

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/313672385>

Métodos de simulação de processos epidêmicos Markovianos em redes complexas

Research · February 2017

DOI: 10.13140/RG.2.2.19340.72325

CITATIONS

0

READS

434

1 author:



Wesley Cota

Universidade Federal de Viçosa (UFV)

18 PUBLICATIONS 84 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Investigation of the effects of network disorder on various dynamical systems [View project](#)



Simulation of Markovian epidemic processes on complex networks [View project](#)



UNIVERSIDADE FEDERAL DE VIÇOSA
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
DEPARTAMENTO DE FÍSICA

Métodos de simulação de processos epidêmicos Markovianos em redes complexas

Wesley Cota

Prof. Silvio C. Ferreira (Orientador)

Monografia apresentada ao Departamento de Física da Universidade Federal de Viçosa como parte das exigências para a Conclusão da Disciplina *Monografia e Seminários* (FIS 497).

VIÇOSA
MINAS GERAIS – BRASIL
DEZEMBRO DE 2016

À vida, ao Universo e a tudo mais.

Agradecimentos

À Universidade Federal de Viçosa, e aos seus funcionários. Ao Grupo de Investigação de Sistemas Complexos (GISC) e ao Núcleo de Simulação de Sistemas Complexos (NSSC) pelo suporte computacional.

À CAPES, ao CNPq e à FAPEMIG pelo apoio financeiro nos projetos de pesquisa, nas bolsas do Programa Institucional de Bolsas de Iniciação Científica (PIBIC), e pelos auxílios para participação em eventos e visitas científicas.

Também ao Romualdo Pastor-Satorras, pelas discussões durante suas visitas à UFV pelo programa Ciência sem Fronteiras/CAPES pelo projeto nº 88881.030375/2013-01.

Índice

Resumo	1
Abstract	2
I Introdução	3
II Grafos e complexidade	6
2.1 Redes complicadas ou complexas?	6
2.2 Propriedades e medidas	7
2.2.1 Ligações e matriz de adjacência	7
2.2.2 Coeficiente de agregação	8
2.2.3 Distâncias e caminhos	9
2.2.4 Estudo estatístico e distribuições de conectividade	10
2.2.5 Correlações estruturais	12
2.3 Redes reais: propriedades e modelos	14
III Processos dinâmicos	16
3.1 Processos Markovianos e equações mestras	16
3.2 Processos de Poisson	18
3.3 Modelos epidêmicos clássicos	19
3.3.1 Modelo PC - Processo de Contato	19
3.3.2 Modelo SIS - Suscetível-Infectado-Suscetível	20
3.3.3 Modelo SIR - Suscetível-Infectado-Recuperado	20
3.4 Criticalidade e transições de fase	21
3.4.1 Universalidade e expoentes críticos	22
3.5 Estados e distribuições quase-estacionárias (QS)	26
IV Simulação de processos epidêmicos	28
4.1 Construção de redes complexas	28
4.1.1 Outras redes utilizadas neste trabalho	30
4.2 Algoritmo de Gillespie	30
4.3 Simulação do estado quase-estacionário	32

ÍNDICE	iv
4.4 Tipos de análises e medidas	33
4.4.1 Análise de espalhamento	33
4.4.2 Análise de decaimento	33
4.4.3 Análise quase-estacionária	34
V Otimização de algoritmos e aplicações	36
5.1 Algoritmos gerais	36
5.1.1 Algoritmo padrão	36
5.1.2 Algoritmo otimizado	37
5.2 Aplicação no modelo SIS	39
5.3 Aplicação no Processo de Contato	41
5.4 Aplicação no modelo SIR	42
VI Comparações numéricas e resultados	43
6.1 Efeito do gerador de números aleatórios	43
6.2 Aplicação dos algoritmos para diferentes análises	44
6.2.1 Modelo SIS	45
6.2.2 Processo de contato	47
6.2.3 Modelo SIR	48
VII Considerações finais	50
7.1 Outros trabalhos desenvolvidos e perspectivas	51
7.2 Disponibilização de códigos e artigo	51
Referências Bibliográficas	52

Resumo

O mundo atual é formado por diversos sistemas em rede: a infraestrutura de transporte, as conexões entre neurônios no cérebro, as relações em teias alimentares, e até mesmo as citações de artigos científicos formam redes. Com o surgimento da *Internet* e das redes sociais, como o *Facebook* e o *Twitter*, ficou ainda mais evidente que o mundo atual é altamente conectado, com a possibilidade de fenômenos inimagináveis até pouco tempo, como a disponibilidade de conhecimento quase ilimitado a apenas um clique. Se por um lado a interconectividade foi positiva, por outro permitiu surtos epidêmicos violentos e quase incontroláveis, como da Gripe Suína e do vírus Ebola nos primeiros anos do século XXI. Neste contexto, a ciência de redes complexas aglutina áreas do conhecimento distintas, como as ciências biológicas, as engenharias, a ciência da computação e as ciências sociais, para buscar compreender como essas redes são estruturadas e a relevância disso para processos dinâmicos como de difusão de informação ou de doenças. A Física também faz parte de tais estudos, sendo responsável por uma maneira particular de ver e de modelar o mundo. Em especial, a Física Estatística contribui com ideias de suas teorias de transições de fase e de outros fenômenos físicos. Para lidar com tantos dados e informações presentes nas redes reais, análises numéricas e simulações computacionais são necessárias e contam como importantes aliadas no entendimento de processos dinâmicos em redes complexas. Este trabalho tem como objetivo fazer uma revisão sobre topologias de redes e processos de propagação de epidemias, e também propor algoritmos de simulação eficientes para contribuir para o estudo de fenômenos ubíquos no mundo contemporâneo.

Abstract

The contemporary world is made up of several networked systems: the transport infrastructure, the connections between neurons in the brain, the relationships in food webs, and even the citations of scientific papers form networks. With the advent of the Internet and social networks, e.g. Facebook and Twitter, it has become even more evident that the current world is highly connected, with the possibility of phenomena which were unthinkable until recently, as the availability of almost unlimited knowledge just a click away. On the one hand the interconnectivity was positive, but on the other hand it has allowed violent and almost uncontrolled epidemic outbreaks, such as the Swine influenza and Ebola in the early twenty-first century. In this context, the science of complex networks has joined different fields of knowledge, such as the biological sciences, engineering, computer science and the social sciences, to try to understand how these networks are structured and their relevance for the dynamical processes of information or disease spreading occurring on them. Physics is also part of this study, being responsible for a particular way of seeing and modeling the world. In particular, the Statistical Physics contributes with ideas from phase transitions and other physical phenomena. To deal with such a large amount of data and information coming from real networks, numerical analysis and computer simulations are required and are important partners in the understanding of dynamical processes on complex networks. This work aims at making a review of network topologies and epidemics spreading processes, and also to propose efficient simulation algorithms to contribute to the study of phenomena ubiquitous in the contemporary world.

Capítulo I

Introdução

Diversos sistemas podem ser representados por meio de redes, que consistem em elementos conectados entre si estabelecendo algum tipo de relação. Com o crescimento do poder computacional, temos à nossa disposição uma grande quantidade de dados, possuindo um grande impacto no estudo e na caracterização de sistemas de interação entre proteínas, ou de neurônios no cérebro, ou de pessoas na *Internet*.

Os meios de comunicação, de interações sociais, de transporte, entre outros, estão todos envolvidos em fenômenos muito importantes, tais como a propagação de informação, de epidemias e crises financeiras, afetando todo o mundo. Todos esses sistemas possuem relações que não são nada triviais de serem analisadas, sendo representados por **redes complexas**.

Cientistas de diversas áreas têm se unido para entender como esses sistemas funcionam e como são responsáveis por fenômenos muitas vezes impressionantes – uma única pessoa infectada poderia ser responsável pela propagação de uma epidemia pelo mundo inteiro em uma curta escala de tempo, por exemplo. Cada área contribui com suas ferramentas, e esta nova área de estudo é tema também da Física Estatística, com o uso de seu ferramental para lidar com sistemas com um grande número de elementos.

Diferentemente de uma rede regular, uma rede complexa é formada por vértices que possuem números diferentes de conexões com os demais, e essas ligações não são triviais e sequer precisam ser de proximidade física, como na representação de átomos em um cristal. Um exemplo de rede complexa é a rede social *Facebook*, como apresentada na Figura 1.1. Note que uma rede regular jamais poderia representar esse tipo de rede, já que cada perfil possui um número diferente de amizades realizadas de forma heterogênea.

Fenômenos como a propagação de doenças em pessoas, ou de vírus em computadores conectados pela *Internet*, ou mesmo de informações em uma determinada rede social, são todos processos dinâmicos que acontecem na natureza e na sociedade e que utilizam as conexões entre os elementos constituintes para propagar alguma atividade. Há três propriedades encontradas em redes complexas reais que são particularmente importantes nessas dinâmicas, que serão comentadas no próximo capítulo:

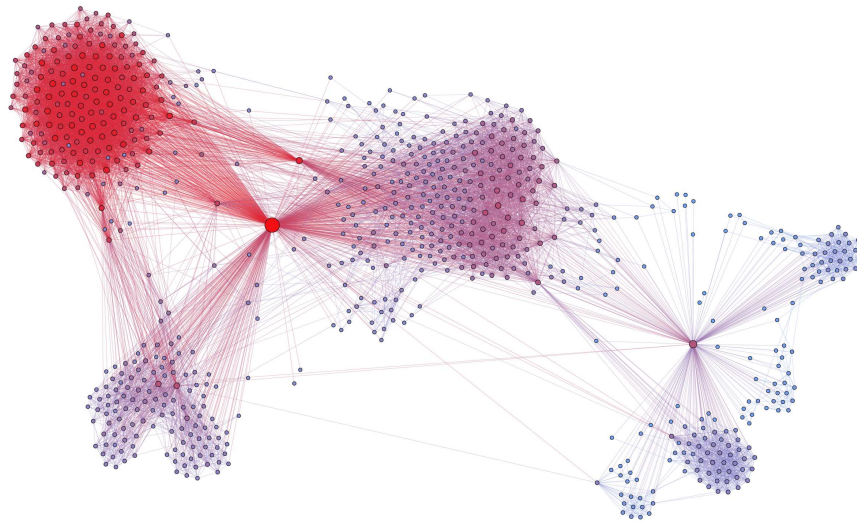


Figura 1.1: Rede de amigos e amigos em comum de dois perfis da rede social *Facebook*, na qual os vértices representam os usuários e as ligações as relações de amizade. Tal imagem foi gerada utilizando o *software Gephi* e os dados foram obtidos pela ferramenta *Visualizador de rede de amizades no Facebook*, disponível em <http://fb.wesleycota.com/>, desenvolvida pelo autor.

- *Agrupamento:* Os vértices possuem a tendência de formar *triângulos*, criando comunidades.
- *Livre de escala:* A distribuição do número de ligações de cada vértice é muito heterogênea, com grandes flutuações.
- *Mundo pequeno:* A *distância* entre *quaisquer* dois vértices da rede é relativamente pequena.

Modelos clássicos já conhecidos e estudados em Física Estatística utilizando estruturas regulares, como os passeios aleatórios, o modelo de Ising, e a percolação direcionada, podem ser naturalmente estendidos para as redes complexas. Neles podem ocorrer transições de fase, que são mudanças macroscópicas singulares em função da variação de um parâmetro de controle. Não obstante, conceitos de transições de fase são utilizados até mesmo na teoria de populações biológicas, por exemplo.

Em todos esses modelos, as interações microscópicas são responsáveis por uma mudança macroscópica no sistema e por isso é importante compreender como as características estruturais de redes complexas afetam essas propriedades dinâmicas. Neste trabalho, é dado destaque às transições de fase para estados absorventes, a partir dos quais não é possível mudar de estado. A propagação de uma epidemia pode ser descrita por essa transição, tendo como interesse a extinção de uma doença.

Como nem sempre os modelos matemáticos para tais dinâmicas possuem soluções exatas, é importante a utilização de resultados numéricos e de simulações computacionais. Para isso, métodos devem ser aperfeiçoados e explicados para otimizar essas análises, sem perder os aspectos relevantes de tais modelos. Uma das dificuldades encontradas no

estudo de processos epidêmicos em redes complexas é a falta de referências que justifiquem os métodos computacionais utilizados, e este trabalho tem como objetivo sintetizá-los e generalizá-los de forma clara.

O restante desta monografia está organizada da seguinte forma: nos Capítulos II e III são feitas revisões de conceitos estruturais de redes e de processos dinâmicos, respectivamente. No Capítulo IV os métodos de simulação computacional são descritos. Como objeto de estudo e desenvolvimento desta monografia, no Capítulo V um modelo geral de otimização de algoritmos é descrito, e os resultados são comparados para diferentes dinâmicas no Capítulo VI. Conclusões finais são feitas no Capítulo VII.

Capítulo II

Grafos e complexidade

Diante da não trivialidade das relações entre elementos, novas medidas e métodos para lidar com a enorme quantidade de dados envolvidos em redes complexas são necessários: surge a **Ciência de Redes Complexas**. Neste contexto, as redes representam um *background* teórico que permite a representação matemática de interações entre elementos de sistemas complexos em geral a partir da teoria dos grafos. Tais conceitos são definidos e discutidos neste capítulo, tendo como base a referência [1]. Antes é discutido brevemente o conceito de *complexidade*.

2.1 Redes complicadas ou complexas?

Com a revolução industrial e tecnológica, a interconectividade entre diversos elementos aumentou ainda mais. O efeito de mundo pequeno, já citado anteriormente, é dito ser um fenômeno moderno [2], com implicações importantes na propagação de doenças no mundo atual¹. A *Internet*, outra manifestação moderna, é uma rede enorme, que possui tanto elementos físicos quanto virtuais — roteadores, servidores, provedores de acesso, computadores, *sites*, *blogs*, redes sociais, e até mesmo *coisas*² —, que representa um substrato altamente complexo e robusto de comunicação humana.

Além dela, as redes biológicas, de infraestrutura e redes sociais apresentam um número relativamente grande de elementos e as relações entre eles não é óbvia. A partir do momento em que as conexões entre elementos são feitas de forma não trivial e aspectos microscópicos são responsáveis por manifestações macroscópicas, o sistema é dito ser *complexo*. Com isso, o comportamento coletivo não é linear: frações independentes do sistema não seriam capazes sozinhas de sustentar uma atividade.

Embora haja a *complexidade*, isso não significa que as estruturas complexas são *complicadas*. As medidas apresentadas neste capítulo possuem formas relativamente simples:

¹ A *peste negra* no século XIV na Europa teria se espalhado por todo o mundo muito mais rapidamente no mundo moderno, devido à facilidade de se transportar de um lugar ao outro.

² *Internet das coisas* é um termo que descreve a interconectividade entre dispositivos físicos, como geladeiras, televisores, câmeras de segurança e até mesmo carros.

leis de potência, comportamentos hierárquicos na conectividade, dentre outras. Modelos simples de propagação de doenças, como os descritos no Capítulo III, conseguem reproduzir comportamentos nada triviais que ocorrem em sistemas reais, e é neste contexto que a Física Estatística se apresenta para ajudar a compreender fenômenos nessas estruturas.

2.2 Propriedades e medidas

As redes complexas representam matematicamente sistemas que podem ser descritos por meio de **grafos**. Um grafo $G = (\mathcal{V}, \mathcal{E})$ é definido por meio de dois conjuntos: \mathcal{V} , cujos elementos são chamados de **vértices**, e \mathcal{E} , um conjunto não ordenado de pares (i, j) , chamados de **arestas** ou **ligações**, que estabelecem relações entre os vértices i e j . Neste caso, i e j são ditos *conectados*, *adjacentes* ou *vizinhos*, como na Figura 2.1.

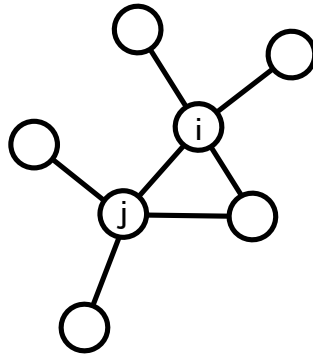


Figura 2.1: Exemplo de uma rede. Cada círculo representa um vértice e as arestas, representadas pelas linhas, estabelecem conexões entre os vértices.

O número total de vértices em uma rede é denotado por N e o número total de ligações é denotado por E . Embora em teoria dos grafos o tamanho do sistema seja identificado por E , aqui este será identificado como N . A cada vértice pode ser associado um número k de ligações, sendo denominado como o *grau* do vértice: k_i representa o grau do vértice i . Quando dois vértices diferentes são conectados, estes se tornam vizinhos, e os vértices vizinhos do vértice i são denotados por $V(i)$.

Neste trabalho, serão consideradas redes não direcionadas e não ponderadas: uma ligação (i, j) é equivalente a (j, i) e tais elementos são únicos em \mathcal{E} . Sendo assim, cada ligação possui dois *stubs* ou pontas de aresta, uma correspondente ao vértice i e outra ao vértice j .

2.2.1 Ligações e matriz de adjacência

Todas as conexões de uma rede podem ser descritas pela **matriz de adjacência** $\tilde{A} = \{A_{ij}\}$, quadrada de tamanho $N \times N$, com seus elementos definidos como:

$$A_{ij} = \begin{cases} 1, & \text{se } (i, j) \in \mathcal{E} \\ 0, & \text{se } (i, j) \notin \mathcal{E} \end{cases}. \quad (2.1)$$

Para uma rede não direcionada as ligações são recíprocas: $A_{ij} = A_{ji}$. Desta forma, o grau de um vértice i é simplesmente $k_i = \sum_j A_{ij}$.

Em uma rede de tamanho N , o número máximo de ligações possíveis é $\binom{N}{2} = [N(N-1)/2]$, o que corresponderia a todos os vértices estarem conectados com todos os vértices da rede. Para uma rede conectada, isto é, sem partes isoladas, pode-se definir a **densidade** de ligações como sendo $\mathcal{D} = E/[N(N-1)/2]$. Se $\mathcal{D} \ll 1$, a rede é dita ser *esparsa*. Sendo assim, a matriz \tilde{A} será também esparsa, contendo muitos elementos nulos.

Na aplicação computacional, se considerarmos cada elemento de \tilde{A} como sendo binário, cada um deles demandaria 1 *bit* de informação. Em uma rede de tamanho $N = 10^6$, seriam N^2 elementos e, assim, seriam necessários $1 \text{ bit} \times 10^6 \times 10^6 = 10^{12} \text{ bits} = 125 \text{ gigabytes}$ de memória computacional, já que cada *byte* corresponde a 8 *bits*, e *giga* representa 10^9 . Tal quantidade de memória é proibitiva para aplicações computacionais, sendo conveniente trabalhar com a **lista de adjacência** ℓ , que contém apenas a informação não nula de \tilde{A} e será definida na Seção 4.1.

2.2.2 Coeficiente de agregação

O coeficiente de agregação de um vértice i é denotado por c_i e pode ser definido como a razão entre o número de ligações entre seus vizinhos e o número máximo de ligações possíveis entre esses vértices [3]. Sendo e_i o número de ligações entre esses vizinhos,

$$c_i = \frac{e_i}{k_i(k_i - 1)/2}, \text{ se } k_i > 1. \quad (2.2)$$

Podendo e_i ser escrito em termos da matriz de adjacência \tilde{A} :

$$e_i = \frac{1}{2} \sum_{j,l} A_{ij} A_{jl} A_{li}. \quad (2.3)$$

Para $k_i \leq 1$, definimos $c_i \equiv 0$. Sendo assim, esse coeficiente define localmente a quantidade de triângulos envolvendo o vértice i e os vizinhos $V(i)$. Redes reais de amizades, por exemplo, possuem coeficiente de agregação altos. Considere um mecanismo de conexão entre vértices no qual há preferência por conexões entre vértices já conectados entre si. Com esse mecanismo, na Figura 2.2, o vértice 1 possui uma tendência maior a fazer amizade com 3 do que com 4 ou 5: as pessoas tendem a ser amigas de pessoas que já são amigas entre si [4].

