



Exame de Qualificação

Fenômenos de Propagação em Redes Complexas

Aluno: Marlon Ramos

Orientadora: Celia Anteneodo

Departamento de Física

Pontifícia Universidade Católica do Rio de Janeiro

Junho 2012

Pontifícia Universidade Católica do Rio de Janeiro

MARLON RAMOS

EXAME DE QUALIFICAÇÃO

FENÔMENOS DE PROPAGAÇÃO EM REDES COMPLEXAS

RESUMO

Sumário

1	Introdução e Metodologia	1
1.1	Introdução e Motivação	1
1.2	Metologia	1
2	Revisão Bibliográfica	2
2.1	Modelos	2
2.2	Modelo SIR	4
2.2.1	Exemplos SIR	5
2.3	Redes Complexas	6
2.3.1	Modelo Configuracional	8
2.4	Modelo SIR - Comportamento assintótico	9
2.5	Aproximação por conectividade	12
2.6	Modelos de Rumores	14
2.7	Resultados Numéricos	15
3	Propostas e Conclusões	19
	Referências Bibliográficas	20

Capítulo 1

Introdução e Metodologia

1.1 Introdução e Motivação

CONTAGIO: (biologia e sociologia) keywords PRx rumor+spreading

FOCO: fenomenos de propagacao em meios sociais:

- difusao de rumores, informacoes

- surgimento de consenso na adopcao de inovacoes tecnologicas, formacao de opinioes, da linguagem, disseminacao de ideias, da corrupcao, praticas sociais, etc.

- epidemias de doencas infecciosas

diversos fenomenos governados por interacoes sociais que podem evoluir para estados finais fragmentados desordenados ou de consenso, em que todos os individuos sao atingidos por informacoes ou doencas (consenso de opinioes, epidemias, respectivamente) ou adoptam um mesmo padrao comportamental (opinioes, linguagem) [1]

a pergunta fundamental eh como esses estados estacionarios emergem, quais os mecanismos e como exercer controle sobre os mesmos

a topologia da rede em que as interacoes ocorrem tambem eh crucial para determinar as propriedades emergentes

1.2 Metodologia

analogia com problemas de fisica estatistica, tr. de fase, percolacao, difusao, etc

fisica est: analogia com sistemas fisicos (magneticos tipo Ising, 1/2 ou 1)

uso de redes complexas

Capítulo 2

Revisão Bibliográfica

2.1 Modelos

Uma das maiores motivações para o estudo de redes sociais é o seu papel no espalhamento de doenças. Doenças se espalham através da rede de contatos entre indivíduos, por exemplo: doenças respiratórias se espalham quando pessoas que estão no mesmo ambiente e respiram o mesmo ar, certas doenças se transmitem quando as pessoas se tocam, etc. O padrão de contatos entre as pessoas pode ser representado como uma rede e por isso o esforço para se caracterizar bem a estrutura das redes sociais. Um tópico relacionado é o espalhamento de vírus de computador e o espalhamento de rumores e notícias. Nesse trabalho, vamos descrever os modelos clássicos de epidemias e espalhamento de rumores.

A biologia do que acontece dentro de uma pessoa quando ela é infectada com uma doença é complicada. Há uma luta do patógeno com o sistema imunológico, o que pode levar a cura, uma infecção permanente ou a morte. Para estudar o espalhamento de uma doença teríamos que levar tudo isso em conta, o que não é possível na prática.

Na formulação matemática usual de uma epidemia a dinâmica da doença é reduzida à mudança entre alguns estados. Na versão mais simples de todas, trabalhamos com dois estados somente: susceptíveis e infectados. Um indivíduo no estado susceptível não tem a doença mas pode vir a ser infectado. O indivíduo no estado infectado é aquele que tem a doença e pode transmiti-la caso tenha contato com um alguém susceptível. Essa é uma simplificação forte, mas captura as características básicas da doença e é útil quando queremos saber o que acontece com uma dada população e não o que acontece com cada um dos membros da população.

A abordagem usual não leva em conta a estrutura da rede de contatos. Fazemos uso de uma aproximação chamada *homogeneous mixing hypothesis* em que supomos que cada um dos indivíduos tem a mesma probabilidade, por unidade de tempo, de entrar em contato com qualquer outro. As pessoas se encontram completamente ao acaso nessa abordagem. É claro que essa não é uma hipótese realística porque as pessoas tem contato direto com uma fração pequena da população e esses contatos não são ao acaso. Exatamente por isso o estudo de redes é importante.

Consideremos uma doença que se espalha em uma dada população. Vamos denotar por $S(t)$ e $I(t)$ o número de indivíduos susceptíveis e infectados no tempo t respectivamente. O número de infectados aumenta quando indivíduos susceptíveis entram em contato com infectados. Suponha que as pessoas se encontram, e façam contato por um tempo suficiente para que haja contágio, ao acaso com uma taxa β , isso quer dizer que cada membro da população faz em média β contatos com outras pessoas escolhidas ao acaso por unidade de tempo.

Um contágio acontece só quando um susceptível encontra com um infectado. Se o número total de pessoas é n , a probabilidade de encontrar uma pessoa ao acaso e ela ser

susceptível é S/n . Daí, cada infectado encontra em média $\beta S/n$ indivíduos susceptíveis por unidade de tempo. Como temos I pessoas infectadas, a taxa de novas infeções é $\beta SI/n$ e ficamos com a seguinte equação diferencial para o número de infectados:

$$\frac{dI}{dt} = \beta SI/n. \quad (2.1)$$

O número de susceptíveis decresce com a mesma taxa

$$\frac{dS}{dt} = -\beta SI/n \quad (2.2)$$

Esse é o modelo conhecido como *fully mixed susceptible-infected model* ou SI. É comum trabalhar com frações de indivíduos, vamos definir novas variáveis $s = S/n$ e $i = I/n$, com isso as equações 2.1 e 2.2 ficam na forma:

$$\frac{di}{dt} = \beta si. \quad (2.3)$$

$$\frac{ds}{dt} = -\beta si \quad (2.4)$$

Nesse modelo a população é constante, então, na verdade podemos reduzir as equações à uma única através da relação $s + i = 1$, o que nos dá

$$\frac{di}{dt} = \beta i(i - 1) \quad (2.5)$$

Essa equação é a conhecida equação logística que aparece em muitas outras situações e tem como solução

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}} \quad (2.6)$$

onde $i_0 = i(0)$. A expressão acima tem a forma mostrada em 2.1. Ela cresce exponencialmente para tempos curtos, o que corresponde a fase inicial da doença onde há muitos indivíduos susceptíveis, mas satura eventualmente, á medida em que a epidemia avança e fica mais difícil encontrar novas vítimas.

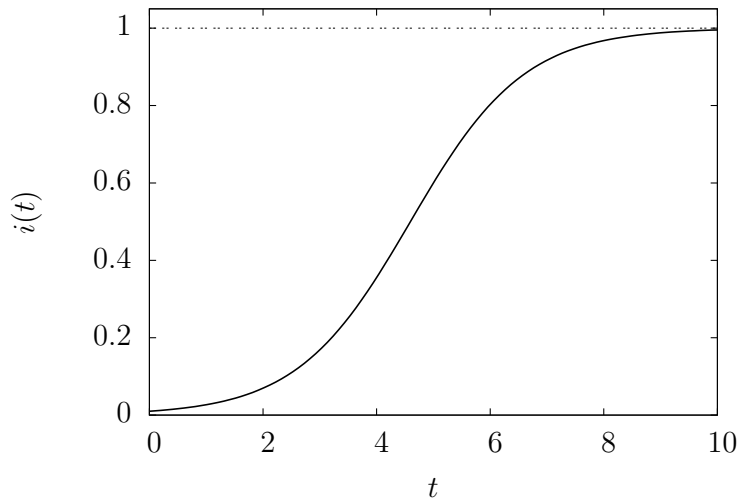


Figura 2.1: Curva clássica da equação logística. Usei $i_0 = 1\%$.

