 4, que dá origem ao GV mostrado na Fig.4.2(b):

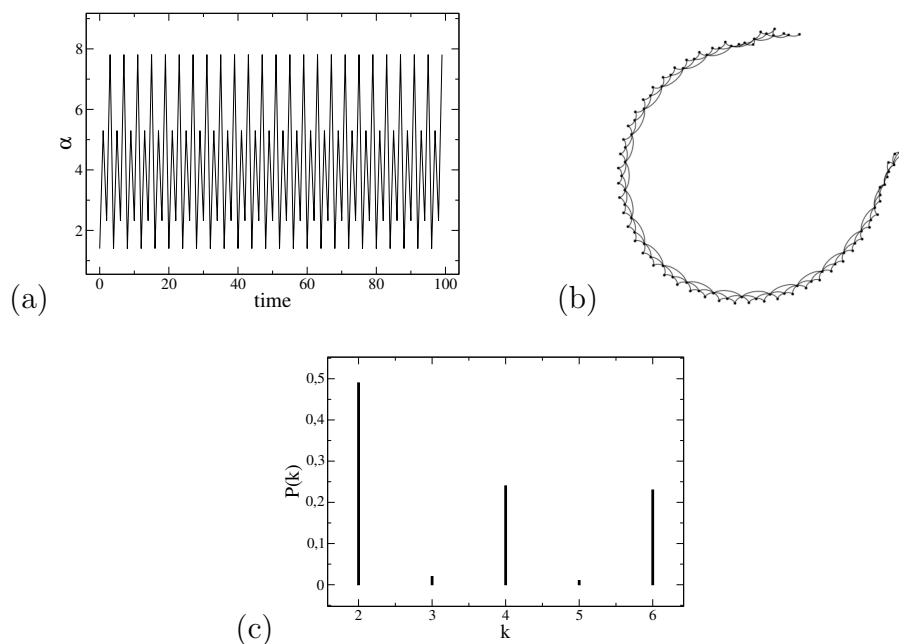


Figura 4.2 – (a) Exemplo de série periódica, (b) GV gerado (regular) a partir dela e a (c) distribuição de grau do GV.

Outro tipo de série temporal é o ruído branco (RB), gerada a partir de uma distribuição uniforme com valores entre 0 e 1 [57] e mostrada na Fig.4.3(a). Poderíamos esperar que essas séries mapeassem em redes aleatórias sem nenhum tipo de correlação, pois a própria série não possui qualquer correlação entre pontos subsequentes. Porém, o fato da série ser um conjunto ordenado de valores aleatórios gera uma correlação intrínseca, pois um pico muito provavelmente não será seguido por outro valor alto.

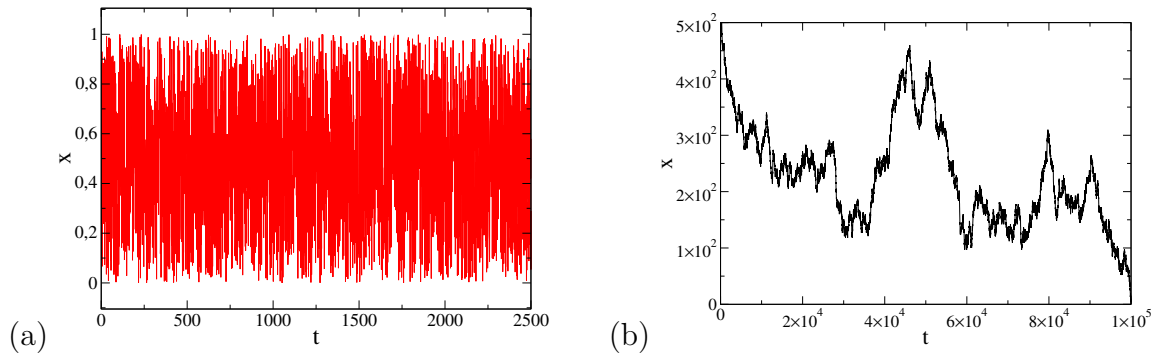


Figura 4.3 – Exemplos de série (a) do tipo ruído branco extraída de uma distribuição uniforme de intervalo $[0, 1]$ e (b) gerada a partir do movimento Browniano unidimensional.

Essa correlação estará presente no GV gerado pois a presença de valores intermediários na série compromete a visibilidade dos muito altos fazendo com que valores baixos se conectem mais entre si e o mesmo para aqueles muito altos, explicando o comportamento associativo mostrado na Fig.4.4(b).

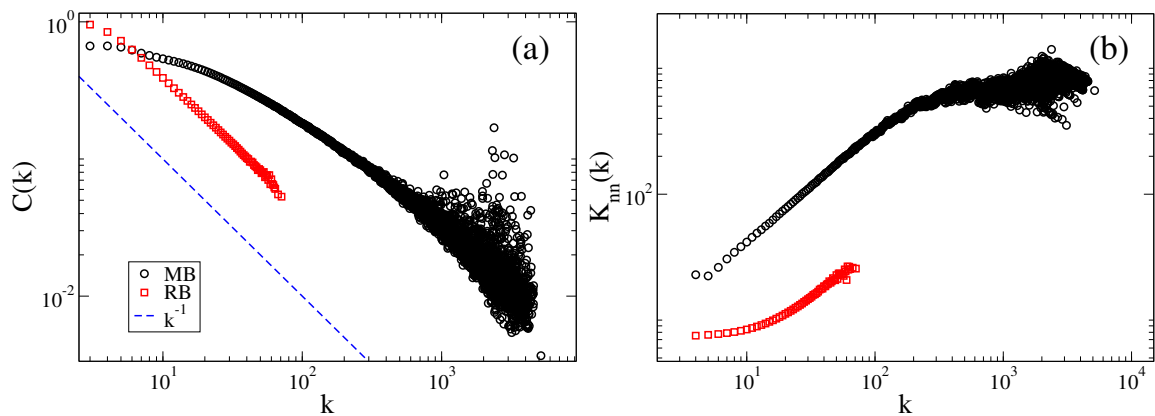


Figura 4.4 – (a) Coeficiente de agregação e (b) grau médio dos vizinhos em função do grau para o GV gerado a partir de séries de ruído branco (RB) e movimento Browniano (MB), com tamanho $t = 10^6$. As curvas mostradas são a média para 10 amostras de séries de cada tipo.

Por outro lado, como a série é limitada devido a variância finita, existirá um corte na visibilidade e isso irá impor um limite superior para o número de conexões que um ponto da série pode exercer. Em comparação com o modelo Erdős-Renyi com conexões completamente aleatórias entre os nós, a distribuição de grau deve ter uma cauda

mais pesada que a distribuição de Poisson (cuja cauda é Gaussiana), porém decaindo mais rápido do que aquela para uma rede livre de escala, como as geradas por conexão preferencial. O que se observa para o RB é uma distribuição exponencial que, de fato, possui as características relacionadas acima, como mostrado na Fig. 4.5.

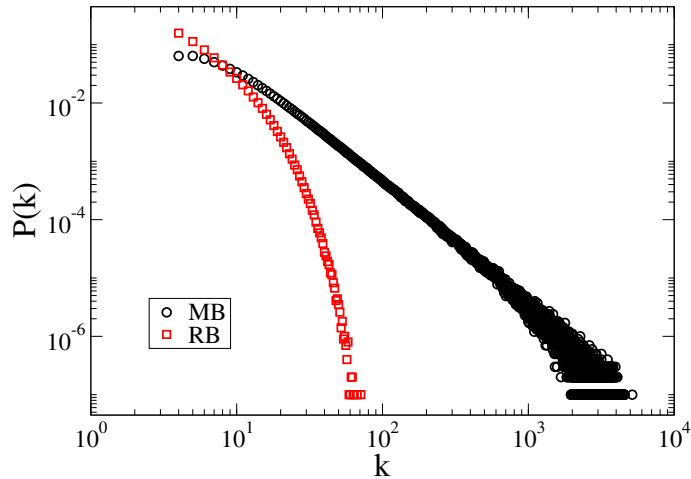


Figura 4.5 – Comparação entre distribuições de grau do GV (tamanho $t = 10^6$) gerado a partir de duas séries diferentes: ruído branco (RB) e movimento Browniano (MB). As curvas mostradas são a média para 10 amostras de séries de cada tipo.

Séries temporais fractais estão presentes em muitas áreas de estudo, caracterizando um objeto importante de investigação. Uma das séries fractais mais conhecidas é aquela gerada por um movimento Browniano (MB) unidimensional [49]. O MB ocorre para uma partícula se movimentando aleatoriamente [57]. Esse processo gera séries de posição da partícula como mostrada na Fig. 4.3(b). A princípio essas séries terão valores negativos. Esses, porém, podem comprometer a implementação do critério de visibilidade, sendo necessário fazer uma translação vertical dos dados tomando como referência o menor valor da série, de forma que os valores fiquem todos positivos.

Um processo estocástico contínuo, $X(t) = \{X(t), t > 0\}$, é dito *auto-similar* se

$$X(\epsilon t) \equiv \epsilon^H X(t), \quad t \geq 0, \quad (4.4)$$

para $\epsilon > 0$, e $0 < H < 1$. Em que H é o parâmetro de auto-similaridade ou parâmetro de Hurst [49]. O MB é um tipo de processo auto-similar, com $H = 0.5$.

A relação entre auto-similaridade e redes livres de escala tem sido bastante investigada [58] e isso pode levantar o seguinte questionamento: o GV captura a característica de auto-similaridade presente em uma série fractal? Aparentemente sim. Pois como mostrado na Fig. 4.5, e discutido em [12], a distribuição de grau do GV para o MB é do tipo lei de potência com expoente $\gamma \approx 2$, caracterizando uma rede livre de escala. Note na Fig. 4.5 para o GV do RB, que a distribuição de grau possui decaimento exponencial, como já discutido anteriormente. E para o GV gerado pela série de MB, a distribuição decai bem

mais lentamente, mostrando a grande variedade de nós existentes na rede e evidenciando os *hubs* na cauda pesada.

Na Fig. 4.4 temos o cálculo de duas quantidades de interesse já discutidas no capítulo II deste trabalho: o coeficiente de agregação $C(k)$ e o grau médio dos vizinhos $K_{nn}(k)$ para as séries de RB e MB. É possível notar na Fig.4.4(a) que os GV gerados por ambas as séries possuem a característica de modularidade hierárquica, $C(k) \sim k^{-1}$, comentada na seção 2.3, mostrando que os nós com grau maior estão mais ligados à comunidades diferentes da qual pertencem, enquanto que nós com graus menores estão mais conectados em suas comunidades que são mais densas. De certa forma, isso está relacionado ao comportamento associativo do $K_{nn}(k)$, exibido na Fig. 4.4(b), ou seja, nós tendem a se conectar com outros de grau semelhante. Porém, os valores extremos da série de MB, de fato, enxergam todo tipo de ponto e por isso a associatividade fica atenuada para graus $k \approx 500$, passando a exibir comportamento mais neutro, não correlacionado. Assim, as análises de $C(k)$ e $K_{nn}(k)$ também confirmam que o RB gera um GV com alto coeficiente de agregação, estrutura hierárquica e correlações de grau. Até onde sabemos, os resultados da Fig. 4.4 não foram relatados na literatura ainda.

Outra análise interessante a se fazer é a do menor caminho médio $\langle d \rangle$ em função do tamanho para os GV gerados a partir das duas séries, como mostrado na Fig. 4.6. Para ambas, temos um comportamento de $\langle d \rangle \sim \ln(t)$, que caracteriza o GV gerado como sendo do tipo *small-world*, ou seja, com dimensão infinita¹. Note que a para a série de MB a inclinação é um pouco menos acentuada, mas ainda sim é possível dizer que a dimensão do GV não é finita. Desta forma, ainda que a série de MB mapeie em uma rede cuja distribuição de grau é do tipo livre de escala, a rede não é fractal, e sim *small-world*.

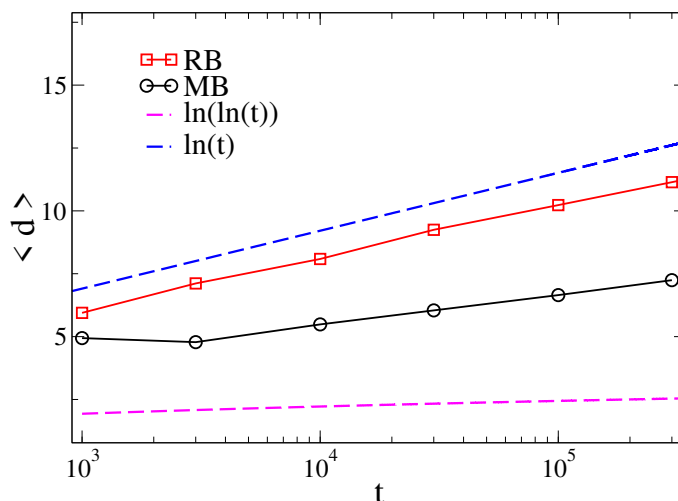


Figura 4.6 – Menor caminho médio em função do tamanho do GV gerados a partir de séries de ruído branco (RB) e movimento Browniano (MB).