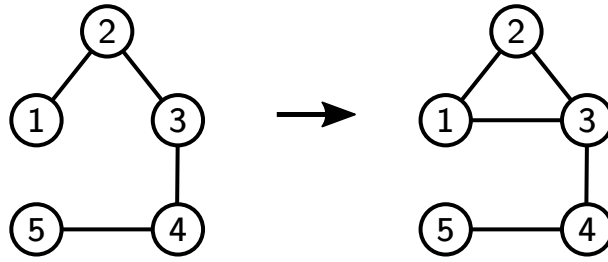


Figura 2.2: Para o lado esquerdo, $c_i = 0$ para $i = 1, \dots, 5$. No lado direito, após conexão de 1 com 3, $c_1 = c_2 = 1$, $c_3 = 1/3$ e $c_4 = c_5 = 0$.

2.2.3 Distâncias e caminhos

Um caminho é uma sequência ordenada de $p - 1$ arestas $(i_1, i_2), (i_2, i_3), \dots, (i_{p-1}, i_p)$ conectando os vértices i_1 e i_p através dos vértices i . O número de arestas de um caminho é definido como sendo o seu tamanho. Se em um conjunto de vértices existe pelo menos um caminho que conecte *cada par* destes, é dito que eles pertencem à mesma **componente**. As diferentes componentes e um caminho estão representados na Figura 2.3.

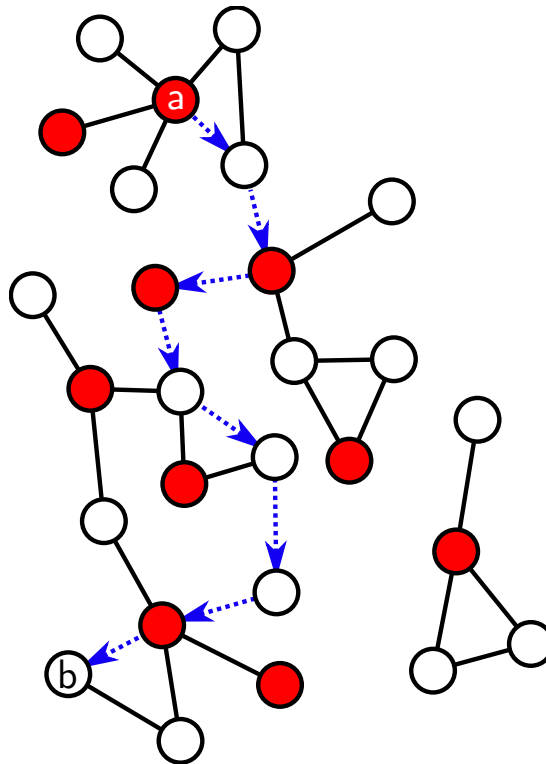


Figura 2.3: Arestas podem transmitir uma doença entre os vértices a e b por um caminho, e cada tom pode representar um estado (infectado ou suscetível, por exemplo). Note que há vértices que não podem ser acessados por todos os demais da rede, definindo componentes.

Uma propriedade de redes aleatórias é a presença de uma **componente gigante**, definida como a componente cujo tamanho escala com N , divergindo juntamente com $N \rightarrow \infty$. A sua presença implica que uma fração macroscópica da rede é conectada.

A **distância** entre vértices i e j pertencentes à mesma componente é definida como o

menor tamanho entre todos os caminhos existentes entre i e j . O caminho com esse menor tamanho é dito ser o *menor caminho*, e a distância é denotada por ℓ_{ij} . Exceto quando dito o contrário, redes com uma única componente serão consideradas neste trabalho. O **diâmetro** de uma rede é definido como a maior distância existente na rede:

$$d_G = \max_{i,j} \{\ell_{ij}\}. \quad (2.4)$$

Outra definição importante é a de **menor distância média** como

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{i,j} \ell_{ij}, \quad (2.5)$$

sendo o fator $N(N-1)$ no denominador o número de pares possíveis na rede. A *centralidade de proximidade* de um vértice i aos demais da rede é definida como

$$g_i = \frac{1}{\sum_{j \neq i} \ell_{ij}}, \quad (2.6)$$

já que quanto menor a distância de um vértice aos outros vértices da rede, maior seria sua proximidade.

No grafo completo, $\langle \ell \rangle = 1$, e em rede hipercúbica de d dimensões com N vértices, $\langle \ell \rangle \sim N^{1/d}$. Em uma rede aleatória, a menor distância média tem um comportamento logarítmico com o tamanho N : $\langle \ell \rangle \sim \ln N$. Como tal crescimento é muito mais lento do que em uma rede hipercúbica, é como se tal rede tivesse *dimensão infinita*, apresentando o *efeito de mundo pequeno*.

A existência, o tamanho dos caminhos, e a forma pelas quais os vértices estão conectados são determinantes para a propagação de epidemias, surgindo fenômenos interessantes como fases de *Griffiths* em módulos desconectados [5].

2.2.4 Estudo estatístico e distribuições de conectividade

A *Internet*, redes biológicas e redes sociais costumam ter alto coeficiente de agregação e a propriedade de mundo pequeno. Uma forma de estudar tais propriedades é associando a essas redes uma distribuição de conectividade $P(k)$, geralmente em **lei de potência**.

Nessas redes, os vértices com grau de conectividade muito grande são chamados de *hubs*. Os *hubs* com graus muito maiores que a média entre eles são chamados de *outliers*. Tais vértices atuam como super espalhadores em processos epidêmicos, sendo documentados em doenças infecciosas desde varíola até AIDS [6]: mesmo que a rede não tenha um coeficiente de agregação alto, os *hubs* são responsáveis por fazer a intermediação entre dois vértices da rede, reduzindo a menor distância média da rede.

Eles também estão relacionados com a propriedade das redes reais serem *livres de escala*, nas quais a variância das distribuições de conectividade é muito grande, como na Figura 2.4: os *hubs* aparecem nas *caudas pesadas*. Assim, a distribuição de conectividade

$P(k)$ pode ser analisada para determinar a importância de um vértice da rede para um processo epidêmico [5, 7], já que a transmissão da doença pode depender do grau k .

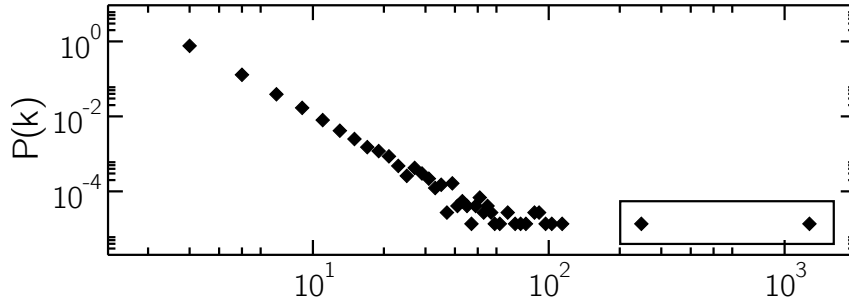


Figura 2.4: Exemplo de uma distribuição de conectividade $P(k) \sim k^{-3,5}$, com destaque para os *outliers*.

Geralmente, o interesse é em medidas que possam representar uma característica geral de toda a rede. Assim, são definidos o grau médio $\langle k \rangle$ e o coeficiente de agregação médio $\langle c \rangle$ de uma rede como:

$$\langle k \rangle = \frac{1}{N} \sum_i k_i, \quad \langle c \rangle = \frac{1}{N} \sum_i c_i, \quad (2.7)$$

sendo os somatórios realizados sobre todos os vértices i da rede. Note que, em termos da distribuição de grau e do número de arestas da rede, podemos escrever o grau médio para uma rede não direcionada como

$$\langle k \rangle = \sum_k k P(k) \equiv \frac{2E}{N}, \quad (2.8)$$

possuindo o fator 2 devido ao fato de cada aresta contribuir para o grau de cada um dos vértices conectados, como mostrado na Figura 2.5.

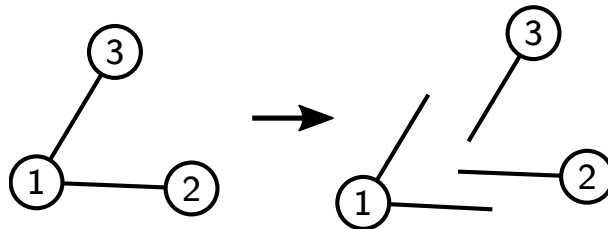


Figura 2.5: Exemplo da identificação de cada vértice e seu respectivo grau. Neste caso, são $N = 3$ elementos, $E = 2$ ligações, $k_1 = 2$, $k_2 = k_3 = 1$ e o grau médio é $4/3$. Note que o número de *stubs* é $2E = 4$.

Sendo conhecida a distribuição de grau $P(k)$, com ela pode ser analisada a **heterogeneidade** da rede, e um parâmetro para isso é

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}, \quad (2.9)$$

sendo $\langle k^n \rangle$ o n -ésimo momento da distribuição de grau

$$\langle k^n \rangle = \sum_k k^n P(k). \quad (2.10)$$

No caso de distribuições em lei de potência, $\langle k^2 \rangle$ diverge para $2 < \gamma < 3$ e, assim, $\kappa \gg \langle k \rangle$. Neste caso, a rede é dita heterogênea. Já para as distribuições Gaussiana e de Poisson, por exemplo, $\kappa \sim \langle k \rangle$, sendo a rede considerada homogênea. Sendo assim, distribuições de grau de redes heterogêneas possuem caudas pesadas — o que é comum para distribuições em lei de potência, como na Figura 2.4 —, enquanto em redes homogêneas $P(k)$ decai rapidamente, como na Figura 2.6.

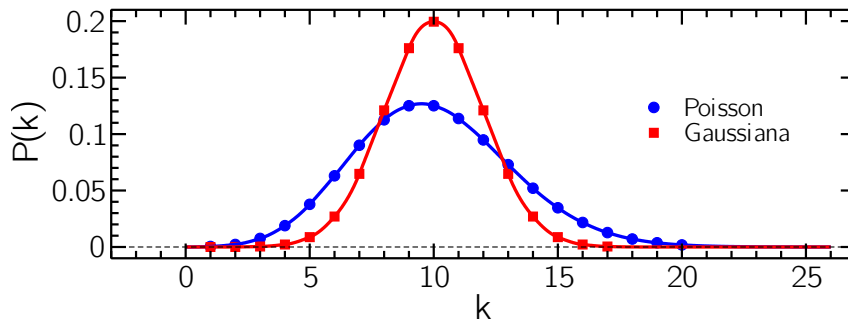


Figura 2.6: Distribuições de conectividade do tipo Poisson e do tipo Gaussiana, com mesmo valor médio $\langle k \rangle = 10$. Note que as caudas decaem rapidamente, ao contrário da Figura 2.4 com lei de potência.

2.2.5 Correlações estruturais

Pode-se definir também o grau médio dos vizinhos de um vértice i , denotado por

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j, \quad (2.11)$$

com a soma sendo feita sobre todos os vizinhos j do vértice i . Determinando o $k_{nn}(k)$ médio de todos os vértices cujo grau é k , a correlação entre os graus pode ser investigada. Assim, $k_{nn}(k)$ é definido fazendo-se o somatório sobre todos os vértices i cujo grau é k e, sendo N_k o número total de vértices de grau k na rede, temos:

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn,i}. \quad (2.12)$$

Quando tal função é crescente, a rede é dita *associativa*. Nos casos de possuir valores aproximadamente constantes, a rede é *não correlacionada* ou *neutra*. Se é decrescente, é

desassociativa, como demonstrado qualitativamente na Figura 2.7. Esta função mostra a correlação entre os vértices: eles tendem a se conectar com vértices de mesmo grau, sem distinção ou com vértices de grau distinto, respectivamente para os três casos.

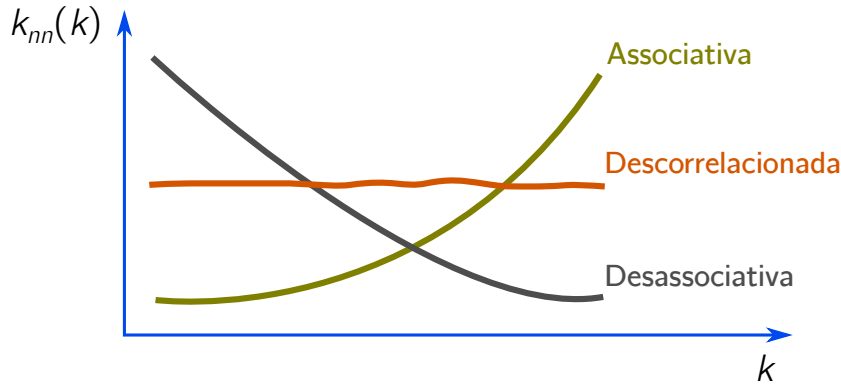


Figura 2.7: Comportamento qualitativo da função $k_{nn}(k)$ e classificações.

Tais classificações estão relacionadas ao fato de a função $k_{nn}(k)$ poder ser escrita como

$$k_{nn}(k) = \sum_{k'} k' P(k'|k), \quad (2.13)$$

em que $P(k'|k)$ é a probabilidade condicional de um vértice de grau k estar conectado a um de grau k' . Uma rede é não correlacionada, portanto, se $P(k'|k)$ é função apenas de k' e, com isso, $k_{nn}(k)$ é constante.

Em uma rede não correlacionada, qual a probabilidade de que um vértice de grau k qualquer esteja ligado com um de grau k' ? Já que ela não deve depender de k , basta dividir o número de *stubs* de vértices de grau k' pelo número total de *stubs*, como mostrado na Figura 2.8. Considerando $P(k) = N_k/N$, pode-se escrever:

$$P_{nc}(k'|k) = \frac{k' N_{k'}}{\sum_{k''} k'' N_{k''}} = \frac{k' (N_{k'}/N)}{\sum_{k''} k'' (N_{k''}/N)} = \frac{1}{\langle k \rangle} k' P(k'), \quad (2.14)$$

na qual o índice *nc* significa *não correlacionada*. Sendo assim, utilizando a equação (2.13):

$$k_{nn}^{nc}(k) = \sum_{k'} k' P_{nc}(k'|k) = \sum_{k'} \frac{k' k' P(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} = \kappa, \quad (2.15)$$

sendo κ a constante que define o grau de heterogeneidade da rede na equação (2.9).

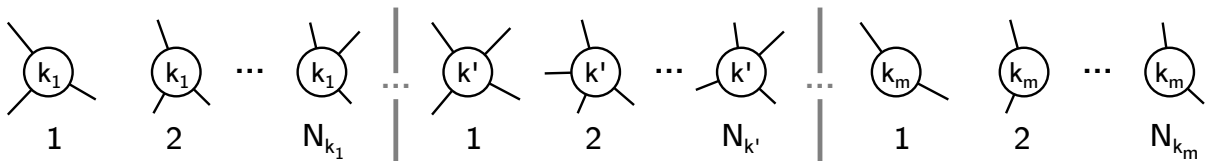


Figura 2.8: Conjuntos de todos os N_k vértices de grau k e seus respectivos *stubs*. Note que $\sum_{k''} k'' N_{k''} = 2E$ e que $k'' = k_1, \dots, k', \dots, k_m$.

Uma outra medida de correlação é o coeficiente de agregação médio em função do grau k :

$$c(k) = \frac{1}{N_k} \sum_{i/k_i=k} c_i. \quad (2.16)$$

Esta medida está diretamente ligada à *hierarquia* do padrão de conectividade da rede, como na rede de provedores e organizações na *Internet* [8], já que está associada a formação de comunidades na rede.

Há, claro, diversas outras medidas importantes que podem ser feitas e definidas em redes complexas [3], porém não são necessárias para os propósitos do presente trabalho. Na próxima seção, algumas propriedades de redes reais são brevemente discutidas.

2.3 Redes reais: propriedades e modelos

Como já citado anteriormente, as redes reais possuem as propriedades de serem livres de escala, possuírem o efeito de mundo pequeno e alto coeficiente de agregação. Diferentes modelos de redes aleatórias podem ser definidos, possuindo ou não tais características.

O mais simples de todos é o de **rede aleatória**, proposto por Erdős e Rényi, sendo possível na seguinte formulação: a rede possui N vértices e cada uma das $N(N-1)/2$ ligações possíveis é feita com probabilidade p , sendo o número total de ligações caracterizado por uma distribuição binomial. Como resultado, a rede possui grau médio $\langle k \rangle \approx Np$ para $N \gg 1$ [1] e, mantendo $\langle k \rangle$ constante, a distribuição de conectividade é do tipo Poisson: a rede é *homogênea*. Além disso, o coeficiente de agregação tende a zero para $N \rightarrow \infty$. No entanto, a rede resultante possui o efeito de mundo pequeno.

Um outro modelo é o de Watts-Strogatz [4]: a configuração inicial é um anel de N vértices simetricamente conectados a seus $2m$ vizinhos mais próximos, sendo m um parâmetro. Com probabilidade p e evitando autoconexões, uma ligação é redirecionada para os demais vértices, construindo *atalhos* na rede. O resultado é uma distribuição de conectividade do tipo Poisson com $\langle k \rangle = 2m$. Dependendo dos valores de p , o coeficiente de agregação é alto e a menor distância média possui valores baixos. Porém, ainda é considerada homogênea.

Com o objetivo de modelar a rede *World Wide Web* (WWW), surgiu o modelo de Barabási-Albert, que possui os seguintes mecanismos:

- **Crescimento:** a partir de m_0 vértices interconectados, um vértice é adicionado e se conecta com m outros vértices já existentes na rede, com $m < m_0$.
- **Ligação preferencial:** na escolha de quais vértices irá se conectar, a probabilidade de se conectar com um vértice de grau i é diretamente proporcional a k_i .

Como resultado, a rede possui uma distribuição de conectividade em lei de potência, característica de redes como de computadores, de relacionamentos sexuais, de pessoas em redes sociais, de colaboração entre pesquisadores, dentre outras. A rede é livre de escala e possui o efeito de mundo pequeno, embora o coeficiente de agregação tenda a zero para $N \rightarrow \infty$. Da forma em que está, o expoente γ de $P(k) \sim k^{-\gamma}$ é 3, mas modificações nas probabilidades das ligações preferenciais podem ajustá-lo para outros valores controlados.

Neste trabalho, as redes são construídas utilizando um algoritmo que preserva uma distribuição de grau $P(k)$, assumindo um grau mínimo k_0 e um máximo k_c : é o **modelo de configurações**. As conexões são feitas ao acaso, ligando os *stubs* desconectados de um vértice ao outro de forma aleatória, impedindo múltiplas e autoconexões, indesejáveis na modelagem de sistemas reais. Com isso, podem surgir correlações na rede, principalmente para aquelas nas quais $\langle k^2 \rangle$ diverge. Uma rede regular gerada por este modelo é chamada de *Random Regular Network*³ (RRN). Uma rede é chamada *regular* de grau k se todos os vértices da rede possuem o mesmo grau k .

Para as redes em lei de potência, dependendo do corte k_c do grau máximo da rede $k_{\text{máx}}$, podemos ter redes não correlacionadas, nas quais $k_{nn}(k) \approx \text{constante}$: é o *Uncorrelated Configuration Model* (UCM)⁴ [9]. Isso pode ocorrer pois, assumindo que k_c é ilimitado, o grau máximo médio gerado por tal distribuição é da forma $\langle k_{\text{máx}} \rangle \sim N^{1/(\gamma-1)}$ [10], introduzindo um corte natural no sistema. Colocando o grau máximo permitido como $k_c = \sqrt{N}$, as correlações de graus são eliminadas, conforme mostrado em [9]. Este modelo será retomado no Capítulo IV.

³ Rede regular aleatória

⁴ Modelo de configurações descorrelacionado.