2.2 Modelo SIR

No modelo SI, uma vez que a pessoa seja infectada, ela permanece nesse estado para sempre, mas em muitas doenças as pessoas conseguem se recuperar da infecção depois de um tempo e uma vez recuperadas adquirem imunidade. Essa observação motiva a introdução de um novo estado: os recuperados R . Em outras doenças, as pessoas não se recuperam, pelo contrário, elas morrem. Do ponto de vista do indivíduo isso é exatamente o contrário de ficar curado, mas para a doença tanto faz uma pessoa estar morta ou imune, de qualquer maneira o número de hospedeiros potenciais é reduzido. Por isso, a letra R algumas vezes significa removidos. Nesse modelo, as duas situações são equivalentes. O acrônimo SIR significa *susceptible-infected-removed*.

O modelo tem duas etapas. Primeiro, uma pessoa se infecta quando entra em contato com alguém que tem a doença, os contatos ocorrem com a mesma taxa β do modelo SI. Segundo, as pessoas infectadas se recuperam (ou morrem) com uma taxa γ . Em termos das densidades s , i e r , o modelo SIR tem as equações:

$$\frac{ds}{dt} = -\beta si. \quad (2.7)$$

$$\frac{di}{dt} = \beta si - \gamma i \quad (2.8)$$

$$\frac{dr}{dt} = \gamma i \quad (2.9)$$

Completamos a formulação com as condições iniciais

$$s(0) = s_0 > 0, \quad i(0) = i_0 > 0, \quad r(0) = 0. \quad (2.10)$$

Temos também a equação de conservação $s + r + i = 1$, que é automaticamente satisfeita pelas equações acima.

A primeira questão relevante a ser respondida é se existirá uma epidemia ou não e, se houver, como será a evolução temporal.

Pela equação 2.8, temos para $t = 0$

$$\left. \frac{di}{dt} \right|_{t=0} = i_0(\beta s_0 - \gamma) \geq 0 \quad \text{se} \quad s_0 \geq \frac{\gamma}{\beta} \equiv \rho \quad (2.11)$$

Dado que $ds/dt < 0$, então temos que $s < s_0$ sempre, portanto se $s_0 < \rho$, resulta

$$\frac{di}{dt} \leq 0 \quad \forall \quad t \geq 0. \quad (2.12)$$

Nessa situação, a fração de pessoas infectadas é sempre menor que i_0 e tende a 0 para tempos longos. Portanto, não temos uma epidemia.

Agora, se $s_0 > \rho$, então di/dt aumenta inicialmente e dizemos que a uma epidemia. Portanto, existe um limiar que separa esses dois regimes.

Alguma nomenclatura: ρ é a taxa de remoção relativa. $R_0 \equiv \rho s_0 / \gamma$ é a taxa de remoção relativa e $1/\gamma$ é o período médio de duração da infecção.

Dividindo 2.8 por 2.7, temos uma equação que relaciona i e s :

$$\frac{di}{ds} = \frac{-i(\beta s - \gamma)}{\beta si} = -1 + \rho/s \quad (2.13)$$

integrando 2.13, temos:

$$i + s - \rho \ln(s) = i_0 + s_0 - \rho \ln(s_0) = \text{cte} \quad (2.14)$$

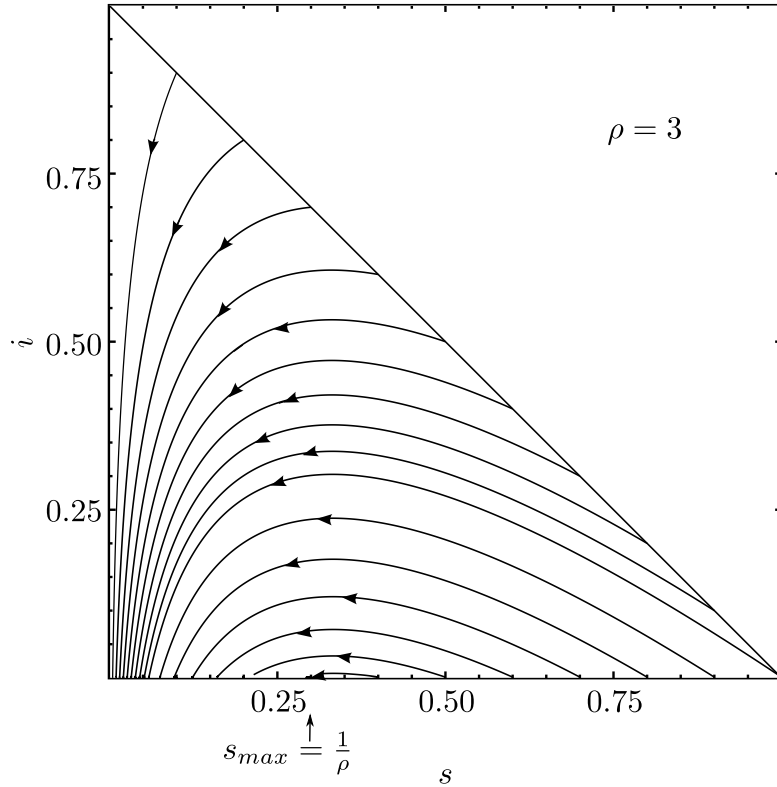


Figura 2.2: Plano de fases para o modelo SIR. Aqui a variável σ faz o papel de ρ .

Dado que $r = 1 - s - i > 0$, vemos que as trajetórias são limitadas pelo triângulo mostrado na figura 2.3.

Em uma epidemia real pode ser mais fácil conhecer o número de pessoas recuperadas do que de infectadas. Nesse caso, é importante conhecer a expressão para dr/dt . A partir de 2.13 e 2.9, temos:

$$\frac{dr}{dt} = \gamma(1 - r - s_0 e^{-r/\rho}) \quad (2.15)$$

É fácil calcular essa expressão numericamente se conhecemos todos os parâmetros, o que, infelizmente, não é possível em muitos casos. No entanto, podemos fitar os dados se acreditamos que o modelo descreve bem a epidemia. Nos casos em que a epidemia não é muito grande podemos também fazer a aproximação de que $\rho \ll 1$.

2.2.1 Exemplos SIR

A peste de Bombay foi uma epidemia que aconteceu entre 1905 e 1906. Quase todos os infectados morreram e daí podemos ajustar a taxa de mortalidade com a taxa de remoção, podemos também usar aproximação mencionada acima. Este caso foi analisado no trabalho original que propôs o modelo SIR, por Kermack e McKendrick em 1927.