¹ Em redes regulares, $\langle d \rangle \sim N^{1/d}$ com d sendo a dimensão da rede [25]. Se $\langle d \rangle \sim \ln(N)$ é como se o expoente $1/d$ se aproximasse de zero, logo $d \rightarrow \infty$.

4.2 GV para séries de prevalência epidêmica - exemplo em rede quadrada

Na seção anterior foi explicitado como é possível transformar uma série temporal em uma rede complexa. As aplicações desse procedimento são gerais, podendo-se utilizar a princípio qualquer série de dados reais ou sintéticos. A aplicação escolhida para este trabalho foi a de séries de prevalência epidêmica, já mencionadas na seção 3.2. Tomando novamente o exemplo do modelo SIS em uma rede quadrada podemos mapear as séries de ρ mostradas na Fig. 4.7 em grafos de visibilidade com distribuições de grau dados pela Fig. 4.8.

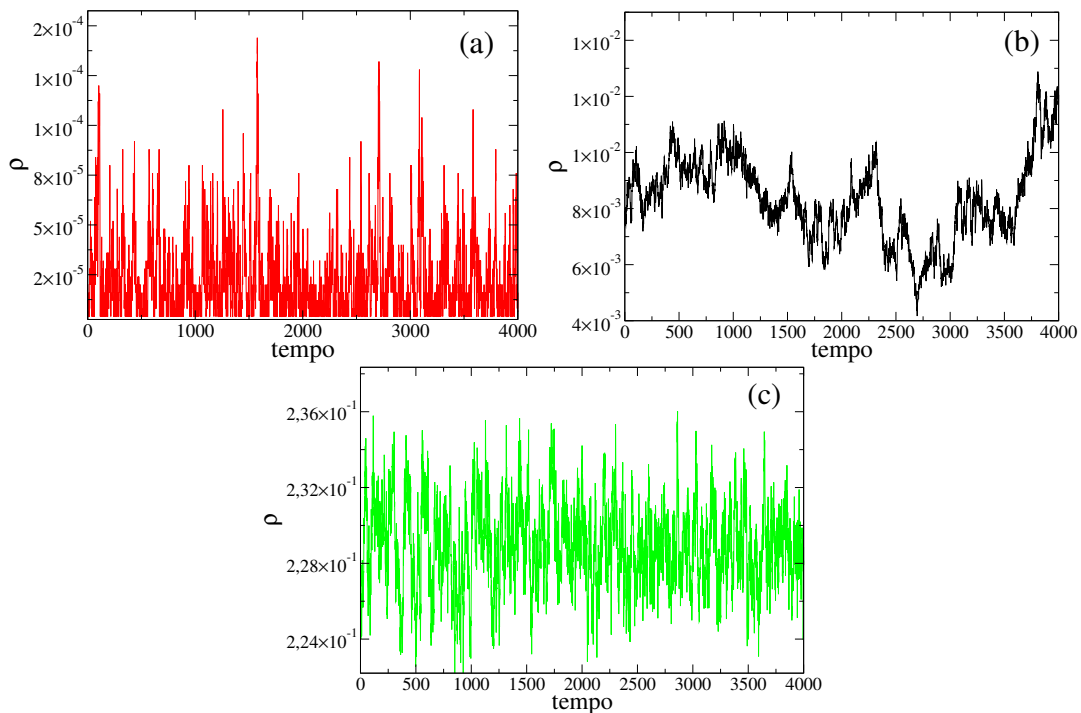


Figura 4.7 – Séries de prevalência epidêmica do modelo SIS numa rede quadrada, de tamanho $N = L^2$ com $L = 500$, para as fases (a) subcrítica $\lambda_c = 0.37$, (b) crítica $\lambda_c = 0.41$ e (c) supercrítica $\lambda_c = 0.44$.

É possível notar que, devido a diferença na escala característica das flutuações nas séries, as interfaces de cada uma se diferenciam. Sendo que as séries da fase sub e supercrítica são mais semelhantes enquanto a da fase crítica possui um padrão diferente, com picos e vales mais espaçados. Essa diferença é capturada pela distribuição de grau do GV como mostrado da Fig. 4.8.

Na Fig. 4.9 é mostrado o cálculo de $K_{nn}(k)$ e $C(k)$ para as diferentes fases epidêmicas. Mais uma vez é possível notar a diferença entre as curvas sub e super da crítica. O comportamento é bastante semelhante para o coeficiente de agregação, tendo apenas um peso maior na cauda da curva da fase crítica, evidenciando mais uma vez a presença de *hubs* no GV gerado.

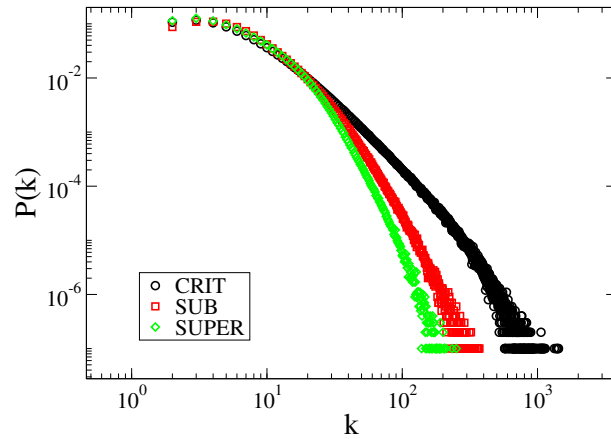


Figura 4.8 – Distribuição de grau para os GV gerados a partir das séries de prevalência da Fig. 4.7 para as diferentes fases epidêmicas.

Já para a correlação de grau a diferença é um pouco mais relevante já que as curvas das fases sub e supercrítica tem padrão puramente associativo enquanto a fase crítica apresenta mudança de regimes, mesmo que suave, de associativa para neutra. Note que ao compararmos as séries de prevalência epidêmica com as de RB e MB, algumas semelhanças são evidentes nessas quantidades. Como esses padrões interessantes também ocorrem em outros tipos de rede, será feita uma melhor discussão sobre isso no capítulo VI.

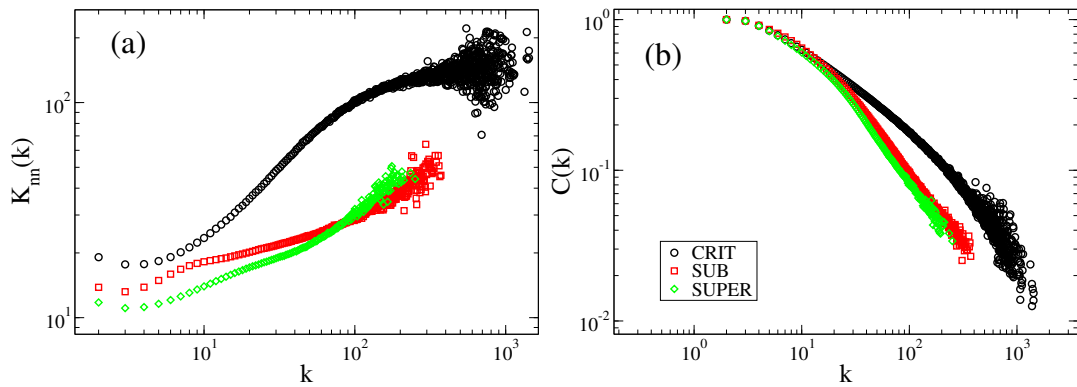


Figura 4.9 – (a) Grau médio dos vizinhos e (b) coeficiente de agregação como funções do grau para os GV gerados a partir das séries de prevalência epidêmica da Fig. 4.7.

V Metodologia

Muitas redes reais possuem um número muito grande de ligações e nós. Redes simuladas computacionalmente também podem ter tamanhos extremamente grandes, já que para muitas situações é interessante comparar comportamentos com o limite termodinâmico, em que $N \rightarrow \infty$. Assim, a maneira como são armazenados os dados da rede deve ser implementada cuidadosamente. Um exemplo disso é o uso da matriz de adjacência para identificar as ligações na rede - muitas delas são esparsas, ou seja, o número de ligações que de fato existe é muito menor do que seria comportado para o tamanho da rede. Isso faz com que a matriz de adjacência possua muitas entradas zeradas, gerando um grande volume de dados não utilizados. Então, o que se faz é substituir essa matriz por listas, que contenham apenas informações sobre as ligações que de fato existem.

Por exemplo, na Fig. 2.6 temos uma rede não direcionada e sua matriz de adjacência. Podemos construir três vetores que juntos permitem acessar toda a informação da matriz A_{ij} , são eles:

1. O vetor de grau ($g[i]$): que armazena o grau de cada nó i em suas entradas, logo, tem tamanho igual ao da rede. Para o exemplo da Fig. 2.6, temos $g = \{2, 3, 2, 4, 1\}$.
2. O vetor de adjacência ($a[i]$): nele estão contidas as ligações de cada nó. Esse vetor tem tamanho igual ao dobro do número de ligações da rede, já que contamos duas vezes a mesma ligação. No exemplo da Fig. 2.6, temos

$$a = \left\{ \underbrace{2, 4}_{\text{nó 1}}, \underbrace{1, 3, 4}_{\text{nó 2}}, \underbrace{2, 4}_{\text{nó 3}}, \underbrace{1, 2, 3, 5}_{\text{nó 4}}, \underbrace{4}_{\text{nó 5}} \right\}. \quad (5.1)$$

3. O vetor auxiliar ($x[i]$): que marca o começo da informação correspondente a cada nó no vetor de adjacência, ou seja as ligações do nó i começam em $a[x[i]]$ e terminam em $a[x[i] + g[i]]$. Para o exemplo da Fig. 2.6 seria $x = \{0, 2, 5, 7, 11\}$.

Dessa maneira é possível armazenar toda a informação das ligações da rede com utilização de memória acessível, o que não ocorre quando armazenamos matrizes $N \times N$.

5.1 Gerando redes complexas

Como foi mostrado em seções anteriores redes reais possuem correlação de grau. Porém, sob o ponto de vista teórico muitas vezes é interessante desconsiderar essa propriedade e gerar redes não correlacionadas. Uma forma de fazer isso é utilizando o modelo de configurações não-correlacionado, UCM (do inglês *Uncorrelated Configuration Model*) [28]. A implementação desse modelo segue os seguintes passos:

- a) Sortear os graus de todos os N nós seguindo uma distribuição de grau preestabelecida e atribuir a cada nó uma quantidade de pontas desconectadas (ou *stubs*) de acordo com o grau. Para isso, deve-se converter uma sequência de números aleatórios ϵ_i (distribuição uniforme entre 0 e 1) na sequência de graus desejada k_i ; Fig. 5.1(a). Uma forma de fazer isso é calculando a seguinte expressão [57]

$$\epsilon = \int_{k_0}^k P(k') dk' = F(k), \quad (5.2)$$

em que $F(k)$ é a distribuição de probabilidades cumulativa associada a variável k , k_0 é o grau mínimo (para esse trabalho $k_0 = 3$). k_c é o corte superior da distribuição, ou seja, o maior grau que os nós podem ter. Note que não necessariamente $k_c = k_{\max}$ mas sempre $k_{\max} \leq k_c$. Assim, para obter a sequência de graus usamos $k = F^{-1}(\epsilon)$. Para que a rede não possua correlações de grau usamos $k_c = \sqrt{N}$, como explicado na referência [59]. Isso garantirá que a chance de ocorrerem auto ou múltiplas conexões (Fig. 5.1(c)) é desprezível. Além disso, a soma de todos os graus deve ser par, para que não sobre nenhum *stub* desconectado.

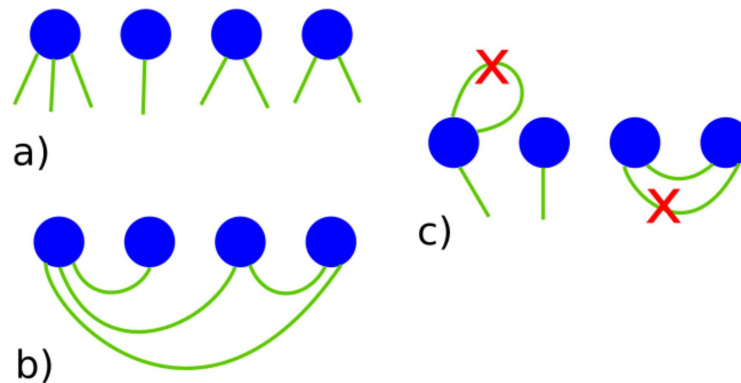


Figura 5.1 – Esquematização dos passos de implementação do modelo UCM. (a) *Stubs* distribuídos segundo uma sequência de graus - $g = 3, 1, 2, 2$. (b) Conexões entre dois nós aleatoriamente escolhidos. (c) Exemplos de auto e múltipla conexões que são proibidas nesse modelo.

