



**INSTITUTO DE FÍSICA**  
Universidade Federal Fluminense

**MODELOS PARA OTIMIZAÇÃO DE  
TRÁFEGO E PARA A FORMAÇÃO DE  
ESTRUTURAS MODULARES NA REDE DE  
DEPENDÊNCIAS DO DEBIAN GNU/LINUX**

**por Orahcio Felício de Sousa**

**Niterói-RJ**

**2013**

Orahcio Felício de Sousa

MODELOS PARA OTIMIZAÇÃO DE  
TRÁFEGO E PARA A FORMAÇÃO DE  
ESTRUTURAS MODULARES NA REDE DE  
DEPENDÊNCIAS DO DEBIAN GNU/LINUX

Tese apresentada ao programa de Pós-graduação em Física da Universidade Federal Fluminense como requisito parcial para obtenção do grau de doutor em Física.

**Orientador:** Thadeu Josino Pereira Penna

Niterói-RJ, Junho de 2013

Tese de doutorado sob o título “*Modelos para otimização de tráfego e para a formação de estruturas modulares na rede de dependências do Debian GNU/Linux*”, defendida por Orahcio Felício de Sousa e aprovada em oito de julho de 2013, em Niterói, estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Thadeu Josino Pereira Penna  
Orientador  
Universidade Federal Fluminense

Prof. Dr. Sílvio Manuel Duarte Queirós  
Centro Brasileiro de Pesquisa Físicas

Prof. Dr. Nuno Miguel Melo Crokidakis  
Peregrino  
Pontifícia Universidade Católica-RJ

Prof. Dr. Jürgen Fritz Stilck  
Universidade Federal Fluminense

Prof. Dr. Antonio Tavares da Costa Junior  
Universidade Federal Fluminense

# *Agradecimentos*

Agradeço aos meus orientadores Thadeu Josino Pereira Penna e Sergey Dorogovtsev – orientador em Portugal – pelas discussões durante esse trabalho de pesquisa. Agradeço também as discussões no Brasil ocorridas graças aos professores Márcio Argollo, Jorge Sá e Paulo Murilo. Agradeço todo apoio, amizade e ótimas discussões com os colegas Alexandre Lima, Florência Noriega, Vitor Lara, Marlon Ramos, Nuno Miguel, Víctor Becerra e Thiago Caramês.

Durante a estada em Portugal agradeço aos professores e colegas José Fernando Mendes, KyoungEun Lee, Sooyeon Yoon, Nahid Azimi, Detlef Holstein, Marinho Lopes, Rui Costa por todo apoio, amizade e discussões. Agradeço também aos amigos Fabrício Forgerini, Grasiely Borges e Arash Sohrabi por todo apoio.

Minha mãe e meu pai, Maria Gilvanda Felício de Sousa e José Benício de Sousa agradeço por todo apoio e carinho. Obrigado também a Erahsto, Mariana e Matehus Felícios de Sousas, que sempre estiveram na melhor torcida onde quer que eu fosse.

Agradeço às amigas e amigos que mantiveram-me espiritualmente em Niterói, Jaqueline Siqueira Rocha, Maria Eugênia, Vladimir Gonçalves Miranda, Carlos Eduardo Rodrigues e o Sh. Abd al-Rashyd.

“(...) Tolerância às cheias é outra coisa que estão a patentear. Culturas resistentes ao clima são invenções suas como se fosse produtos da engenharia genética, quando na verdade são produto de pirataria, uma forma usual de colonização.”

*Vandana Shiva*  
O Tempo e o Modo, RTP2 2012

# *Resumo*

A tese apresenta em sua primeira parte o estudo de uma dinâmica aplicada a fluxos em uma rede regular quadrada, que tem características semelhantes a um processo estocástico de reações de partícula única, como reações de difusão, coalescência e criação de partículas. Utilizamos uma aproximação de campo médio para obtermos a solução analítica das distribuições de fluxos estacionárias. Essa solução concorda em boa aproximação com nossas simulações numéricas para o mesmo modelo.

Na segunda parte é apresentada a estrutura de comunidades de um sistema real definido pelo sistema operacional livre Debian GNU/Linux. Estudamos a estrutura geral da rede de pacotes da última versão estável desse sistema e a estrutura de comunidades, comparando essas subestruturas com as divisões pré-estabelecidas pelos desenvolvedores desse sistema. Propusemos um método para agrupar desenvolvedores individuais em grupos que maximizam as relações de dependência entre os pacotes por eles desenvolvidos.

# *Abstract*

The first part of this thesis shows a study of flow dynamics applied on directed regular network, which presents similar features of single particle reaction process, with diffusion, coalescence and branching reactions. We obtained a solution by mean field approach for the steady distribution of flows, that presents a good agreement with our numerical simulations.

In the second part, it is presented the community structure of the real system defined by the Debian GNU/Linux free operating system. We studied the general structure of the packages network on the last steady release and the substructures of communities, the emergence of these ones and the relation between the maintainers groups of these packages. We propose a method for grouping individual maintainers that maximizes the number of dependency relations per group.

# *Sumário*

## **Lista de Figuras**

<b>1</b>	<b>Introdução</b>	p. 15
<b>2</b>	<b>Recursos preliminares</b>	p. 18
2.1	Soma de variáveis aleatórias . . . . .	p. 18
2.2	Distribuição uniforme . . . . .	p. 19
2.3	Distribuição exponencial . . . . .	p. 21
2.4	Transformada de Laplace . . . . .	p. 21
<b>3</b>	<b>Aproximação de campo médio para otimização de fluxos numa rede regular direcionada</b>	p. 24
3.1	Introdução . . . . .	p. 24
3.2	Otimização local de fluxos . . . . .	p. 25
3.3	Modelo Finito . . . . .	p. 29
3.3.1	Método de intervalos vazios para processos de reação-difusão . . . . .	p. 32
3.4	Aproximação de campo médio . . . . .	p. 35
3.4.1	Coalescência. . . . .	p. 40
3.5	Custos uniformemente distribuídos . . . . .	p. 42
3.6	Limite para grandes correntes . . . . .	p. 44

3.7	Limite para baixas correntes . . . . .	p. 49
3.7.1	Tempo de relaxação . . . . .	p. 52
3.7.2	Número arbitrário de canais de entrada e saída . . . . .	p. 53
3.7.3	Correlação entre dois pontos . . . . .	p. 54
3.8	Conclusões . . . . .	p. 55
<b>4</b>	<b>Formação de estruturas modulares na rede de dependência do Debian GNU/Linux</b>	<b>p. 57</b>
4.1	Rede de dependências entre pacotes Debian . . . . .	p. 57
4.2	Propriedades gerais . . . . .	p. 59
4.2.1	Distribuição de conectividade . . . . .	p. 60
4.2.2	Menor caminho médio e centralidade de entroncamento, <i>betweenness</i>	p. 61
4.2.3	Coefficiente de agrupamento . . . . .	p. 63
4.3	Estrutura de comunidades ou modularidade . . . . .	p. 67
4.3.1	Modularidade em redes não-direcionadas . . . . .	p. 67
4.3.2	Modularidade em redes direcionadas . . . . .	p. 69
4.4	Comunidades na rede Debian . . . . .	p. 70
4.5	Maximização da modularidade . . . . .	p. 71
4.5.1	Método espectral . . . . .	p. 72
4.5.2	Arrefecimento simulado . . . . .	p. 75
4.5.3	Algoritmo guloso . . . . .	p. 77
4.6	Resultados . . . . .	p. 78
4.7	Conclusão . . . . .	p. 83

**5 Conclusão**

p. 85

**Referências Bibliográficas**

p. 87

## *Lista de Figuras*

- 2.1 Distribuição resultante da soma de duas variáveis aleatórias. . . . . p. 20
- 3.1 Bifurcação de um canal com fluxo  $i$ , cada caminho possui um custo ligado ao tempo para gasto para concluir o caminho escolhido  $j$  ou  $k$ . . . . . p. 25
- 3.2 Diagrama fundamental do tráfego, fluxo ( $F$ ) de veículos *versus* densidade ( $D$ ). Dados obtidos no endereço eletrônico do departamento de trânsito do estado de Minnesota, vide referências bibliográficas. A reta preta indica a relação linear  $F \propto D$  que se adequa bem aos dados para valores baixos da densidade e fluxo. Figura retirada de Peregrino (2011, página 28). . . . . p. 26
- 3.3 Função para os fluxos ótimos  $j$  em dois canais, dado um fluxo de entrada  $i$ , antes e depois do valor crítico de fluxo definido entre os canais. A curva azul é a solução ótima para o canal de menor custo e a vermelha para segundo de maior custo, a soma das duas curvas deve ser sempre igual a  $i$ . . . . . p. 28
- 3.4 Representação geométrica do modelo. É injetada uma corrente  $j$  em cada sítio da linha 0 distribuída ao longo da cadeia conforme uma distribuição  $Q(0, j)$ . Esta gera a distribuição  $P(0, j)$  nas arestas subsequentes, tais correntes são somadas nos sítios da linha 1 gerando a distribuição de correntes  $Q(1, j)$ . Os passos acima descritos são repetidos ao longo da direção  $t$ . . . . . p. 30
- 3.5 Eventos de divisão e entrada de fluxo num sítio e processos consequentes. A escala de cinza indica o tempo, do inicial (branco) até o final (preto). Linhas pontilhadas indicam a possibilidade de processos adicionais. . . . . p. 31
- 3.6 Termos para fluxos provenientes de sítios com fluxo não-nulo. . . . . p. 36