Capítulo III

Processos dinâmicos

Há uma diversidade de fenômenos que podem ser descritos por processos epidêmicos em redes, como [6]:

Biológicos: Propagação de patógenos a partir do contato físico de pessoas, como a gripe, Ebola, e outras doenças infecciosas.

Digitais: Um vírus pode se propagar por uma rede de computadores, se reproduzindo indefinidamente. Os padrões de infecção são semelhantes aos biológicos [11].

Sociais: A transmissão de conhecimentos, comportamentos, boatos, notícias e até mesmo *memes*¹ na *Internet*.

Todos estes fenômenos envolvem agentes que espalham algo utilizando alguma estrutura como base. Um vírus da gripe, por exemplo, utiliza da rede de contatos de proximidade física de um indivíduo para sua propagação. Já um boato ou informação, utiliza da rede de comunicação de indivíduos, como as redes sociais *on-line*. Embora possuam mecanismos totalmente diferentes de espalhamento, tais processos obedecem padrões comuns e podem ser analisados por meio de uma modelagem geral. Neste trabalho utilizamos a modelagem *compartimental*, classificando os indivíduos de um sistema em compartimentos ou estados. Usando o jargão de epidemias, consideraremos:

Suscetíveis (S): vértices que não estão em contato com o patógeno e podem se infectar.

Infectados (I): vértices contagiosos, que podem transmitir o patógeno aos seus vizinhos.

Recuperados (R): vértices que já foram infectados antes, porém não suscetíveis.

3.1 Processos Markovianos e equações mestras

Um dado processo epidêmico irá modificar o estado de cada indivíduo a partir de determinadas regras dinâmicas. Para cada vértice i , podemos adotar estados σ_i , definindo

¹ Mídias digitais que se propagam rapidamente pela *internet*, a partir da replicação pelos seus usuários, como fotos e vídeos engraçados.

o estado microscópico dinâmico por meio do conjunto $\sigma(t) = (\sigma_1(t), \dots, \sigma_N(t))$, para certo tempo t . Porém, devido ao grande número de variáveis envolvidas em uma dinâmica e à estocasticidade² dos processos, uma das formas de análise é por meio de **taxas de transição** $W_{\sigma \rightarrow \sigma'} \geq 0$, de um estado σ para σ' . Com uma condição inicial, o conjunto de todas as configurações possíveis e as taxas de transição, definimos um modelo estocástico.

A probabilidade de um estado σ ser encontrado em um instante de tempo t é denotado por $P(\sigma, t)$. Se a evolução para $t' > t$ pode ser determinada completamente pelo estado no tempo t , a dinâmica é dita ser **Markoviana**, não dependendo da história do sistema. Por meio de uma **equação mestra**, considerando $W_{\sigma \rightarrow \sigma'}$ como uma taxa por unidade de tempo, a dinâmica de $P(\sigma, t)$ pode ser descrita como

$$\frac{\partial P(\sigma, t)}{\partial t} = \sum_{\sigma'} W_{\sigma' \rightarrow \sigma} P(\sigma', t) - \sum_{\sigma'} W_{\sigma \rightarrow \sigma'} P(\sigma, t). \quad (3.1)$$

Note que o primeiro termo do lado direito da equação (3.1) define as *entradas*, isto é, transições para o estado σ , enquanto o segundo termo, negativo, define as *saídas* do estado σ , sendo satisfeita a seguinte condição de normalização:

$$\sum_{\sigma} P(\sigma, t) = 1. \quad (3.2)$$

Da equação mestra podem ser extraídos os valores médios no tempo t de qualquer função de estado $f(\sigma)$:

$$\langle f(t) \rangle = \sum_{\sigma} f(\sigma) P(\sigma, t). \quad (3.3)$$

Alguns sistemas possuem um limite assintótico bem definido, denotado por $P_{\infty}(\sigma) = \lim_{t \rightarrow \infty} P(\sigma, t)$, chamado de *estado estacionário*, no qual a média temporal representa o sistema após um tempo transiente. Quando o sistema possui um número finito de estados, a característica de o processo não possuir memória intrínseca garante a unicidade do estado estacionário [12].

Dentre as formas de encontrar soluções aproximadas para as equações mestras, estão a **aproximação de campo médio** (CM), que consiste em utilizar os valores médios das variáveis microscópicas a partir de algumas hipóteses, e as **simulações computacionais**. Utilizando geradores de números aleatórios, associação de probabilidades às taxas de transição e *passos de Monte Carlo*³, ao atingir um estado estacionário a distribuição de estados deve convergir para a solução estacionária do processo investigado.

A condição de *balanço detalhado* $W_{\sigma' \rightarrow \sigma} P(\sigma') = W_{\sigma \rightarrow \sigma'} P(\sigma)$ leva à equação (3.1) ser nula, sendo condição suficiente para o equilíbrio. Porém, existem processos fora do equilíbrio, que não satisfazem tal relação e, assim, a distribuição estacionária fora do

² Um processo é dito *estocástico* se seu comportamento depende de variáveis aleatórias.

³ Um passo de Monte Carlo é cada ocorrência de algum procedimento que envolva números aleatórios.

equilíbrio pode ser desconhecida.

Ao contrário de fenômenos de equilíbrio, nos quais basta conhecer a função de partição do sistema a partir de uma Hamiltoniana [13], a maioria dos processos dinâmicos de interesse para o mundo real são processos fora do equilíbrio, não sendo possível utilizar a mesma formulação da Mecânica Estatística de equilíbrio. No entanto, podem ser utilizados conceitos análogos, como na análise das transições de fase e nos conceitos de criticalidade. Antes de discuti-los, é importante definir os Processos de Poisson, que servem como base para a modelagem de processos epidêmicos Markovianos.

3.2 Processos de Poisson

Suponha que um evento ocorra a taxa α , o que significa que tal evento ocorre com probabilidade αdt em um certo intervalo de tempo dt . A probabilidade do evento ocorrer duas vezes no intervalo dt é de ordem $(dt)^2$, que se torna irrelevante quando $dt \rightarrow 0$. Assumindo que este seja um processo Markoviano, tal probabilidade não deve depender de t e muito menos do número de eventos que já ocorreram, denotado por n . A probabilidade de que ocorra n eventos até o instante de tempo t pode ser obtido a partir da seguinte equação mestra [14, 15]:

$$\frac{dP(n,t)}{dt} = \begin{cases} -\alpha P(0,t), & \text{se } n = 0; \\ \alpha P(n-1,t) - \alpha P(n,t), & \text{caso contrário,} \end{cases} \quad (3.4)$$

com $n \geq 0 \in \mathbb{Z}$. Com a condição inicial $P(0,0) = 1$ e $P(n,0) = 0$ para $n \neq 0$, que significa que nenhum evento aconteceu até o tempo $t = 0$, a equação é facilmente resolvida. Tal processo definido de tal forma é um *processo de Poisson*. Note que ele é equivalente a um passeio aleatório completamente assimétrico [15], já que existe possibilidade de transição para a direita, mas não para a esquerda.

Uma forma de solucionar equações é definir a *função geratriz*

$$g(z,t) = \sum_{n=0}^{\infty} P(n,t) z^n. \quad (3.5)$$

Temos que a condição inicial para $P(n,0)$ torna-se $g(z,0) = 1$. Derivando a equação (3.5) em relação ao tempo, e utilizando a equação (3.4), obtemos

$$\frac{dg(z,t)}{dt} = -\alpha(1-z)g(z,t), \quad (3.6)$$

cuja solução para a condição inicial $g(z,0) = 1$ é

$$g(z,t) = e^{-\alpha(1-z)t} = e^{-\alpha t} e^{\alpha z t} = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} z^n. \quad (3.7)$$

Comparando a equação anterior com a equação (3.5), é obtida a *Distribuição de*

Poisson:

$$P(n, t) = e^{-\alpha t} \frac{(\alpha t)^n}{n!}, \quad (3.8)$$

cujo valor médio de n é $\langle n \rangle = \alpha t$ e a variância $\langle n^2 \rangle - \langle n \rangle^2 = \alpha t$. Portanto, α é o número de eventos esperados por unidade de tempo. Uma distribuição muito importante para o desenvolvimento deste trabalho é a do *tempo entre eventos*. Considere que em t ocorra um evento e que sua duração seja τ . Como os eventos são independentes e não há superposição, entre t e $t + \tau$ não ocorre outro evento e a probabilidade disso acontecer é equivalente a considerar t como uma nova origem da distribuição descrita na equação (3.8) até um tempo $t = \tau$ com $n = 0$: $e^{-\alpha \tau}$. Logo após um intervalo dt , um evento ocorre, e a probabilidade de que o tempo entre eventos seja τ será

$$W(\tau)dt = P(0, \tau)\alpha dt = (e^{-\alpha \tau}\alpha)dt. \quad (3.9)$$

Portanto, a distribuição de probabilidade do tempo entre eventos possui uma forma exponencial, e o resultado é uma base para a construção do Algoritmo de Gillespie, estudado na Seção 4.2:

$$W(\tau) = \alpha e^{-\alpha \tau}. \quad (3.10)$$

Em resumo, um processo de Poisson é aquele tal que eventos ocorram a uma taxa fixa, mas de forma aleatória. Nele, cada evento é independente e diferentes eventos não ocorrem durante o mesmo intervalo de tempo.

3.3 Modelos epidêmicos clássicos

Neste trabalho, foram simulados processos que apresentam transições de fase para estados absorventes, a partir dos quais a dinâmica não evolui. Todos eles podem ter um valor *crítico* λ_c de um parâmetro de controle λ , de modo que para $\lambda > \lambda_c$ o estado estacionário é ativo, com flutuações, enquanto para valores menores a dinâmica alcança o estado absorvente. São eles:

3.3.1 Modelo PC - Processo de Contato

O processo de contato é um modelo simples de propagação de epidemia [16]. Nele, cada vértice i da rede pode assumir dois estados: infectado ($\sigma_i = 1$) e suscetível ($\sigma_i = 0$). Neste modelo, os vértices infectados são capazes de transmitir a infecção a partir do contato com seus vizinhos mais próximos.

Assim, um vértice suscetível pode se tornar infectado apenas se houver algum vizinho infectado. Os vértices infectados ficam suscetíveis a uma taxa μ , enquanto que um vértice

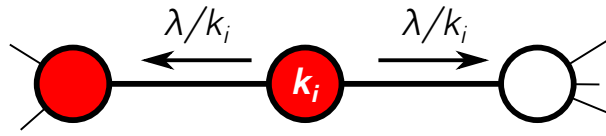


Figura 3.1: No processo de contato, um vértice de grau k_i transmite a infecção para cada vizinho com taxa λ/k_i .

infectado i transmite a infecção, criando uma cópia em cada vizinho, a uma taxa λ/k_i , sendo k_i seu grau e λ a taxa de infecção: **a taxa λ é dividida** entre os vizinhos, como na Figura 3.1. Note que os vértices que se tornam suscetíveis podem ser infectados novamente e que o estado $\sigma_i = 0, \forall i$ é um estado absorvente.

3.3.2 Modelo SIS - Suscetível-Infectado-Suscetível

O modelo SIS é semelhante ao processo de contato. Nele, dois estados podem ser adotados pelos vértices: infectado ($\sigma_i = 1$) e suscetível ($\sigma_i = 0$). Da mesma forma, somente os vértices infectados transmitem a infecção aos seus vizinhos. Os vértices suscetíveis tornam-se infectados a uma taxa λn_i , sendo n_i o número total de vizinhos infectados do vértice i e λ a taxa de infecção: cada vértice infectado transmite a infecção para cada um de seus vizinhos com **taxa λ , sem divisão por k_i** , como na Figura 3.2. Já os vértices infectados ficam suscetíveis novamente com taxa μ . Note que não há imunização dos vértices previamente infectados e que o estado $\sigma_i = 0, \forall i$ é um estado absorvente.

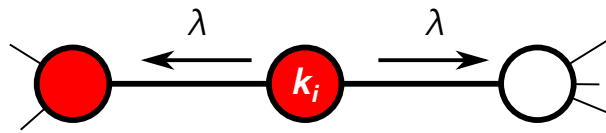


Figura 3.2: No modelo SIS, um vértice de grau k_i transmite a infecção para cada vizinho com taxa λ .

Para redes regulares de grau k , o modelo SIS é equivalente ao PC, sendo a taxa de infecção apenas reescalada em termos do grau k da rede: $\lambda_{\text{SIS}} = \lambda_{\text{PC}}/k$. Já para redes heterogêneas, o papel dos *hubs* é evidenciado, já que no modelo SIS a taxa de infecção efetiva de um *hub* é muito maior que dos demais vértices: cada vértice de grau k possui uma taxa de infecção efetiva λk .

3.3.3 Modelo SIR - Suscetível-Infectado-Recuperado

Além dos estados S ($\sigma_i = 0$) e I ($\sigma_i = 1$), há o estado recuperado R ($\sigma_i = 2$). A única mudança para o modelo anterior, o modelo SIS, é que a transição $I \rightarrow S$ é substituída por $I \rightarrow R$, não mais para o estado suscetível. Portanto, somente os vértices infectados transmitem a infecção aos seus vizinhos, com os suscetíveis tornando-se infectados com taxa λn_i e os infectados sendo recuperados com taxa μ . Desta forma, não há estado

estacionário ativo, já que todo vértice infectado ficará recuperado, não tendo mais evolução da dinâmica, caindo no estado absorvente $\sigma_i \neq 1, \forall i$.

Todos esses modelos podem apresentar transições de fase, a serem tratadas na próxima seção. Outros processos epidêmicos parecidos também podem ser elaborados, assumindo diferentes estados e taxas. Vale salientar que fazendo $\mu = 1$, fixamos a escala de tempo. Todos os resultados a serem apresentados a partir daqui serão com $\mu = 1$.

3.4 Criticalidade e transições de fase

Os modelos descritos anteriormente apresentam um *parâmetro de controle* λ , que é a taxa de infecção. Há também um *parâmetro de ordem*, que pode ser tanto a densidade média de vértices infectados $\rho_{inf} \equiv \rho = n/N$, para os modelos SIS e PC, quanto a de vértices recuperados $\rho_{rec} = n_{rec}/N$, sendo n e n_{rec} o número de vértices infectados e recuperados, respectivamente. Uma transição de fase ocorre para um determinado valor crítico do parâmetro de controle: no *limiar* ou *ponto crítico* λ_c , como na Figura 3.3.

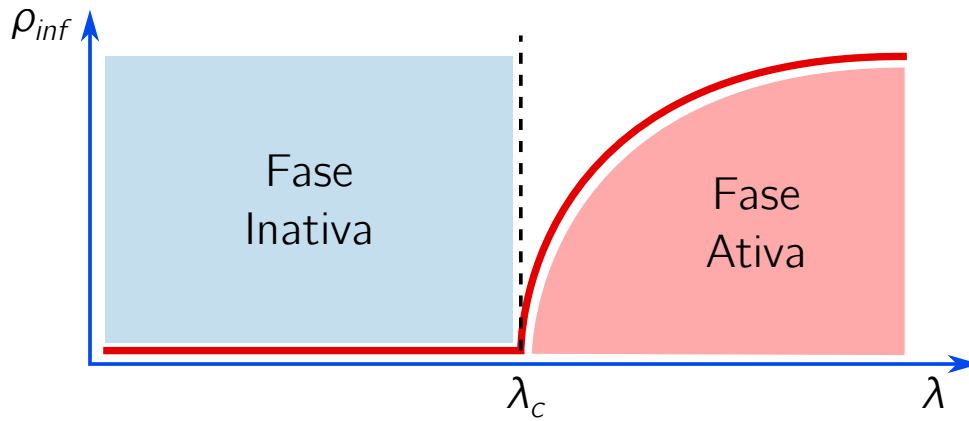


Figura 3.3: Curva típica para o valor quase-estacionário (a ser definido na Seção 3.5) da densidade média de vértices infectados em função do parâmetro de controle λ : para $\lambda < \lambda_c$, não há vértices infectados para $t \rightarrow \infty$, enquanto que a partir de λ_c os vértices infectados correspondem a uma fração macroscópica do sistema.

Em sistemas de equilíbrio, as transições de fase podem ser caracterizadas pela não analiticidade dos potenciais termodinâmicos, e nos pontos críticos as flutuações estatísticas resultam em mudanças macroscópicas por meio de um comportamento coletivo entre os constituintes do sistema [13]. Processos fora do equilíbrio também podem sofrer transições de fase, embora não sejam relacionadas a potenciais termodinâmicos.

Como já dito, um processo fora do equilíbrio possui uma dinâmica que viola o balanço detalhado, não podendo ser aplicadas, portanto, as mesmas hipóteses de processos em equilíbrio às equações mestras. Em particular, existem os sistemas com *estados absorventes*, nos quais a dinâmica de um sistema pode ficar *congelada* ao estarem em determinadas configurações. Em um modelo epidêmico, por exemplo, a infecção só pode ser transmitida se há algum vetor infectado: se tais vetores tornam-se inexistentes, o processo de infecção

é interrompido e a dinâmica não evolui. Os estados *transientes* podem ser acessados reversivamente, enquanto que a mudança de um estado $\sigma \rightarrow \sigma'$ no qual σ' é um estado absorvente não pode ser revertida.

Sistemas com esses estados absorventes podem apresentar estados estacionários ativos além do absorvente, podendo existir **transições para estados absorventes**.

3.4.1 Universalidade e expoentes críticos

Em Física Estatística, o conceito de universalidade diz respeito à observação de propriedades similares em uma classe de modelos diferentes. Neles, o comportamento na transição de fase está associado à uma *classe de universalidade*, na qual as quantidades de interesse são descritas por um mesmo conjunto de *expoentes críticos* e de outras quantidades universais [17].

A interação entre um sistema magnético de *spins* e um campo externo \vec{H} não nulo pode ser estudada a partir de um modelo simples: o **Modelo de Ising**, sendo um dos melhores exemplos de transição de fase contínua, na qual um parâmetro de ordem varia continuamente, porém com uma divergência na sua derivada no ponto crítico, que também se manifesta na divergência de funções resposta, como a suscetibilidade e calor específico. A hamiltoniana desse sistema pode ser definida como [13]:

$$\mathcal{H} = -J \sum_{ij} A_{ij} \sigma_i \sigma_j - H \sum_{i=1}^N \sigma_i, \quad (3.11)$$

na qual J é a constante de acoplamento dos spins, H é o módulo do campo magnético, e $\sigma_i = \pm 1$ representa o momento magnético de cada vértice, como no caso de *spin* 1/2. Note que A_{ij} corresponde aos elementos da matriz de adjacência, sendo a interação entre *spins* efetivamente apenas entre vizinhos.

Na ausência de um campo magnético ($H = 0$) e abaixo de uma certa temperatura crítica T_c , a magnetização M é não nula para um ferromagneto ideal. Tal temperatura é conhecida como *temperatura de Curie*. Na fase ordenada, $T < T_c$, à medida que T se aproxima de T_c , a magnetização tende a zero seguindo uma lei de potência

$$|M| \sim |T - T_c|^\beta. \quad (3.12)$$

O calor específico C , e a suscetibilidade magnética χ também apresentam comportamentos em lei de potência na criticalidade:

$$C \sim |T_c - T|^{-\alpha}, \quad (3.13)$$

e

$$\chi \sim |T_c - T|^{-\gamma}. \quad (3.14)$$

Os expoentes β , α e γ são os *expoentes críticos*. Uma função de correlação $c(r)$ também pode ser definida como $c(r) = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$, sendo r o espaçamento entre i e j . Assim, é possível a análise de como o estado de um vértice correlaciona-se com o de outros vértices. Longe do ponto crítico,

$$c(r) \sim e^{-r/\xi}, \quad (3.15)$$

em que ξ é o *comprimento de correlação do sistema*. Próximo de T_c , as correlações são de longo alcance, resultando em $\xi \rightarrow \infty$ e a função de correlação tem um decaimento dado por

$$c(r) \sim \frac{1}{r^{d-2-\eta}}, \quad (3.16)$$

sendo d a dimensão do sistema e η um expoente crítico. Tais relações são demonstradas analiticamente para o modelo de Ising [13]. Na criticalidade, ξ diverge como

$$\xi \sim |T - T_c|^{-\nu}, \quad (3.17)$$

com outro expoente crítico ν .