Em uma escola da Inglaterra com 763 alunos aconteceu uma epidemia entre os meses de janeiro e fevereiro de 1978. No total 512 alunos foram infectados durante o surto que parece ter se originado de um único caso. Temos os dados para $i(t)$ e com eles podemos ajustar a curva mostrada na figura 2.2.1

Existe na literatura muitos outros modelos de campo médio mais relistas que foram extensivamente estudados. Os modelos que vimos nessa seção são suficientes para os

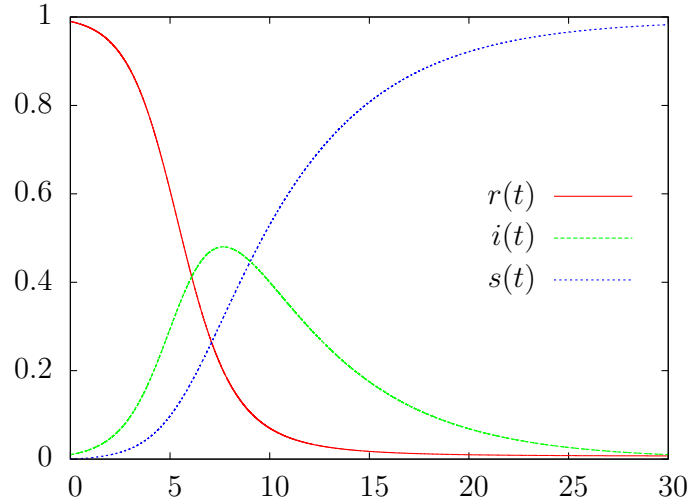


Figura 2.3: Evolução temporal para as variáveis do modelo SIR.

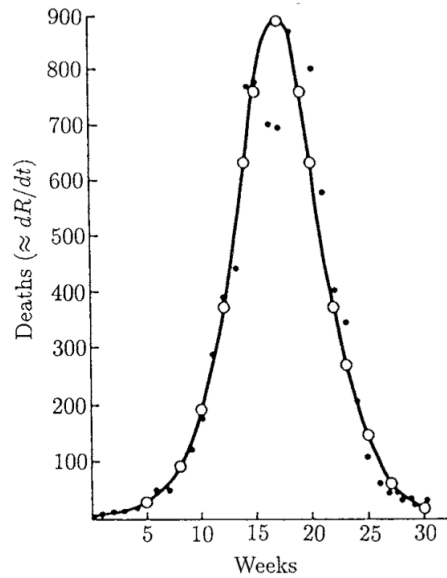


Figura 2.4: Comparação entre os dados (círculos fechados) e o modelo SIR (círculos abertos) para a peste de Bombay.

nossos propósitos. Vamos ver como a inclusão de uma estrutura de rede pode modificar nossos cálculos.

2.3 Redes Complexas

Os modelos que vimos excluem a discussão sobre a rede fazendo uso da hipótese da aproximação de campo médio. Essa é uma aproximação ruim porque no mundo real a probabilidade de que duas pessoas escolhidas ao acaso é desprezível. As pessoas se relacionam apenas com um pequeno número de conhecidos com regularidade. Esse conjunto de contatos em potencial pode ser representado através de uma rede e o tipo de rede pode ter efeitos significantes para a propagação de uma doença.

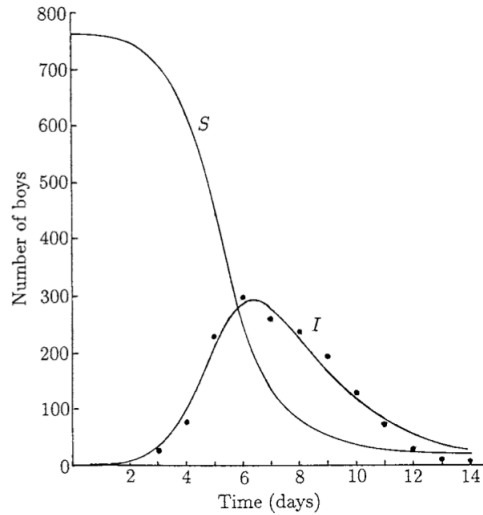


Figura 2.5: Dados para a epidemia de gripe na escola inglesa discutida no texto. Os círculos fechados representam os dados coletados no período.

A definição mais simples para uma rede é um conjunto de pontos conectados por linhas. No jargão do estudo de redes os pontos são chamados de vértices ou nós. Existem muitos sistemas de interesse físico e tecnológico que podem ser modelados com redes. Exemplos típicos de sistemas modelados como redes são a internet, cadeias alimentares e redes sociais.

Tradicionalmente tais sistemas são modelados através do grafo aleatório de Erdős e Rényi (ER). Começa-se considerando uma rede com n nós. Depois escolhemos cada par de nós e os conectamos com uma certa probabilidade p . A partir da última década novos modelos passaram a ser considerados, por duas razões principalmente: o grande aumento de dados relacionados as redes, e o aparecimento de computadores mais potentes e mais baratos. Com isso, se descobriu que as redes reais tem muitas características que diferem do grafo de ER. Algumas grandezas são o coeficiente de clustering, o surgimento de correlações e a distribuição de conectividade.

A distribuição de conectividade $P(k)$ é definida como a probabilidade de escolher um sítio da rede ao acaso e esse sítio ter conectividade igual a k . Também é de interesse é a distribuição condicional $P(k'|k)$, que a probabilidade de que um sítio com conectividade k esteja conectado a outro com conectividade k' , essa última distribuição caracteriza as correlações na rede.

Em muitos cálculos que envolvem redes é conveniente definir a função geratriz da distribuição

$$g(z) = \sum_{k=0}^{\infty} P(k) z^k \quad (2.16)$$

os momentos da distribuição de conectividade pode ser obtido de $g(z)$, usando:

$$\langle k^n \rangle = \left. \frac{d^n g}{d(\ln(z))^n} \right|_{z=1} \quad (2.17)$$

Outra definição importante é a de componentes, um componente é definido como um subconjunto de vértices tal que existe pelo menos um caminho de cada membro do conjunto á cada um dos outros.

O grafo de ER têm para n grande, uma distribuição de Poisson

$$P(k) = \frac{e^{-k} \langle k \rangle}{k!} \quad (2.18)$$

Essa é uma das maiores diferenças entre o grafo de ER e as redes reais, que tem como distribuição uma lei de potência:

$$P(k) = Ak^{-\gamma} \quad (2.19)$$

onde A é o fator de normalização e k_{min} é a menor conectividade da distribuição. Redes reais tem $2 \leq \gamma \leq 3$, tais redes são chamadas livres de escala. O modelo de Barabasi se tornou clássico por gerar uma rede livre de escala com expoente igual a 3, generalizações podem ser usadas obter outros expoentes. Na próxima seção descrevemos um algoritmo simples para se gerar uma rede com a distribuição de conectividade que desejarmos.

2.3.1 Modelo Configuracional

O modelo configuracional é um dos modelos de grafo aleatório generalizado mais famosos. Nesse modelo, não especificamos a distribuição $P(k)$, e sim a sequência de conectividade que é o conjunto de $\{k_1, k_2, \dots\}$ de conectividade de cada vértice. Note que assim também iremos fixar o número de ligações do grafo $m = \frac{1}{2} \sum k_i$. A rede é construída da seguinte maneira: suponha que especificamos a conectividade de cada um dos vértices, atribuímos a cada vértice um número k_i de ligações abertas, depois escolhemos duas ligações abertas e as ligamos da maneira ilustrada na figura (2.6). Esse processo é repetido até que cada um dos vértices tenha a conectividade atribuída na primeira etapa.