3.7	Termos para fluxos resultantes não-nulos. . . . .	p. 37
3.8	Termos para a distribuição de correntes nos vértices, equação 3.28. . . . .	p. 38
3.9	Gráfico para o tempo de coalescência. Simulação com $2L = 20000$ canais, $\langle j \rangle = 10^{-9}$ e 25 amostras, usamos uma distribuição uniforme de custos para os vértices da rede que define a corrente crítica máxima $\beta = 1$ . A curva vermelha contínua representa uma lei de potência com expoente $1/2$ . . . . .	p. 41
3.10	Iteração numérica das equações 3.28 e 3.44 para $t_{\text{máx}} = 250$ passos. Os gráficos correspondem aos valores para corrente média $\langle j \rangle = 5 \cdot 10^{-4}$ , $10^{-2}$ , e $1,5$ respectivamente. A largura da distribuição uniforme foi escolhida igual a unidade ( $\beta=1$ ), por simplicidade. A linha vermelha para o primeiro gráfico corresponde a um decaimento exponencial do tipo $\sim \exp\left(-\frac{2}{\sqrt{\langle j \rangle}}j\right)$ . . . . .	p. 45
3.11	Iteração final para a fração de canais usados, foi obtido um comportamento monótono para o parâmetro de ordem com o tempo em torno de 250 iterações partindo de uma distribuição inicial uniforme. Usamos $\beta = 1$ por simplicidade na implementação. . . . .	p. 46
3.12	Solução numérica para distribuição estacionária de fluxos nos canais, eq. 3.54, no limite de grandes fluxos, $B = 0$ (pontos). Parâmetros utilizados: $\langle j \rangle = 10^4$ , $\beta = 1$ . A fim de comparação a curva contínua corresponde a uma gaussiana de variância $1/12$ . . . . .	p. 48
3.13	Determinação do coeficiente da variância para a distribuição estacionária para altas correntes. . . . .	p. 48
3.14	Simulação para uma rede com custos uniformemente distribuído no intervalo $(0, 1]$ , $\langle j \rangle = 1$ , $L = 1000$ e 100 amostras. A gaussiana ajustada possui média $\mu = 1,01$ e variância $\sigma^2 = 1/2$ . Figura retirada de Forgerini (2013, página 89). . . . .	p. 49

3.15	Distribuição de correntes para o limite de baixa correntes. Parâmetros utilizados: $\beta = 1$ , $\langle j \rangle = 5 \cdot 10^{-5}$ , $5 \cdot 10^{-6}$ , $5 \cdot 10^{-7}$ , respectivamente. Os ajustes seguem a equação 3.60. Figura retirada de Forgerini (2013, página 89). . . . .	p. 51
4.1	Exemplo de uma rede de dependências a partir do pacote <i>iceweasel</i> , vermelho escuro. As cores ficam mais claras conforme o grau de vizinhança aumenta do sítio mais escuro. . . . .	p. 59
4.2	Distribuição cumulativa de conectividade $P(k)$ da rede de dependências entre pacotes Debian <i>squeeze</i> . A linha contínua corresponde a lei de potência com $\alpha = 1$ e a linha tracejada ao coeficiente de decaimento $\beta = 0,15$ ; os erros para cada coeficiente estão estimados em $6,8 \cdot 10^{-4}$ e $1,4 \cdot 10^{-3}$ respectivamente, conforme ajuste. A distribuição de tamanhos de <i>clusters</i> entre pacotes de mesmo colaborador $P(s)$ são os pontos em cruz. . . . .	p. 61
4.3	Distribuição cumulativa de <i>betweenness</i> $P(C_B)$ da rede de dependências entre pacotes Debian <i>squeeze</i> . A linha contínua corresponde a lei de potência com $\gamma = 0,74$ . O erro para o valor do expoente foi estimado em $5,6 \cdot 10^{-4}$ , conforme ajuste. O <i>betweenness</i> foi normalizado pela quantidade de ligações na rede. . . . .	p. 63
4.4	Centralidade de entrocamento acumulada nos sítios em função da conectividade. A linha contínua apresenta a lei de potência com expoente $\beta_{fit} = 1,3$ e erro estimado de $9,6 \cdot 10^{-3}$ . Esse apresenta uma boa concordância coma equação 4.5 usando os expoentes obtidos nas figuras 4.2 e 4.3, $\beta = 1/0,74 \simeq 1,4$ . . . . .	p. 64
4.5	Tipos de agrupamento em torno de um sítio $i$ , de cima para baixo as duas possibilidades de cada um: ciclo, entrada, saída e <i>middleman</i> . Ilustração construída com base no trabalho do Fagiolo (2007). . . . .	p. 65
4.6	Exemplo de ciclo. Os pacotes em vermelhos somente poderão ser instalados sem levar em consideração a relação de dependências entre o ciclo. . . . .	p. 66

- 4.7 Dendrograma para o histórico de divisões do método espectral, o eixo ordenado corresponde à distância logarítmica até o estado de maior modularidade neste método,  $Q_{spec} \simeq 0,58$  com 70 módulos. Os rótulos de cada comunidade correspondem ao pacote de maior conectividade de saída, ou ao tamanho da comunidade  $s$ , caso  $s \leq 10$ . . . . . p. 78
- 4.8 Subgrafo obtido com o método espectral, contém pacotes como o compilador de linguagem C (o *gcc*) e *software* de análise estatística (o *R*). As cores indicam os times de desenvolvedores contidos neste subgrafo. Na legenda constam os times ou colaborador que responsáveis por mais pacotes dentro desse módulo. . . . . p. 79
- 4.9 Matriz de frequências entre pacotes tanto nos módulos obtidos pelo método espectral (abscissas) quanto nos agrupamentos definidos pelos times de desenvolvedores (ordenadas), normalizadas pelo tamanho dos agrupamentos por colaboradores. Blocos mais próximos do preto indicam agrupamentos de pacotes dos colaboradores Debian praticamente inseridos numa única comunidade pelo método espectral. . . . . p. 80
- 4.10 Componente gigante para a rede de dependência entre pacotes Debian 6.0,  $n = 26149$  pacotes e  $m = 121833$  ligações direcionadas. A componente em (a) é colorida pelos agrupamentos de pacotes por colaboradores,  $Q_{dev} \simeq 0,23$  e a mesma componente em (b) é colorida conforme o método espectral,  $Q_{spec} \simeq 0,58$ . A região em destaque é a comunidade dos pacotes mantidos pelo time *Perl* e seu módulo correspondente obtido com o método espectral. . . . . p. 81
- 4.11 Distribuição de tamanhos de comunidades, os valores dos expoentes são:  $\alpha \simeq 0,289 \pm 0,013$ ;  $\beta \simeq 0,5821 \pm 0,0028$ ; e  $\gamma \simeq 1,433 \pm 0,014$ . Temos o coeficiente de decaimento exponencial dado por  $s_0 = \langle s \rangle = 461$ . . . . . p. 82

4.12 Subgrafo correspondente aos dois maiores aglomerados de time de colaboradores. Sítios de cores azuis correspondem ao segundo maior aglomerado de 213 grupos Debian com 3449 pacotes, em tons mais escuros os pacotes mantidos pelo *Debian Perl Group* (48,4% dos pacotes no grafo). Sítios de cor vermelha e amarela constituem o maior aglomerado de 341 times responsáveis por 4606 pacotes, o *Debian Python Modules Team* corresponde aos sítios amarelos (5,6% dos pacotes no grafo). . . . . p. 83

# 1 *Introdução*

Os resultados da física, em sua grande parte, descrevem sistemas dinâmicos que representam o movimento dos corpos na natureza, seja na descrição de partícula única, livre ou sob a ação de campos externos (GRIFFITHS, 2004), até sistemas de muitas partículas, usando métodos da mecânica estatística (PATHRIA, 1996). Porém uma nova área apareceu, na qual o estudo do movimento dos corpos é apenas uma base para métodos que podem ser empregue na análise de sistemas de muitos agentes. A interação local não é descrita por leis de interação à distância ou espalhamento, mas sim por meio de regras pré-estabelecidas no modelo em questão ou por meio do conhecimento do sistema. Sistemas que se comportam desta maneira são os chamados de sistemas complexos, visto que suas regras de interação nem sempre podem ser facilmente descritas, desde a interação elemento-a-elemento bem como a distribuição dessas.