Para as transições de fase para estados absorventes, existem dois comprimentos de correlação relevantes: o *comprimento de correlação espacial* ξ_{\perp} e o *comprimento de correlação temporal* ξ_{\parallel} , com os respectivos expoentes críticos ν_{\perp} e ν_{\parallel} . O *expoente dinâmico* z é definido como a razão entre eles: $z = \nu_{\parallel}/\nu_{\perp}$. Note que \parallel corresponde a uma evolução temporal, no mesmo sentido da dimensão dinâmica, enquanto \perp corresponde à evolução espacial, perpendicular à direção de evolução do sistema, que é temporal.

Rigorosamente falando, esse expoente z diz o quão rápido uma perturbação se espalha, desde $z = 1$ (espalhamento em “cone de luz”) a $1 < z < 2$ (superdifusivo) e $z = 2$ (difusivo), até $z > 2$ (subdifusivo) [17]. Em vários modelos com interações de curta distância, $1 < z \leq 2$. Como os outros expoentes críticos, z é uma quantidade universal e, assim, determina a classe de universalidade e dimensionalidade do sistema.

O processo de contato, definido na Seção 3.3.1, apresenta esse tipo de transição: quando $\sigma_i = 0$, $\forall i$, o estado é absorvente. A partir de uma condição inicial transiente, a infecção dos vértices se espalha seguindo uma influência de curto alcance, mas se este espalhamento não for rápido o suficiente, a infecção do sistema pode desaparecer, já que os vértices se tornam suscetíveis com taxa unitária. A transição de fase para um estado estacionário ativo é dada por um limiar crítico λ_c . Vale dizer que o PC é o “modelo de Ising” dos sistemas com transições de fase para estados absorventes.

A Figura 3.4 mostra o significado físico dos comprimentos de correlação ξ visualmente, na análise do processo de contato. No lado esquerdo, correspondente uma rede regular unidimensional, nos quais os vértices estão sequencialmente ligados em forma de um anel, é notória a correlação tanto espaciais quanto temporais, e inclusive a presença

de *clusters* vazios. Os tamanhos deles são amplamente distribuídos e se estendem até as escalas típicas em tempo e em espaço, sendo interpretadas como os comprimentos de correlação temporal e espacial ξ_{\parallel} e ξ_{\perp} , respectivamente [17].

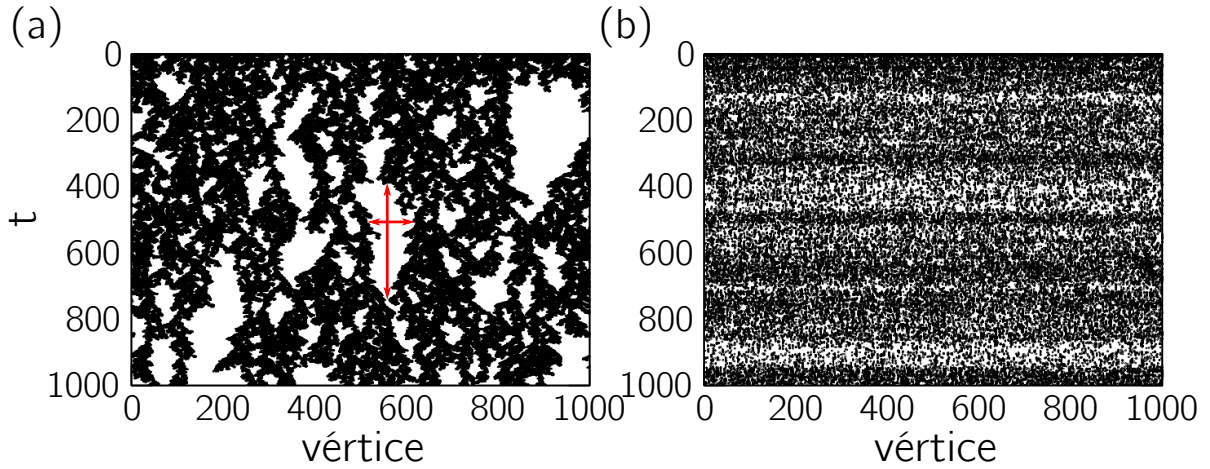


Figura 3.4: Padrões espaçotemporais de apenas uma evolução dinâmica do processo de contato para (a) uma rede regular unidimensional com condições de contorno periódicas e (b) uma rede aleatória regular de grau $k = 4$. O tamanho das redes é $N = 10^3$ e λ ligeiramente acima do valor crítico.

É importante notar na Figura 3.4 que em redes complexas o comprimento de correlação espacial não é facilmente observado em tal gráfico, já que os vértices estão conectados de uma forma não trivial. Porém, a correlação temporal é vista, dadas as “franjas” horizontais ao longo da evolução em t .

Na proximidade de λ_c , é observado que a densidade média $\bar{\rho}$ no estado estacionário segue uma lei de potência

$$\bar{\rho} \sim \Delta^\beta, \text{ com } \lambda > \lambda_c, \quad (3.18)$$

em que $\bar{\rho} \equiv \rho(t \rightarrow \infty)$, $\Delta \equiv |\lambda - \lambda_c|$ e β o expoente crítico associado. A partir de uma configuração quase absorvente, como apenas um vértice infectado, o número médio de vértices infectados $n(t)$ e a probabilidade de sobrevivência $P_S(t)$ podem ser analisados, sendo a última a probabilidade de que exista pelo menos um vértice infectado no tempo t . Na criticalidade, seguem leis de potência com expoentes críticos δ e η :

$$P_S(t) \sim t^{-\delta} \quad (3.19)$$

e

$$n(t) \sim t^\eta. \quad (3.20)$$

Uma forma incrivelmente simples, mas não trivial, de obter tais relações é por meio de uma *teoria de escala*. Ela se baseia na hipótese de que as propriedades assintóticas do sistema próximo da criticalidade são invariantes sob transformações do tipo

$$\Delta \rightarrow a\Delta, \quad (3.21)$$

$$t \rightarrow a^{-\nu_{\parallel}} t, \quad (3.22)$$

$$P_S \rightarrow a^{\beta'} P_S, \quad (3.23)$$

por exemplo, com $a > 0$ chamado de fator de escala. Isso faz com que haja restrições às funções que governam o comportamento crítico. Essa é uma abordagem proposta por Grassberger e de la Torre [18], assumindo um único vértice infectado como condição inicial. A probabilidade de sobrevivência $P_S(t)$ sofre tal transformação:

$$P_S(t) = a^{-\beta'} P_S(a^{-\nu_{\parallel}} t). \quad (3.24)$$

Como o valor de a pode ser escolhido arbitrariamente de forma que $a^{-\nu_{\parallel}} t = 1$, tem-se $a = t^{1/\nu_{\parallel}}$ e, assim,

$$P_S(t) = t^{-\beta'/\nu_{\parallel}} P_S(1). \quad (3.25)$$

Definindo-se $\delta = \beta'/\nu_{\parallel}$, no ponto crítico tem-se que $P_S(t) \sim t^{-\delta}$. Seguindo a mesma ideia, é possível obter outras relações de escala em torno do ponto crítico [18]. Esses resultados são confirmados por análises numéricas e simulações computacionais.

Para $\lambda < \lambda_c$, tanto $P_S(t)$ quanto $n(t)$ decaem exponencialmente, enquanto que para $\lambda = \lambda_c$, $n(t)$ cresce com t^{η} . Com isso, o limiar crítico pode ser estimado pelas curvas $P_S(t) \times t$ e de $n(t) \times t$, como será mostrado na Seção 4.4.1.

Analogamente ao caso ferromagnético, os comprimentos de correlação espacial e temporal seguem as respectivas leis de potência:

$$\xi_{\perp} \sim \Delta^{-\nu_{\perp}} \quad (3.26)$$

e

$$\xi_{\parallel} \sim \Delta^{-\nu_{\parallel}}. \quad (3.27)$$

Todos os expoentes obtidos para o processo de contato correspondem à classe de universalidade de *percolação direcionada* (PD), que é um modelo simples de sistema sem termodinâmica que apresenta transição de fase. Nessa modelagem, simula-se a passagem de um líquido através ou de poros (vértices) ou de arestas (ligações), sendo cada um deles “molhados” com probabilidade p e existe uma direção preferencial para tal passagem. Se apenas considerando tais vértices ou ligações, forma-se uma componente conectada suficientemente grande, é dito que há a percolação do fluido. Para um valor crítico $p = p_c$, há uma transição de uma fase impermeável para uma permeável. É observado que os expoentes críticos do PC são os mesmos deste modelo, caracterizando, assim, uma classe de universalidade PD [16].

3.5 Estados e distribuições quase-estacionárias (QS)

Os estados nos quais a doença é erradicada são chamados de absorventes, já que uma vez que o sistema o tenha visitado, não ocorrem mudanças de estado. Para redes de tamanho finito, existe uma probabilidade não nula de visitar um desses estados em algum tempo, já que o número de configurações acessíveis é finito e, assim, o único estado verdadeiramente estacionário nesses sistemas é o absorvente.

Simulações só podem ser feitas em tamanhos finitos, e os efeitos dessas estruturas devem ser calculados. Próximo do limiar epidêmico, as flutuações dinâmicas se tornam muito grandes e esses efeitos tornam-se computacionalmente impossíveis de serem ignorados. Para contornar tal dificuldade, são utilizados os valores quase-estacionários (Figura 3.5): as médias dos parâmetros de ordem até um tempo t são feitas apenas em amostras que não visitaram a configuração absorvente até tal tempo, por meio das *amostras sobreviventes*.

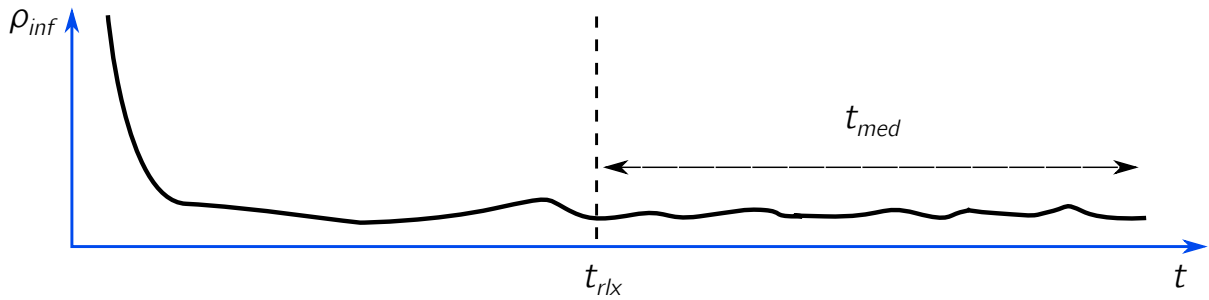


Figura 3.5: Após um certo tempo de relaxação t_{rlx} , $\bar{\rho}$ é calculado pela média de ρ ao longo de um tempo t_{med} .

Esta análise converge para o estado estacionário real no limite termodinâmico tanto para os estados super e subcríticos, já que para o primeiro caso não há visita ao estado absorvente, e no último a densidade média de vértices infectados converge para 0. Porém, limitar às amostras sobreviventes é ineficiente computacionalmente, obtendo dados muito ruidosos e, assim, os regimes estacionários não são facilmente identificados. Um método alternativo e eficiente foi proposto por de Oliveira e Dickman [19], e é aqui utilizado.

Para introduzir a análise QS, considere um processo Markoviano X_t^* , no qual o símbolo $*$ é usado para deixar claro que os estados absorventes estão presentes, e descrito por uma equação mestra como a da equação (3.1). Representando o conjunto dos estados absorventes por \circledast , $W_{\sigma \rightarrow \sigma'} = 0$ para $\sigma \in \circledast$. A probabilidade de sobrevivência $P_s(t)$, de que o sistema não visite o estado absorvente até um tempo t é dada por

$$P_s(t) = \sum_{\sigma \notin \circledast} P(\sigma, t) = 1 - \sum_{\sigma \in \circledast} P(\sigma, t), \quad (3.28)$$

e a distribuição QS para esse processo é definida por

$$\bar{P}(\sigma) = \lim_{t \rightarrow \infty} \frac{P(\sigma, t)}{P_s(t)}, \quad \sigma \notin \circledast. \quad (3.29)$$

É possível notar que a distribuição QS é normalizada. Somando a equação (3.1) sobre $\sigma \in \otimes$, para o qual $W_{\sigma \rightarrow \sigma'} = 0$, e usando a equação (3.28), temos:

$$\frac{dP_s}{dt} = - \sum_{\sigma \in \otimes} \sum_{\sigma' \notin \otimes} W_{\sigma' \rightarrow \sigma} P(\sigma', t) \equiv -W_*. \quad (3.30)$$

Note que W_* é a taxa total de transição para estados absorventes. Assumindo que a distribuição QS exista e é única, e substituindo $P(\sigma, t) = P_s(t) \bar{P}(\sigma)$ na equação (3.30):

$$\frac{1}{P_s} \frac{dP_s}{dt} = - \sum_{\sigma \in \otimes} \sum_{\sigma' \notin \otimes} W_{\sigma' \rightarrow \sigma} \bar{P}(\sigma') = -\bar{W}_*, \quad (3.31)$$

cuja solução é uma probabilidade de sobrevivência decaindo exponencialmente como $P_s(t) \sim e^{-t/\tau}$, com tempo de meia vida ou *lifespan*

$$\tau = \frac{1}{\bar{W}_*}. \quad (3.32)$$

Considere agora um processo estocástico X_t com todas as transições idênticas às de X_t^* , exceto para aquelas que envolvam os estados absorventes. Assim, a equação mestra é modificada para $\sigma \notin \otimes$ como:

$$\frac{dQ(\sigma, t)}{dt} = \sum_{\sigma' \notin \otimes} [W_{\sigma' \rightarrow \sigma} Q(\sigma', t) - W_{\sigma \rightarrow \sigma'} Q(\sigma, t)] + \widetilde{W}_*(t) Q(\sigma, t), \quad (3.33)$$

na qual $\widetilde{W}_*(t)$ é dada pela equação (3.30) com P substituído por Q . A solução estacionária da equação mestra (3.1) é uma solução estacionária da equação (3.33), afirmação que pode ser verificada substituindo $P(\sigma, t) = P_s(t) \bar{P}(\sigma)$, $\sigma \notin \otimes$, nas equações (3.1) e (3.30) para encontrar a seguinte equação autoconsistente:

$$\bar{P}(\sigma) = \frac{\sum_{\sigma' \notin \otimes} W_{\sigma' \rightarrow \sigma} \bar{P}(\sigma')}{W_\sigma - \bar{W}_*}, \quad (3.34)$$

com $W_\sigma = \sum_{\sigma'} W_{\sigma \rightarrow \sigma'}$. Igualando o lado direito da equação (3.33) a zero, correspondendo a uma configuração estacionária, encontramos:

$$\bar{Q}(\sigma) = \frac{\sum_{\sigma' \notin \otimes} W_{\sigma' \rightarrow \sigma} \bar{Q}(\sigma')}{W_\sigma - \widetilde{W}_*}, \quad (3.35)$$

que possui a mesma forma da equação (3.34), mostrando que a solução QS da equação mestra original é uma solução estacionária da equação (3.33) do processo modificado.

Outros termos de fonte podem ser aplicados à equação mestra, implicando em outros tipos de análises QS [20], porém neste trabalho utilizaremos apenas o método explicado acima. Note que a equação (3.33) não é linear e, portanto, não é uma equação mestra propriamente dita. Um método de solução de tal equação será explicado na Seção 4.3.

Capítulo IV

Simulação de processos epidêmicos

Na análise computacional de processos epidêmicos em redes complexas, é necessário definir *algoritmos* para gerar tanto as redes quanto para simular os processos dinâmicos. Este capítulo é dedicado a isso, estabelecendo alguns conceitos e métodos computacionais necessários para este trabalho. As otimizações de algoritmos para simulação de processos Markovianos gerais serão feitas no próximo capítulo, além da particularização para os modelos epidêmicos definidos no Capítulo III.

4.1 Construção de redes complexas

Para redes esparsas, que é o caso da maior parte das redes reais, nas quais os vértices estão bem longe de serem conectados com todos os demais entre si, a matriz \tilde{A} também é esparsa, algo completamente ineficiente de ser armazenado na memória de um computador, como exemplificado na Seção 2.2.1. Uma forma de lidar computacionalmente com essas redes é em termos de uma *lista de adjacência* [3]. Conhecendo o conjunto k_i de graus dos vértices i , a lista de adjacência é definida como um *array*¹ unidimensional $\ell = \{\ell_s\}$ com $s = 1, \dots, \sum_{i=1}^N k_i$ sendo o índice dos elementos de ℓ . Os vizinhos do vértice i são armazenados entre s_i e $s_i + k_i - 1$, com $s_1 = 1$ e $s_i = 1 + \sum_{j=1}^{i-1} k_j$ para $i > 1$. Um exemplo está mostrado na Figura 4.1.

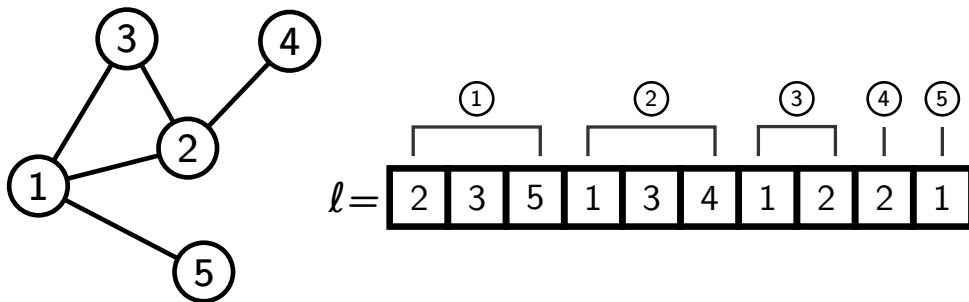


Figura 4.1: Exemplo de uma rede simples e sua respectiva lista de adjacência. Os elementos correspondentes a cada vértice estão indicados.

¹ Em computação, um *array* é um objeto que contém elementos indexáveis de mesmo tamanho e tipo.

Por conveniência, um *array* $s = \{s_i\}$ também é salva ao longo da simulação, cujos elementos são os valores s_i citados anteriormente, para $i = 1, \dots, N$. O algoritmo utilizado neste trabalho para construção de redes complexas é o **modelo de configurações**, que segue as seguintes regras:

1. Estabelecido o tamanho N da rede desejada, o grau de cada vértice é distribuído aleatoriamente seguindo uma distribuição $P(k)$, de forma que $\sum_{i=1}^N k_i$ seja par: não podem sobrar pontas de arestas desconectadas.
2. A cada passo, são sorteados dois *stubs* ainda não ligados e os respectivos vértices são conectados, ignorando *autoconexões* (um vértice não pode se conectar a ele mesmo) e *múltiplas conexões* (um vértice não pode ter mais de uma aresta conectada com o mesmo vizinho).
3. Quando todos os vértices estiverem conectados, a rede está formada.

Qualquer rede com tamanho finito N apresenta um corte estrutural k_c na sua distribuição de conectividade. Assim, os requisitos de não haver ligações “soltas”, nem autoconexão e nem múltiplas conexões pode fazer com que o algoritmo não consiga conectar toda a rede. Isso ocorre para redes com $\langle k^2 \rangle$ que diverge para $N \rightarrow \infty$, como em distribuições em lei de potência para $2 < \gamma < 3$.

Há algumas formas de contornar isso, como fazer as conexões a partir de *stubs* dos vértices de maior grau primeiramente, em vez de ser em sequência aleatória. Normalmente, tentamos um número finito de vezes fazer as conexões com a mesma sequência $\{k_i\}$. Caso não haja sucesso, uma nova sequência é gerada a partir de $P(k)$, até que gere uma rede conectada. Outra forma é simplesmente descartar os *stubs* não conectados, e utilizar apenas a componente gigante.