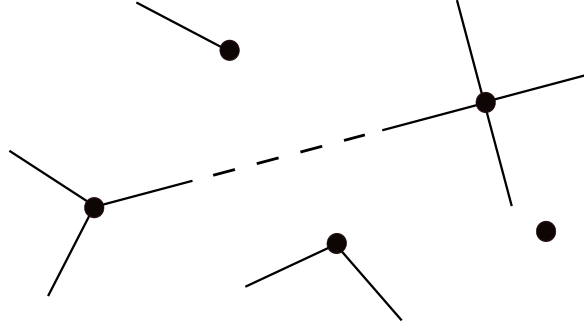


Figura 2.6: Each vertex is given a number of stubs of edges equal to its desired degree. Then pairs of stubs are chosen at random and connected together to form edges (dotted line).

Esse modelo fornece uma forma de gerar redes no computador com uma dada distribuição e permite fazer alguns cálculos analíticos. Por exemplo, suponha que escolhemos um dos vértices ao acaso e seguimos uma de suas ligações até o vértice na outra ponta. Qual a probabilidade de que esse vértice tenha conectividade k ? No modelo configuracional a probabilidade de uma ligação do vértice que escolhemos terminar em qualquer uma das outras pontas é a mesma. Excluindo a ponta que escolhemos existem $2m - 1$ outras pontas na rede e k delas estão conectadas a qualquer vértice com conectividade k , daí a ligação que escolhemos tem probabilidade $k/2m$, no limite de m grande, de terminar em um determinado vértice com conectividade k . Dado que existem no total $nP(k)$ vértices com conectividade k a probabilidade de uma ligação estar em um qualquer um dos vértices com conectividade k é:

$$\frac{k}{2m} \times np_k = \frac{k p_k}{\langle k \rangle} \quad (2.20)$$

onde $\langle k \rangle$ é a conectividade média da rede e usamos que $2m = n\langle k \rangle$.

A probabilidade de alcançar um vértice com $k + 1$ ligações é chamada de distribuição de conectividade em excesso e é definida por

$$q_k = \frac{(k + 1)p_{k+1}}{\langle k \rangle} \quad (2.21)$$

2.4 Modelo SIR - Comportamento assintótico

Agora, todo o processo de transmissão da doença acontece usando a rede de contatos. Definimos a taxa de transmissão ou infecção como sendo a probabilidade por unidade de tempo que a doença seja transmitida de um indivíduo infectado para um susceptível que estejam conectados por uma ligação da rede. Essa taxa é denotada por β como nos modelos de campo médio. Note, que as definições exatamente as mesmas dado que nos modelos de campo médio corresponde a taxa com que indivíduos infectados se encontram com qualquer outro membro da população e aqui diz respeito a taxa de contato entre dois indivíduos.

A taxa de infecção é uma propriedade da doença, certas doenças são mais infecciosas que outras, mas também depende do comportamento da população ao que se está estudando.

Para um dado valor de β , podemos definir modelos para epidemias na rede. Os modelos SI e SIR que vimos anteriormente tem suas versões na rede. Vamos começar com o modelo SI que é mais simples. Cada sítio de uma rede com n nós corresponde a um membro da população. Inicialmente, em $t = 0$, uma fração i_0 desses vértices estão infectados, inclusive podemos usar a condição inicial de que só um vértice está infectado, que nessa situação chamamos de semente. Com uma probabilidade com unidade de tempo β esses indivíduos espalham a doença para seus vizinhos e a doença se espalha pela rede.

É muito complicado resolver esse modelo de maneira geral, mas com relação às propriedades assintóticas o modelo é trivial. Dado que uma vez no estado infectado o indivíduo permanece infectado para sempre. No limite em que $t \rightarrow \infty$ todos os indivíduos que podem ser infectados serão infectados, para qualquer valor de β e i_0 . Só temos uma condição: cada vértice tem que estar conectado com pelo menos um vértice infectado por pelo menos um caminho.

Dito isso, todos os vértices que pertencem ao mesmo componente que o vértice que deu origem a epidemia serão eventualmente infectados. A maioria das redes sociais reais e também aquelas feitas no computador, tem uma componente que contém a grande maioria dos vértices e alguns componentes menores. Se escolhermos a semente ao acaso dentre todos os vértices da rede, a probabilidade dessa semente pertencer a maior componente é igual a S , o fração de rede ocupada pela componente maior, e nesse caso temos um grande número de indivíduos infectados, caso contrário a epidemia fica restrita a uma fração pequena da população. Não podemos prever com exatidão o tamanho da epidemia, a menos que saibamos em qual componente caiu a semente. Esse comportamento é diferente dos modelos de campo médio, que prediz os mesmos resultados dado um conjunto de parâmetros. Temos a introdução de um elemento estocástico. O comportamento do sistema depende da escolha da semente e da estrutura da rede e podemos ter uma saída diferente mesmo usando a mesma rede e os mesmos parâmetros.

A situação é bem mais interessante quando consideramos o modelo SIR numa rede. No modelo SIR, os indivíduos permanecem infectados somente durante um intervalo finito de tempo após o qual eles se recuperam. Não é verdade que um vértice susceptível com um vizinho infectado acabe sendo infectado. Pode ser que ele nunca venha a pegar a doença. A probabilidade de que haja infecção após qualquer intervalo de tempo dt é βdt e de que não haja é $(1 - \beta dt)$. Portanto, a probabilidade de que uma infecção não ocorra em um intervalo de tempo τ finito é igual a

$$\lim_{dt \rightarrow 0} (1 - \beta dt)^{\tau/dt} = e^{-\beta\tau} \quad (2.22)$$

Por fim, a probabilidade de que a doença seja transmitida é

$$\phi = 1 - e^{-\beta\tau} \quad (2.23)$$

Por simplicidade vamos supor que ϕ é o mesmo para toda população. Cada indivíduo susceptível tem a mesma probabilidade de pegar a doença de um de seus vizinhos infectados.

Para estudar as propriedades assintóticas do modelo, vamos usar a formulação proposta em [2], onde cada uma das ligações é considerada “colorida”(ou ocupada) com probabilidade ϕ ou vazia com probabilidade $1 - \phi$. Esse problema é análogo ao problema de percolação por ligações, onde uma fração ϕ das ligações escolhida ao acaso está ocupada. O grupo de vértices conectados por ligações coloridas é chamado de *cluster*.

As ligações fechadas representam contatos pelos quais a doença pode ser transmitida se alcançar uma das pontas dessa ligação, se a doença não alcançar uma dessas pontas não ocorrerá transmissão da doença.

Para valores pequenos de ϕ existem poucas ligações coloridas as quais se agrupam em pequenos clusters desconexos. À medida que ϕ aumenta esses cluster menores se agrupam formando um cluster que tem o tamanho da ordem de n , chamado de cluster gigante. Se ϕ aumenta ainda mais o cluster gigante continua aumentando e atinge seu valor máximo quando $\phi = 1$. Contudo, não necessariamente esse cluster vai englobar todos os elementos da rede porque é limitado pelo tamanho da maior componente, que por sua vez pode ser menor que n . A transição em que aparece o cluster gigante é chamada de transição de percolação.

Voltando para o espalhamento de doenças: para ϕ pequeno, a semente pertence a um dos clusters pequenos e não há epidemia. Suponha agora que $\phi > \phi_c$ e existe um cluster percolante, nesse caso existe a possibilidade de uma epidemia. O cluster gigante ocupa uma fração S da rede e portanto existe uma probabilidade S que a semente caia nesse cluster dando origem a uma epidemia. A medida em que ϕ aumenta, aumenta também a probabilidade de haver uma epidemia. Portanto a transição de percolação por ligações correspondente ao limiar epidêmico para uma doença ocorrendo na mesma rede.