Um exemplo típico dessas interações está no modelo do votante (LIGGETT, 1999), no qual os elementos de uma rede ordenada podem copiar o estado de um de seus vizinhos não tendo, portanto, uma interação derivada de um potencial, por exemplo. É possível estender este conceito a organizações humanas definindo algum tipo de interação social para identificar a estrutura: rede de colaboração científica (ligações entre pesquisadores com artigos de autoria compartilhada) (NEWMAN, 2001), rede de atores (atores que trabalharam num mesmo projeto cinematográfico) (BARABÁSI; ALBERT, 1999), rede de contatos sexuais (LILJEROS *et al.*, 2001), etc. Exemplos de estudos destas organizações são modelos para observar características como a formação de consenso ou propagação de epidemia dentro de estruturas sociais (ROBINSON; COHEN; COLIJN, 2012).

Outra característica que não está relacionada com uma dinâmica na rede, mas apenas com

sua estrutura de ligações é a existência de agrupamentos altamente conectados entre si e com uma densidade menor de ligações entre sítios de grupos diferentes. Estas são as chamadas estruturas de comunidades, que constituem ainda um problema em aberto na matemática (BRANDES *et al.*, 2008), mas que a partir das características do sistema utilizado podemos encontrar soluções aceitáveis.

O conceito de rede – conjunto de vértices conectados entre si por um conjunto de ligações – é usado para representar uma variedade de sistemas reais, quando estes afastam-se da descrição usual de uma estrutura regular, como sólidos cristalinos (ASHCROFT; MERMIN, 1976), ou de uma estrutura completamente aleatória, grafos aleatórios clássicos (DOROGVTSEV, 2010), denominamos redes complexas a estrutura em grafo que descrevem tais sistemas. Estes podem entrar numa subcategoria de grafos aleatórios que apresentam distribuições com características livre-de-escala ou de mundo-pequeno. Exemplos de sistemas reais descritos com o conceito de redes complexas são: redes sociais, *world wide web*, rede de subestações elétricas, cadeia alimentar, etc (DOROGOVTSEV; MENDES, 2003).

Este trabalho aborda duas vertentes: uma é o estudo de modelo com interações estabelecidas por meio de uma regra local, com base em otimização de fluxos ao longo de uma rede direcionada, mostrando-se um sistema que está de acordo com modelos de reação de partículas únicas, com criação e coalescência de partículas (ALIMOHAMMADI; KHORRAMI; AGHAMOHAMMADI, 2001; KRAPIVSKY; REDNER; BEN-NAIM, 2010), substituindo a existência ou não destas por uma corrente nula ou não-nula passando pelos vértices da rede. O outro sistema estudado foi composta a partir das relações entre os elementos constituintes de um sistema real, criado por uma organização humana, o sistema operacional Debian GNU/Linux, estudando a formação dessa estrutura com base na relação dos indivíduos que desenvolveram tal sistema com os elementos constituintes deste.

Iniciaremos por um capítulo para discutir o formalismo matemático utilizado em nossa solução para o modelo de otimização de tráfego, a ser apresentado no segundo capítulo. O terceiro capítulo aborda um problema de otimização de fluxos numa rede direcionada, e a pro-

posta desta tese para obter as distribuições estacionárias desse modelo, fornecendo uma nova interpretação para as taxas de criação de partículas nos modelos de reação de difusão. O capítulo 4 trata da rede de dependências entre pacotes Debian, no qual discutiremos as propriedades gerais dessa rede e sua estrutura de comunidades. Parte desse capítulo foi compilado no trabalho ainda não publicado (SOUSA; PENNA, 2013).

Durante meu doutorado estudei também um problema de interferência entre elétrons de um substrato polarizado e uma impureza magnética com um acoplamento antiferromagnético com esse substrato, conhecido como problema Kondo. O que culminou nos trabalhos (SERIDÔNIO *et al.*, 2012) e (SERIDONIO *et al.*, 2013), realizados em colaboração com os professores Antônio Carlos Ferreira Seridônio, na época em pós-doutoramento no nosso grupo, e Marcos Sérgio Figueira, do IFUFF, não será abordado nesta tese, por não estar relacionado com as técnicas e problemas apresentados neste trabalho.

## 2 *Recursos preliminares*

Ao longo deste trabalho faremos uso de variáveis aleatórias, definindo suas distribuições iniciais e obtendo as finais ao analisar as características do modelo no segundo capítulo, ou obtendo a distribuição de um sistema real, como no terceiro capítulo. Para isso alguns recursos da estatística como a distribuição resultante de uma função de variáveis aleatórias, transformada de Laplace e distribuições cumulativas serão empregados. Procuraremos, neste capítulo, nos restringir aos pontos que serão necessários para o entendimento do trabalho. Admitiremos que conceitos básicos de Estatística, como variáveis aleatórias e funções distribuição sejam de conhecimento prévio do leitor.

### 2.1 Soma de variáveis aleatórias

Podemos determinar a distribuição de uma dada variável aleatória  $Z$  definida pela soma de duas outras variáveis aleatórias  $X$  e  $Y$ , isto é,  $Z = X + Y$ . As distribuições das variáveis  $X$  e  $Y$  são consideradas conhecidas. A distribuição cumulativa de  $Z$  pode ser escrita como

$$F_Z(z) = \iint_{x+y \leq z} f_{XY}(x,y) dx dy$$

onde  $f_{XY}(x,y)$  é a distribuição conjunta das variáveis  $X$  e  $Y$ . Para obtermos a distribuição de  $Z$  desejada, basta derivarmos a distribuição cumulativa com ajuda da regra de Leibniz (KAPLAN, 1991):

$$\frac{d}{dt} \int_{a(t)}^{b(t)} \phi(x,t) dx = \int_{a(t)}^{b(t)} \frac{\partial \phi(x,t)}{\partial t} dx + \phi(b(t),t) \frac{db(t)}{dt} - \phi(a(t),t) \frac{da(t)}{dt}.$$

Dessa maneira

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \frac{d}{dz} \int_{-\infty}^{+\infty} \int_{-\infty}^{z-y} f_{XY}(x,y) dx dy = \int_{-\infty}^{+\infty} f_{XY}(z-y,y) dy.$$

Se  $X$  e  $Y$  são variáveis estatisticamente independentes, a distribuição resultante é a convolução das duas primeiras (ROUSSAS, 1997, página 220)

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y) dy. \quad (2.1)$$

Neste caso  $f_{XY}(x,y) = f_X(x)f_Y(y)$ , pois a probabilidade de ocorrer a variável  $Y = y$  dado ocorrido  $X = x$  não é afetada por essa, isto é,  $f_{Y|X}(y|x) = f_Y(y)$ .

Essa é uma consequência direta do teorema de Bayes, equação 2.2, para as probabilidades condicionais, que trata em verificar a probabilidade condicional de um evento  $A$  dada a ocorrência de um outro evento  $B$  (ROUSSAS, 1997, página 24). Essa é a razão entre a probabilidade de ocorrência dos dois eventos conjuntamente  $P(A,B)$  e a probabilidade de ocorrência de  $B$ ,

$$P(A|B) = \frac{P(A,B)}{P(B)}, \text{ (lei de Bayes).} \quad (2.2)$$

Logo se os eventos são independentes a probabilidade conjunta é a multiplicação das probabilidades individuais, justificando assim o resultado acima da convolução.

Algumas distribuições serão de interesse nos capítulos que seguem, portanto calcularemos a distribuição entre variáveis igualmente distribuídas e estatisticamente independentes nas seções seguintes.

## 2.2 Distribuição uniforme

Sejam  $X$  e  $Y$  distribuídas uniformemente no intervalo  $[\alpha, \beta]$ :

$$f_X(x) = f_Y(x) = \frac{1}{\beta - \alpha} \Theta(\beta - x) \Theta(x - \alpha), \quad (2.3)$$

sendo  $\Theta$  uma distribuição degrau. Temos a distribuição de  $Z = X - Y$ , segundo a equação 2.1, dada por:

$$f_Z(z) = \frac{1}{(\beta - \alpha)^2} \int_{-\infty}^{+\infty} \Theta(\beta - z + y) \Theta(z - y - \alpha) \Theta(\beta + y) \Theta(-y - \alpha) dy$$

Podemos definir a região em que a distribuição não é nula, enumerando as inequações fornecidas pelas funções degrau do integrando  $y \leq -\alpha$ ,  $y \geq -\beta$ ,  $z - y - \alpha \geq 0$  e  $\beta - z + y \geq 0$ . Para  $\alpha > 0$ , a integração reduz-se para

$$f_Z(z) = \frac{1}{(\beta - \alpha)^2} \left[ \Theta(z) \int_{z-\beta}^{-\alpha} dy + \Theta(-z) \int_{-\beta}^{z-\alpha} dy \right] \quad (2.4)$$

De forma simplificada:

$$f_Z(z) = \frac{1}{(\beta - \alpha)^2} \begin{cases} \beta - \alpha - z, & z \geq 0 \\ z - (\beta - \alpha), & z < 0 \end{cases} \quad (2.5)$$

A figura 2.1 ilustra essa distribuição resultante.