Porém, tudo isso leva a correlações de grau no sistema. Para contornar este problema, é utilizado o modelo UCM, já citado na Seção 2.3. Nele, um corte $k_c = \sqrt{N}$ é estabelecido na distribuição $P(k)$, gerando redes não correlacionadas e completamente conectadas, isto é, com apenas uma componente.

Neste trabalho, este modelo é utilizado para os seguintes tipos de redes aleatórias:

- **Rede UCM:** A distribuição de grau é em lei de potência $P(k) \sim k^{-\gamma}$, com $k = (k_0 = 3), \dots, (k_c = \sqrt{N})$, gerando redes altamente heterogêneas e não correlacionadas.
- **Rede exponencial:** O grau é dado por $k = k_0 + k'$, com $k_0 = 3$ e k' dado por uma distribuição $P(k') = a \exp(-k'/a)$, com $a = 3$ e $k' \geq 0$. É importante notar que $a = \langle k' \rangle$ e, portanto, $\langle k \rangle = k_0 + a = 6$.
- **Random Regular Network (RRN) de grau k :** Os graus de todos os vértices são iguais, de mesmo valor k . Este algoritmo equivale a uma rede completamente aleatória, com a única condição de todos os vértices terem o mesmo número de ligações, como na Figura 4.2 (a).

4.1.1 Outras redes utilizadas neste trabalho

Outros tipos de redes são utilizadas neste trabalho, apresentadas nas Figura 4.2 (b), (c) e (d), são elas:

- **Grafo estrela:** Um único vértice está conectado aos outros $N - 1$ vértices, chamados de *folhas*. O vértice central é chamado de estrela e é claramente um *hub*.
- **Grafo roda:** $N - 1$ vértices formam um anel unidimensional, enquanto um vértice central é conectado com todos esses $N - 1$ vértices. Enquanto o grafo estrela possui coeficiente de agregação nulo, este grafo possui um valor alto, devido à formação de triângulos.
- **Grafo completo:** Todos os N vértices estão conectados aos outros $N - 1$ vértices.

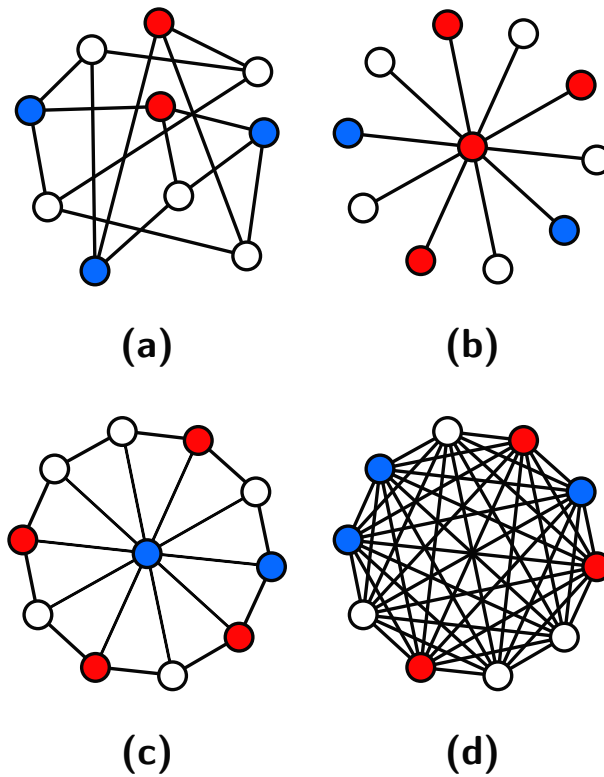


Figura 4.2: Alguns grafos simples: (a) RRN com $k = 3$, (b) estrela, (c) roda, e (d) completo.

4.2 Algoritmo de Gillespie

Geralmente as simulações de dinâmicas Markovianas são feitas utilizando atualizações *síncronas*, nas quais todos os vértices são alterados simultaneamente. Por outro lado, podem também ser feitas *assincronicamente*, sendo um vértice atualizado a cada passo de tempo. De forma geral, processos Markovianos podem ser decompostos em um conjunto de L processos de Poisson independentes $p = 1, 2, \dots, L$, com taxas $\alpha_1, \alpha_2, \dots, \alpha_L$. Tais processos foram definidos na Seção 3.2.

No modelo SIS, por exemplo, cada vértice i se torna suscetível espontaneamente com taxa $\mu_i = \mu$, enquanto um vértice infectado transmite a doença para cada um de seus vizinhos com taxa λ . Então, um processo de Poisson pode ser definido para cada vértice tal que $\alpha_i = \mu$ se o vértice i está infectado e $\alpha_i = \lambda n_i$, se i está suscetível, sendo n_i o número de vizinhos infectados de i . Note que há processos redundantes com $n_i = 0$, para os quais a infecção não ocorre.

Para um intervalo infinitesimal dt , a probabilidade de que um vértice i mude seu estado é $P_{\alpha_i} = \alpha_i dt$. Em uma simulação síncrona, o tempo é discretizado em intervalos finitos δt e, com isso, há grande discrepância com a versão contínua se esses valores não são suficientemente pequenos [21]. Já para o caso de atualização assíncrona, a simulação é inspirada no *Algoritmo de Gillespie*, que gera uma trajetória estatisticamente correta de uma equação estocástica. Popularizado em dois artigos publicados em 1976 [22] e 1977 [23], foi inicialmente proposto para resolver equações de reações químicas, utilizando procedimentos de Monte Carlo² para simular a evolução temporal de um sistema químico.

Nessas simulações, cada vértice pode estar em um dos possíveis estados e a cada passo de tempo a evolução dinâmica depende da interação microscópica do vértice com seus vizinhos. Seja $\tilde{P}(\tau)$ a probabilidade de que nenhum evento ocorra até o tempo τ . Como mostrado na Seção 3.2, a probabilidade que não ocorra o evento p no tempo τ é

$$\tilde{P}_p(\tau) = P_p(0, \tau) = e^{-\alpha_p \tau}. \quad (4.1)$$

Seja $W(\tau, p)dt$ a probabilidade de que o evento p seja o próximo a ocorrer em um intervalo de tempo infinitesimal $(t + \tau, t + \tau + dt)$. Ela é dada pela probabilidade de nenhum evento ocorrer durante o tempo τ , e de ocorrer o evento p após um intervalo infinitesimal dt :

$$W(\tau, p)dt = \tilde{P}(\tau)P_{\alpha_p} = \tilde{P}(\tau)\alpha_p dt. \quad (4.2)$$

Como *nenhum* evento pode ocorrer, e estes são independentes entre si, a probabilidade $\tilde{P}(\tau)$ é dada pelo produto de $\tilde{P}_p(\tau)$ ao longo de todos os processos p :

$$\tilde{P}(\tau) = \prod_{p=1}^L \tilde{P}_p(\tau) = \prod_{p=1}^L e^{-\alpha_p \tau} = e^{-\sum_{p=1}^L \alpha_p \tau} = e^{-R\tau}, \quad (4.3)$$

na qual $R = \sum_{p=1}^L \alpha_p$ é a taxa total de transições. Assim,

$$W(\tau, p)dt = e^{-R\tau} \alpha_p dt = (R e^{-R\tau}) \cdot \left(\frac{\alpha_p}{R} \right) dt \equiv P(\tau) \cdot \Pi(p) dt, \quad (4.4)$$

sendo

$$P(\tau) = R e^{-R\tau} \quad (4.5)$$

² O método de Monte Carlo consiste em obter repetidas amostragens de uma dinâmica a partir de sequências de números aleatórios.

a densidade de probabilidade do tamanho do passo de tempo ser τ , e

$$\Pi(p) = \alpha_p / R \quad (4.6)$$

a probabilidade desse evento ser o p . Desta forma, a ideia central do algoritmo de Gillespie é que a cada passo de tempo, basta escolher um dos eventos p com probabilidade $\Pi(p)$ e o passo de tempo τ a partir da distribuição $P(\tau)$.

Em princípio, tal algoritmo é aplicável para sistemas Markovianos que possam ser decompostos em L processos de Poisson independentes. Porém, toda vez que o estado de um vértice muda, a lista de tais processos também deve ser atualizada, o que é computacionalmente ineficiente para sistemas muito grandes.

4.3 Simulação do estado quase-estacionário

Como já dito na Seção 3.5, a equação (3.33) não é uma equação mestra propriamente dita. Porém, o método proposto por de Oliveira e Dickman [19] é capaz de solucioná-la numericamente. O último termo do lado direito da equação (3.33) representa a transição total para estados absorventes na dinâmica original que são redirecionadas para estados ativos proporcionalmente à probabilidade de que eles ocorram na dinâmica modificada. Uma interpretação está apresentada na Figura 4.3. Neste caso, o estado ativo $\sigma \notin *$ é escolhido com probabilidade $Q(\sigma, t)$, mas Q não é conhecido a priori.

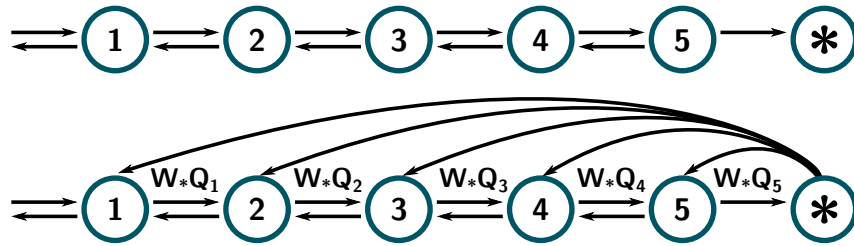


Figura 4.3: Representação do método de simulação QS. Na dinâmica original, mostrada na figura do topo, o sistema sequencialmente visita os estados representados por $1, 2, 3, \dots, *$, onde $*$ é um estado absorvente. As transições $i \rightarrow i \pm 1$ são reversíveis, mas não aquelas para $*$. Na dinâmica modificada, entretanto, os estados absorventes são redirecionados para configurações ativas, visitados durante a evolução temporal proporcionalmente à probabilidade Q_i de que o sistema esteja no estado i .

Uma forma de contornar tal problema é utilizar o próprio histórico da simulação para estimar $Q(\sigma, t)$. Isso é possível ao construir uma lista com M configurações ativas visitadas ao longo da simulação, atualizada de forma conveniente. Escolhendo aleatoriamente uma dessas configurações e a substituindo pela configuração atual com probabilidade $p_r \Delta t$, na qual p_r é a probabilidade por unidade de tempo de substituição das configurações e Δt o passo de tempo unitário. Quando o sistema visita um estado absorvente, ele adquire aleatoriamente uma das M configurações. Tal método reproduz com muita precisão os

resultados para o PC utilizando o método padrão [24, 25], e a convergência para o estado QS real foi investigada analiticamente na referência [26].

No início da simulação, as primeiras M configurações podem ser arbitrariamente escolhidas entre as configurações ativas visitadas nos primeiros M passos da simulação. Um valor grande de p_r pode ser utilizado nesse regime para acelerar o “esquecimento” da condição inicial. Depois de certo tempo, a taxa p_r deve depender da frequência com a qual o sistema visita o estado absorvente, sendo uma forma simples e eficiente assumir que seja igual à taxa de transição para estados pré-absorventes, estimada ao longo do processo, mas sempre maior que um valor mínimo, garantindo a troca de configurações. Após um tempo de relaxação t_{rlx} , as médias quase-estacionárias são computadas por um tempo t_{med} . O número M é restrito à memória computacional: para $M = 100$ e $N = 10^8$, em uma rede em lei de potência com $\gamma = 3,5$, uma simulação típica do modelo SIS pelos algoritmos II e III, a serem descritos na Seção 5.2, demandam 4 Gb e 6 Gb, respectivamente, incluindo a alocação da lista de adjacência.

4.4 Tipos de análises e medidas

Finalmente, podem ser definidos tipos básicos de simulação para a análise da propagação de epidemias em redes complexas. Análises de espalhamento e de decaimento são feitas em função do tempo t , e as médias dinâmicas são feitas sobre *todas* as amostras, independente de terem atingido ou não o estado absorvente. Já na análise do estado quase-estacionário, apenas as amostras sobreviventes são utilizadas. Como já apresentado na seção anterior, o método QS é utilizado para o último caso.

4.4.1 Análise de espalhamento

Neste tipo de simulação, a condição inicial é um único vértice infectado, enquanto os demais estão suscetíveis. Para $\lambda < \lambda_c$, o regime é **subcrítico**: o número médio de vértices infectados $n(t)$ decai exponencialmente após um ligeiro crescimento devido à difusão. No regime **supercrítico**, no qual $\lambda > \lambda_c$, as amostras sobrevivem indefinidamente. Já no limiar crítico $\lambda = \lambda_c$, o processo tem probabilidade 1 de alcançar o estado absorvente em algum instante, porém o tempo de meia vida diverge. Com isso, há um crescimento em lei de potência com $n(t) \sim t^\eta$, como na Figura 4.4.

4.4.2 Análise de decaimento

O decaimento pode ser analisado ao considerar uma condição inicial com uma fração ou todos os vértices inicialmente infectados e observando a evolução temporal da densidade média de vértices infectados ρ . No regime subcrítico, o decaimento assintótico é exponencial. Já no regime supercrítico, ρ atinge valores estacionários após um longo tempo

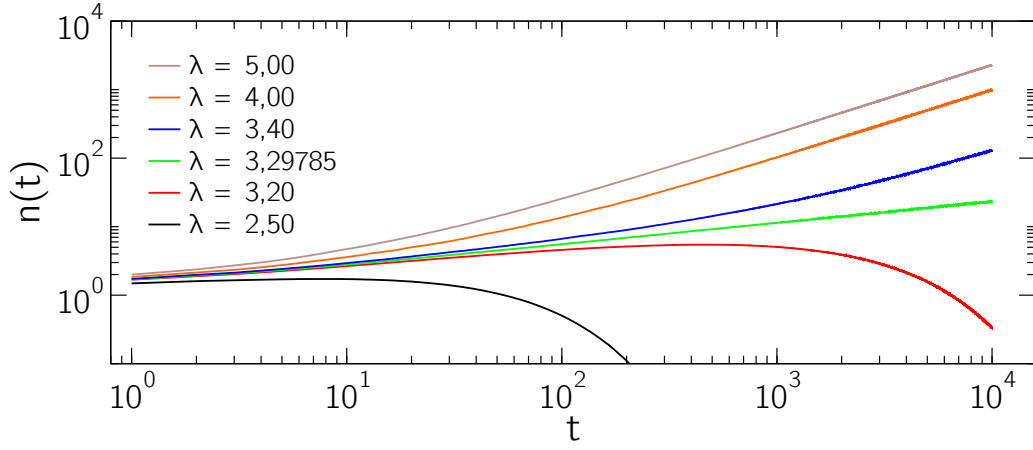


Figura 4.4: Evolução temporal do número médio de vértices infectados em uma análise de espalhamento para o Processo de Contato em uma rede regular unidimensional com $N = 10^4$, para várias taxas de infecção λ . Nesse modelo, $\lambda_c \approx 3,29785$, e para $\lambda = \lambda_c$ a curva cresce com $\sim t^\eta$ com $\eta = 0,31368(4)$ [16]. O número de amostras utilizadas foi de 3000 a 10^6 , para valores de λ em ordem decrescente.

quando $N \rightarrow \infty$, enquanto que em $\lambda = \lambda_c$ assume lei de potência $\rho(t) \sim t^{-\delta}$ até cair para o estado absorvente. Redes de tamanho finito apresentam apenas o estado absorvente como verdadeiramente estacionário, sendo apropriada uma análise quase-estacionária.

4.4.3 Análise quase-estacionária

A forma padrão de estudar o estado quase-estacionário é utilizando as análises de decaimento nas quais as amostras não sobreviventes são ignoradas para cada tempo t . No entanto, neste trabalho utilizamos o método QS. As médias são feitas após um tempo de relaxação t_{rlx} por um intervalo t_{med} , como exemplificado na Figura 3.5. Os gráficos obtidos são das quantidades em função do parâmetro de controle λ . É importante salientar que em gráficos em função de λ , ρ denota a densidade quase-estacionária média.

Muitas das quantidades de interesse dependem do número de vértices infectados, sendo útil definir a probabilidade QS \bar{P}_n de que haja n vértices infectados na rede. Em muitos modelos epidêmicos, os estados pré-absorventes possuem apenas um vértice ativo, implicando que a taxa que o sistema visita o estado absorvente é $\bar{W}_* = \mu \bar{P}_1$. Com $\mu = 1$, temos $\tau = 1/\bar{P}_1$. Uma vez computada a probabilidade QS, as quantidades básicas são calculadas em termos das distribuições, como:

$$\langle \rho^s \rangle = \frac{1}{N^s} \sum_n n^s \bar{P}_n. \quad (4.7)$$

Além disso, a suscetibilidade [27] é definida em termos da variância relativa de ρ :

$$\chi = N \frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle}. \quad (4.8)$$

Junto com $\langle \rho \rangle$ e χ , o tempo de meia vida τ da equação (3.32) são as quantidades

básicas medidas nesta análise e são utilizadas para validar os métodos de simulação aqui apresentados.

Para redes de tamanho finito, os limiares epidêmicos são geralmente dependentes de N , como no modelo SIS [28], ou possuem fortes correções como no modelo PC [27]. As simulações quase-estacionárias são utilizadas para estimar o *limiar efetivo* utilizando o máximo da suscetibilidade χ [27, 28]. A Figura 4.5 mostra a suscetibilidade no modelo SIS para diferentes tipos de rede, sendo possível observar que posição dos picos de χ correspondem ao ponto a partir do qual ρ começa a crescer partindo do zero. A suscetibilidade também pode apresentar mais de um pico, associados a múltiplas transições [7]. O tempo de meia vida diverge com o tamanho muito rápido na fase ativa e, assim, é utilizado para definir qual dos picos de χ corresponde à transição de fase global do sistema.

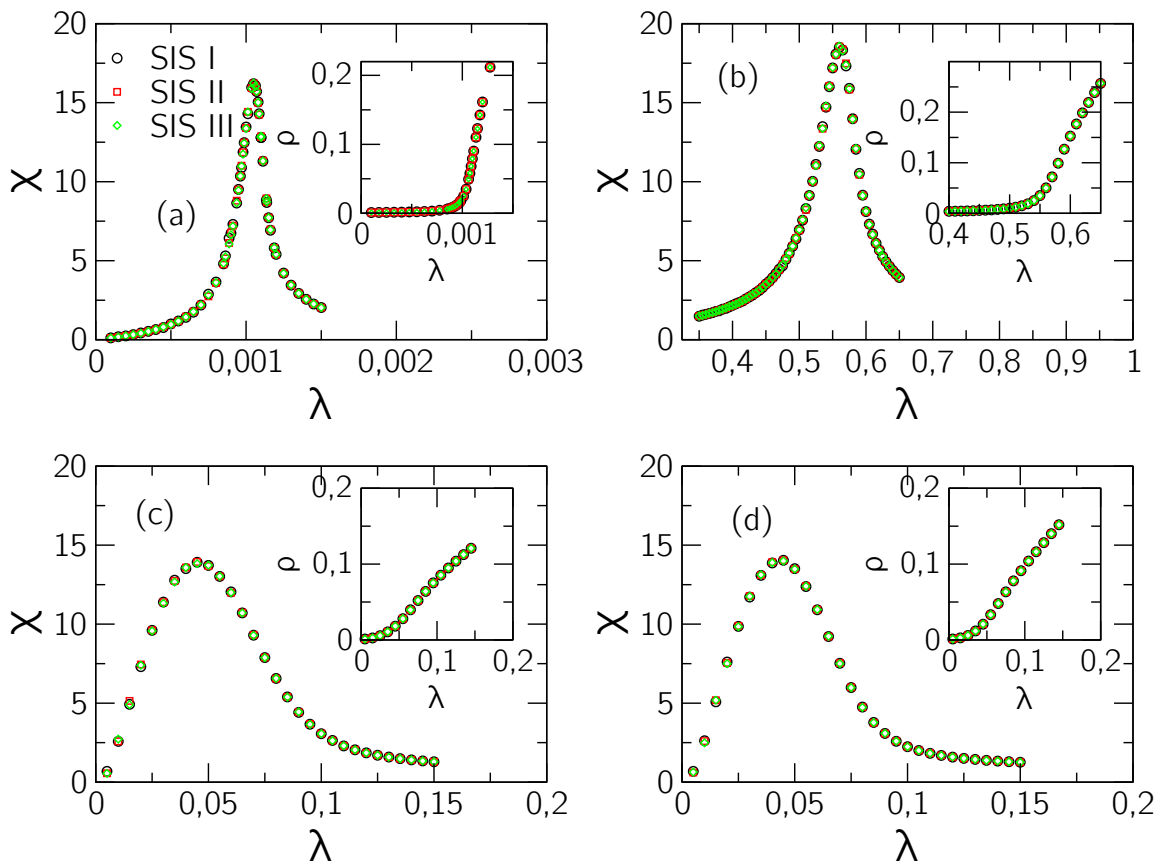


Figura 4.5: Suscetibilidade versus taxa de infecção para o modelo SIS em diferentes tipos de rede com $N = 10^3$: (a) grafo completo, (b) RRN com $k = 3$, (c) grafo estrela e (d) grafo roda. As inserções correspondem às densidades quase-estacionárias. Os algoritmos SIS-I, SIS-II e SIS-III serão definidos no Capítulo V.