Considere então o modelo SIR ocorrendo em uma rede montada com o modelo configuracional com probabilidade $P(k)$. Considere que u é a probabilidade média de que um dado vértice não esteja conectado ao cluster percolante por uma de suas ligações. Existem duas possibilidades: ou essa ligação está vazia, o que acontece com probabilidade $1 - \phi$, ou está ocupada mas não pertence ao cluster percolante. A última opção será verdade se o vértice não está conectado ao cluster percolante por nenhuma de suas ligações, o que tem probabilidade u^k , no caso em que o vértice tem k ligações. Somando as duas possibilidades temos que a probabilidade total será $(1 - \phi + \phi u^k)$, onde valor de k tem distribuição 2.21. Fazendo a média sobre todos os valores de k , ficamos com:

$$u = 1 - \phi + \phi \sum_{k=0}^{\infty} q_k u^k = 1 - \phi + \phi g_1(u) \quad (2.24)$$

na equação acima $g_1(u)$ é a função geratriz de q_k .

The probability that a vertex of total degree k does not belong to the giant cluster is now simply u^k , and the average such probability over the whole network, which is equal to $1 - S$, is calculated by averaging u^k over the degree distribution p_k giving

$$S = 1 - \sum_{k=0}^{\infty} p_k u^k = 1 - g_0(u) \quad (2.25)$$

As equações 2.24 e 2.25 fornecem a solução para o nosso problema. No entanto, não conhecemos a forma exata para $g_1(k)$, mas podemos encontrar a solução ainda assim usando um método gráfico.

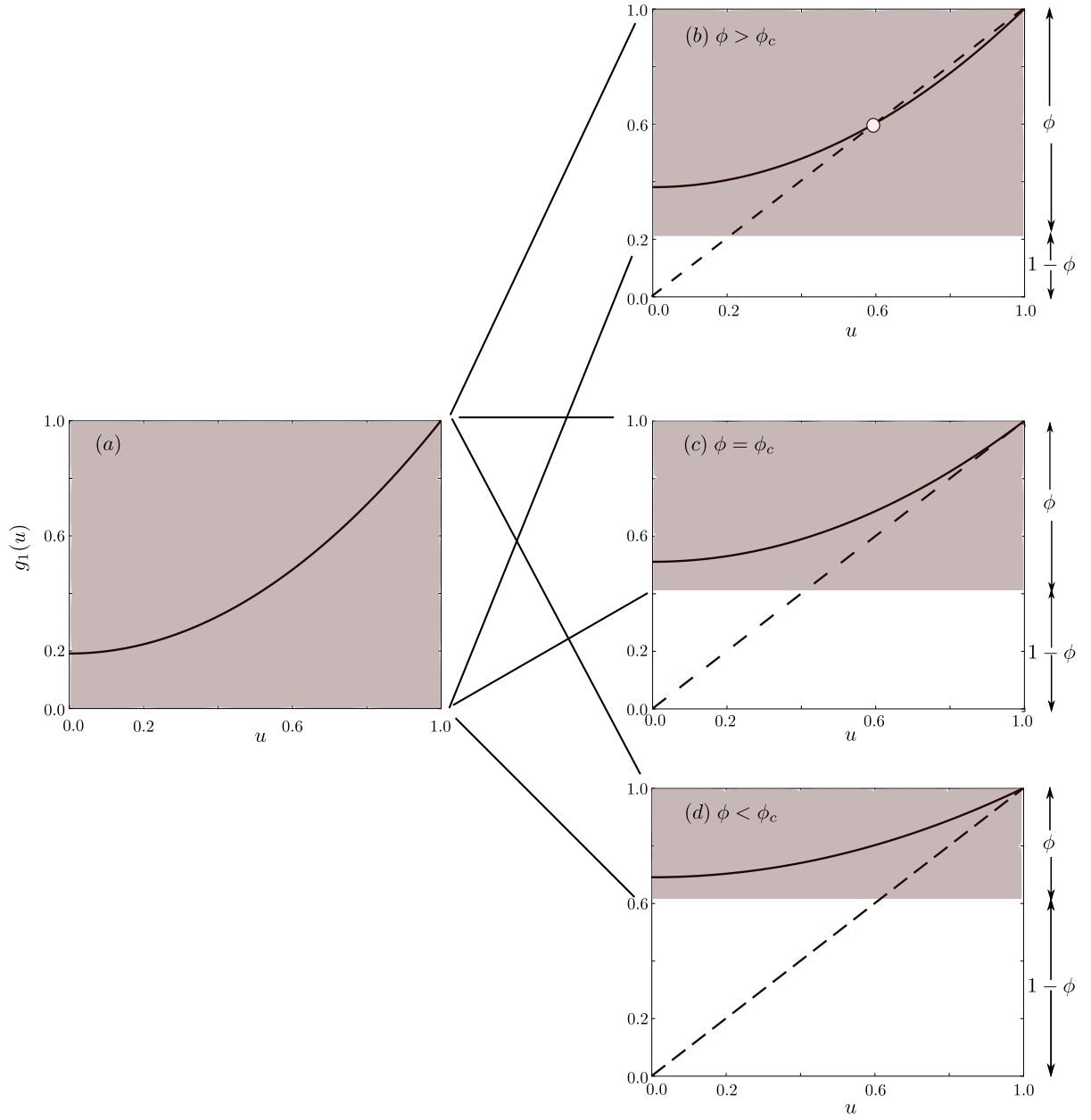


Figura 2.7: Método gráfico usado para resolver a equação 2.24.

A função característica $g_1(x)$ é um polinômio com todos os coeficientes não negativos, e daí é uma função crescente com todas as derivadas positivas como ilustrado na figura (2.4). A função 2.24 pode ser obtida comprimindo g_1 por um fator ϕ e depois somando $1 - \phi$. As interseções com a curva $y = u$ nos fornece as soluções.

Para um valor grande de ϕ existem duas soluções, uma delas, $u = 1$ é trivial e sempre acontece. A outra solução nos dá o tamanho do maior cluster.

Se u é muito pequeno só existe a solução trivial. O ponto de transição acontece quando a curva $y = 1 - \phi + \phi g_1(u)$ é tangente a reta $y = u$, ou seja, quando

$$\left[\frac{d}{du}(1 - \phi + \phi g_1(u)) \right]_{u=1} = 1 \quad (2.26)$$

Performing the derivative we then find that the value of \ddot{I}^\dagger at the transition, which we call the critical value, denoted \ddot{I}^\dagger_c , is

$$\phi_c = \frac{1}{g'_1(1)} \quad (2.27)$$

Usando as equações 2.16 e 2.21:

$$\begin{aligned} g'_1(1) &= \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k+1)p_{k+1} = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k-1)p_k \\ &= \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \end{aligned} \quad (2.28)$$

substituindo em 2.27, ficamos com

$$\phi_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (2.29)$$

ou, em termos das variáveis originais, β e τ :

$$\beta\tau = -\ln(1 - \phi_c) = \ln \left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle} \right) \quad (2.30)$$

abaixo de ϕ_c qual a prevalência de epidemia é nula, e acima do qual ela atinge uma população finita. Um caso importante a ser considerado são as distribuições com lei de potência. Se temos uma rede livre de escala, ou seja, com expoente entre 2 e 3, então $\phi_c = 0$, porque o segundo momento diverge quando $N \rightarrow \infty$, mas a média permanece finita. Portanto nesse caso sempre temos uma epidemia, mesmo se a taxa de transmissão da doença for muito pequena. Por outro lado para redes reais que tem N finito, existe um limiar que embora pequeno é diferente de zero.