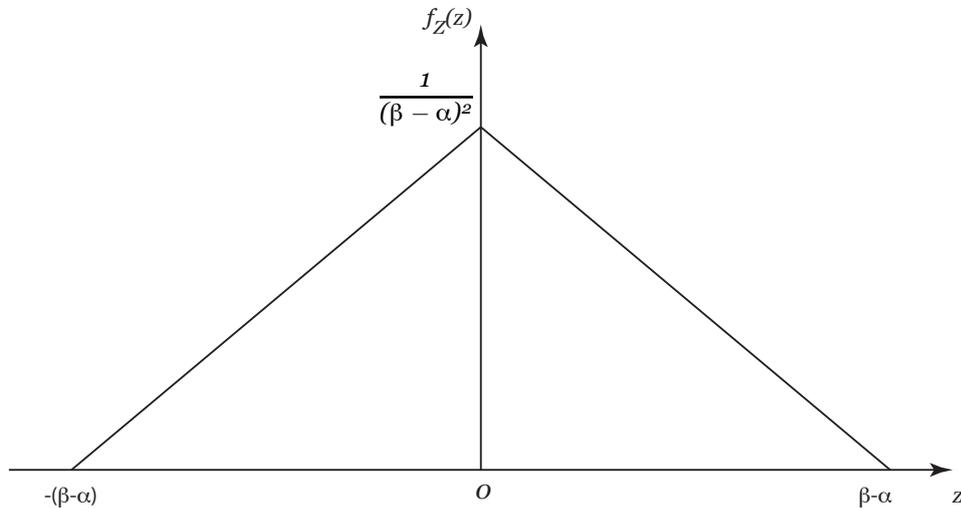


Figura 2.1: Distribuição resultante da soma de duas variáveis aleatórias.

## 2.3 Distribuição exponencial

Sejam  $X$  e  $Y$  distribuídas conforme o decaimento exponencial:

$$f_X(x) = f_Y(x) = \gamma e^{-\gamma x} \Theta(x). \quad (2.6)$$

Com o mesmo protocolo do item anterior a distribuição da nova variável  $Z=X-Y$  será:

$$\begin{aligned} f_Z(z) &= \gamma^2 \int_{-\infty}^{\infty} e^{-\gamma(z-x)} \Theta(z-x) e^{+\gamma x} \Theta(-x) dx \\ &= \gamma^2 e^{-\gamma z} \int_{-\infty}^0 e^{2\gamma x} \Theta(z-x) dx \end{aligned} \quad (2.7)$$

para o caso  $y \leq 0$  temos,

$$f_Z(z) = \gamma^2 e^{-\gamma z} \int_{-\infty}^y e^{2\gamma x} dx = \frac{\gamma}{2} e^{\gamma z},$$

e caso contrário

$$f_Z(z) = \gamma^2 e^{-\gamma z} \int_{-\infty}^0 e^{2\gamma x} dx = \frac{\gamma}{2} e^{-\gamma z}.$$

Resumindo,

$$f_Z(z) = \frac{\gamma}{2} \begin{cases} e^{\gamma z}, & z \leq 0 \\ e^{-\gamma z}, & z > 0 \end{cases}. \quad (2.8)$$

Tais distribuições serão úteis em nossas análises pois possuem um parâmetro característico que define sua largura  $\beta - \alpha$ , no caso uniforme, e uma escala característica  $1/\gamma$  no caso exponencial.

## 2.4 Transformada de Laplace

Para problemas envolvendo distribuições  $f(t)$ , definidas num domínio semi-infinito  $[0, \infty)$ , é comum o uso das transformadas de Laplace  $\mathcal{L}\{f(t)\}$ , no tratamento desses. A transformada de Laplace de uma distribuição  $f(t)$  (SCHIFF, 1999, página 2) é definida como :

$$\mathcal{L}\{f(t)\} = F(z) = \int_0^{\infty} f(t) e^{-zt} dt. \quad (2.9)$$

Sendo  $f(t)$  uma distribuição, podemos definir uma função geradora, ou geratriz, dos momentos dessa distribuição (ROSS, 1995, página 15):

$$\psi(z) = E [e^{zT}] = \int e^{zt} f(t) dt, \quad (2.10)$$

onde  $E[\ ]$  representa o valor esperado da função entre colchetes. A derivada de  $n$ -ésima ordem da função geradora corresponde ao  $n$ -ésimo momento desta distribuição, por exemplo, o momento de ordem 1 corresponde ao valor médio da distribuição. Em distribuições com o intervalo definido para valores não negativos, a função geradora dos momentos corresponde à transformada de Laplace (ROSS, 1995), como segue:

$$\begin{aligned} \left. \frac{dF(z)}{dz} \right|_{z=0} &= - \int_0^{\infty} t f(t) e^{-(z=0)t} dt = - \langle t \rangle \\ \left. \frac{d^2 F(z)}{dz^2} \right|_{z=0} &= \int_0^{\infty} t^2 f(t) e^{-(z=0)t} dt = \langle t^2 \rangle \\ &\vdots \\ \left. \frac{d^n F(z)}{dz^n} \right|_{z=0} &= \int_0^{\infty} t^n f(t) e^{-(z=0)t} dt = (-1)^n \langle t^n \rangle. \end{aligned}$$

Neste trabalho, trataremos um modelo de fluxos direcionados e não consideraremos seus valores negativos, que correspondem a um fluxo contrário ao sentido das ligações na rede. Como veremos adiante, o cálculo da convolução de distribuições irá surgir ao longo de nosso formalismo, e para isto faremos uso de algumas propriedades das transformadas de Laplace que facilitam esse tipo de cálculo.

Uma outra propriedade muito útil das transformadas de Laplace é que podemos tratar a convolução das distribuições como apenas um produto das respectivas transformadas. A convolução 2.1 pode ser também definida para variáveis aleatórias positivas, como:

$$(f * g)(y) = \int_0^y f(y-x)g(x)dx. \quad (2.11)$$

A transformada de Laplace será:

$$\begin{aligned}\mathcal{L}\{(f * g)(y)\}(z) &= \int_0^{\infty} \left\{ \int_0^y f(y-x)g(x)dx \right\} e^{-zy} dy \\ &= \int_0^{\infty} \left\{ \int_{-v}^{\infty} f(u)g(v)du \right\} e^{-z(u+v)} dv,\end{aligned}$$

como  $f$  e  $g$  são nulas para valores menores que 0 obtemos o resultado:

$$\mathcal{L}\{(f * g)(y)\}(z) = \int_0^{\infty} f(u)e^{-zu} du \int_0^{\infty} g(v)e^{-zv} dv = \mathcal{L}\{f(t)\}\mathcal{L}\{g(t)\}. \quad (2.12)$$

Essas duas propriedades serão novamente usadas ao longo da definição do formalismo apresentado neste trabalho, juntamente com a inversão dessa transformada (ROUSSAS, 1997) para obtenção da distribuição desejada:

$$f(t) = \mathcal{L}^{-1}\{F(z)\} = \frac{1}{2\pi i} \lim_{y \rightarrow \infty} \int_{x-iy}^{x+iy} F(z)e^{tz} dz, \quad (2.13)$$

na qual deve ser integrada ao longo de uma linha vertical  $x = \Re(z) > \alpha$  tal que a integral abaixo seja convergente

$$\int_0^{\infty} |f(t)e^{-xt}| dt = \int_0^{\infty} |f(t)| e^{-xt} dt < \infty, x > \alpha.$$

esta é a condição de existência da inversa da transformada de Laplace.

O conteúdo deste capítulo apresenta, resumidamente, as bases matemáticas necessárias para podemos desenvolver o modelo de tráfego proposto no terceiro capítulo, além de introduzir as distribuições empíricas encontradas no quarto capítulo.

### ***3 Aproximação de campo médio para otimização de fluxos numa rede regular direcionada***

Neste capítulo estudaremos um problema de fluxos por uma rede bidimensional direcionada que possui regras de otimização local para a corrente que irá fluir por cada canal (ligações da rede). Verifica-se, com este formalismo, que o modelo é independente da quantidade de canais que saem de cada sítio e apresenta um comportamento de ponto crítico trivial na condição de correntes não-nulas.