- b) Sortear aleatoriamente dois *stubs* e conectar os nós aos quais eles pertencem; Fig. 5.1(b). Verificar se forma auto ou múltipla conexão, Fig. 5.1(c), e implementar somente se não for o caso.
- c) Repetir o passo b) até que a rede esteja conectada. Se ao fim do processo sobrarem *stubs* desconectados, descarta-se a rede e os passos são refeitos.

O algoritmo explicitado acima foi usado para gerar redes com distribuição $P(k) \sim k^{-\gamma}$ nesse trabalho. Na Fig. 5.2(a) temos um exemplo de distribuição de grau de uma rede gerada pelo modelo UCM, com tamanho $N = 10^6$ e $\gamma = 2.75$. Veja na Fig. 5.2(b) que essa rede é não correlacionada.

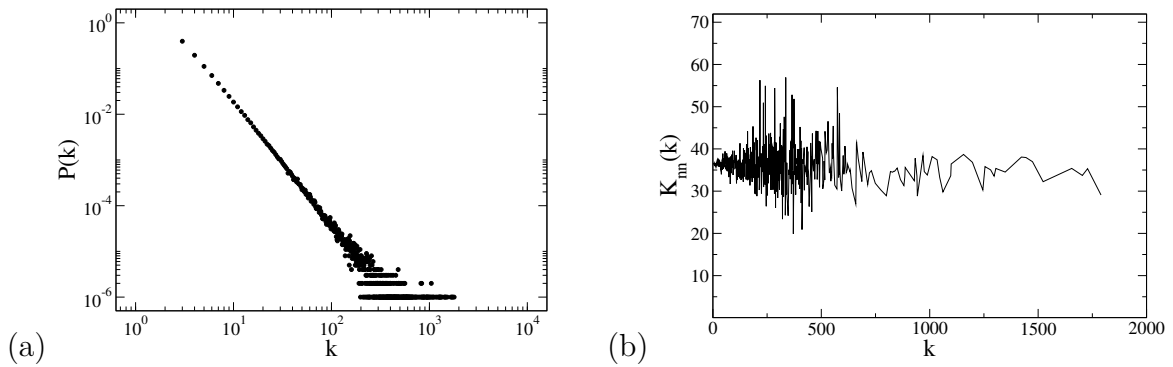


Figura 5.2 – (a) Distribuição de grau em lei de potência para uma rede gerada pelo modelo UCM com $N = 10^6$ nós e $\gamma = 2.75$. (b) Grau médio dos vizinhos para essa mesma rede. Note que $K_{nn}(k) \sim k^\mu$ com $\mu \approx 0$, mostrando que a rede é do tipo neutra ou não correlacionada (seção 2.4).

5.2 Algoritmo BFS

Na seção 2.1 os conceitos de diâmetro (d_{\max}) e menor caminho médio ($\langle d \rangle$) foram apresentados. Saber como $\langle d \rangle$ varia com N é importante para avaliar se a rede apresenta a característica de mundo pequeno, em que $\langle d \rangle \sim \ln(N)$. O cálculo de distâncias em redes complexas pode ser uma tarefa extremamente trabalhosa para redes de tamanho considerável, sendo necessária a utilização de um método que meça o número de caminhos entre todos os nós da rede em pares. Uma das formas de se fazer isso é utilizando algoritmos de busca em largura, que percorrem toda a extensão da rede. Nesse trabalho utilizamos o chamado BFS, do inglês *Breadth-First Search* [25].

O algoritmo BFS é frequentemente usado em redes complexas. Através dele é possível encontrar a distância de um dado nó inicial s até outro qualquer que esteja na mesma componente¹ de s [11]. Para sua implementação deve-se seguir os seguintes passos:

- a) Começa-se com um nó i escolhido aleatoriamente. Esse nó recebe o rótulo 0. Será utilizado um vetor $d[a]$, com $a = 1, \dots, N$ para guardar as distâncias de todos os nós j do nó i . Deve-se atribuir inicialmente um valor negativo qualquer, -1 por exemplo, a todas as entradas desse vetor. Além disso, $d[i] = 0$, ou seja a distância do primeiro nó até ele mesmo é zero; Fig. 5.3(a).
- b) Encontrar os vizinhos j do nó i e verificar se a distância deles é desconhecida ($d[j] = -1$). Se sim, a eles é atribuída uma distância $d[j] = 1$ e são rotulados com 1; Fig. 5.3(b). Esses nós são colocados em um outro vetor $Q[a]$ que será uma espécie de fila.

¹ Uma componente é um sub-conjunto de nós da rede para o qual existe pelo menos um caminho entre quaisquer pares nós que a formam.

- c) Repete-se o passo b) para cada elemento do vetor $Q[a]$, colocando os vizinhos desses elementos também na fila com auxílio de ponteiros para fazer a leitura. A esses nós é atribuída a distância $d[j] = 2$; Fig. 5.3 (c). E assim por diante.
- d) O item c) deve ser repetido até que se atinja o nó j para o qual se quer calcular a distância d_{ij} ou até que não restem mais distâncias desconhecidas.

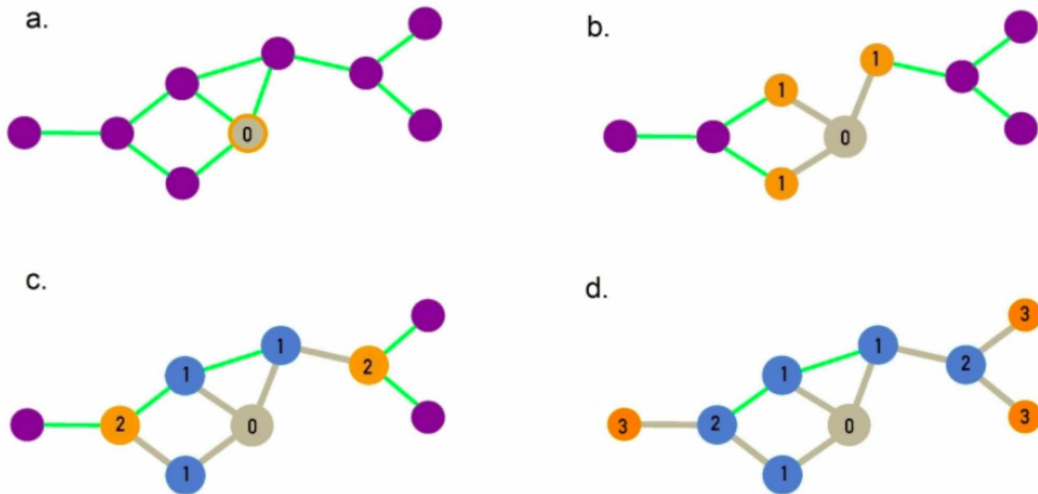


Figura 5.3 – Ilustração do algoritmo BFS. Cada figura corresponde à uma rotulação de vizinhos. Em (a) temos o nó inicial. Pretende-se calcular a distâncias dos outros até ele. O rótulo sempre significa a distância até esse nó. (b) Os vizinhos do nó com rótulo 0 recebem rótulo 1, já que essa é a distância deles até o nó inicial. (c) Os nós com rótulo 2 são os vizinhos dos com rótulo 1 que ainda não tinham distância marcada. (d) Os nós com rótulo 3 indicam que essa é a distância deles até o nó inicial. Retirado de [25].

Fazendo esse processo para todos os nós da rede teremos todas as distâncias d_{ij} possibilitando o cálculo de $\langle d \rangle$, através da Eq.(2.1). Note que é possível também calcular d_{\max} usando o algoritmo BFS, para isso é preciso encontrar a maior d_{ij} que será o diâmetro da rede.

5.3 Algoritmo de Gillespie otimizado

Na seção 3.1, consideramos o modelo SIS, que é um tipo de modelo compartimental para estudo de processos de disseminação. Esse tipo de sistema, em que σ_i é o estado de cada nó e $\sigma(t) = (\sigma_1(t), \dots, \sigma_N(t))$ o estado microscópico de N nós no instante t , pode ser descrito como um processo *Markoviano*, para o qual a probabilidade $P(\sigma, t)$ de um estado σ ocorrer é dada por

$$\frac{\partial P(\sigma, t)}{\partial t} = \sum_{\sigma'} [W_{\sigma' \rightarrow \sigma} P(\sigma', t) - W_{\sigma \rightarrow \sigma'} P(\sigma, t)], \quad (5.3)$$

em que $W_{\sigma' \rightarrow \sigma}$ e $W_{\sigma \rightarrow \sigma'}$ são taxas de transição entre os estados σ e σ' e a evolução pode ser determinada totalmente a partir estado no instante t , caracterizando um processo Markoviano [57].

Uma das primeiras abordagens para simulação de processos Markovianos de tempo contínuo foi proposta por Gillespie em 1976 [60]. Porém na formulação de seu algoritmo, a cada mudança de estado, uma lista com todos os processos espontâneos possíveis (cura, infecção, perda de imunidade, etc) deveria ser atualizada, gerando um aumento na complexidade² e se fazendo inviável para redes grandes.

Na referência [8], foi descrita uma modificação no algoritmo de Gillespie, que reduz consideravelmente sua complexidade. Isso se deu pela introdução dos *processos fantasma* para os casos em que as transições não levam a mudanças de estado mas contam para incrementos no tempo. Esse novo algoritmo, chamado Algoritmo de Gillespie Otimizado OGA (do inglês *Optimized Gillespie algorithm*), foi utilizado neste trabalho para simular o modelo SIS e sua implementação deve seguir as seguintes etapas:

- a) Estabelecer uma condição inicial para o número de nós infectados N_i (neste trabalho utilizamos $N_i = N$ como condição inicial) e a soma dos graus desses nós infectados dada por N_n . Também é preciso criar um vetor $I[N_i]$ e armazenar os rótulos dos nós infectados;
- b) Escolher qual evento irá ocorrer. O evento de cura, em que um nó infectado torna-se suscetível ocorre com probabilidade

$$p_{\text{cura}} = \frac{N_i}{N_i + \lambda N_n}. \quad (5.4)$$

Uma nova infecção, em que um nó infectado é escolhido proporcionalmente ao seu grau e um dos vizinhos desse nó é escolhido aleatoriamente para ser infectado, ocorre com a probabilidade

$$p_{\text{infec}} = \frac{\lambda N_n}{N_i + \lambda N_n}. \quad (5.5)$$

Caso esse vizinho seja suscetível, será infectado. Porém, se for o contrário a simulação prossegue e apenas o tempo será incrementado, o que define um processo fantasma. Note que $p_{\text{cura}} + p_{\text{infec}} = 1$, como deve ser;

- c) Atualizar N_i , N_n e $I[N_i]$;
- d) Incrementar o tempo t como $t \rightarrow t + \Delta t$, com $\Delta t = \frac{1}{N_i + \lambda N_n}$;
- e) Iterar os passos b), c) e d) até um tempo máximo t_{max} . Caso o estado absorvente $N_i = 0$ ocorra é preciso utilizar métodos quase estacionários para que o sistema volte a uma configuração ativa. Esses métodos serão comentados na próxima seção.

² Complexidade computacional ou apenas complexidade é a medida de tempo de execução do algoritmo.

5.4 Método quase estacionário

Existe uma grande quantidade de processos dinâmicos que apresentam o chamado estado absorvente, no qual a dinâmica do sistema fica presa e não evolui [61]. O modelo SIS é um desses sistemas, em que existe um estado estacionário ativo no limite termodinâmico, mas devido a finitude do sistema real, sua dinâmica sempre visitará o estado absorvente, em que todos os indivíduos então suscetíveis.

A fim de contornar esse problema utiliza-se as simulações quase estacionárias, em que a dinâmica do sistema é perturbada para que o estado absorvente não ocorra, sem que isso afete as quantidades intensivas no limite termodinâmico. Para entender como essas perturbações funcionam serão seguidos os passos da referência [24]: pode-se definir um processo estocástico X_t para o qual ocorrem transições do tipo $n \rightarrow n \pm 1$ com $n = 0, 1, 2, \dots$ e sendo $n = 0$ o estado absorvente. Definimos também um processo semelhante X_t^* para o qual $n = 0$ não é mais o estado absorvente. Para $n > 0$, a dinâmica de X_t^* é a mesma de X_t exceto por algumas perturbações que podem ser desprezadas no limite termodinâmico. Sejam P_n e P_n^* as probabilidades do sistema se encontrar no estado n nas dinâmicas original e modificada, respectivamente. Para a dinâmica original temos

$$\frac{dP_n}{dt} = \sum_m (W_{nm}P_m - W_{mn}P_n), \quad (5.6)$$

em que W_{nm} é a taxa de transição do estado m para n . Para a dinâmica modificada introduz-se uma fonte perturbativa de atividade de forma a remover o estado absorvente

$$\frac{dP_n^*}{dt} = \sum_m [W_{nm}P_m^* - W_{mn}P_n^*] + F(P_0^*, P_1^*, \dots), \quad (5.7)$$

em que F é um funcional que depende do método quase estacionário utilizado. Para esse trabalho, dois desses métodos foram utilizados - para redes heterogêneas usamos a Reativação de *Hub* (RH) e para as homogêneas a Condição de Contorno Refletora (CCR). Ambos os métodos serão brevemente descritos a seguir.

a) Método da Condição de Contorno Refletora

Um das formas de se evitar o estado absorvente é deixar que ele ocorra no sistema, mas em seguida retorne à configuração ativa imediatamente anterior, representando uma condição de contorno refletora [42]. Assim, o termo de fonte perturbativa da Eq.(5.7) é $F = (\delta_{1,n} - \delta_{0,n})P_0^*$ de modo que a equação dinâmica seja

$$\frac{dP_0^*}{dt} = P_1^* - P_0^* \quad (5.8)$$

logo, no estado quase estacionário teremos $\overline{P}_1 = \overline{P}_0$ já que $\overline{P}_n = \lim_{t \rightarrow \infty} P_n^*$. Assim, o sistema retorna sempre à configuração pré-absorvente com a mesma taxa que o visita, possibilitando o cálculo da média de algumas quantidades quase estacionárias de interesse, como por exemplo a densidade de infectados.

b) Método da Reativação de *Hub*

A interpretação desse método é semelhante àquela para CCR, porém agora ao invés de retornar a configuração pré-absorvente o que deve ser feito é reativar o nó de maior grau da rede, ou seja um *hub* [42]. Caso exista mais de um *hub* com mesmo grau, escolhe-se aleatoriamente entre eles. A razão para se escolher os *hubs* como forma de evitar o estado absorvente está relacionada ao seu papel nos mecanismos de ativação da epidemia, como já foi discutido na seção 3.3.