2.5 Aproximação por conectividade

A aproximação introduzida na referência [3] assume que cada vértice com a mesma distribuição tem a mesma probabilidade de estar infectado em um dado instante de tempo. Vamos definir $s_k(t)$, $i_k(t)$ e $r_k(t)$ como sendo a probabilidade de que um sítio com conectividade k , esteja no estado susceptível, infectado ou removido, respectivamente, no tempo t .

Considere um vértice qualquer da rede, para se tornar infectado esse vértice tem que ter pego a doença de um dos seus vizinhos, ou seja, esse sítio deve ter uma ligação que aponte para um vértice infectado. Vimos na seção que a probabilidade de que uma dada ligação aponte para um sítio com conectividade k é proporcional á $kP(k)$. Portanto a probabilidade de que uma dada ligação aponte para um sítio infectado será:

$$\Theta(t) = \frac{\sum_k kP(k)i_k(t)}{\sum_k kP(k)} = \frac{\sum_k kP(k)i_k(t)}{\langle k \rangle} \quad (2.31)$$

Se o vizinho de um dado vértice está infectado, então a probabilidade de que a doença seja transmitida para esse vértice em um intervalo de tempo dt é βdt . A probabilidade

de pegar a doença de um dos vizinhos será $\beta\Theta(t)dt$ e por fim a probabilidade total de transmissão da doença por qualquer um dos vizinhos é $\beta k\Theta(t)dt$. Em adição exigimos que o próprio sítio em consideração seja susceptível o que tem probabilidade s_k de acontecer. Portanto a taxa de mudança em s_k é dada por

$$\frac{ds_k}{dt} = -\beta k s_k \Theta(t) \quad (2.32)$$

as outras probabilidades evoluem segundo as equações:

$$\frac{di_k}{dt} = -\beta k s_k \Theta(t) - \gamma i_k \quad (2.33)$$

e

$$\frac{dr_k}{dt} = \gamma i_k \quad (2.34)$$

as condições iniciais são $r_k(0) = 0$, $i_k(0) = i_k^0$ e $s_k(0) = 1 - i_k^0$. Podemos fazer também $s(0) = 1$ assumindo que o tamanho do sistema é grande e a fracção de infectados é pequena. Vamos definir a probabilidade de que um vizinho esteja no estado r .

$$\theta = \frac{\sum_k k P(k) r_k(t)}{\langle k \rangle} \quad (2.35)$$

derivando essa expressão em relação ao tempo, encontramos:

$$\frac{d\theta}{dt} = \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{dr_k}{dt} = \gamma \Theta, \quad (2.36)$$

substituindo essa expressão em 2.32 e integrando:

$$s_k = e^{-\beta k \theta / \gamma} \quad (2.37)$$

e voltando para a expressão 2.36

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{\gamma}{\langle k \rangle} \sum_k k P(k) i_k \\ &= \frac{\gamma}{\langle k \rangle} \sum_k k P(k) (1 - r_k - s_k) \\ &= \gamma - \gamma \theta - \frac{\gamma}{\langle k \rangle} \sum_k k P(k) e^{-\beta k \theta / \gamma} \end{aligned} \quad (2.38)$$

A solução da equação acima pode ser usada para obter a prevalência da doença r_∞ como função de $\theta_\infty = \lim_{t \rightarrow \infty} \theta(t)$. Dado que $r_\infty = 1 - s_\infty$, temos

$$r_\infty = \sum_k P(k) (1 - e^{-\beta k \theta_\infty / \gamma}) \quad (2.39)$$

Não podemos obter uma expressão fechada para a equação ??, mas podemos encontrar uma expressão assintótica para a equação acima, no limite em que $t \rightarrow \infty$, ou seja, no final da epidemia. Nesse limite temos $i(t \rightarrow \infty) = 0$ e daí $\lim_{t \rightarrow \infty} d\theta/dt = 0$ e obtemos a seguinte expressão para θ_∞

$$\theta_\infty = 1 - \frac{1}{\langle k \rangle} \sum_k k P(k) e^{-\beta k \theta_\infty / \gamma} \quad (2.40)$$

$\theta = 0$ será sempre uma solução. Para obter uma solução não trivial podemos usar um método gráfico, como descrito na figura [XXXXX], parecido com o que usamos na seção

(2.4). A curva $y = 1 - \frac{1}{\langle k \rangle} \sum_k kP(k)e^{-\beta k \theta_\infty / \gamma}$ está ilustrada na figura, vemos que outra solução só existe se

$$\left. \frac{d}{d\theta_\infty} \left(1 - \frac{1}{\langle k \rangle} \sum_k kP(k)e^{-\beta k \theta_\infty / \gamma} \right) \right|_{\theta_\infty=0} > 1, \quad (2.41)$$

essa relação implica

$$\frac{1}{\langle k \rangle} \sum_k kP(k) \left(\frac{\beta}{\gamma} k \right) = \beta \frac{\langle k^2 \rangle}{\langle k \rangle} > 1 \quad (2.42)$$

Esse condição estabelece o limiar epidêmico

$$\beta_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \quad (2.43)$$

A diferença em relação ao resultado anterior pode ser atribuída as diferenças entre as definições dos modelos.

2.6 Modelos de Rumores

Como vimos nas seções anteriores, no estudo de epidemias queremos modelar a dinâmica de uma certa doença para eventualmente conseguir desenvolver estratégias para controlar o avanço da doença. No entanto, o objetivo é exatamente o oposto queremos que o rumor se espalhe o mais rápido possível e atinja o maior número de pessoas. Algumas aplicações dos modelos de rumores são a disseminação de protocolos para disseminação de dados na internet e estratégias em campanhas de marketing. Uma diferença entre essas duas classes, é que em modelos para o espalhamento de rumores temos a liberdade de criar as regras que melhor atentem os objetivos.

O modelo mais famoso de rumores é conhecido como modelo de Daley-Kendall (DK). Cada um dos n elementos de rede pode estar em 3 estados distintos: ignorantes I , informantes S ou contidos R . Os indivíduos ignorantes são aqueles que não ouviram o rumor, ou seja, não foram atingidos pela informação que se espalha. A classe S diz respeito aos elementos que estão espalhando o rumor pela rede e a última classe são aqueles elementos que ouviram o rumor mas perderam o interesse em espalhá-lo.

A dinâmica do modelo, em cada instante de tempo, pode ser resumido através do esquema:

$$I(i) + S(j) \xrightarrow{\lambda} S(i) + S(j) \quad (2.44)$$

$$S(i) + S(j) \xrightarrow{\alpha} R(i) + S(j) \quad (2.45)$$

$$S(i) + R(j) \xrightarrow{\alpha} R(i) + R(j) \quad (2.46)$$

$$(2.47)$$

onde i e j são vizinhos. Através do contato entre ignorantes e informantes se dá o espalhamento do rumor. Os elementos da classe I se tornam informantes com uma taxa λ . Há dois processos que podem levar o informante a passar para a classe dos indivíduos contidos: ou ele esquece o rumor e essa transição é espontânea, ou ele se dá conta de que o rumor perdeu o valor de novidade, isso acontece com uma taxa α quando tal elemento encontra outro informante ou um elemento da classe R . Por ora assumimos que só vale a última opção.