É importante salientar que em todos esses tipos de simulação, um tempo $t_{\text{máx}}$ é definido, já que há regimes nos quais a dinâmica pode evoluir indefinidamente — no método QS, por exemplo. Todos esses tipos de análises foram feitos e serão apresentados resultados para diferentes redes no Capítulo VI.

Capítulo V

Otimização de algoritmos e aplicações

Considere uma dinâmica com os três estados epidemiológicos: suscetível ($\sigma_i = 0$), que pode ser infectado por outro vértice; infectado ($\sigma_i = 1$), que transmite a doença para um vértice suscetível; e recuperado ($\sigma_i = 2$), o qual não pode nem transmitir nem adquirir a doença, embora possa se tornar suscetível novamente. A generalização para outros estados é direta, mas nos limitaremos aos modelos definidos no Capítulo III.

5.1 Algoritmos gerais

Considere que σ represente os estados de cada vértice. Vamos assumir o caso geral no qual um vértice infectado i torna-se espontaneamente recuperado com taxa μ_i , transmite a infecção para um vértice j com taxa λ_{ij} e, se recuperado, torna-se suscetível com taxa α_i . Por convenção, $\lambda_{ij} = 0$ para $A_{ij} = 0$. Como exemplo, se $\mu_i = \mu$, $\lambda_{ij} = \lambda A_{ij}$, temos os modelos clássicos SIS e SIR, para $\alpha_i = \infty$ e $\alpha_i = 0$, respectivamente.

5.1.1 Algoritmo padrão

Ao fazer simulações estocásticas, precisamos converter as taxas em probabilidades, preservando o número de acontecimentos por unidade de tempo, agora discretizado. Neste trabalho, é desenvolvido um algoritmo de atualização assíncrona equivalente à síncrona quando os intervalos de tempo são infinitesimalmente pequenos.

Uma dinâmica Markoviana em uma rede de tamanho N pode ser simulada de maneira simples: selecione um vértice i qualquer, com igual chance. Se i está infectado, torna-se recuperado com probabilidade

$$p_i = \mu_i \zeta. \quad (5.1)$$

Caso esteja no estado recuperado, fica suscetível com probabilidade

$$r_i = \alpha_i \zeta. \quad (5.2)$$

E finalmente, se o vértice está suscetível, torna-se infectado com probabilidade

$$q_i = \sum_{j=1}^N \lambda_{ji} \delta(\sigma_j, 1) \zeta, \quad (5.3)$$

na qual $\delta(a, b)$ é a delta de Kronecker. O tempo é incrementado aleatoriamente por um intervalo de tempo com uma distribuição exponencial com média δt . Escolhemos $\delta t = \zeta/N$, no qual o fator $1/N$ conta a escolha de um único vértice por passo de tempo. Se a rede é suficientemente grande, com $\delta t \ll 1$, independente de ζ , pode ser escolhido o maior valor possível de ζ garantindo que $p_i, r_i, q_i \leq 1$. Uma boa escolha, mas nem sempre eficiente é

$$\zeta = \frac{1}{\max\{\lambda_{ij} + \alpha_i + \mu_i\} k_{\max}}, \quad (5.4)$$

na qual $\max\{\lambda_{ij} + \alpha_i + \mu_i\}$ é o valor máximo de $\lambda_{ij} + \alpha_i + \mu_i$ entre todos os vértices i e j da rede. No caso de uma taxa infinita, basta retirá-la da relação anterior, sendo tal transição descartada da simulação (é instantânea). Um exemplo é quando não existe o estado recuperado ($\alpha_i = \infty$):

$$\zeta = \frac{1}{\max\{\lambda_{ij} + \mu_i\} k_{\max}}, \quad (5.5)$$

Note que tal algoritmo não é otimizado, já que ocorrem casos em que a dinâmica não evolui: um vértice suscetível pode ser selecionado e não possuir algum vizinho infectado. Portanto, este não é um algoritmo de Gillespie, que considera apenas as transições que de fato modificam os estados. Este algoritmo é mais fácil de implementar e é estatisticamente exato, já que basta assumir que existem pseudotransições que não mudam os estados. Porém, é possível otimizar ainda mais os algoritmos ao nos restringirmos aos vértices que estão ativamente participando do processo epidêmico.

Além disso, as probabilidades podem ser bem pequenas se k_{\max} é muito grande e as taxas muito pequenas. Isso ocorre para o modelo SIS em um grafo completo, por exemplo. Neste caso, o gerador de números aleatórios deve ser capaz de gerar números próximos de zero. O efeito disso também foi investigado, e será brevemente tratado na Seção 6.1.

5.1.2 Algoritmo otimizado

Uma forma eficiente de simular tais modelos é manter listas que mostrem as regiões passíveis de mudança dinâmica, fazendo a aplicação de tais regras mais diretas. Para o nosso propósito, guardemos duas listas: $\mathcal{P}^{(\text{inf})} = \{\mathcal{P}_j^{(\text{inf})}\}$ e $\mathcal{P}^{(\text{rec})} = \{\mathcal{P}_j^{(\text{rec})}\}$, com as posições dos N_{inf} vértices infectados ($j = 1, \dots, N_{\text{inf}}$) e dos N_{rec} vértices recuperados ($j = 1, \dots, N_{\text{rec}}$), respectivamente.

A atualização dessas listas deve ser simples e eficiente. Se um vértice é infectado,

este deve ser acrescentado à lista $\mathcal{P}^{(\text{inf})}$ na posição $N_{\text{inf}} + 1$, e $N_{\text{inf}} \leftarrow N_{\text{inf}} + 1$, no qual o símbolo \leftarrow significa que a variável N_{inf} recebe o valor $N_{\text{inf}} + 1$. Se um vértice se torna suscetível, sua posição na lista é substituída pelo elemento em N_{inf} , e $N_{\text{inf}} \leftarrow N_{\text{inf}} - 1$. O mesmo procedimento vale para $\mathcal{P}^{(\text{rec})}$, de forma análoga.

A taxa total P com que vértices infectados são recuperados, e a taxa total R com que os recuperados ficam suscetíveis são:

$$P = \sum_i \mu_i \delta(\sigma_i, 1), \quad (5.6)$$

e

$$R = \sum_i \alpha_i \delta(\sigma_i, 2) \quad (5.7)$$

respectivamente. Analogamente, a taxa total com que os suscetíveis se tornam infectados é

$$L = \sum_{ij} \delta(\sigma_i, 1) \delta(\sigma_j, 0) \lambda_{ij} = Q - F, \quad (5.8)$$

na qual

$$Q = \sum_{ij} \lambda_{ij} \delta(\sigma_i, 1) \quad (5.9)$$

é a taxa total de infecção sendo transmitida por vértices infectados ao longo das ligações, enquanto

$$F = \sum_{ij} \lambda_{ij} \delta(\sigma_i, 1) [\delta(\sigma_j, 1) + \delta(\sigma_j, 2)] \quad (5.10)$$

é a taxa total com a qual as tentativas de infecção são direcionadas a vértices infectados e recuperados, que são frustradas. Assim, o algoritmo otimizado é implementado da seguinte forma: com o passo de tempo sendo dado por

$$\delta t = \frac{1}{P + R + Q}, \quad (5.11)$$

com probabilidade

$$p = \frac{P}{P + R + Q} \quad (5.12)$$

um vértice infectado i é selecionado com probabilidade proporcional a μ_i e fica recuperado. Já com probabilidade

$$r = \frac{R}{P + R + Q} \quad (5.13)$$

um vértice recuperado i é escolhido com probabilidade proporcional a α_i e torna-se suscetível. Com probabilidade complementar

$$q = \frac{Q}{P + R + Q}, \quad (5.14)$$

um vértice infectado i é selecionado com probabilidade proporcional a sua taxa de infecção

$$w_i = \sum_j \lambda_{ij} \quad (5.15)$$

e um de seus k_i vizinhos é selecionado ao acaso com probabilidade proporcional a λ_{ij} , dentre todos os λ_{ij} com i fixo, sendo $j \in V(i)$. Se suscetível, torna-se infectado e, caso contrário, o tempo apenas evolui, sem ocorrer qualquer transição. O tempo é incrementado por τ seguindo uma distribuição $P(\tau) = (1/\delta t) \exp(-\tau/\delta t)$, semelhante à equação (4.5), já que δt é o inverso da soma das taxas de transições envolvidas. As infecções frustradas já estão contabilizadas na taxa F . Da mesma forma que no algoritmo padrão, caso uma das taxas seja infinita, basta removê-la das relações anteriores e fazer as transições necessárias serem instantâneas.

Note que os incrementos são comumente assumidos como sendo o valor esperado δt no lugar da distribuição exponencial com média δt . Isso foi verificado e a diferença é irrelevante tanto para análise QS quanto para as simulações de espalhamento com várias amostras.

Os modelos epidêmicos clássicos CP, SIS e SIR foram simulados utilizando diferentes algoritmos, baseados nos anteriormente apresentados, com as taxas iguais para todos os vértices.

5.2 Aplicação no modelo SIS

Denominamos SIS-I o algoritmo padrão modificado para as taxas deste modelo, e SIS-II e SIS-III os algoritmos baseados na versão otimizada.

SIS-I

Selecionamos um vértice i aleatoriamente e as probabilidades de cura e de infecção tornam-se, respectivamente,

$$p_i = \frac{\mu}{(\mu + \lambda)k_{\max}} \quad (5.16)$$

e

$$q_i = \frac{\lambda n_i}{(\mu + \lambda)k_{\text{máx}}}, \quad (5.17)$$

onde $n_i = \sum_j A_{ij}\delta(\sigma_j, 1)$ é o número de vizinhos infectados do vértice i , computado toda vez que tal vértice é selecionado. O passo de tempo médio é

$$\delta t = \frac{1}{(\mu + \lambda)Nk_{\text{máx}}}. \quad (5.18)$$

SIS-II

Nesta versão, as taxas totais tornam-se $P = \mu N_{\text{inf}}$, $R = \infty$ e $Q = \lambda N_e$, onde $N_e = \sum_{ij} A_{ij}\delta(\sigma_i, 1) = \sum_i k_i\delta(\sigma_i, 1)$ é o número de pontas de arestas *saindo* de vértices infectados e $N_{\text{inf}} = \sum_i \delta(\sigma_i, 1)$ é o número total de vértices infectados na rede. Então, com probabilidade

$$p = \frac{\mu N_{\text{inf}}}{\mu N_{\text{inf}} + \lambda N_e} \quad (5.19)$$

um vértice infectado i é selecionado com igual chance usando a lista $\mathcal{P}^{(\text{inf})}$ e torna-se suscetível diretamente ($\alpha_i = \infty$). Com probabilidade complementar

$$q = 1 - p = \frac{\lambda N_e}{\mu N_{\text{inf}} + \lambda N_e} \quad (5.20)$$

um vértice infectado i é selecionado aleatoriamente da mesma lista, mas aceito com probabilidade $k_i/k_{\text{máx}}$, já que $w_i = \lambda k_i$. A tentativa de seleção é feita até obter sucesso. Como a taxa de infecção é a mesma para todas as arestas, um vizinho de i é selecionado com igual chance e torna-se infectado se suscetível. As listas são atualizadas e o passo de tempo médio é

$$\delta t = \frac{1}{\mu N_{\text{inf}} + \lambda N_e}. \quad (5.21)$$

SIS-III

O algoritmo é o mesmo do SIS-II, com a diferença no modo de fazer a seleção de um vértice infectado i para a tentativa de infecção. No lugar de aceitar com probabilidade $k_i/k_{\text{máx}}$, é mantida uma lista $\mathcal{L} = \{\mathcal{L}_j\}$ com k_i cópias de cada vértice i infectado, e é a partir dela que a seleção é feita. É necessário tomar certo cuidado na implementação

desta lista, já que em cada processo de infecção e de cura devemos remover todas as k_i cópias ou garantir que há k_i cópias na lista.

Para fazer isso, uma lista auxiliar $c = \{c_i\}$ é mantida com o número de cópias c_i de cada vértice i salvos entre as posições $j = 1, \dots, N_{\text{cópias}}$ de \mathcal{L} , sendo¹ $N_{\text{cópias}} = \sum_i c_i$. Então, para a cura, um vértice infectado é escolhido a partir da lista $\mathcal{P}^{(\text{inf})}$, como no algoritmo SIS-II. O número de arestas N_e é atualizado como $N_e \leftarrow N_e - k_i$, mas \mathcal{L} não é alterada. Assim, \mathcal{L} ainda contém k_i cópias de um vértice i que acabou de ser curado, isto é, contém vértices que estão suscetíveis.

No processo de infecção, um vértice i é obtido a partir de um elemento de índice j em \mathcal{L} , e não de $\mathcal{P}^{(\text{inf})}$. Assim, se na posição j há um vértice a suscetível, este é removido de \mathcal{L} , sendo substituído pelo elemento na posição $j = N_{\text{cópias}}$ de \mathcal{L} . Assim, $N_{\text{cópias}} \leftarrow N_{\text{cópias}} - 1$ e $c_a \leftarrow c_a - 1$. O processo é realizado até que uma posição j selecionada seja correspondente a um vértice infectado. No caso de sucesso de infecção de um vértice i , são adicionadas $k_i - c_i$ cópias ao final da lista \mathcal{L} , e $c_i \leftarrow k_i$. Note que assim sempre haverá k_i cópias de vértices infectados, sendo o único fator limitante o processo de remoção das cópias de vértices suscetíveis. O restante segue como no SIS-II.

5.3 Aplicação no Processo de Contato

No processo de contato, cada vértice infectado tenta transmitir a infecção para um vizinho selecionado aleatoriamente com taxa λ , o que significa que a taxa de transmissão de um vértice i para um vértice j é $\lambda_{ij} = \lambda A_{ij}/k_i$.

Denominamos CP-I o algoritmo padrão modificado para tal modelo, e CP-II o algoritmo otimizado, adotando a abreviação do nome em inglês do modelo: *Contact Process* (CP). São eles:

CP-I

Um vértice é selecionado aleatoriamente com igual chance. Se o vértice está infectado, com probabilidade $p_i = \mu\zeta$ fica suscetível. Caso contrário, isto é, se é suscetível, fica infectado com probabilidade

$$q_i = \left[\lambda \sum_{j=1}^N \frac{\delta(\sigma_j, 1)}{k_j} A_{ij} \right] \zeta. \quad (5.22)$$

E o passo de tempo médio é $\delta t = \zeta/N$, sendo

$$\zeta = \frac{1}{\mu + \lambda} \quad (5.23)$$

¹ Não confundir c_i com o coeficiente de agregação definido no Capítulo II.

uma boa escolha, sem necessidade do fator k_{\max} no denominador.

CP-II

Neste caso, temos $P = \mu N_{\text{inf}}$. A taxa total de criação é

$$L = \sum_{ij} (1 - \sigma_j) A_{ij} \sigma_i \frac{\lambda}{k_i} = Q - F, \quad (5.24)$$

onde $Q = \lambda N_{\text{inf}}$ é a taxa total de infecção a partir de arestas emanando de vértices infectados, e $F = \lambda \sum_{ij} \sigma_i \sigma_j A_{ij} / k_i$ é a taxa total de infecções frustradas. Como o estado recuperado é ausente, $R = \infty$ e note que $w_i = \lambda$ e que λ_{ij} é independente do alvo j : a probabilidade de aceite de um vértice infectado e de um de seus vizinhos é 1, resultando no algoritmo comumente utilizado na literatura [16, 29]. Utilizando a lista $\mathcal{P}^{(\text{inf})}$, um vértice infectado i é selecionado aleatoriamente e, com probabilidade

$$p = \frac{\mu}{\mu + \lambda}, \quad (5.25)$$

o vértice selecionado é curado e, assim, fica suscetível. Com probabilidade complementar

$$q = \frac{\lambda}{\mu + \lambda} \quad (5.26)$$

um dos k_i vizinhos de i é aleatoriamente selecionado com igual chance e, se suscetível, torna-se infectado. O passo de tempo médio é

$$\delta t = \frac{1}{N_{\text{inf}}(\mu + \lambda)}. \quad (5.27)$$

5.4 Aplicação no modelo SIR

Escolhendo $\alpha_i = 0$, $\lambda_{ij} = \lambda$ e $\mu_i = \mu$ para todos os vértices, temos o modelo clássico SIR. Os algoritmos SIR-I, SIR-II e SIR-III são equivalentes às versões correspondentes SIS, com a única diferença de que a transição do estado infectado para o suscetível torna-se do infectado ($\sigma_i = 1$) para o recuperado ($\sigma_i = 2$) e permanece neste estado para sempre. Note que a lista $\mathcal{P}^{(\text{rec})}$ é desnecessária para os algoritmos SIR-II e SIR-III, já que os vértices recuperados não possuem dinâmica.

Todos esses algoritmos foram utilizados e comparados para avaliar a eficiência computacional de tais métodos, e no próximo capítulo tais resultados são apresentados.

Capítulo VI

Comparações numéricas e resultados

Alguns resultados dos algoritmos definidos no capítulo anterior já foram apresentados nas Figuras 3.4, 4.4 e 4.5 para apresentar conceitos de processos dinâmicos em redes. Este capítulo é dedicado à comparação entre os diferentes métodos e análise da eficiência computacional. Todas as simulações foram feitas em um computador com processador Intel Xeon com 6 núcleos de 3,2 GHz e 24 Gb de memória RAM.

6.1 Efeito do gerador de números aleatórios

Simulações estocásticas fazem uso de *geradores de números aleatórios*. Na verdade, os números obtidos por tais geradores não são totalmente aleatórios, mas *pseudoaleatórios*, sendo possível recuperar sempre a mesma sequência de números a partir de uma *semente*, que nada mais é do que um número que o gerador recebe para que possa iniciar o cálculo das sequências.