#### **3.1 Introdução**

Sistemas de transporte eficiente são essenciais para o funcionamento de um sociedade moderna e industrializada. Como escrito por Greenberg, em 1959 (HELBING, 2001, referência contida):

“The volume of vehicular traffic in the past several years has rapidly outstripped the capacities of the nation’s highways. It has become increasingly necessary to understand the dynamics of traffic flow and obtain a mathematical description of the process.”

Podemos estender tal pensamento ao volume de informação que é compartilhado em todo o mundo por redes como a *internet* ou redes de telecomunicações. Assim, os sistemas de transporte eficiente são representados tanto pelo tráfego de veículos no mapa rodoviário como pelo

tráfego de informação, i.e., rede entre servidores *web*, rede de colaboração científica, etc.

O problema pode ser tratado de forma a considerar cada agente do tráfego com suas regras locais, tal como um *automatum* celular representando os veículos (MAERIVOET; MOOR, 2005). Outra abordagem é o estudo da influência da topologia no fluxo de informação com modelos minimalistas em redes complexas (MARTINO *et al.*, 2009a, 2009b). Nossa abordagem será um modelo de otimização local cujos agentes são as bifurcações dos canais de fluxo que formam uma estrutura regular.

## 3.2 Otimização local de fluxos

Um conceito importante nesse tipo de problema é o conceito de equilíbrio do tráfego, que consiste no princípio pelo qual o agente do tráfego (motorista por exemplo) tende a escolher sempre a rota que minimize o tempo de seu itinerário (NAGURNEY, 2000). Peregrino define em 2011 um modelo de otimização local do tempo a partir de um conjunto de rotas  $i$ , que bifurcam como na figura 3.1, descritas por custos diretamente relacionado ao tempo gasto em cada uma delas.

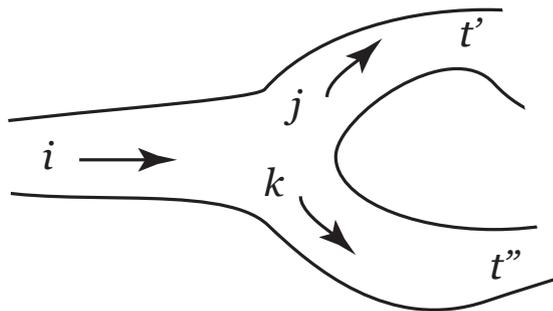


Figura 3.1: Bifurcação de um canal com fluxo  $i$ , cada caminho possui um custo ligado ao tempo para gasto para concluir o caminho escolhido  $j$  ou  $k$ .

Portanto podemos supor que o tempo gasto num dado itinerário será o tempo característico dessa rota  $\bar{t}_i$  mais uma função linear do fluxo  $j_i$  caso este não seja tão intenso, mais contribuições

não-lineares com o fluxo numa situação de uso intenso da rota.

$$t_i = \bar{t}_i + a j_i + (b j_i^2 + c j_i^3 + \dots) \Theta(j_i - j_c). \quad (3.1)$$

Tal comportamento não-linear pode ser verificado no chamado diagrama fundamental do tráfego, introduzido nos anos de 1950 por jornais de pesquisa em engenharia (HELBING, 2001), o qual relaciona o uso de uma autoestrada com o fluxo na mesma. Na condição de uso intenso a relação puramente linear deixa de ser válida, conforme evidencia a figura 3.2.

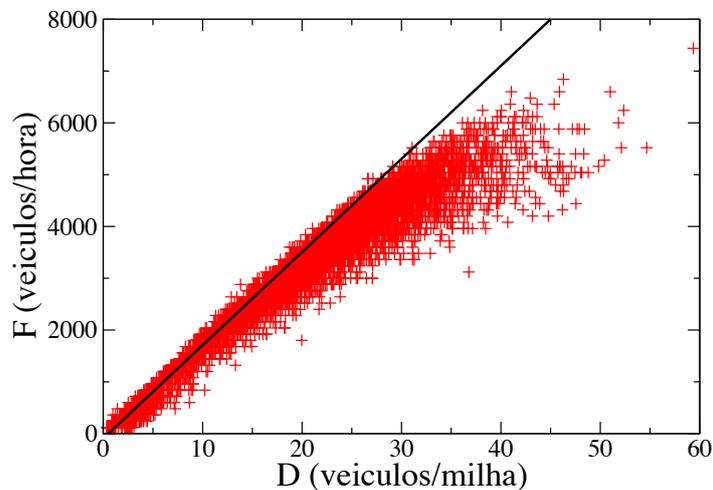


Figura 3.2: Diagrama fundamental do tráfego, fluxo ( $F$ ) de veículos *versus* densidade ( $D$ ). Dados obtidos no endereço eletrônico do departamento de trânsito do estado de Minnesota, vide referências bibliográficas. A reta preta indica a relação linear  $F \propto D$  que se adequa bem aos dados para valores baixos da densidade e fluxo. Figura retirada de Peregrino (2011, página 28).

A partir da definição da equação 3.1 podemos definir a função custo  $C(\mathbf{j})$  a partir de um conjunto de estradas que ligam a mesma origem  $A$  e destino  $B$ , que produzem diferentes itinerários:

$$C(\mathbf{j}) = \sum_{i=1}^n (a_i j_i + b_i j_i^2), \quad (3.2)$$

onde  $\mathbf{j} = (j_1, j_2, \dots, j_n)^T$  é o vetor de fluxos dos  $n$  caminhos que ligam os pontos  $A$  e  $B$ . Por simplicidade truncamos a série da equação 3.1 até o termo de ordem dois no fluxo, ou primeiro termo não-linear da expansão. Os coeficientes  $a_i$  e  $b_i$  são as contribuições lineares e não-lineares

para o custo de cada canal  $i$ . Para o formalismo descrito posteriormente usaremos  $b_i = 1/2$  fazendo com que os coeficientes do custo  $a_i$  tenham a dimensão de fluxo. Escolhemos para a função custo uma grandeza que tenha a dimensão do quadrado do fluxo:

$$C(\mathbf{j}) = \sum_{i=1}^n \left( a_i j_i + \frac{1}{2} j_i^2 \right), \quad (3.3)$$

Dada a condição de equilíbrio do tráfego – a qual cada agente do tráfego escolhe a rota que minimize sua a função custo  $C$  – podemos minimizar a função 3.2 na condição que o fluxo de entrada é superior ao fluxo crítico, impondo o vínculo da conservação do fluxo de entrada  $\sum_{i=1}^n j_i - J = 0$ . Para tanto, empregaremos a técnica de multiplicadores de Lagrange (REIF, 1965):

$$\begin{aligned} L(\lambda) &= \sum_i \left( a_i j_i + \frac{j_i^2}{2} \right) + \lambda \left( \sum_i j_i - J \right) \\ \frac{\partial L}{\partial j_i} &= a_i + j_i + \lambda = 0 \\ \Rightarrow n\lambda &= -\sum_i a_i - J \end{aligned}$$

onde  $\lambda$  é o multiplicador atribuído ao vínculo. A corrente ótima para um conjunto de  $n$  canais é portanto:

$$j_i = -a_i + \frac{\sum_k a_k + J}{n}.$$

No caso mais simples, de dois canais, a equação acima torna-se

$$\begin{aligned} j_1 &= \frac{J - (a_1 - a_2)}{2} \\ j_2 &= \frac{J - (a_2 - a_1)}{2} \end{aligned} \quad (3.4)$$

Como o comportamento não-linear ocorre apenas para valores altos de fluxo, podemos usar essa condição para definir o fluxo crítico a partir do qual os dois canais passam a ser usados com fluxos dados pela equação 3.4. Tal condição é escolhida para que tenhamos apenas valores

positivos de fluxo. Definimos portanto o fluxo crítico entre dois canais como sendo

$$j_c = |a_1 - a_2|$$

$$\Rightarrow j_{\pm}(J) = \begin{cases} \frac{J \pm j_c}{2}, & \text{se } J \geq j_c \\ J, 0, & \text{se } J < j_c. \end{cases} \quad (3.5)$$

Essa solução indica o uso do canal de menor custo quando o fluxo de entrada é menor ou igual ao fluxo crítico. Na condição de fluxo de entrada maior temos os fluxos dividindo-se entre os canais, sempre com o canal de menor custo recebendo o maior fluxo, como esquematicamente mostrado na figura 3.3.