Temos que tomar um certo cuidado com a nova definição para não confundir com as usadas nos modelos anteriores. Vamos usar essa definição porque é a notação padrão usada no modelo de Daley e Kendal.

Na hipótese de mistura homogênea o modelo DK pode ser descrito usando a densidades de ignorantes $i(t)$, informantes $s(t)$ e contidos $r(t)$. As equações de campo médio são:

$$\frac{di}{dt} = -\lambda i s \quad (2.48)$$

$$\frac{ds}{dt} = \lambda i s - \alpha s[s + r] \quad (2.49)$$

$$\frac{dr}{dt} = \alpha s[s + r] \quad (2.50)$$

com as condições iniciais $i(0) = (n-1)/n$, $s(0) = 1/n$ e $r(0) = 0$, onde n é o número de indivíduos na população. Temos também a equação de vínculo,

$$s + r + i = 1, \quad (2.51)$$

que assume que a população é constante. As equações acima tem solução no limite em que $t \rightarrow \infty$. Consideramos que $s(t \rightarrow \infty) \equiv s_\infty = 0$, $r(t \rightarrow \infty) \equiv r_\infty$ e consequentemente $i(t \rightarrow \infty) = 1 - r_\infty$. Dividindo a equação (2.49) por (2.48) e usando a equação (2.51), temos:

$$\frac{ds}{di} = -1 + \frac{\alpha}{\lambda} \frac{1-i}{i}, \quad (2.52)$$

integrando de $[t_0, t]$,

$$s(t) - s(t_0) = -\left(1 + \frac{\alpha}{\lambda}\right) (i(t) - i(t_0)) + \frac{\alpha}{\lambda} \ln\left(\frac{i(t)}{i(t_0)}\right) \quad (2.53)$$

tomando o limite de $t \rightarrow \infty$, $t_0 \rightarrow 0$ e de $n \gg 1$

$$\ln(1 - r_\infty) = -\frac{\lambda}{\alpha} \left(1 + \frac{\alpha}{\lambda}\right) r_\infty \Rightarrow r_\infty = 1 - e^{-\beta r_\infty}, \quad (2.54)$$

onde $\beta = 1 + \lambda/\alpha$. Usando o mesmo truque da solução gráfica, temos que

$$\beta > 1 \Rightarrow \lambda/\alpha > 0 \quad (2.55)$$

mas essa condição sempre é atendida, o que indica que não existe um “limiar epidêmico”. Esse efeito é causado pelo novo termo não-linear na equação de Daley-Kendall, não presente no modelo SIR. Na próxima seção, vamos voltar nosso foco para o que acontece quando a estrutura de rede é incluída.

2.7 Resultados Numéricos

Para topologias mais complexas, temos que nos apoiar em simulações e cálculos numéricos. Vamos usar a aproximação por conectividade usada no modelo SIR para levar em conta a heterogeneidade da rede. Agora além dos três estados anteriores, cada nó pertence a uma classe com conectividade diferente. Se denotarmos por $i_k(t)$, $s_k(t)$ e $r_k(t)$, a densidade de ignorantes, informantes e contidos com conectividade k e usando $r_k + s_k + i_k = 1$ as equações para a taxa de variação dessas variáveis se escrevem como:

$$\frac{di_k}{dt} = -\lambda k i_k \frac{\sum_k k P(k) s_k}{\langle k \rangle} \quad (2.56)$$

$$\frac{ds_k}{dt} = \lambda k i_k \frac{\sum_k k P(k) s_k}{\langle k \rangle} - \alpha k s_k \frac{\sum_k k P(k) (s_k + r_k)}{\langle k \rangle} \quad (2.57)$$

$$\frac{dr_k}{dt} = \alpha k s_k \frac{\sum_k k P(k) (s_k + r_k)}{\langle k \rangle} \quad (2.58)$$

Para ilustrar resultados de cálculos numéricos típicos para o esse caso [4]. Duas tipos de estrutura foram consideradas: Erdős e Rényi e uma distribuição livre de escala, nessa última se considerou $\gamma = 3$. As equações foram resolvidas usando um esquema de diferenças finitas. Nessa referência também é incorporado um mecanismo de “esquecimento”, onde os informantes espontaneamente param de espalhar o rumor com uma taxa δ . Isso introduz os termos δs_k e $-\delta s_k$ nas equações 2.58 e 2.57 respectivamente.

O primeiro conjunto de dados mostra o tamanho final do rumor, ou seja, o número de pessoas atingidas:

$$R = \sum_k r_k(\infty) \quad (2.59)$$

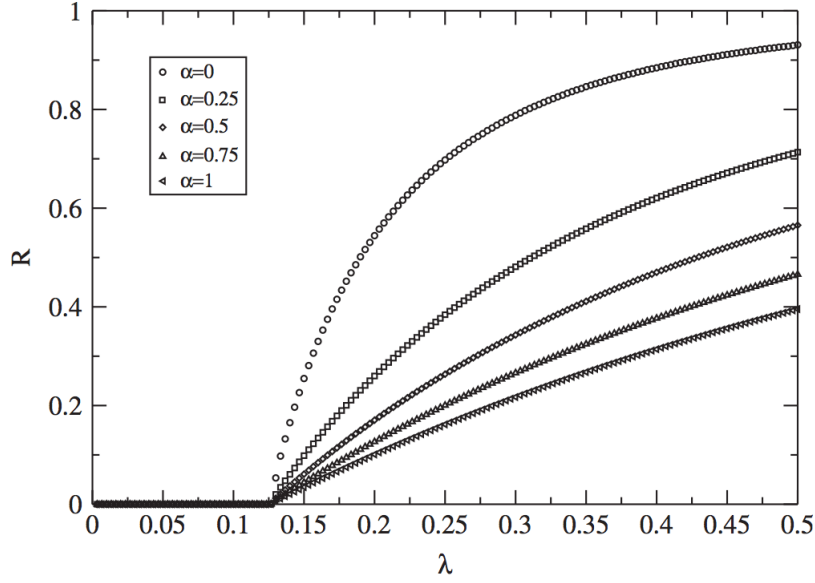


Figura 2.8: Tamanho final do rumor para rede ER, aqui R é função da taxa de espalhamento λ . A rede considerada tem tamanho N^6 .

Para a rede ER, o tamanho do rumor exibe uma transição de fase para um certo valor de λ_c estimado em 0.12507. Esse resultado parece ser independente de α . Nas vizinhanças do ponto crítico R se comporta na forma (veja 2.9):

$$R = A(\lambda - \lambda_c) \quad (2.60)$$

Os resultados para a rede livre de escala são mostrados na figura 2.10. Outra vez temos um valor limite de λ para o qual ocorre epidemia. Note no entanto que esse valor é bem menor que para a rede ER. Como o valor de λ_c não depende de α podemos usar o resultado que encontramos para o modelo SIR. As equações para DK com esquecimento se reduzem as do modelo SIR no caso em que $\alpha = 0$. Daí, concluímos que no limite de uma rede infinita, o valor do limiar tende a zero.

Para caracterizar ainda mais o comportamento nessa rede, pode-se fitar os resultados com uma exponencial (figura ??):

$$R \sim \exp(-C/\lambda) \quad (2.61)$$

e o fator C depende varia pouco com α .