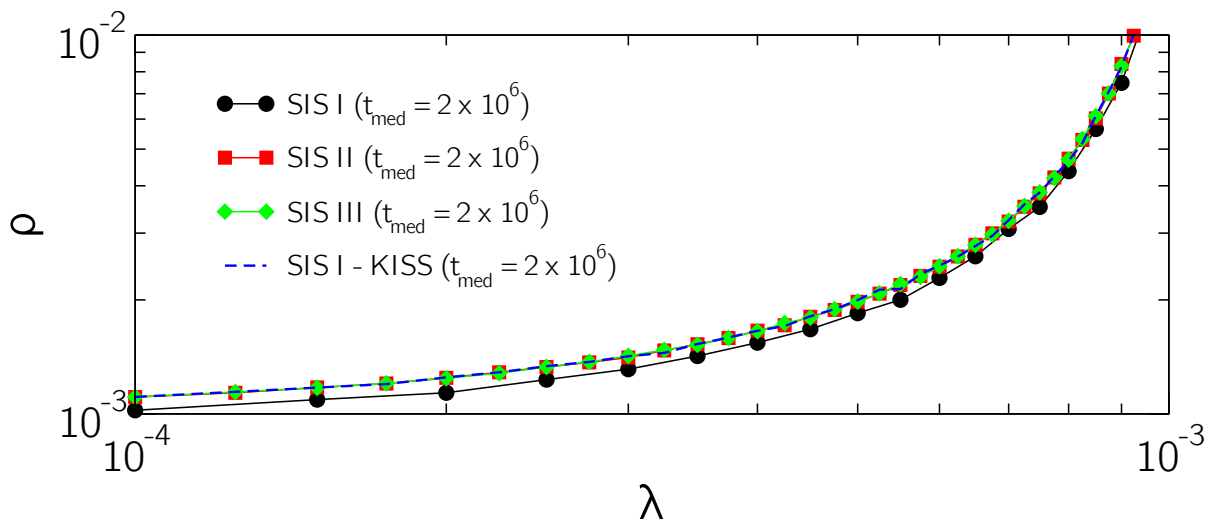


Figura 6.1: Comparação entre diferentes algoritmos para uma simulação QS do modelo SIS em um grafo completo de tamanho $N = 1000$. Para os algoritmos SIS-II e SIS-III, a diferença entre os geradores *ran2* e *KISS* é irrelevante.

Quando lidamos com probabilidades muito baixas, é necessário que o gerador seja

capaz de gerar números próximos de zero, o que pode não acontecer dependendo da precisão numérica utilizada. Dois geradores foram comparados, sendo eles o *ran2*, apresentado na referência [30], e o *KISS* – *Keep It Simple, Stupid*, proposto por G. Marsaglia [31]. A precisão numérica do último é maior que a do primeiro.

Foi verificado que os algoritmos do tipo II e III praticamente independem do gerador utilizado. Já os algoritmos do tipo I são muito sensíveis, principalmente à medida que o tamanho da rede N é aumentado. A Figura 6.1 mostra uma comparação entre os algoritmos SIS-I, SIS-II e SIS-III, todos utilizando o *ran2*, e o SIS-I utilizando o *KISS*. Note que realmente o SIS-I com *ran2* apresenta uma diferença considerável em relação aos SIS-II e SIS-III. Já com o *KISS*, a diferença é imperceptível, e é coerente com o fato de que esses três algoritmos devam gerar os mesmos resultados estatísticos. Para as simulações deste trabalho, o *KISS* foi um gerador de números aleatórios adequado.

6.2 Aplicação dos algoritmos para diferentes análises

As análises foram feitas para os três modelos e a eficiência foi comparada em termos do tempo computacional para cada simulação.

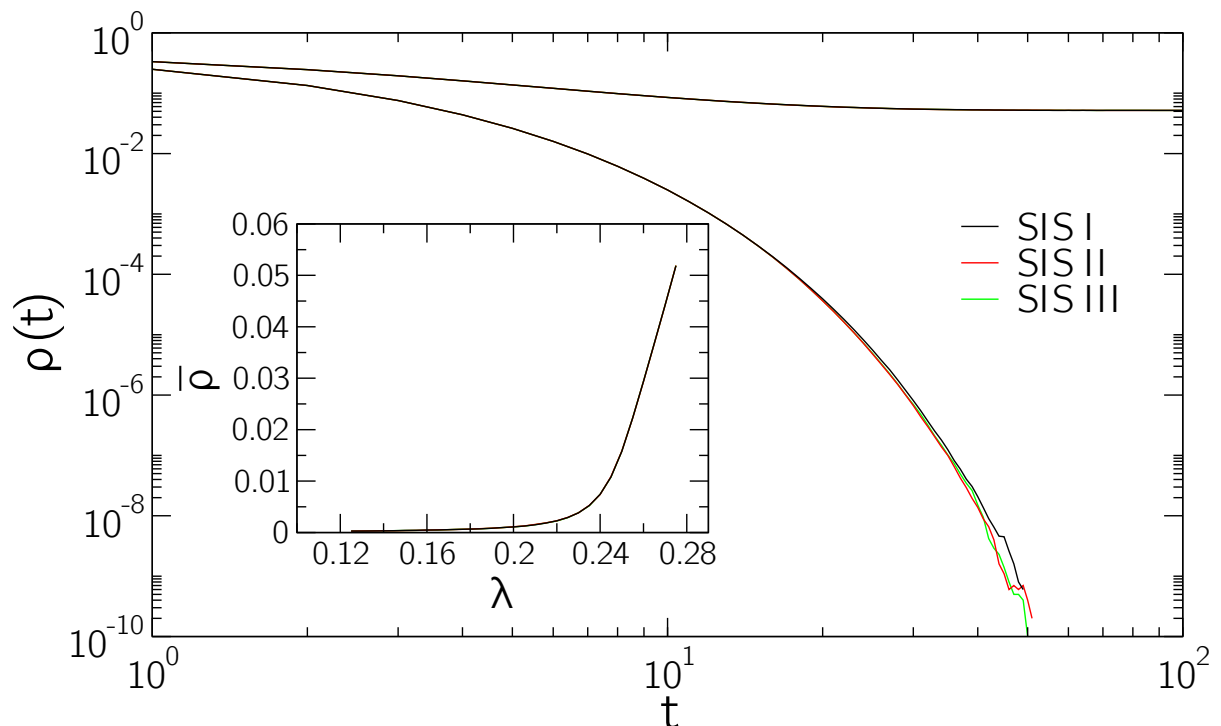


Figura 6.2: Densidade média de vértices infectados em função do tempo para o modelo SIS com $\mu = 1$ em uma única amostra de rede com $\gamma = 4,00$, $k_0 = 3$, $N = 10^4$, e $k_c = \sqrt{N}$. A taxa de infecção é $\lambda = 0,125$ (subcrítico) e $\lambda = 0,275$ (supercrítico), respectivamente para as curvas inferiores e as superiores. Note que há três curvas para cada caso, conforme legenda. As médias foram realizadas para 10^6 e 10^4 amostras dinâmicas independentes, para o primeiro e o segundo caso, respectivamente, com condição inicial $\rho(0) = 1$. O gráfico interior mostra a densidade quase-estacionária média em função da taxa λ .

6.2.1 Modelo SIS

Os três algoritmos para o modelo SIS foram comparados na Figura 6.2 para uma única rede UCM com $\gamma = 4,00$ e $N = 10^4$. A densidade média de vértices infectados em função do tempo é apresentada, fazendo médias comuns¹ sobre amostras dinâmicas independentes com condição inicial $\rho(0) = 1$. As diferenças entre as curvas para os diferentes algoritmos são imperceptíveis, exceto para densidades menores, devido ao menor número de amostras.

A concordância é excelente na análise QS, inclusive para a distribuição quase-estacionária tanto para as fases sub e supercríticas, como apresentado na Figura 6.3. Novamente, as diferenças obtidas são devido às flutuações presentes.

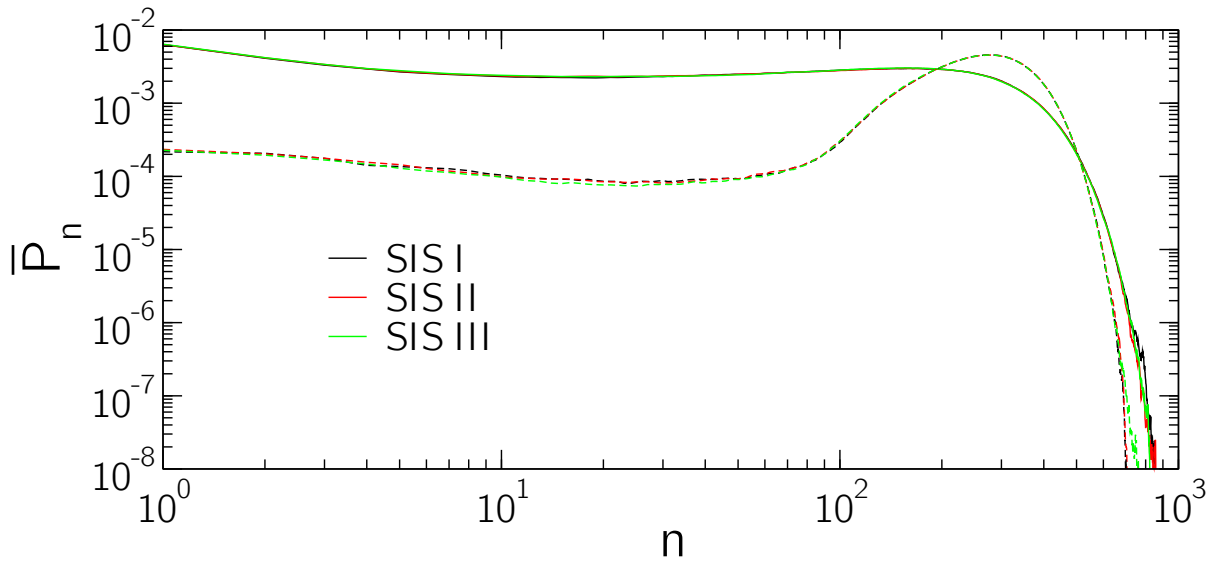


Figura 6.3: Distribuição quase-estacionária para o modelo SIS em redes UCM utilizando os diferentes algoritmos. Os parâmetros das redes são $N = 10^5$, $k_0 = 3$, $k_c = N^{1/2}$, $\gamma = 2,30$ (linhas sólidas) e $\gamma = 4,00$ (linhas pontilhadas). As taxas de infecção são $\lambda_c = 0,02617$ e $0,18505$ para $\gamma = 2,30$ e $4,00$, respectivamente, cujos valores representam a estimativa pela análise da suscetibilidade χ .

A análise de eficiência dos três métodos foram realizadas a partir de simulações QS no limiar crítico e 5% acima dele, com valor estimado pela análise da suscetibilidade χ . No caso de múltiplos picos, foram selecionados aqueles nos quais o tempo de meia vida τ diverge. Para todos os casos, todos os vértices começaram inicialmente infectados e a dinâmica foi realizada até um tempo máximo $t_{rlx} + t_{med} = 3 \times 10^6$. As redes definidas na Seção 2.3 foram utilizadas.

Para o grafo estrela foi possível definir um algoritmo específico: o SIS-star. Funciona praticamente da mesma forma que o SIS-II, com a diferença no evento de infecção. Em tal evento há duas possibilidades: se o vértice central (a estrela propriamente dita) está suscetível, ele deve ser infectado com probabilidade 1, já que certamente a infecção vem de uma de suas folhas. Caso esteja infectado, com probabilidade $(N - 1)/(N - 1 + N_{inf})$,

¹ As amostras que visitaram os estados absorventes estão incluídas

uma folha é escolhida ao acaso, tenta-se infectá-la e o tempo é incrementado pelo passo de tempo correspondente. Esta probabilidade é justamente a de que o vértice central seja selecionado para tentar infectar uma das folhas, e estas não podem infectar a estrela pois ela já está infectada. Como mostrado na Figura 6.4, tal algoritmo reduz consideravelmente o tempo computacional utilizado.

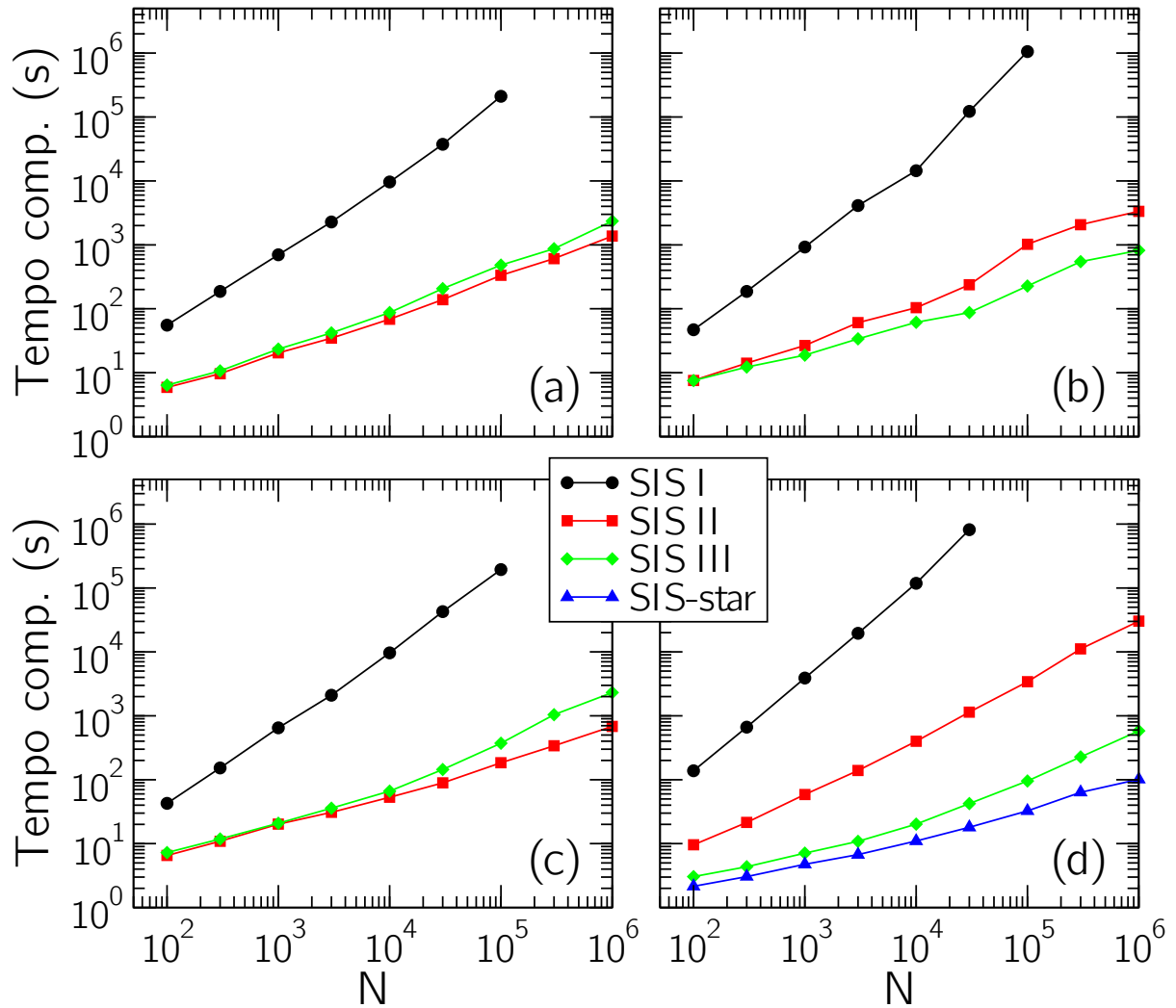


Figura 6.4: Tempo computacional para uma simulação QS do modelo SIS no limiar crítico para redes de diferentes níveis de heterogeneidade: (a) exponencial, (b) UCM com $\gamma = 4,00$, (c) $\gamma = 2,30$, e (d) grafo estrela. Para o último, um algoritmo específico SIS-star também é apresentado. O tempo total de simulação é $t_{\text{med}} + t_{\text{rlx}} = 3 \times 10^6$.

Ligeiramente acima do ponto crítico, a performance tanto do SIS-II quanto do SIS-III não é consideravelmente alterada em relação às simulações no limiar crítico. Para regiões muito supercríticas, a diferença entre os métodos não é significativa. A Tabela 6.1 apresenta o tempo computacional para os três algoritmos no ponto crítico das redes analisadas.

Em síntese, os algoritmos otimizados certamente foram os mais *rápidos* computacionalmente. Entre os modelos SIS-II/SIR-II e SIS-III/SIR-III, para $\gamma > 3$ o III foi o mais eficiente, já que para tais valores de γ os *outliers* são mais frequentes, dificultando

	SIS-I	SIS-II		SIS-III	
Tipo de rede	$N = 10^5$	$N = 10^5$	$N = 10^6$	$N = 10^5$	$N = 10^6$
Exponencial	3520	5,50	23,3	8,00	40,0
UCM $\gamma = 4,00$	17600	16,7	55,0	3,83	13,7
UCM $\gamma = 2,30$	3240	3,00	11,3	6,33	38,3
Estrela	-	56,7	505	1,60	9,67

Tabela 6.1: Tempo computacional em minutos para as simulações QS do modelo SIS no limiar crítico λ_c para redes UCM, com $\gamma = 2,30$ e $\gamma = 4,00$. O tempo total de simulação é de $t_{\text{med}} + t_{\text{rlx}} = 3 \times 10^6$ passos.

a seleção pela probabilidade $k_i/k_{\text{máx}}$. Para $\gamma < 3$, naturalmente a versão II venceu, já que o tempo para manter a lista de k_i cópias de vértices infectados torna-se maior que fazer o método de aceitação por probabilidades: há poucos *outliers*. Tal resultado foi confirmado para testes feitos em redes com taxas subcríticas $\lambda < \lambda_c$, supercríticas $\lambda > \lambda_c$ e críticas $\lambda = \lambda_c$.

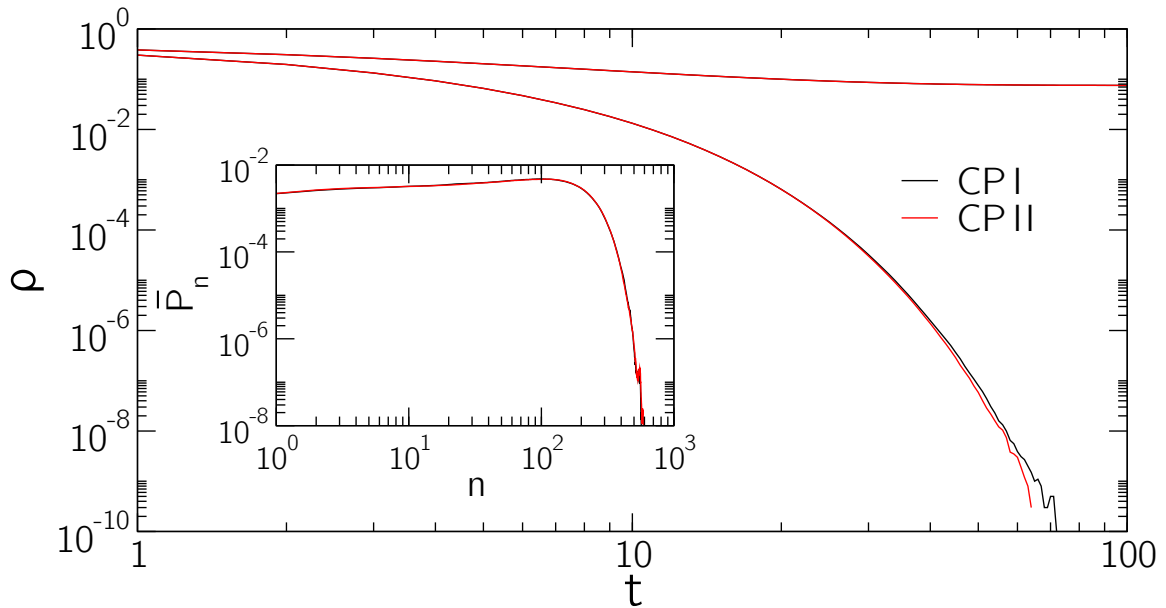


Figura 6.5: Comparação entre os algoritmos CP-I e CP-II para o processo de contato em redes UCM com $\gamma = 4,00$, $N = 10^4$, $k_0 = 3$ usando taxas de infecção $\lambda = 0,90$ (subcrítica), e $\lambda = 1,50$ (supercrítica). As médias foram feitas para 10^6 e 10^4 amostras para dinâmicas sub e supercríticas, respectivamente. A inserção mostra a distribuição quase-estacionária para o PC no limiar crítico em redes UCM com os mesmos parâmetros do gráfico principal e apenas uma simulação QS, com $t_{\text{máx}} = t_{\text{med}} + t_{\text{rlx}} = 3 \times 10^6$.