Assim, temos que é possível contornar o estado absorvente, evitando que a dinâmica do sistema seja congelada. Desta forma podemos fazer, por exemplo, médias de interesse do modelo SIS. Uma delas é $\langle \rho \rangle$ no estado quase estacionário, necessária para análise de transição entre as fases inativa e ativa. Para isso, deixamos que o processo descrito na seção 5.3 seja iterado por um tempo de relaxação t_{rlx} grande e calculamos a média de ρ durante um tempo t_{med} , com $t_{rlx} + t_{med} = t_{max}$ como ilustrado na Fig. 5.4.

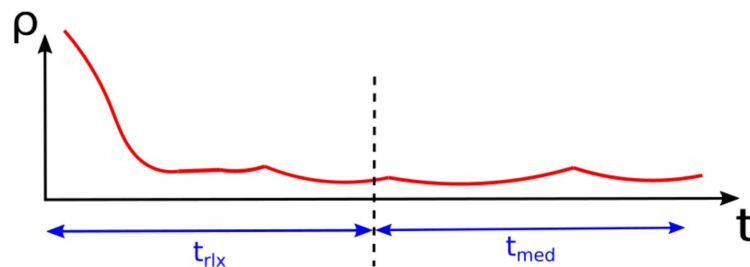


Figura 5.4 – Ilustração para série de densidade de infectados (ρ), evidenciando os tempos de relaxação e de média. A densidade quase estacionária é a média de ρ ao longo de um intervalo t_{med} .

VI Resultados e discussões

No capítulo V foram discutidos os métodos e algoritmos para gerar redes complexas e implementar o modelo SIS. A partir disso, é possível construir séries temporais quase estacionárias de densidade de sítios infectados na rede (prevalência epidêmica). Foi mostrado também no capítulo III que entender os diferentes regimes epidêmicos é fundamental para que se possa compreender a transição de fase para o estado absorvente, que ocorre no modelo SIS. Uma forma de estudar as fases epidêmicas é analisando as séries temporais de ρ para cada uma delas. Para isso, utilizamos a ferramenta apresentada no capítulo IV - o grafo de visibilidade. Os objetivos deste trabalho são verificar se o GV é sensível a natureza das séries temporais para os diferentes regimes epidêmicos e, além disso, saber se o GV difere quanto ao tipo de rede original em que o processo de disseminação ocorre. Essas análises são discutidas neste capítulo.

6.1 Escolha das redes utilizadas

Sabe-se que o tipo de rede em que o processo de disseminação acontece influencia [17], como é o caso das redes livres de escala, para as quais o limiar epidêmico do modelo SIS e outros similares tende a zero no limite termodinâmico, indicando que independentemente da taxa de infecção a epidemia pode se espalhar. Assim, é de particular interesse conhecer características da rede que suporta o processo epidêmico. Desta forma, para analisar o GV das séries de ρ foram escolhidos dois tipos de rede complexa já mencionados no capítulo II. Uma delas é muito heterogênea, gerada pelo modelo UCM com distribuição de grau $P(k) \sim k^{-\gamma}$ e parâmetros $\gamma = 2.75$, $k_0 = 3$ e $k_c = 2\sqrt{N}$. A outra é quase homogênea, sendo uma RRA mista com uma mistura de nós com graus 6 e 7 de maneira que o $\langle k \rangle$ é ajustável fazendo

$$\langle k \rangle = 6f + (1 - f)7 \quad (6.1)$$

com $f \in [0, 1]$. De acordo com o $\langle k \rangle$ desejado escolhe-se f . Isso foi feito para que pudessemos comparar os dois tipos de redes mas com o mesmo $\langle k \rangle$ nas seções 6.2 e 6.3. Também serão utilizadas RRAs com $m = 4$ e $m = 8$, além de outro tipo de rede heterogênea também gerada pelo modelo UCM, com $k_0 = 3$, $k_c = 2\sqrt{N}$ e $\gamma = 2.25$. Uma razão para se analisar redes homogêneas e heterogêneas é a diferença quanto aos mecanismos de ativação (seção 3.3) de cada uma das redes, como será discutido na seção 6.4.

Para determinar o limiar epidêmico do modelo SIS em cada uma das redes, utilizamos o método do cálculo da suscetibilidade discutido na seção 3.1. Assim, devido a um

aumento das flutuações em ρ nas proximidades da transição em $\lambda = \lambda_c$, a suscetibilidade apresenta um pico nesse ponto. Na Fig. 6.1 temos essa simulação para diferentes tamanhos de rede UCM com $\gamma = 2.75$. É interessante notar que conforme aumenta-se N o limiar epidêmico diminui, em acordo qualitativo com a previsão teórica QMF para redes em lei de potência com $\gamma > 5/2$, em que $\lambda_c \simeq 1/\sqrt{k_{\max}}$ e k_{\max} diverge com o tamanho da rede [21].

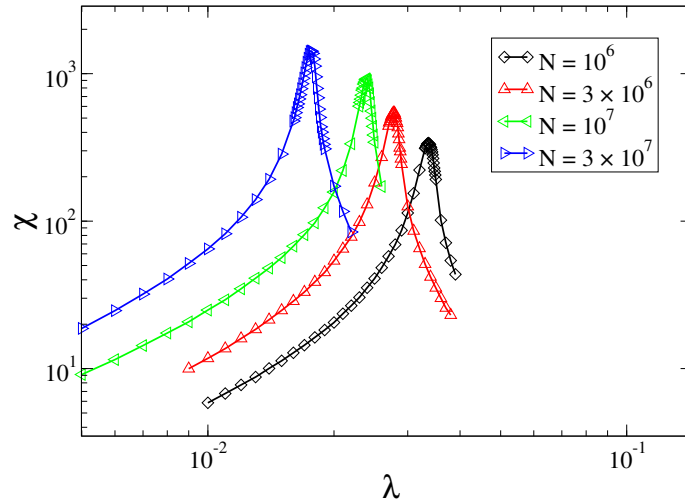


Figura 6.1 – Suscetibilidade dinâmica para o modelo SIS em redes geradas pelo modelo UCM com distribuição de grau em lei de potência e $\gamma = 2.75$ com tamanhos diferentes.

O mesmo procedimento descrito acima foi feito para a RRA mista, mostrado na Fig. 6.2, na qual temos um comportamento diferente do exibido para as redes heterogêneas: o pico da suscetibilidade converge para um valor não nulo ao aumentarmos N . Isso porque para esse tipo de rede não há *hubs* que promovem atividade para $\lambda \ll 1$ e a propagação da epidemia é praticamente uniforme em toda estrutura.

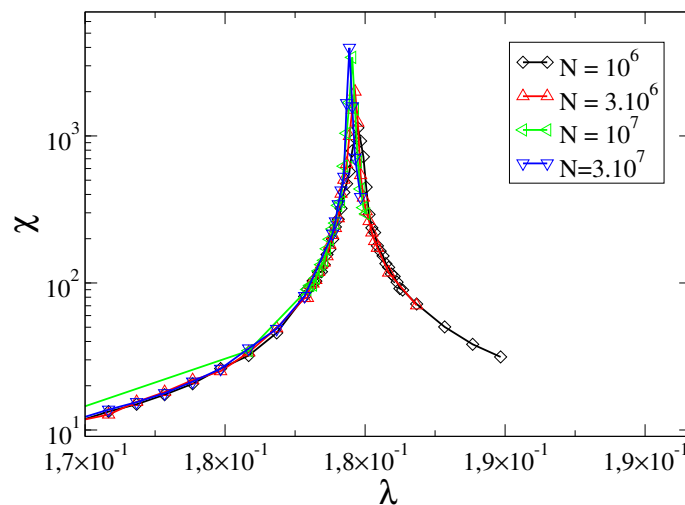


Figura 6.2 – Suscetibilidade dinâmica para RRA mista de graus $m = 6$ e 7 com tamanhos diferentes.

6.2 Análise de tamanho finito para o GV na fase crítica

Tendo escolhido quais serão as redes originais para a implementação do modelo SIS, dois parâmetros principais devem ser estabelecidos: o tamanho da rede original (N) e o número de dados da série temporal (t), que é o tamanho do GV gerado. Sabemos que alguns efeitos em redes complexas só ocorrem para sistemas de tamanho infinito [11]. Portanto é necessário entender como a modificação desses dois valores influencia nas características do GV a medida que $N \rightarrow \infty$ e $t \rightarrow \infty$.

Para esse estudo utilizamos três quantidades de interesse em análise de redes complexas: distribuição de grau $P(k)$ (seção 2.2), coeficiente de agregação $C(k)$ (seção 2.3) e grau dos vizinhos $K_{nn}(k)$ (seção 2.4). Na Fig. 6.3, mostramos essas quantidades para a rede UCM com $\gamma = 2.75$ e para a RRA com $m = 6$ e 7 , todas com tamanhos de rede original indo de $N = 10^6$ até $N = 3 \cdot 10^7$. Todas as quantidades exibidas na Fig. 6.3 foram calculadas para os GV gerados a partir das séries de prevalência na transição com tamanho $t = 10^6$. Isso porque as flutuações nessa série são mais relevantes do que para as outras fases epidêmicas. Logo, uma possível influência do tamanho da rede original poderia ser melhor notada na transição.

É possível notar que para todas essas imagens as curvas de $N = 10^7$ e $N = 3 \times 10^7$ coincidem, mesmo para $K_{nn}(k)$ em que há uma diferença perceptível para $N < 10^7$. Isso permite inferir que o GV de tamanho $t = 10^6$ é pouco sensível a diferenças no tamanho da rede original a partir de $N = 10^7$, sendo esse o valor a ser utilizado nas análises a seguir.

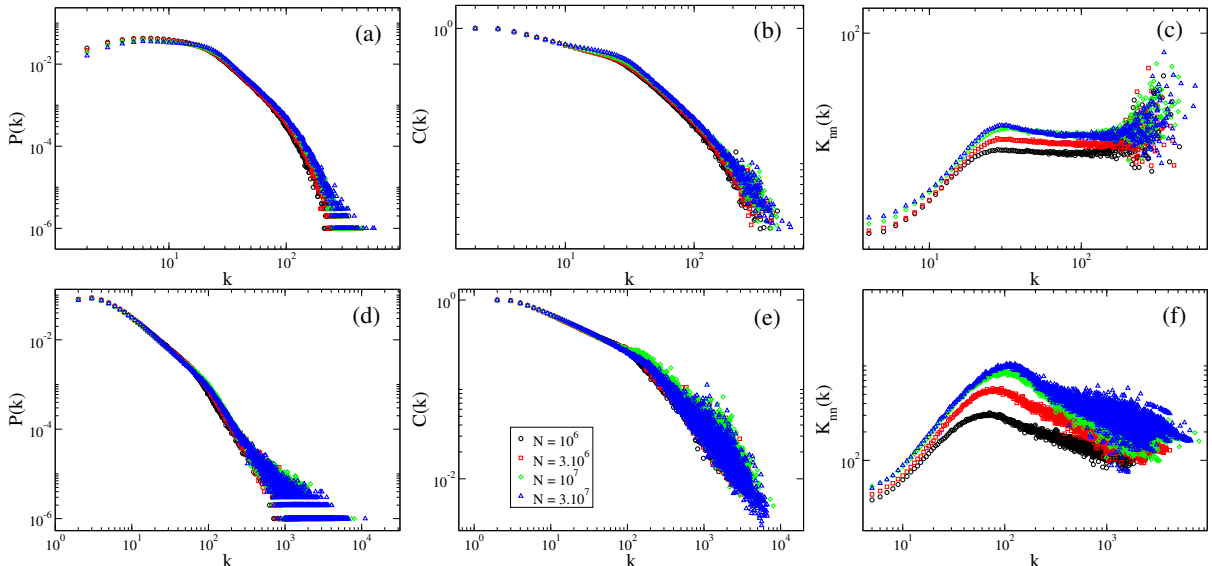


Figura 6.3 – Análise de tamanho finito para as redes originais na distribuição de grau $P(k)$, no coeficiente de agregação $C(k)$ e no grau dos vizinhos $K_{nn}(k)$ do GV. Imagens (a), (b) e (c) referem à rede UCM com $\gamma = 2.75$ e (d), (e) e (f) à RRA com $m = 6$ e 7 . Para ambas foi utilizada a série na transição com $t = 10^6$.