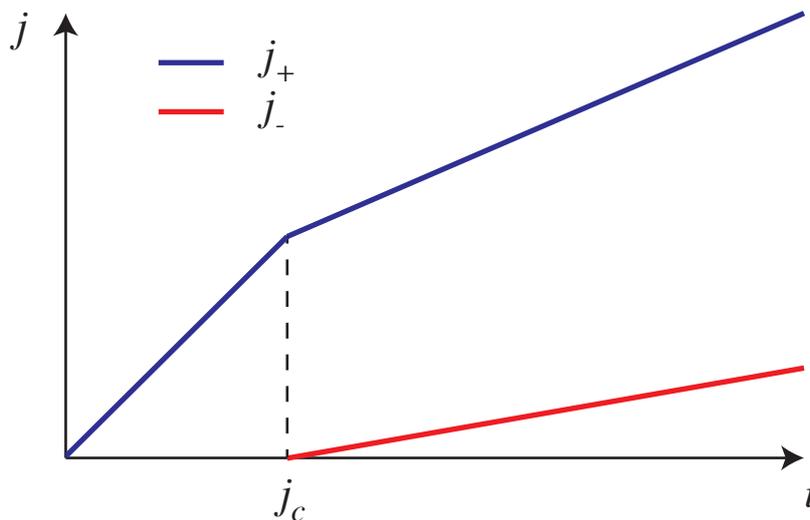


Figura 3.3: Função para os fluxos ótimos  $j$  em dois canais, dado um fluxo de entrada  $i$ , antes e depois do valor crítico de fluxo definido entre os canais. A curva azul é a solução ótima para o canal de menor custo e a vermelha para segundo de maior custo, a soma das duas curvas deve ser sempre igual a  $i$ .

O modelo estudado aqui levará em consideração tal solução para a otimização local do fluxo na rede de tráfego considerada, que será descrita a seguir.

### 3.3 Modelo Finito

Consideramos uma cadeia linear de tamanho  $L$  em que cada sítio contém duas ligações de entrada de fluxos e duas de saída a qual é repetida ao longo de sua direção perpendicular, ligando cada ponto da cadeia anterior a dois outros da seguinte. Forma-se assim uma rede quadrada direcionada na qual é injetado um fluxo total  $J$  na primeira cadeia e este segue para as cadeias abaixo segundo a regra de otimização da equação 3.5.

Para cada aresta é atribuído um valor de custo  $c_i$  que a cada par ligado a um sítio com fluxo de entrada forma um fluxo crítico  $j_c$ . A dinâmica se dá injetando correntes nos sítios da cadeia superior e estas seguem pelas arestas obedecendo a regra de otimização da função custo definida em cada sítio pelo valor dos coeficientes  $a_i$  de cada caminho que o fluxos pode seguir a partir de cada sítio, após esse passo, os fluxos são aglutinados nos sítios da cadeia inferior, retomando o processo até que um estado estacionário seja observado. A figura 3.4 mostra uma representação das correntes ao longo desse sistema.

Essa dinâmica impõe no modelo alguns eventos comuns em modelos de partículas interagentes, ao supormos partículas como sendo um sítio da rede que está presente caso tenha um fluxo não-nulo passando por ele. Os processos são:

- **Difusão:** Um sítio possui uma corrente baixa comparada ao valor da corrente crítica gerada pelos canais que saem dele. Os sítios imediatamente abaixo formam um par de sítio vazio e outro cheio, a corrente deste por sua vez é distribuída para um único sítio novamente<sup>1</sup>;
- **Coalescência:** Duas correntes coalescem num sítio e este, por sua vez, possui corrente crítica grande o suficiente, portanto fluxo insuficiente, para que distribua a corrente em dois canais;
- **Ramificação:** O oposto da coalescência. Um sítio possui corrente menor que seu valor

---

<sup>1</sup>Esse termo está sendo empregado ao observarmos o que ocorre no nível de ocupação ou desocupação de um sítio da rede, ao nível do fluxo propagando na rede, teríamos um processo de transmissão, sendo o sítio completamente transparente a esse processo

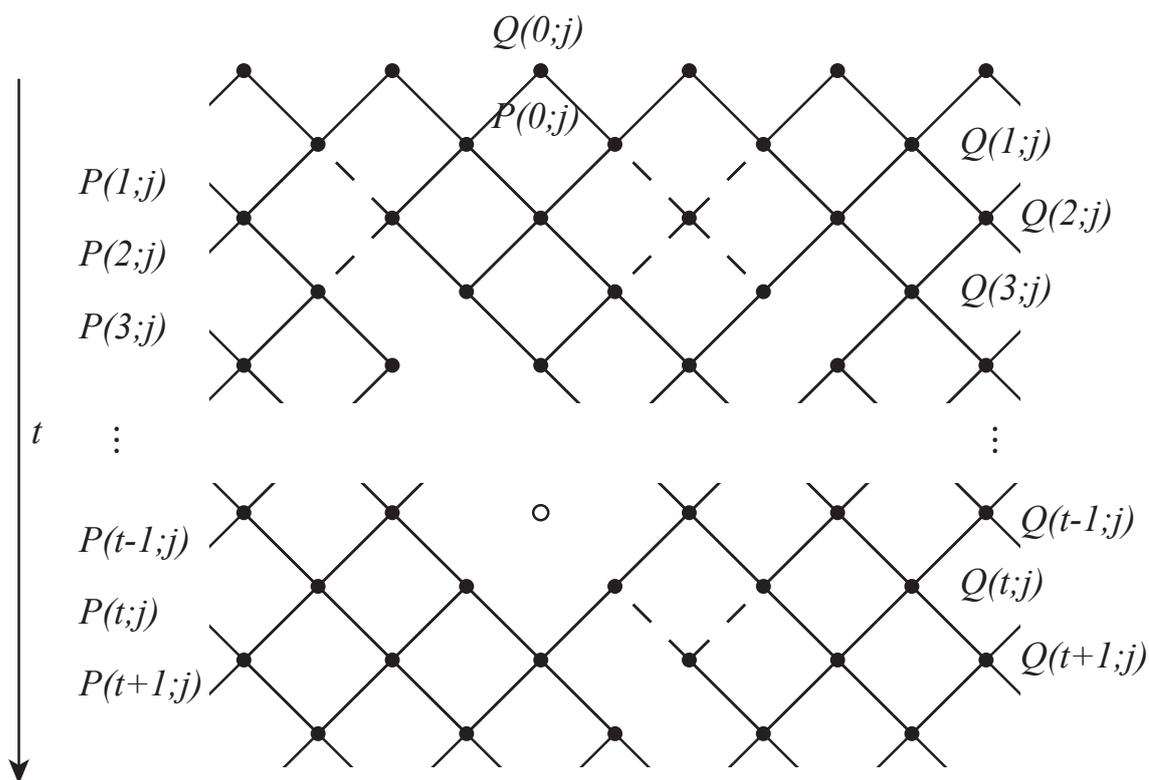


Figura 3.4: Representação geométrica do modelo. É injetada uma corrente  $j$  em cada sítio da linha 0 distribuída ao longo da cadeia conforme uma distribuição  $Q(0, j)$ . Esta gera a distribuição  $P(0, j)$  nas arestas subsequentes, tais correntes são somadas nos sítios da linha 1 gerando a distribuição de correntes  $Q(1, j)$ . Os passos acima descritos são repetidos ao longo da direção  $t$ .

crítico e distribui o fluxo para outro com corrente crítica mais baixa, dividindo o fluxo para dois sítios subsequentes.

A figura 3.5 ilustra tais processos. Podemos enumerar um terceiro processo que ocorre quando dois canais coalescem num sítio e este distribui o fluxo por seus dois canais de saída, uma coalescência seguida de ramificação. Este processo conjunto não altera o número de ocupação de partículas.

Como descrito na figura 3.4 podemos definir um passo de Monte-Carlo em nossa simulação como o processo de aglutinação de corrente nos sítios da cadeia seguido da divisão da mesma por esses sítios. Ao longo do tempo, a corrente total é conservada, portanto podemos definir

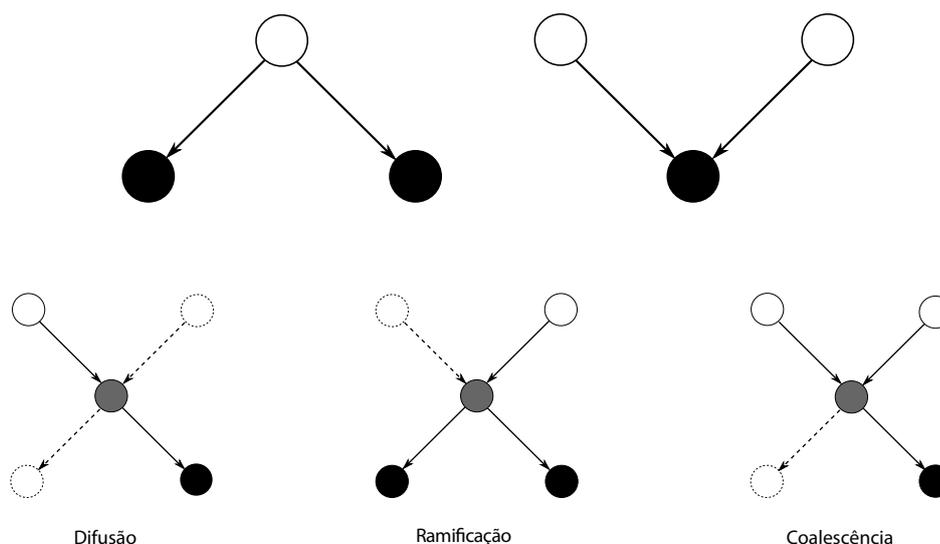


Figura 3.5: Eventos de divisão e entrada de fluxo num sítio e processos consequentes. A escala de cinza indica o tempo, do inicial (branco) até o final (preto). Linhas pontilhadas indicam a possibilidade de processos adicionais.

seu valor médio como:

$$\langle j \rangle = \frac{J}{2L} = \frac{1}{2L} \sum_i^{2L} j_i. \quad (3.6)$$

Dado que o número de canais é duas vezes maior que o de sítios, a corrente média nos canais será duas vezes menor que nos sítios.