Usando a simulação numérica também obter a evolução temporal do tamanho do rumor, ilustrado na figura 2.12.

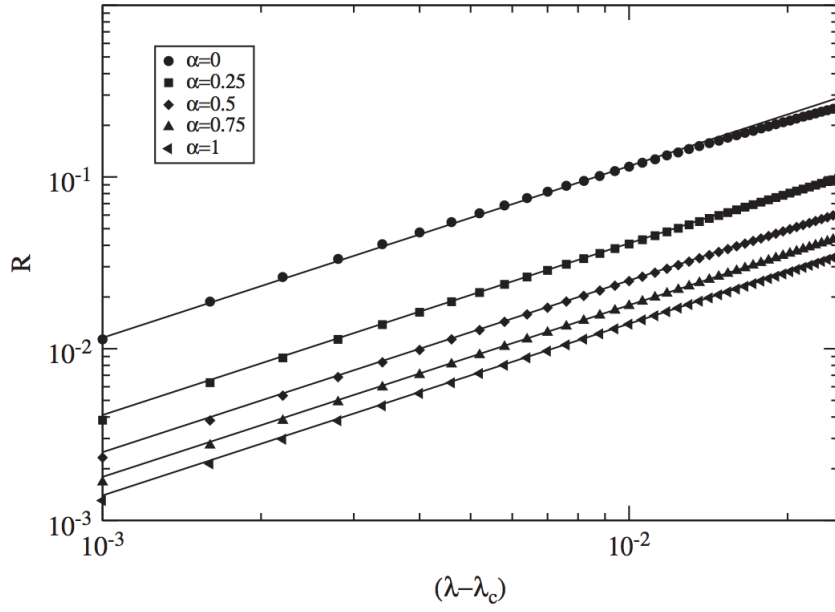


Figura 2.9: Ajuste da equação 2.60.

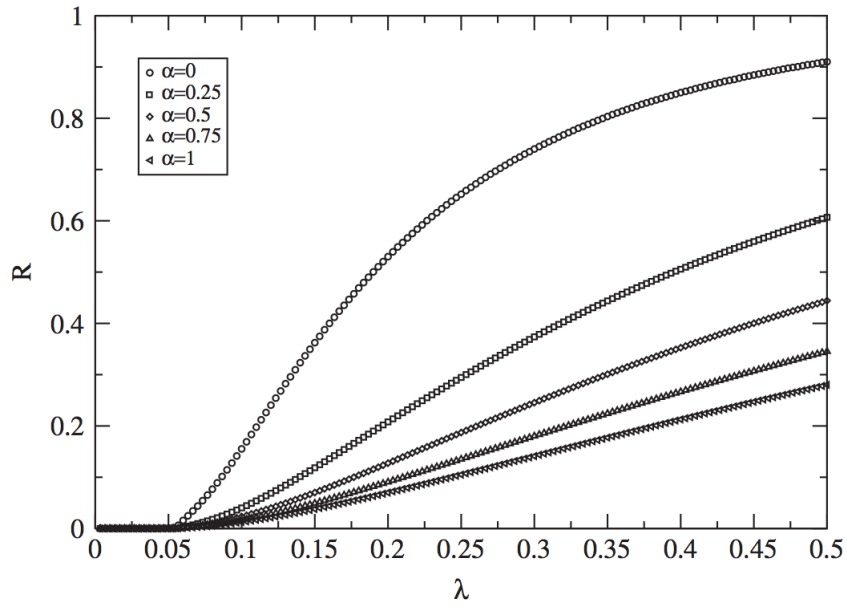


Figura 2.10: Tamanho final do rumor para rede SF com tamanho $N = 10^6$.

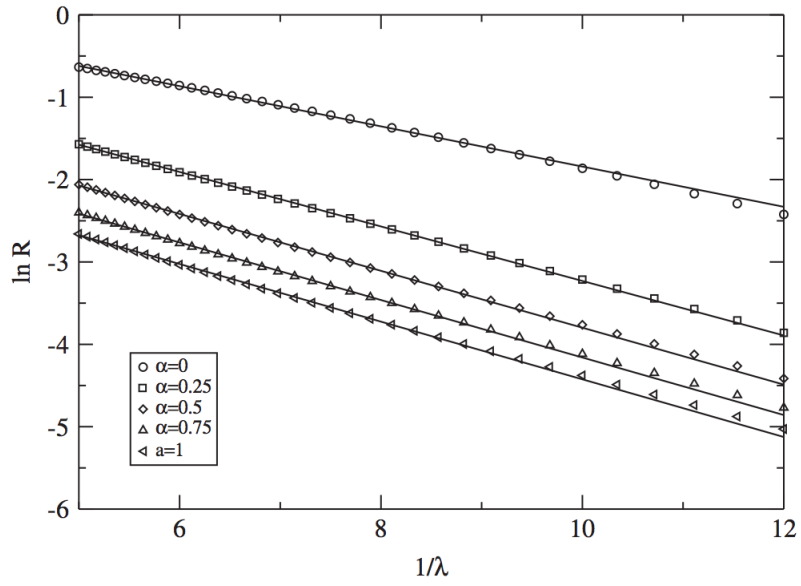


Figura 2.11: Ajuste da equação 2.61.

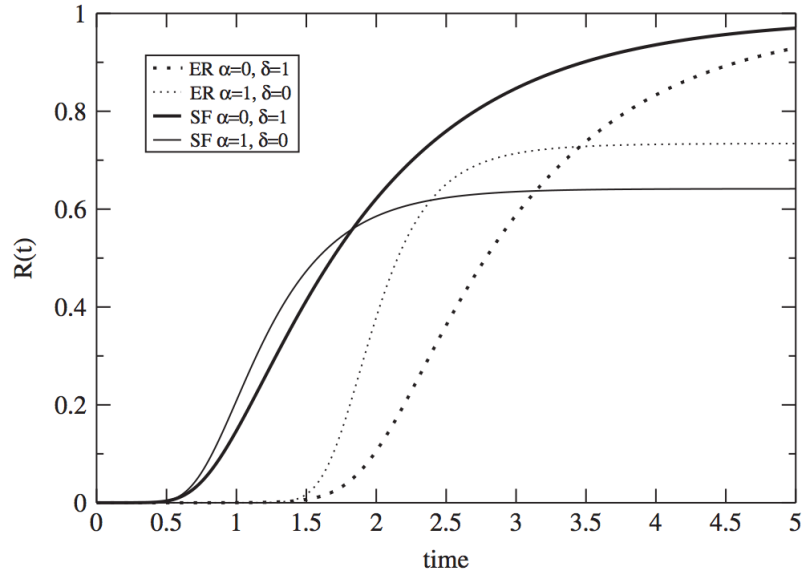


Figura 2.12: Evolução temporal da densidade de indivíduos contidos para as redes consideradas no texto.

Capítulo 3

Propostas e Conclusões

- outros detalhes em modelos já estudados além das propriedades estacionárias, é importante caracterizar os transientes, quando leva em chegar no SS
- novos
- difusão
- rumores \rightarrow generalizações

Referências Bibliográficas

- [1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. *Physics Reports*, 424(4-5):175–308, February 2006.
- [2] P. Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157 – 172, 1983.
- [3] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 26:521–529, 2002. 10.1140/epjb/e20020122.
- [4] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 374(1):457 – 470, 2007.