Uma análise de tamanho finito semelhante a anterior pode ser feita para o tamanho do GV gerado t . O estudo da influencia de t na forma do GV é de extrema relevância por razão da complexidade computacional. Temos que os algoritmos UCM e OGA possuem baixo tempo de processamento ($O(N)$) [8] de maneira que é possível utilizar tamanhos de rede original grandes. Já o algoritmo de mapeamento da série no GV, proposto a partir do critério de visibilidade dado pela Eq.(4.2), possui complexidade algorítmica $O(N^2)$. O tempo de processamento desse algoritmo para uma série de tamanho $t = 10^5$ é cerca de 30 segundos, já para $t = 10^6$ é aproximadamente 50 minutos. Mas quando utilizamos uma série de tamanho $t = 10^7$ esse tempo vai para quase 80 horas. Portanto, mesmo fazendo algumas otimizações no algoritmo, o custo computacional se torna muito alto para séries maiores que $t = 10^7$.

Porém, um resultado que ameniza essa dificuldade é mostrado na Fig. 6.4, na qual é feita a análise de tamanho finito do GV para as quantidades de interesse já mencionadas. Pode-se notar que as curvas coincidem a partir de $t = 10^5$, mostrando que o tamanho do GV não influencia no cálculo de $P(k)$, $C(k)$ e $K_{nn}(k)$ a partir desse valor. Como esse tamanho difere suavemente para a correlação de grau da RRA com $m = 6$ e 7 (Fig. 6.4(f)) utilizaremos $t = 10^6$ para as análises a seguir.

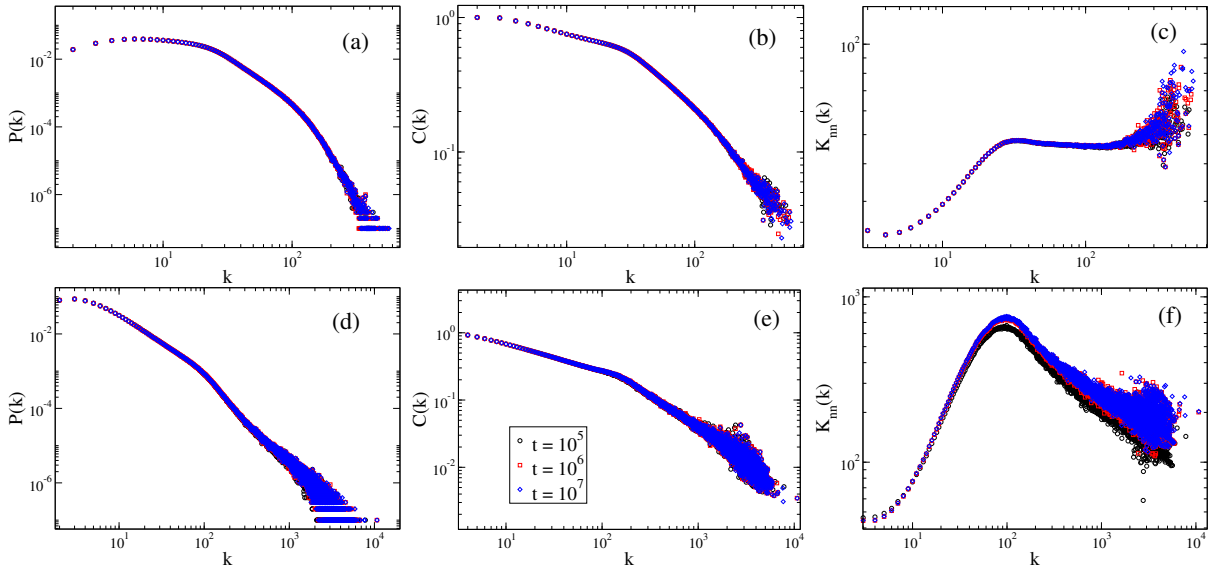


Figura 6.4 – Análise de tamanho finito para o GV gerado a partir da série na fase crítica utilizando a distribuição de grau $P(k)$, o coeficiente de agregação $C(k)$ e o grau médio dos vizinhos $K_{nn}(k)$ do GV. Imagens (a), (b) e (c) referentes à rede UCM com $\gamma = 2.75$ e (d), (e) e (f) à RRA com $m = 6$ e 7 . Para ambas, o tamanho da rede original utilizado foi $N = 10^7$.

6.3 Análise do GV para diferentes fases epidêmicas

Na seção 3.1 foi discutida a transição que ocorre no parâmetro de ordem ($\langle \rho \rangle$) ao variarmos a taxa de infecção (λ). Para cada valor de λ é possível gerar uma série temporal de prevalência utilizando o método quase estacionário (seção 5.4). Assim, para diferentes fases epidêmicas pode se escolher um valor de taxa de infecção e analisar a série temporal correspondente.

Uma das principais questões que motivaram este trabalho é a seguinte: seria o mapeamento dessas séries temporais utilizando o GV capaz de diferenciar as fases epidêmicas no modelo SIS? Para respondê-la, calculamos para cada uma das fases as quantidades de interesse já citadas anteriormente - distribuição de grau, coeficiente de agregação e grau dos vizinhos. Também calculamos para as três fases o menor caminho médio $\langle d \rangle$.

A partir do cálculo da suscetibilidade dinâmica (seção 3.1) foram escolhidos os valores de taxa de infecção λ para cada fase, mostrados na tabela 6.1. Com essas escolhas de λ estamos em regimes suficientemente longe do ponto crítico nas fases super e subcríticas permitindo diferenciá-las com mais clareza da fase crítica.

	UCM $\gamma = 2.75$	RRA $m = 6$ e 7
subcrítica	0.01	0.178
crítica	0.0239	0.1821
supercrítica	0.03	0.1828

Tabela 6.1 – Valores de taxa de infecção λ para cada fase epidêmica do modelo SIS nas redes UCM com $\gamma = 2.75$ e RRA com $m = 6$ e 7 de tamanho $N = 10^7$

a) Distribuição de grau

Na Fig. 6.5 são mostradas as distribuições de grau (seção 2.2) dos GV gerados a partir das séries quase estacionárias de ρ em cada regime epidêmico. Existe uma diferença evidente entre as curvas das fases sub e supercrítica, que possuem comportamento semelhante, e a fase crítica para as duas redes originais.

As curvas de $P(k)$ das fases sub e supercríticas apresentam decaimento mais rápido e cauda leve, indicando que existem poucos *hubs* na rede. Além disso as duas curvas são bastante semelhantes. Já a curva para a fase crítica possui decaimento mais lento e um peso maior na cauda, principalmente para a RRA mostrada na Fig. 6.5(b).

Analisar quantitativamente a distribuição de grau pode ser uma tarefa difícil, já que grande parte das redes reais possuem distribuições de grau não triviais [25]. Ou seja, em geral não se consegue classificar seu comportamento como sendo puramente uma lei de potência, por exemplo.

Uma das distribuições que geralmente ocorre em redes geradas por sistemas reais é a lei de potência com corte exponencial. Essa distribuição possui um regime com decaimento do tipo $P(k) \sim k^{-\gamma}$ mas para $k \geq k_{\text{corte}}$ a curva decai exponencialmente indi-

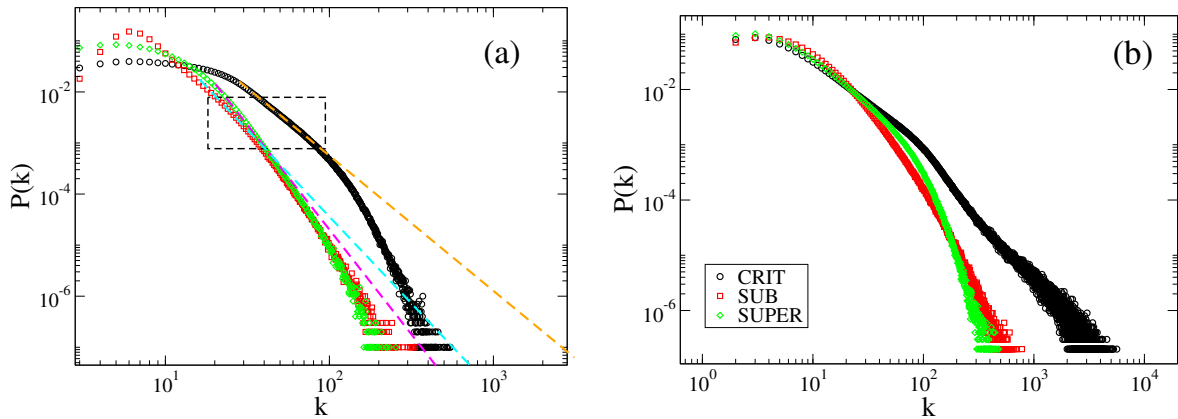


Figura 6.5 – Distribuição de grau do GV ($t = 10^6$) para as diferentes fases epidêmicas do modelo SIS em dois tipos de rede original ($N = 10^7$): (a) Rede UCM com $\gamma = 2.75$ e (b) RRA com $m = 6$ e 7 . As retas tracejadas em (a) são uma ilustração de como seriam os decaimentos das curvas caso seguissem a lei de potência que foi ajustada no retângulo destacado.

cando que existem menos nós de grau grande do que seria esperado para uma distribuição puramente em lei de potência.

Um esquema ilustrativo é mostrado na Fig. 6.5(a), na qual foram traçadas linhas a partir de uma região com possível decaimento em lei de potência. Elas servem para que se possa ver como seriam as caudas das distribuições caso não houvesse o corte exponencial. Pode-se observar também que os expoentes da lei de potência na região destacada são diferentes, sendo maiores para as fases sub e supercrítica do que para a fase crítica.

É preciso destacar que essa análise é qualitativa, servindo apenas para comentar os diferentes regimes na distribuição de grau. Para valores precisos dos expoentes γ e dos graus de corte k_{corte} são necessários o uso de técnicas mais refinadas, como o cálculo da distribuição cumulativa e ajustes mais elaborados que não estão no objetivo desta dissertação, e sim de estudos posteriores.

b) Coeficiente de agregação

Outra quantidade analisada é o coeficiente de agregação em função do grau $C(k)$ (seção 2.3). Para ambas as redes originais, UCM com $\gamma = 2.75$ e RRA com $m = 6$ e 7 mostrado na Fig. 6.6, temos a presença de modularidade hierárquica nos GV gerados, para todas as fases, isto é, as curvas seguem uma lei de potência do tipo $C(k) \sim k^{-1}$ a partir de certo grau. Logo, é possível dizer que para todas as fases, os GV gerados possuem uma estrutura em que nós de grau mais baixo estão mais conectados entre si em comunidades locais mais densas enquanto que aqueles com grau maior conectam-se mais com comunidades diferentes do que dentro daquela que pertencem.

As curvas para a fase crítica, nas duas redes originais, possuem uma mudança de regime em um dado valor de k em que passam de um decaimento mais suave em graus

baixos para $C(k) \sim k^{-1}$, sendo deslocadas em relação as curvas sub e supercrítica. É interessante notar que essa mudança de regime é muito suave nas fases sub e supercrítica, com uma rápida convergência para o decaimento hierárquico. Mas na fase crítica existe um regime inicial intermediário, entre aleatório ($C(k) \sim \text{constante}$) e hierárquico ($C(k) \sim k^{-1}$). Isso pode ser entendido por conta dos valores na série para essa fase possuírem uma visibilidade intermediária, ou seja, os pontos de altura média conseguem enxergar os baixos mas também os picos, de forma que não estarão em comunidades densas no GV e ligados a poucas comunidades diferentes.