O valor de custo nos canais é distribuído uniformemente no intervalo  $(0, 1)$ , o que nos fornece uma distribuição triangular de correntes críticas no intervalo  $(-1, 1)$ , com o valor mais provável centrado em zero. As soluções de corrente crítica negativa vão para o canal de maior custo e as soluções de correntes críticas positivas para o de menor custo, tendo em vista que a corrente tende a fluir pelos canais de menor custo. Uma condição de contorno periódica foi imposta na cadeia linear para que o número de canais de entrada e saída pudesse ser mantido constante na simulação, e para que esta ficasse mais próxima de um comportamento de sistema infinito.

A grandeza de interesse a cada passo de Monte-Carlo é a fração de canais usados  $c(t) = 1 - B(t)$ , sendo  $B$  a fração de canais vazios:

$$1 - B(t) = \frac{1}{2L} \sum_{i=1}^{2L} \Theta(j_i(t) - \delta), \quad (3.7)$$

onde  $t \in \mathbb{N}$  e  $\delta$  um número muito pequeno e maior que zero (podemos usar o menor número real que pode ser representado pela máquina, no caso das simulações).

As flutuações temporais dessa grandeza também foram medidas a fim de comparação com o formalismo desenvolvido nesta tese. A quantidade de interesse, neste caso, é a variância do valor médio de  $B(t)$

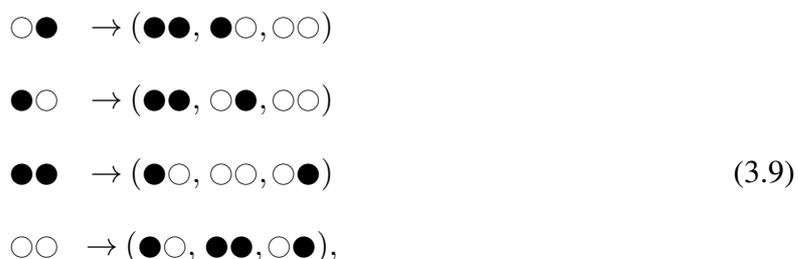
$$\sigma^2 = \langle B^2(t) \rangle - \langle B(t) \rangle^2 \quad (3.8)$$

Um cuidado adicional deve ser tomado ao verificarmos nossas simulações: para um sistema de tamanho finito o número mínimo de canais utilizados é igual a unidade, logo a fração de canais utilizados  $c(t) \in [\frac{1}{2L}, 1] \forall t$ .

### 3.3.1 Método de intervalos vazios para processos de reação-difusão

Os processos existentes para o modelo de fluxos atual em tamanho finito são muito semelhantes aos processos de reação e difusão em uma dimensão, que envolvem difusão de partículas, aglutinação ou coalescência e criação de partículas. A diferença básica é que, no modelo em questão, tais processos são intermediados por uma quantidade  $j_i$  de fluxo em cada sítio, enquanto nos processos de reação-difusão cada processo possui uma taxa de ocorrência que servirá de entrada para sua equação mestra correspondente. Esta equação permite solução exata, conforme descrito por Alimohammadi, Khorrami e Aghamohammadi (2001)

As transições mais gerais que podem existir num modelo de partículas de espécie única são:



assim, uma partícula ( $\bullet$ ) pode dar origem a uma nova, difundir nos espaços vazios ( $\circ$ ) ou morrer, como representado na primeira e segunda linhas. Ainda, duas partículas podem coalescer em uma ou morrerem, como indicado na terceira linha. Finalmente, sítios vazios podem

originar novas partículas, como nos processos representados na última linha.

Consideraremos, neste trabalho, apenas os processo de difusão ( $D$ ), coalescência ( $C$ ) e ramificação ( $R$ ). Desta maneira, podemos definir a equação mestra para  $n$  sítios vazios:

$$E_n = P(\overbrace{\circ \dots \circ}^n)$$

a partir das probabilidades de ocupação de uma cadeia de  $n + 1$  sítios:

$$\begin{aligned} \frac{dE_n}{dt} = & D_e P(\circ \overbrace{\bullet \circ \dots \circ}^n) + D_d P(\overbrace{\circ \dots \circ}^n \bullet \circ) + \\ & -D_d P(\bullet \overbrace{\circ \circ \dots \circ}^n) - D_e P(\overbrace{\circ \dots \circ \circ}^n \bullet) + \\ & -R_d P(\bullet \overbrace{\circ \circ \dots \circ}^n) - R_e P(\overbrace{\circ \dots \circ \circ}^n \bullet) + \\ & +C_e P(\bullet \bullet \overbrace{\circ \dots \circ}^n) + C_d P(\overbrace{\circ \dots \circ}^n \bullet \bullet). \end{aligned} \quad (3.10)$$

As taxas de interação possuem sub-índices que se referem ao sentido  $e$  e  $d$ , esquerda e direita respectivamente, nos sítios que participam da reação. As probabilidades de ocupação podem ser escritas em termos das probabilidades de sítios vazios, ou desocupados:

$$P(\bullet \overbrace{\circ \dots \circ}^n) = P(\overbrace{\circ \dots \circ}^n \bullet) = E_{n-1} - E_n. \quad (3.11)$$

Vamos impor algumas condições que permitam reescrever a equação mestra de uma forma simplificada. As condições impostas são a homogeneidade quanto o sentido da reação ( $e = d$ ); igualdade das taxas de difusão e coalescência, uma vez que o modelo de interesse dois sítios ao se encontrarem coalescem, e usar a relação entre as ocupações de um único sítio mais a de dupla ocupação:

$$P(\overbrace{\circ \circ \dots \bullet \circ}^n) + P(\overbrace{\circ \circ \dots \bullet \bullet}^n) = P(\overbrace{\circ \circ \dots \bullet}^n).$$

Com estas condições, podemos reescrever a equação mestra como

$$\frac{dE_n}{dt} = E_{n-1} - 2E_n + E_{n+1} - \gamma(E_n - E_{n+1}), \quad (3.12)$$

onde  $\gamma = R/D$  é a razão entre ramificação pela taxa de difusão/coalescência. Adicionalmente, o tempo foi reescalado para que a equação fique nessa forma final.

A solução estacionária,  $E_n^*$ , para a equação 3.12 numa cadeia linear de tamanho  $L$  pode ser obtida usando o *ansatz* (ALIMOHAMMADI; KHORRAMI; AGHAMOHAMMADI, 2001; KRAPIVSKY; REDNER; BEN-NAIM, 2010)

$$E_n^* = Az_1^n + Bz_2^n, \quad (3.13)$$

do qual chegamos a equação característica

$$z_i^{-1} - 2 + z_i - \gamma(1 - z_i) = 0.$$

Escolhendo as condições de fronteira

$$E_0(t) = 1, E_{L+1} = 0,$$

obtemos:

$$E_n^* = \frac{1}{1 - (1 + \gamma)^{-L-1}} (1 + \gamma)^{-n} + \frac{-(1 + \gamma)^{-L-1}}{1 - (1 + \gamma)^{-L-1}}. \quad (3.14)$$

A completeza nesta solução pode ser afirmada apenas para os intervalos vazios, esse método quando mapeado num formalismo hamiltoniano, é verificado que o processo de truncamento aparece ao nível da correlação de três pontos (MOBILIA; BARES, 2001). Embora ainda tem-se a vantagem de que podemos usar essa solução para verificar por exemplo a densidade de sítios ativos quando apenas reações entre vizinhos são levadas em consideração:

$$c_L(\gamma) = \langle n \rangle = 1 - E_1 = \frac{\gamma(1 + \gamma)^L}{(1 + \gamma)^L + \gamma(1 + \gamma)^L - 1}, \quad (3.15)$$

ou a correlação de dois pontos, definida como:

$$\sigma(\gamma) = \langle n_i n_{i+1} \rangle = E_2 - 2E_1 + 1 = \frac{\gamma(2 + \gamma)}{(1 + \gamma)(1 + \gamma - (1 + \gamma)^{-L})}. \quad (3.16)$$