6.2.2 Processo de contato

A equivalência entre os algoritmos CP-I e CP-II está ilustrada na Figura 6.5. Como no modelo SIS, as curvas só são distinguíveis para valores baixos de densidade devido à

menor estatística. Os tempos computacionais estão comparados na Figura 6.6, na qual podemos ver que a dinâmica no ponto crítico pode ser ordens de grandeza mais lento que no algoritmo otimizado.

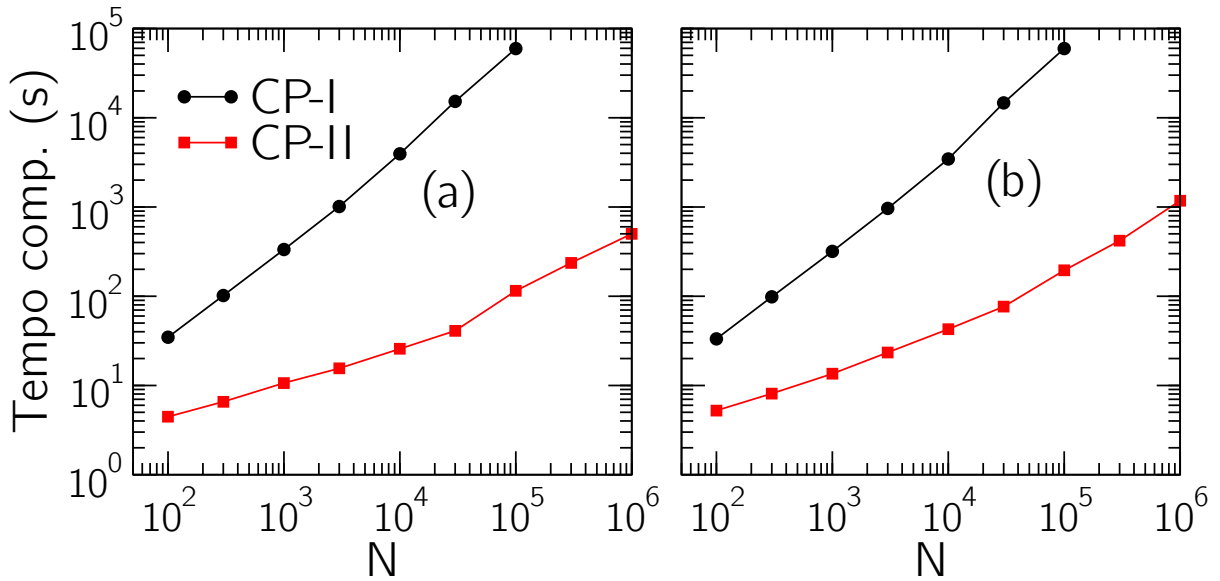


Figura 6.6: Comparação do tempo computacional para uma simulação QS do processo de contato utilizando os algoritmos CP-I e CP-II. As simulações foram realizadas em redes UCM com $N = 10^4$ e grau mínimo $k_0 = 3$, com (a) $\gamma = 2,30$ e (b) $\gamma = 4,00$. O tempo total de simulação é $t_{\text{med}} + t_{\text{rlx}} = 3 \times 10^6$.

6.2.3 Modelo SIR

As simulações foram realizadas assumindo uma configuração inicial na qual apenas o vértice de maior grau da rede está infectado, enquanto os demais estão suscetíveis, de forma que as flutuações estatísticas são reduzidas. A dinâmica é realizada até que o estado absorvente seja atingido, e as médias são feitas em um número grande de repetições da simulação na mesma rede (de 10^5 a 10^6). Um parâmetro de ordem da dinâmica SIR é a densidade final média de vértices recuperados [32], assim que o último vértice infectado se torna recuperado, na iminência da transição para o estado absorvente. É calculado também o tempo médio $\langle t_{\text{máx}} \rangle$ que isso leva para ocorrer.

O limiar epidêmico do modelo SIR é exatamente conhecido como $\lambda_c = \langle k \rangle / (\langle k^2 \rangle - 2\langle k \rangle)$ [32, 33], e é este o valor assumido quando o limiar epidêmico do SIR é mencionado. A equivalência entre os algoritmos SIR-I, SIR-II e SIR-III é mostrada na Figura 6.7 para uma rede UCM com $\gamma = 4,00$. Outros valores de γ foram considerados e a equivalência é corroborada. A eficiência computacional é equivalente à do modelo SIS, como pode ser visto na Figura 6.8.

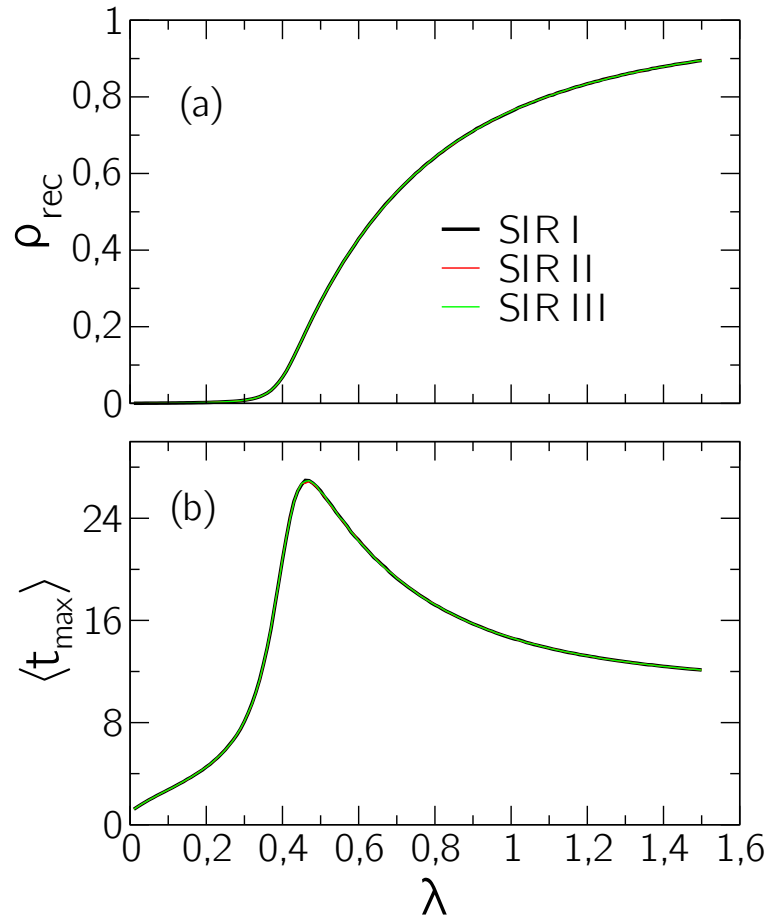


Figura 6.7: Comparação dos algoritmos para o modelo SIR, utilizando uma rede UCM com $\gamma = 4,00$ e $N = 10^4$. (a) A densidade final média de vértices recuperados e (b) o tempo de vida médio da epidemia são mostrados como funções da taxa de infecção λ . As curvas correspondem a médias sobre 10^5 amostras dinâmicas.

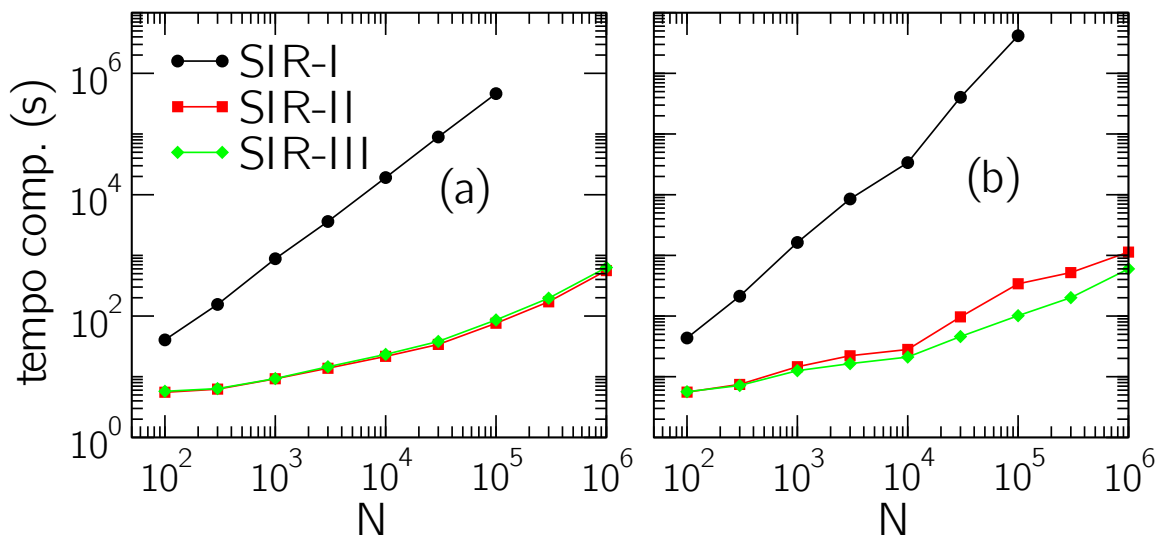


Figura 6.8: Comparação do tempo computacional para simulações de 10^6 amostras independentes da dinâmica SIR em redes UCM com $N = 10^4$, $k_0 = 3$ com: (a) $\gamma = 2,30$ e (b) $\gamma = 4,00$. As simulações foram realizadas para $\lambda = \lambda_c = \langle k \rangle / (\langle k^2 \rangle - 2\langle k \rangle)$.

Capítulo VII

Considerações finais

Neste trabalho, uma revisão de literatura foi realizada tanto para a caracterização estrutural de redes complexas, quanto de processos dinâmicos Markovianos. Nos últimos tempos, a ciência tem avançado significadamente no entendimento da estrutura e funcionamento de redes, mas ainda há desafios, os quais as simulações computacionais ajudam a superá-los.

As dinâmicas em redes grandes e altamente heterogêneas necessitam de simulações numéricas precisas, já que os resultados analíticos têm aplicações restritas e as aproximações teóricas devem ser verificadas. Algoritmos de simulação de processos epidêmicos genéricos foram apresentados nesta monografia, e são estatisticamente exatos, permitindo a análise de redes grandes e heterogêneas com eficiência.

Muitos dos algoritmos aqui propostos já foram utilizados na literatura, seja de forma correta ou não. A proposta deste trabalho foi justamente de reunir e explicar tais algoritmos de maneira clara, apresentando limitações e formas de lidar com desafios computacionais nessas análises. Os algoritmos descritos neste trabalho foram inspirados no algoritmo de Gillespie para simulação de processos estocásticos. O algoritmo padrão é ineficiente, mas a sua otimização pela construção de listas permite o ganho de ordens de grandeza em tempo computacional. Um dos algoritmos otimizado (tipo II) utiliza o método da rejeição para selecionar um vértice infectado, e possui uma performance maior para redes que não apresentam *outliers*. No caso de existirem, o tipo III é mais eficiente, mantendo uma lista de cópias dos vértices ativos.

Os algoritmos do tipo II são mais recomendados para uso geral, já que são fáceis de implementar e possuem ótima performance para vários tipos de redes. A mesma ideia pode ser facilmente aplicada em outros modelos epidêmicos, como o modelo SIS generalizado investigado por Karsai, Juhász e Iglói (KJI) [34]. O modelo SIRS – Suscetível-Infectado-Recuperado-Suscetível também pode ser simulado com o algoritmo geral otimizado, bastando acrescentar a taxa $\alpha_i = \alpha$, possibilitando a transição de um vértice recuperado para suscetível.

7.1 Outros trabalhos desenvolvidos e perspectivas

Os métodos apresentados neste trabalho foram utilizados também para o estudo de fases de Griffiths em redes complexas [5], artigo de minha coautoria publicado na *Physical Review E* em março de 2016. Generalizando o modelo de configurações para construção de redes complexas, foram geradas redes com estrutura modular — redes complexas fracamente ligadas entre si — e análises de decaimento estão sendo realizadas para avançar no entendimento dos resultados desse artigo publicado.

Além disso, um modelo epidêmico generalizado está sendo proposto e investigado por nós, cujas particularidades são o processo de contato e o modelo SIS: cada vértice i transmite a infecção com taxa λ/k_i^α , sendo $\alpha = 0$ o modelo SIS e $\alpha = 1$ o PC. O algoritmo SIS-II é facilmente generalizado para este modelo, bastando modificar as probabilidades de aceite dos vértices infectados nas tentativas de infecção, e os resultados são coerentes com as análises teóricas já desenvolvidas.

7.2 Disponibilização de códigos e artigo

Para ajudar a compreender tais algoritmos, um exemplo foi disponibilizado *online* para *download* livre, sob uma licença GNU GPL, que permite o uso, a distribuição e a modificação desde que utilize a mesma licença e que o código seja também disponibilizado livremente. O exemplo consiste na aplicação do modelo SIS com o algoritmo SIS-II, escrito nas linguagens *Fortran* e *Python*: a primeira é amplamente utilizada no meio científico devido a sua precisão e velocidade de cálculo numérico, enquanto a segunda é bem popularizada e de fácil entendimento, porém ineficiente para esses cálculos. Os códigos estão disponíveis nos seguintes endereços:

- Versão em *Fortran*: <https://github.com/wcota/dynSIS>.
- Versão em *Python* puro: <https://github.com/wcota/dynSIS-py>.
- Versão adaptada para a biblioteca *NetworkX*¹: <https://github.com/wcota/dynSIS-networkx>.

Todos os resultados apresentados nesta monografia estão sendo finalizados em um artigo científico a ser submetido.

¹ Biblioteca de análise de redes em *Python*, disponível em <https://networkx.github.io/>.

Referências Bibliográficas

- [1] BARRAT, A.; BARTHÉLEMY, M.; VESPIGNANI, A. *Dynamical processes on complex networks*. Cambridge University Press, 2008.
- [2] Marvel, S. A.; Martin, T.; Doering, C. R.; Lusseau, D.; Newman, M. E. J. The small-world effect is a modern phenomenon. *ArXiv e-prints*, outubro de 2013.
- [3] NEWMAN, M. *Networks: An introduction*. OUP Oxford, 2010.
- [4] WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *Nature*, London, v. 393, n. 6684, p. 440–442, 1998.
- [5] COTA, W.; FERREIRA, S. C.; ÓDOR, G. Griffiths effects of the susceptible-infected-susceptible epidemic model on random power-law networks. *Physical Review E*, v. 93, n. 3, p. 032322, março de 2016.
- [6] BARABÁSI, A.-L. Spreading Phenomena. In: *Network Science*. Northeastern University, Boston: Cambridge University Press, 2015. Cap. 10.
- [7] MATA, A. S.; FERREIRA, S. C. Multiple transitions of the susceptible-infected-susceptible epidemic model on complex networks. *Physical Review E*, v. 91, n. 1, p. 012816, janeiro de 2015.
- [8] VÁZQUEZ, A.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Large-scale topological and dynamical properties of the internet. *Phys. Rev. E*, v. 65, p. 066130, junho de 2002.
- [9] CATANZARO, M.; BOGUÑÁ, M.; PASTOR-SATORRAS, R. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, v. 71, p. 027103, fevereiro de 2005.
- [10] BOGUÑÁ, M.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B*, v. 38, p. 205–210, 2004.
- [11] PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, Woodbury, v. 86, n. 14, p. 3200–3203, abril de 2001.

- [12] VAN KAMPEN, N. G. *Stochastic processes in physics and chemistry*. Elsevier, 1992. v. 1.
- [13] SALINAS, S. R. A. *Introdução a física estatística vol. 09*. EDUSP, 2005.
- [14] COX, D.; MILLER, H. *The theory of stochastic processes*. Science paperbacks. Taylor & Francis, 1977.
- [15] TOMÉ, T.; DE OLIVEIRA, M. *Dinâmica estocástica e irreversibilidade*. 2. ed. EDUSP, 2015.
- [16] MARRO, J.; DICKMAN, R. *Nonequilibrium phase transitions in lattice models*. Collection Alea-Saclay: Monographs and Texts in Statistical Physics. Cambridge University Press, 2005.
- [17] HENKEL, M.; HINRICHSSEN, H.; LÜBECK, S. *Non-equilibrium phase transitions: Volume 1: Absorbing phase transitions*. Theoretical and Mathematical Physics. Springer Netherlands, 2008.
- [18] GRASSBERGER, P.; DE LA TORRE, A. Reggeon field theory (schlögl's first model) on a lattice: Monte carlo calculations of critical behaviour. *Annals of Physics*, New York, v. 122, n. 2, p. 373 – 396, 1979.
- [19] DICKMAN, R.; MARTINS DE OLIVEIRA, M. Quasi-stationary simulation of the contact process. *Physica A: Statistical Mechanics and its Applications*, v. 357, n. 1, p. 134–141, 2005.
- [20] SANDER, R. S.; COSTA, G. S.; FERREIRA, S. C. Sampling methods for the quasistationary regime of epidemic processes on regular and complex networks. *Phys. Rev. E*, v. 94, p. 042308, outubro de 2016.
- [21] FENNELL, P. G.; MELNIK, S.; GLEESON, J. P. Limitations of discrete-time approaches to continuous-time contagion dynamics. *Phys. Rev. E*, v. 94, p. 052125, novembro de 2016.
- [22] GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, v. 22, n. 4, p. 403 – 434, 1976.
- [23] GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, v. 81, n. 25, p. 2340–2361, 1977.
- [24] DE OLIVEIRA, M. M.; DICKMAN, R. How to simulate the quasistationary state. *Physical Review E*, v. 71, n. 1, p. 016129, 2005.

- [25] DICKMAN, R.; VIDIGAL, R. Quasi-stationary distributions for stochastic processes with an absorbing state. *Journal of Physics A: Mathematical and General*, v. 35, n. 5, p. 1147, 2002.
- [26] BLANCHET, J.; GLYNN, P.; ZHENG, S. *Empirical analysis of a stochastic approximation approach for computing quasi-stationary distributions*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 19–37.
- [27] MATA, A. S.; FERREIRA, R. S.; FERREIRA, S. C. Heterogeneous pair-approximation for the contact process on complex networks. *New Journal of Physics*, v. 16, n. 5, p. 053006, 2014.
- [28] FERREIRA, S. C.; CASTELLANO, C.; PASTOR-SATORRAS, R. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Phys. Rev. E*, v. 86, p. 041125, outubro de 2012.
- [29] CASTELLANO, C.; PASTOR-SATORRAS, R. Non-mean-field behavior of the contact process on scale-free networks. *Phys. Rev. Lett.*, v. 96, p. 038701, janeiro de 2006.
- [30] PRESS, W. H.; FLANNERY, B. P.; TEUKOLSKY, S. A.; VETTERLING, W. T. *Numerical recipes in fortran 77: The art of scientific computing*. Cambridge University Press, 1992.
- [31] Disponível em http://web.mst.edu/~vojttat/class_5403/kiss05/rkiss05.f90.
- [32] NEWMAN, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E*, v. 66, p. 016128, julho de 2002.
- [33] KARRER, B.; NEWMAN, M. E. J. Message passing approach for general epidemic models. *Phys. Rev. E*, v. 82, p. 016101, julho de 2010.
- [34] KARSAI, M.; JUHÁSZ, R.; IGLÓI, F. Nonequilibrium phase transitions and finite-size scaling in weighted scale-free networks. *Phys. Rev. E*, v. 73, p. 036116, março de 2006.