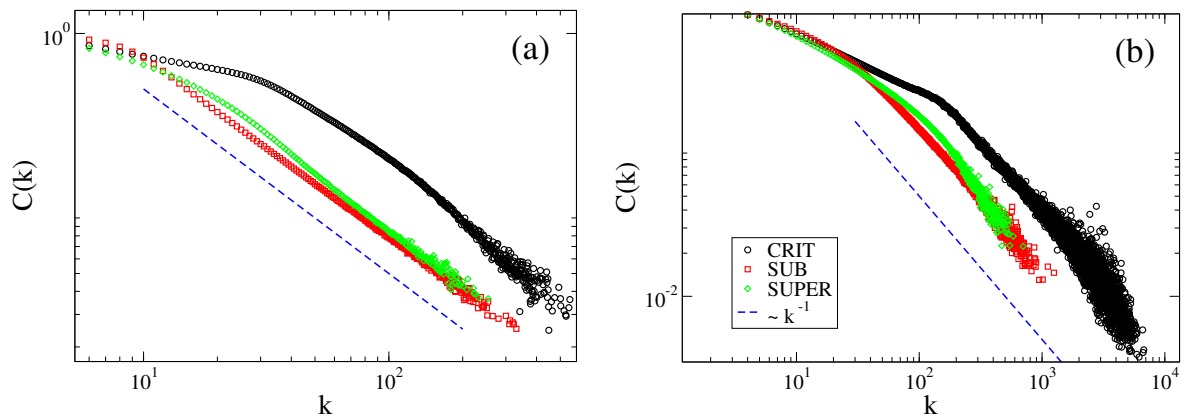


Figura 6.6 – Coeficiente de agregação do GV ($t = 10^6$) para as diferentes fases epidêmicas do modelo SIS. Curvas em preto são referentes à fase crítica, em vermelho à fase subcrítica e em verde à fase supercrítica, para dois tipos de rede original ($N = 10^7$): (a) UCM $\gamma = 2.75$ (b) RRA com $m = 6$ e 7 .

c) Correlação de grau

A diferença entre fases epidêmicas é ainda mais evidente para a correlação de grau e para essa medida existe distinção também entre as redes originais, como pode ser observado na Fig. 6.7. Para a rede UCM com $\gamma = 2.75$ as curvas das fases sub e supercríticas mostram um regime associativo, enquanto que na fase crítica a correlação começa num regime associativo e para determinado valor de k passa a um quase neutro ou fracamente correlacionado. Já para a RRA com $m = 6$ e 7 , as fases sub e supercríticas também são associativas, embora mais suavemente na fase subcrítica, e a fase crítica apresenta um padrão bem diferente, começando em um regime associativo que a partir de um certo grau, passa a ser desassociativo. Esse comportamento será abordado novamente na seção 6.4.

Para todas as quantidade calculadas do GV na fase crítica para a RRA com $m = 6$ e 7 a presença dos *hubs* é mais evidenciada, com curvas de caudas mais pesadas. Enquanto que para as redes UCM com $\gamma = 2.75$, por mais que ainda seja possível identificar os *hubs*, eles ocorrem em menor quantidade e possuem grau menor. Isso será mais discutido na seção 6.4.

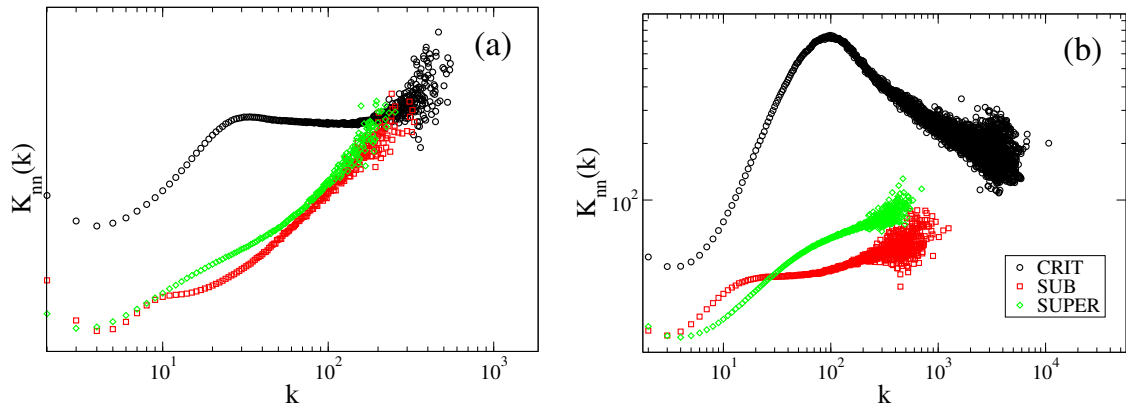


Figura 6.7 – Grau médio dos vizinhos no GV ($t = 10^6$) para as diferentes fases epidêmicas do modelo SIS. Curvas em preto são referentes à fase crítica, em vermelho à fase subcrítica e em verde à fase supercrítica, para dois tipos de rede original de tamanho $N = 10^7$: (a) UCM $\gamma = 2.75$ (b) RRA com $m = 6$ e 7 .

Outra análise que se pode fazer em redes complexas é a respeito das distâncias, como foi discutido na seção 2.1. O menor caminho médio dos GV gerados para as três fases epidêmicas é mostrado na Fig. 6.8 para as duas redes originais. Em ambas é possível notar que a dependência de $\langle d \rangle$ com o tamanho t do GV segue curvas que se encontram entre os regimes $\langle d \rangle \sim \ln(t)$ e $\langle d \rangle \sim \ln(\ln(t))$. Ou seja, o comportamento do menor caminho médio dos GV gerados está entre aquele esperado para redes livres de escala (efeito de mundo ultra-pequeno) e redes Erdős-Renyi (efeito de mundo pequeno). Ainda que o comportamento das curvas seja mais próximo de $\langle d \rangle \sim \ln(t)$, essa evidência concorda qualitativamente com a discussão feita para distribuição de grau do GV, que são do tipo lei de potência com corte exponencial.

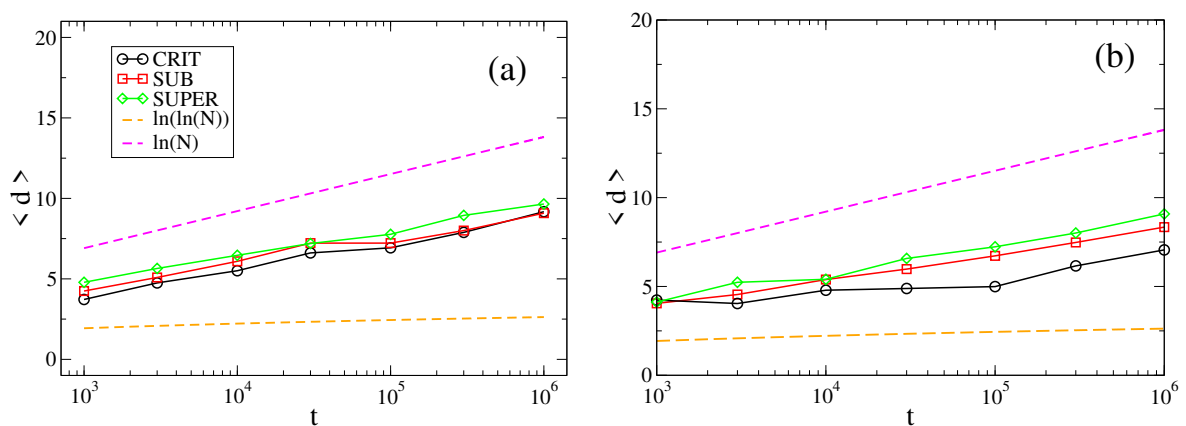


Figura 6.8 – Menor caminho médio em função do tamanho do GV gerados a partir das fases subcrítica, crítica e supercrítica. Redes originais (a) UCM com $\gamma = 2.75$ e (b) RRA com $m = 6$ e 7 , para $N = 10^7$.

6.4 GV para redes com diferentes mecanismos de ativação

Como já foi abordado na seção 3.3, existem diferentes mecanismos de ativação em processos de disseminação que dependem da estrutura da rede. A fim de investigar se o GV é sensível à diferentes mecanismos de ativação da rede original, fizemos as comparações que serão discutidas abaixo, utilizando sempre a fase crítica do modelo SIS.

a) Comparação entre redes homogêneas:

Redes homogêneas são não localizadas, ou seja, em um processo de disseminação acontecendo nessas redes, cada parte delas terá a mesma importância no cálculo de $\langle \rho \rangle$ e não apenas estruturas específicas. Logo a ativação ou não da epidemia depende apenas da taxa de infecção, ou seja, para $\lambda > \lambda_c$ a rede será ativada coletivamente. Redes quadradas e RRAs são exemplos de redes homogêneas já discutidas neste trabalho.

Na Fig. 6.9, temos uma comparação entre RRAs com diferentes valores de grau m na fase crítica. Podemos ver em (a) que a forma das curvas é semelhante, com grandes distâncias entre os picos, gerando visibilidades estendidas. Isso culminará no aparecimento de *hubs* no GV. Assim, vemos que não é possível distinguir entre as quantidades analisadas para os GV gerados em redes originais homogêneas.

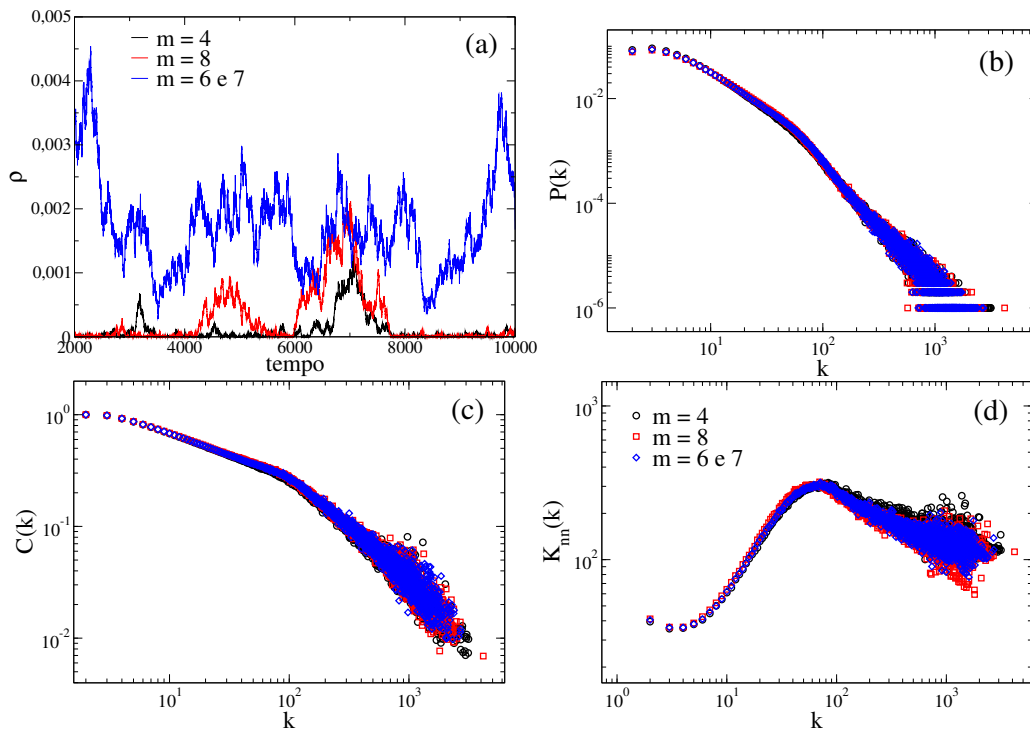


Figura 6.9 – Comparação do GV para a fase crítica do modelo SIS para três redes originais homogêneas, porém com grau m diferentes. (a) Série temporal. (b) Distribuição de grau. (c) Coeficiente de agregação. (d) Grau médio dos vizinhos.

b) Comparação entre uma rede heterogênea e outra homogênea

Podemos também comparar uma rede homogênea (RRA com $m = 6$ e 7) com outra heterogênea (UCM com $\gamma = 2.75$). Esse último tipo de rede possui localização, ou seja, um processo epidêmico ocorrendo nela irá ser sustentado por alguma estrutura específica, não necessariamente na rede toda. No caso da rede UCM com $\gamma = 2.75$ essas estruturas são os *hubs* [62]. Assim, existe um mecanismo de ativação responsável pela manutenção da epidemia. Observando a Fig. 6.10(a), vemos que a diferença entre as redes originais já se manifesta nas séries de prevalência do modelo SIS, tendo a rede homogênea uma amplitude maior e frequência de oscilações menor em relação a série da rede heterogênea.

Essa diferença é evidenciada nas três quantidades calculadas, ficando muito clara no $K_{nn}(k)$. Na Fig. 6.10(d) pode-se observar que o padrão associativo ocorre até certo valor de k , ou seja, os valores com pequena visibilidade enxergam outros com essa mesma característica e da mesma forma pra valores grandes. Porém, ao passar de um certo tamanho, os picos começam a enxergar com maior frequência os valores de menor visibilidade, explicando porque a curva passa a ser desassociativa.

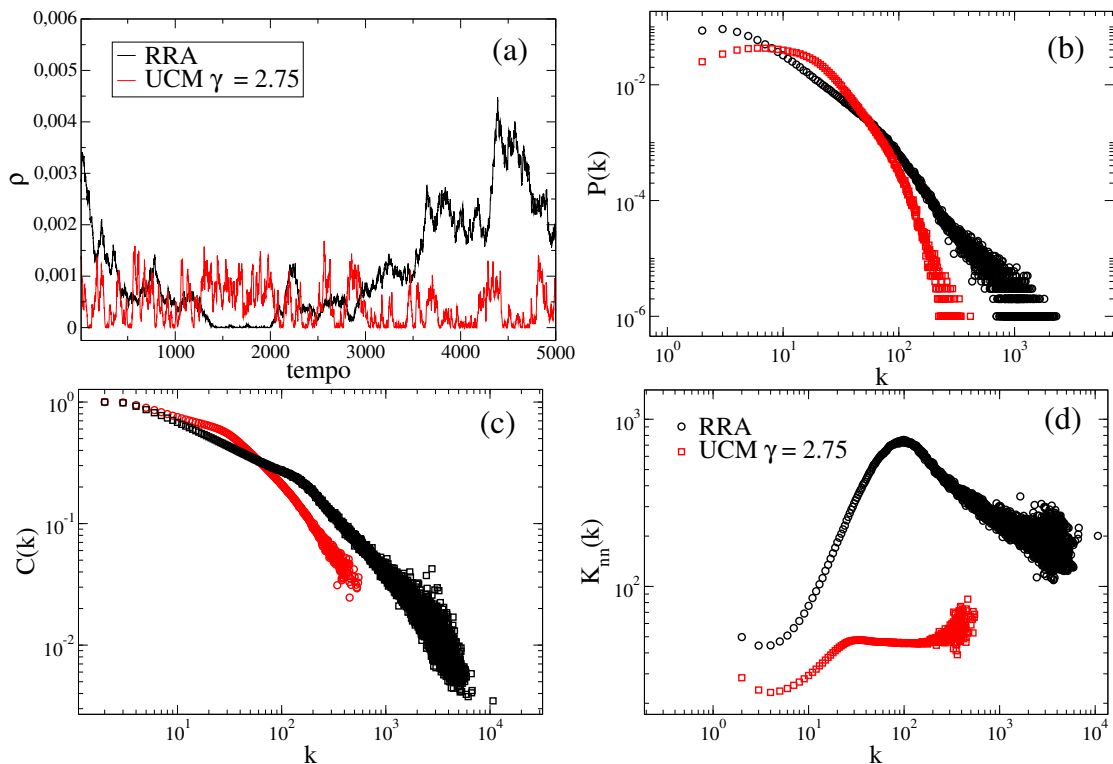


Figura 6.10 – Comparação do GV para a fase crítica do modelo SIS para dois tipos de redes originais: uma delas homogênea (RRA com $m = 6$ e 7) e outra heterogênea (UCM com $\gamma = 2.75$), ambas com mesmo $\langle k \rangle = 6.5180$ e $N = 10^7$. (a) Série temporal. (b) Distribuição de grau. (c) Coeficiente de agregação. (d) Grau médio dos vizinhos.

c) Comparação entre duas redes heterogêneas

Agora será feita a comparação das propriedades do GV para duas redes originais geradas pelo modelo UCM, uma com $\gamma = 2.25$ e outra com $\gamma = 2.75$, ambas com parâmetros $k_0 = 3$, $k_c = 2\sqrt{N}$ e $N = 10^7$. No caso da comparação entre essas duas redes heterogêneas, ambas possuem localização, porém o mecanismo de ativação de cada uma delas é diferente: para redes UCM com $\gamma < 5/2$ a ativação se dá pelo máximo k -core e para $\gamma > 5/2$, pelos *hubs*. É possível notar na Fig. 6.11(a) que as formas das séries são ligeiramente diferentes, tanto na amplitude quando na frequência. Porém, a distribuição de grau e o coeficiente de agregação dos GV gerados não captam tão bem essa diferença quanto o grau médio da vizinhança.

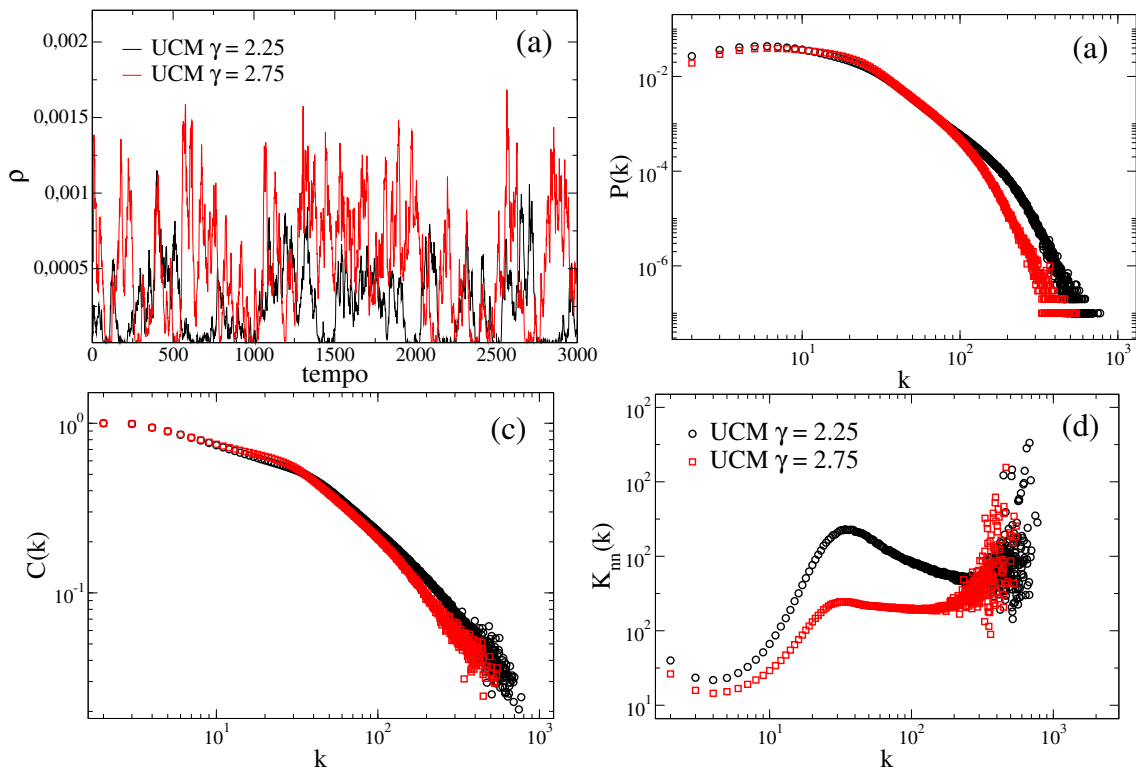


Figura 6.11 – Comparação do GV para a fase crítica do modelo SIS para dois tipos de redes: UCM com $\gamma = 2.25$ e $\gamma = 2.75$. (a) Série temporal. (b) Distribuição de grau. (c) Coeficiente de agregação. (d) Grau médio dos vizinhos.

Ao olharmos especificamente para a Fig. 6.11(d) vemos que a correlação de grau exibe um comportamento bastante diferente para as duas redes originais. Sendo a curva para a rede UCM com $\gamma = 2.25$ mais semelhante àquela encontrada para a RRA, mostrada na Fig. 6.10(d).

Esse comportamento é um ponto interessante, já que indica que a correlação de grau do GV pode ser sensível à diferença nos mecanismos de ativação das redes originais. O motivo é que, por mais que a rede UCM com $\gamma = 2.25$ possua uma estrutura responsável pela sustentação da epidemia, o máximo k -core, ela é menos localizada que a rede UCM

com $\gamma = 2.75$, em que os *hubs* estão mais isolados e a ativação não é simultânea. O nós que compõem o máximo *k-core* são *hubs* densamente conectados entre si mas que também se conectam com muitos nós que não pertencem a esse subgrafo. Assim, quando esses nós são ativados por meio de interações entre eles, vários outros nós são ativados de modo síncrono. Por outro lado, quando os *hubs* são separados, a ativação ocorre localmente em torno deles e seus vizinhos e as interações ocorrem somente por flutuações que não são sincronizadas. Portanto, o caso $\gamma = 2.25$ tem uma atividade muito mais homogênea que o segundo. Isso explica a maior similaridade com o caso homogêneo, apesar de ser mais parecida estruturalmente com o caso $\gamma = 2.75$, já que ambos são redes livres de escala.

Outra observação interessante é que as mudanças de regime que ocorrem em todas as quantidades calculadas na fase crítica acontecem para aproximadamente o mesmo valor de grau, ou seja, o decaimento característico de modularidade hierárquica no coeficiente de agregação e também a mudança no comportamento associativo da correlação de grau para a neutralidade na RRA e desassociatividade na rede UCM com $\gamma = 2.75$. Além disso, o comportamento encontrado para o menor caminho médio em função do tamanho do GV está qualitativamente de acordo com as reflexões feitas a respeito da distribuição de grau. Como foi discutido no capítulo IV, séries fractais mapeiam em GV com distribuição em lei de potência e séries aleatórias em distribuições com decaimento exponencial. Como obtivemos uma espécie de junção dos dois comportamentos para as distribuições de grau das três fases epidêmicas, poderíamos inferir que as séries de prevalência epidêmica estão entre os dois comportamentos.

Portanto, ainda que seja necessário uma análise quantitativa mais rigorosa, há fortes indícios de que o GV é sensível às diferenças nas séries de fases epidêmicas diferentes. E que a correlação de grau pode ser uma boa forma de caracterizar as séries provenientes de redes com mecanismo de ativação diferentes. Essa última evidência é de particular interesse na investigação de processos de disseminação pois, em muitos casos, os dados do sistema de estudo estão disponíveis em forma de série temporal, já que informações sobre a rede original que suporta o processo são escassas. Assim, o fato de se poder obter características dessas redes utilizando a análise das séries temporais via GV pode ser uma ferramenta muito relevante no estudo desses sistemas.

VII Considerações finais

Séries temporais podem ser extraídas de diversos campos de pesquisa: em estudos demográficos pode-se estar interessado em como uma determinada população aumenta ou diminui devido à taxas de natalidade e mortalidade. Em medicina, medidas de pressão arterial podem ser feitas durante certo período de tempo para avaliar a eficácia de medicamentos para controle de hipertensão. Nos estudos atmosféricos pode ser interessante medir a temperatura da superfície do mar a cada dia em vários pontos do oceano, a fim de entender melhor fenômenos como *El Niño*. Na área de neurociência é possível obter dados de atividade elétrica de diferentes setores do cérebro, através de técnicas de imageamento como a ressonância magnética. E em epidemiologia, pode se medir a fração do número de indivíduos capazes de transmitir uma doença em uma população em função do tempo, entre muitos exemplos. De maneira geral, dados reais ordenados em séries temporais estão disponíveis para auxiliar na compreensão dos mais diversos fenômenos [47].

Uma das maneiras de se analisar séries temporais é fazer um mapeamento dos dados em um grafo ou rede. Atualmente, a área da ciência de redes complexas vem ganhando importância pela robustez de seus métodos de análise. Devido à natureza interdisciplinar da área, é possível visualizar um problema por diferentes ângulos, por exemplo, fazendo simulações numéricas ou aplicando conceitos já muito consolidados em Física Estatística. Assim, pode ser vantajoso, em vez de se analisar uma série temporal, fazer a análise da rede complexa gerada a partir dela [1].

Neste trabalho as séries temporais escolhidas foram as de prevalência epidêmica, que são valores da fração de indivíduos infectados em uma população em intervalos de tempo [36]. Para gerar essas séries utilizamos o modelo SIS para processos de disseminação, no qual os indivíduos são divididos em compartimentos de acordo com seu estado em relação a doença. Esses estados podem ser: suscetível (S) ou infectado (I). A doença se propaga na estrutura de rede segundo taxas de infecção e cura preestabelecidas. A evolução do número de indivíduos infectados pode ser representada por meio de métodos analíticos ou simulada numericamente. Neste trabalho utilizamos o algoritmo OGA [8] para simular o modelo SIS em diferentes redes complexas.

Assim, fizemos o mapeamento das séries de prevalência epidêmica em grafos de visibilidade [12] e analisamos as redes geradas. Para isso utilizamos quantidade conhecidas em análise de redes complexas, como a distribuição de grau, o coeficiente de agregação, a correlação de grau e o menor caminho médio em função do tamanho da rede [25, 11]. Utilizamos diferentes redes originais em que a propagação da infecção ocorre: rede quadrada, RRA com diferentes valores de grau e redes com distribuição em lei de potência geradas pelo modelo UCM [28].

Sabe-se que em um processo de disseminação uma transição para o estado absorvente pode ocorrer ao variarmos o parâmetro de controle (taxa de infecção λ), no limite termodinâmico. Logo, para diferentes valores de λ teremos fases epidêmicas distintas. Através de métodos quase estacionários [61, 24] é possível gerar as séries de prevalência epidêmica para cada fase e assim analisar os GVs gerados por elas. Neste trabalho observamos evidências de que o GV consegue diferir os regimes epidêmicos por meio das quantidades acima mencionadas, com uma assinatura distinta para a fase crítica.

Outra análise que fizemos foi utilizar os GV gerados apenas a partir da fase crítica, para redes originais diferentes em que ocorre o processo de disseminação. Foi observado que praticamente não há distinção na distribuição de grau, padrões de correlação de grau e coeficiente de agregação na comparação entre redes homogêneas de graus diferentes. Ao contrário do que ocorre na comparação entre uma rede homogênea e outra heterogênea: todas as quantidades calculadas captam as diferenças entre as redes originais. Para a comparação entre duas redes originais heterogêneas, ambas com distribuição de grau em lei de potência geradas pelo modelo UCM, apenas diferindo no expoente γ , temos um comportamento muito interessante. Essa diferença altera uma característica importante do processo de disseminação que são os mecanismos de ativação [43]. Observamos que a rede UCM com $\gamma = 2.25$ tinha correlação de grau mais semelhante ao encontrado para a RRA com $m = 6$ e 7 do que para uma rede UCM com $\gamma = 2.75$. Uma possível interpretação desse resultado está na análise dos mecanismos de ativação para as duas redes, que são diferentes. Para rede UCM com $\gamma = 2.25$ a sustentação da epidemia é dominada pelo máximo *k-core*, uma estrutura que engloba uma parte dos nós da rede, inclusive os *hubs*, fazendo com que a rede seja menos localizada e os surtos da epidemia aconteçam de forma mais coletiva, uma situação semelhante a que ocorre em uma rede homogênea. Já para a rede UCM com $\gamma = 2.75$ os *hubs* estão mais distantes entre si e a infecção mútua entre eles é muito menor, gerando regiões de atividade muito mais localizada na rede [44, 45]. Assim, vemos que a correlação de grau do GV é uma quantidade sensível a diferença entre redes originais, pelo menos para aquelas em que o mecanismo de ativação é diferente, ainda que ambas sejam redes livres de escala.

O estudo de processo de disseminação sempre foi de grande interesse, por ocorrer em diversos contextos, seja na propagação de informações em redes sociais [4], no estudo de vírus que atacam computadores ou celulares [13, 14] ou na disseminação de doenças infecciosas [34, 63, 64]. Esse último ponto ganhou destaque no ano de 2020 com a pandemia de COVID-19, uma doença respiratória infecciosa transmitida pelo vírus SARS-CoV-2. Sendo uma das piores pandemias que o mundo já vivenciou, a COVID-19 matou mais de um milhão de pessoas em um período de aproximadamente dez meses. Isso revela a grande importância de se buscar compreender os processos de disseminação, a fim de entender melhor o avanço de uma epidemia [16] e elaborar estratégias de imunização da população que sejam mais eficientes [19].

Hoje, sabe-se que a estrutura da rede influencia em um processo de disseminação

ocorrendo nela [38, 17]. Assim, conhecer características das redes originais seria de grande ajuda na compreensão da propagação de uma epidemia. Essa tarefa porém pode ser muito difícil ou até inviável para redes reais, devido a quantidade de variáveis presentes nessas situações. O que se tem muitas vezes são séries temporais de dados, por exemplo o número de indivíduos infectados, por isso a escolha de aplicação deste trabalho. Através das séries de prevalência epidêmica buscamos características da rede original em que o processo de disseminação ocorre utilizando a análise dos grafos de visibilidade gerados.

Para conclusões mais definitivas do ponto de vista quantitativo, algumas análises complementares ainda são necessárias. Como por exemplo, a simulação de outros modelos epidêmicos para obtenção das séries e o uso de outras redes homogêneas e heterogêneas para verificar se elas ainda diferem quanto a correlação de grau do GV. Pretendemos também pesquisar quantidades relacionadas as séries em si e confrontá-las com alguns resultados do GV. Entretanto, é correto afirmar que as evidências obtidas nesse trabalho mostram que o GV pode ser um grande aliado na análise de séries temporais, ajudando a caracterizar os fenômenos descritos por elas.

Referências

- [1] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, “Complex network approaches to nonlinear time series analysis,” tech. rep., 2019.
- [2] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, “Systemic delay propagation in the US airport network,” *Scientific Reports*, vol. 3, 2013.
- [3] R. Albert, I. Albert, and G. L. Nakarado, “Structural vulnerability of the North American power grid,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 2 2, pp. 1–4, 2004.
- [4] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, “Quantifying echo chamber effects in information spreading over political communication networks,” *EPJ Data Science*, vol. 8, no. 1, 2019.
- [5] A. Fornito, A. Zalesky, and E. T. Bullmore, *Fundamentals of Brain Network Analysis*. Academic Press, 2016.
- [6] H. Jeong, B. Tombort, R. Albert, Z. N. Oltvait, and A. L. Barabasi, “The large-scale organization of metabolic networks,” *The Structure and Dynamics of Networks*, vol. 9781400841, no. 1990, pp. 211–214, 2011.
- [7] M. Boguñá, L. F. Lafuerza, R. Toral, and M. Á. Serrano, “Simulating non-Markovian stochastic processes,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 90, no. 4, pp. 1–9, 2014.
- [8] W. Cota and S. C. Ferreira, “Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks,” *Computer Physics Communications*, vol. 219, pp. 303–312, 2017.
- [9] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, “Percolation on correlated networks,” *Physical Review E*, vol. 78, p. 051105, nov 2008.
- [10] E. Bradley and H. Kantz, “Nonlinear time-series analysis revisited,” *Chaos*, vol. 25, no. 9, 2015.
- [11] M. Newman, *Networks - An introduction*. OUP Oxford, 2010.
- [12] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, “From time series to complex networks: The visibility graph,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 13, pp. 4972–4975, 2008.

-
- [13] T. Brett, G. Loukas, Y. Moreno, and N. Perra, “Spreading of computer viruses on time-varying networks,” *Physical Review E*, vol. 99, no. 5, pp. 46–48, 2019.
- [14] P. Wang, M. C. González, C. A. Hidalgo, and A. L. Barabasi, “Understanding the Spreading Patterns of Mobile Phone Viruses,” *Science*, vol. 324, no. 5930, pp. 1071–1076, 2009.
- [15] J. Leskovec, L. A. Adamic, and B. A. Huberman, “The dynamics of viral marketing,” *ACM Transactions on the Web*, vol. 1, no. 1, 2007.
- [16] G. S. Costa, W. Cota, and S. C. Ferreira, “Metapopulation modeling of COVID-19 advancing into the countryside: an analysis of mitigation strategies for Brazil,” *medRxiv*, 2020.
- [17] R. Pastor-Satorras and A. Vespignani, “Epidemic Spreading in Scale-Free Networks,” *Physical Review Letters*, vol. 86, pp. 3200–3203, apr 2001.
- [18] R. Pastor-Satorras and A. Vespignani, “Immunization of complex networks,” *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 65, no. 3, pp. 1–8, 2002.
- [19] G. S. Costa and S. C. Ferreira, “Nonmassive immunization to contain spreading on complex networks,” *Physical Review E*, vol. 101, no. 2, pp. 1–10, 2020.
- [20] H. K. Lee, P. S. Shim, and J. D. Noh, “Epidemic threshold of the susceptible-infected-susceptible model on complex networks,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 87, no. 6, pp. 1–5, 2013.
- [21] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, “Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 86, no. 4, pp. 1–8, 2012.
- [22] W. Cota, A. S. Mata, and S. C. Ferreira, “Robustness and fragility of the susceptible-infected-susceptible epidemic models on complex networks,” *Physical Review E*, vol. 98, no. 1, pp. 1–12, 2018.
- [23] A. S. Mata and S. C. Ferreira, “Multiple transitions of the susceptible-infected-susceptible epidemic model on complex networks,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 91, no. 1, pp. 1–9, 2015.
- [24] R. S. Sander, G. S. Costa, and S. C. Ferreira, “Sampling methods for the quasisationary regime of epidemic processes on regular and complex networks,” *Physical Review E*, vol. 94, no. 4, pp. 1–11, 2016.
- [25] A.-L. Barabási, *Network science*. Cambridge University Press, 2016.

-
- [26] P. Erdos and A. Rényi, “On Random Graphs I,” *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.
- [27] A.-L. Barabási and R. Albert, “Emergence of Scaling in Random Networks,” *Science*, vol. 286, pp. 509–512, oct 1999.
- [28] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, “Generation of uncorrelated random scale-free networks,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 2, pp. 1–4, 2005.
- [29] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [30] M. Latapy, “Main-memory triangle computations for very large (sparse (power-law)) graphs,” *Theoretical Computer Science*, vol. 407, pp. 458–473, nov 2008.
- [31] E. Ravasz and A.-L. Barabási, “Hierarchical organization in complex networks,” *Physical Review E*, vol. 67, no. 2, p. 026112, 2003.
- [32] A. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” *Physical Review E*, vol. 70, p. 066111, dec 2004.
- [33] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, “Dynamical and correlation properties of the internet,” *Physical Review Letters*, vol. 87, no. 25, pp. 258701–1–258701–4, 2001.
- [34] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.
- [35] R. Pastor-Satorras and A. Vespignani, “Epidemic dynamics and endemic states in complex networks,” *Physical Review E*, vol. 63, p. 066117, may 2001.
- [36] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of Modern Physics*, vol. 87, no. 3, pp. 925–979, 2015.
- [37] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*. Cambridge University Press, 2005.
- [38] M. Boguñá and R. Pastor-Satorras, “Epidemic spreading in correlated complex networks,” *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 66, no. 4, p. 4, 2002.
- [39] S. Chatterjee and R. Durrett, “Contact processes on random graphs with power law degree distributions have critical value 0,” *The Annals of Probability*, vol. 37, pp. 2332–2356, nov 2009.

- [40] C. Castellano and R. Pastor-Satorras, “Thresholds for epidemic spreading in networks,” *Physical Review Letters*, vol. 105, no. 21, pp. 1–4, 2010.
- [41] F. Chung, L. Lu, and V. Vu, “The spectra of random graphs with given expected degrees,” *Internet Mathematics*, vol. 1, no. 3, pp. 257–275, 2004.
- [42] R. Dickman, T. Tomé, and M. J. de Oliveira, “Sandpiles with height restrictions,” *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 66, no. 1, pp. 1–8, 2002.
- [43] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras, “Collective versus hub activation of epidemic phases on networks,” *Physical Review E*, vol. 93, no. 3, pp. 1–10, 2016.
- [44] M. Boguñá, C. Castellano, and R. Pastor-Satorras, “Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks,” *Physical Review Letters*, vol. 111, no. 6, pp. 1–5, 2013.
- [45] C. Castellano and R. Pastor-Satorras, “Competing activation mechanisms in epidemics on networks,” *Scientific Reports*, vol. 2, 2012.
- [46] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, “Identification of influential spreaders in complex networks,” *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.
- [47] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*. Springer, 2011.
- [48] T. S. Holger Kantz, *Nonlinear Time Series Analysis*. Cambridge University Press, 2004.
- [49] J. Gao, Y. Cao, W.-w. Tung, and J. Hu, *Multiscale analysis of complex time series*. Wiley-Interscience, 2007.
- [50] R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths, “Recurrence-based time series analysis by means of complex network methods,” *International Journal of Bifurcation and Chaos*, vol. 21, no. 4, pp. 1019–1046, 2011.
- [51] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, “Geometry from a Time Series,” *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.
- [52] F. Takens, “Detecting strange attractors in turbulence,” in *Lecture notes in mathematics*, pp. 366–381, Springer, 1981.
- [53] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, “Determining Lyapunov exponents from a time series,” *Physica D: Nonlinear Phenomena*, vol. 16, pp. 285–317, jul 1985.

-
- [54] P. Grassberger and I. Procaccia, “Characterization of Strange Attractors,” *Physical Review Letters*, vol. 50, pp. 346–349, jan 1983.
- [55] T. Lozano-prez and M. A. Wesley, “An Algorithm for Planning Collision-Free Paths Among Polyhedral Obstacles,” vol. 22, no. 10, 1979.
- [56] G. Nagy, “Terrain visibility,” *Computers & Graphics*, vol. 18, no. 6, pp. 763–773, 1994.
- [57] T. Tomé and M. J. de Oliveira, *Stochastic Dynamics and Irreversibility*. Springer, 2014.
- [58] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [59] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, “Cut-offs and finite size effects in scale-free networks,” *European Physical Journal B*, vol. 38, no. 2, pp. 205–209, 2004.
- [60] D. T. Gillespie, “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions,” *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, 1976.
- [61] M. M. De Oliveira and R. Dickman, “How to simulate the quasistationary state,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 1, pp. 1–5, 2005.
- [62] D. H. Silva and S. C. Ferreira, “Activation thresholds in epidemic spreading with motile infectious agents on scale-free networks,” *Chaos*, vol. 28, no. 12, pp. 1–8, 2018.
- [63] C. T. Bauch, J. O. Lloyd-Smith, M. P. Coffee, and A. P. Galvani, “Dynamically modeling SARS and other newly emerging respiratory illnesses: Past, present, and future,” *Epidemiology*, vol. 16, no. 6, pp. 791–801, 2005.
- [64] V. Colizza, A. Barrat, M. Barthelemy, A. J. Valleron, and A. Vespignani, “Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions,” *PLoS Medicine*, vol. 4, no. 1, pp. 0095–0110, 2007.