Um limite que podemos verificar é o de baixa ramificação, quando os sítios raramente se subdividem. Mais adiante analisaremos o limite de baixas correntes, o que impõe igualmente

uma baixa ramificação de canais usados. Neste caso

$$\begin{aligned}\lim_{\gamma \rightarrow 0} c(\gamma) &= \frac{1}{L+1}, \text{ e} \\ \lim_{\gamma \rightarrow 0} \sigma(\gamma) &= \frac{\gamma}{L+1}.\end{aligned}\quad (3.17)$$

Isso indica que para poucas reações que geram mais partículas, o sistema tende a ocupar apenas um sítio e a flutuação em torno desse número é proporcional a taxa de criação. Em tamanhos muito grandes, estas grandezas se comportam da seguinte forma, em torno de  $\gamma = 0$ ,

$$\begin{aligned}\lim_{L \rightarrow \infty} c_L(\gamma) &= \lim_{L \rightarrow \infty} \left[ \frac{1}{1+L} + \frac{L\gamma}{2(1+L)} + \mathcal{O}(\gamma^2) \right] \simeq \frac{\gamma}{2}, \text{ e} \\ \lim_{L \rightarrow \infty} \sigma(\gamma) &= \lim_{L \rightarrow \infty} \left[ \frac{\gamma}{1+L} + \frac{(L-2)\gamma^2}{2(1+L)} + \mathcal{O}(\gamma^3) \right] \simeq \frac{\gamma^2}{2}.\end{aligned}\quad (3.18)$$

### 3.4 Aproximação de campo médio

Para solucionar esse problema de otimização local, ao invés de calcular exatamente a soma de correntes nas bifurcações na rede, usaremos as distribuições destes valores ao longo da cadeia de bifurcações, como na aproximação de campo médio das bifurcações.

Seja  $\mathcal{P}(j;t)$  a distribuição de fluxos  $j$  nos canais ao longo de uma linha, ou seja, um instante de tempo,  $t$  e  $\mathcal{Q}(j;t)$  a respectiva distribuição nos vértices que precedem tais canais. De acordo com a regra de otimização do fluxo nos vértices a distribuição nos canais da linha subsequente  $\mathcal{P}(j;t+1)$  terá uma dependência com a probabilidade condicional, conforme ilustrada na figura 3.6:

$$\begin{aligned}\mathcal{F}(i|j, j_c) &= \left[ \frac{1}{2}\delta(i) + \frac{1}{2}\delta(i-j) \right] \Theta(j_c - j) + \\ &+ \left[ \frac{1}{2}\delta\left(i - \frac{j-j_c}{2}\right) + \frac{1}{2}\delta\left(i - \frac{j+j_c}{2}\right) \right] \Theta(j - j_c).\end{aligned}\quad (3.19)$$

O primeiro termo é a contribuição para que a corrente  $i$  num determinado canal tenha um valor igual a 0 ou o valor de entrada  $j$ , caso o fluxo de entrada seja menor que o fluxo crítico  $j_c$  (a corrente irá fluir completamente pelo canal de menor custo). Já o segundo termo corresponde à divisão do fluxo de entrada de forma otimizada, ambos os canais são usados mas de forma a

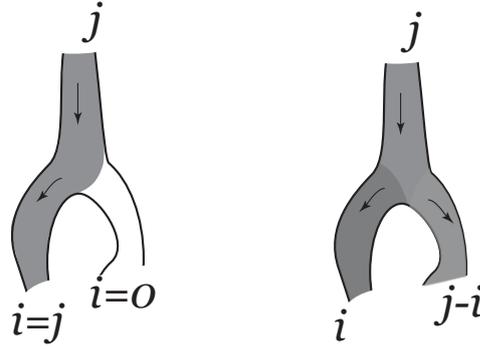


Figura 3.6: Termos para fluxos provenientes de sítios com fluxo não-nulo.

minimizar o custo, caso o fluxo de entrada seja maior que  $j_c$ . Dessa maneira a distribuição de fluxos nos canais, *a posteriori*  $\mathcal{P}(t+1; i)$  dado uma distribuição de fluxos nos vértices, *a priori*  $\mathcal{Q}(t; j)$  será:

$$\mathcal{P}(t+1; i) = 2 \int_{S \in \{c' > c''\}} \int_0^\infty dc' dc'' dj \mathcal{F}(i|j) p(c', c'') \mathcal{Q}(t; j). \quad (3.20)$$

O fator dois aparece devido à ambiguidade na escolha de dois canais. A primeira integração é por toda a superfície de custos, definida pelo elemento de superfície infinitesimal  $dc' dc''$ , em cada canal possível,  $c'$  ou  $c''$ . A segunda integração é sobre todas as possíveis valores de correntes de entrada.

Podemos definir ainda uma distribuição análoga de correntes não-nulas, com valor de integração relacionada à densidade de canais utilizados. Esta distribuição é definida tal qual  $\mathcal{P}(t; i)$ , porém com a densidade de probabilidade condicional  $\mathcal{F}(i|j, j_c)$  subtraída de seu termo de correntes nula, resultando na equação 3.21, correspondendo ao primeiro termo do lado direito da equação 3.19. Como a integral dessa nova distribuição está relacionada com a densidade de canais ativos, precisamos definir para a mesma uma distribuição  $Q(t; i)$  de vértices ativos, análoga à distribuição  $\mathcal{Q}(t; i)$ , na qual sua integração estará relacionada com a densidade de sítios com fluxo não-nulo, esta será analisada com mais detalhes após a completa definição de  $P(t; i)$ . A nova densidade de probabilidade será portanto:

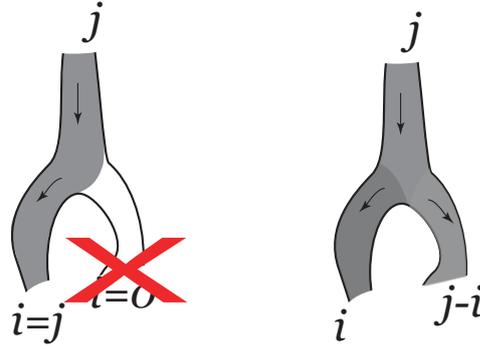


Figura 3.7: Termos para fluxos resultantes não-nulos.

$$F(i|j, j_c) = \frac{1}{2}\delta(i-j)\Theta(j_c-j) + \left[ \frac{1}{2}\delta\left(i - \frac{j-j_c}{2}\right) + \frac{1}{2}\delta\left(i - \frac{j+j_c}{2}\right) \right] \Theta(j-j_c), \quad (3.21)$$

que corresponde ao esquema da figura 3.7.

Vamos considerar a distribuição de custos como estatisticamente independentes, ou seja  $p(c', c'') = p(c')p(c'')$ . Assim, podemos desenvolver a integração sobre os custos e escrever a mesma como

$$P(t+1; i) = 2 \int_0^\infty dj_c p_c(j_c) \int_0^\infty dj F(i|j, j_c) Q(t; j). \quad (3.22)$$

Aqui usamos o fato de que a integração na superfície de custo, com a condição  $c' > c''$ , vai recair na convolução das distribuições das variáveis de custo  $c'$  e  $-c'$ , denotada aqui por  $p_c(j_c)$ , conforme visto na seção 2.2. De fato podemos observar o problema não pela distribuição de custos  $p$  mas por aquela de fluxos críticos  $p_c$ , que neste trabalho será sempre simétrica, já que cada canal possui a mesma natureza de custos.

Resolvendo a integração do termo de correntes abaixo das correntes críticas em 3.21:

$$\begin{aligned} \frac{1}{2} \int_0^\infty 2p_c(j_c) \int_0^\infty dj Q(j) \Theta(j_c-j) dj_c \delta(i-j) &= \frac{Q(i)}{2} 2 \int_0^\infty dj_c p_c(j_c) \Theta(j_c-i) \\ &= \frac{Q(i)}{2} 2 \int_i^\infty dj_c p_c(j_c) \\ &= \frac{Q(i)}{2} \mathcal{P}_c(i). \end{aligned} \quad (3.23)$$

Na segunda linha notamos uma distribuição cumulativa complementar que denotamos aqui por

$$\mathcal{P}_c(i) = 2 \int_i^\infty dj_c p_c(j_c). \quad (3.24)$$

Os termos seguintes, de forma similar, serão:

$$\begin{aligned} \frac{1}{2} \int_0^\infty 2p_c(j_c) \int_0^\infty dj Q(j) \Theta(j-j_c) dj_c \delta\left(i - \frac{j-j_c}{2}\right) &= 2 \int_0^\infty dj_c \int_0^\infty dj p_c(j_c) Q(j) \Theta(j-j_c) \delta(2i-j+j_c) \\ &= 2 \int_0^\infty dj_c p_c(j_c) Q(2i+j_c) \Theta(2i) \\ &= 2 \int_0^\infty dj p_c(j) Q(2i+j), e \end{aligned} \quad (3.25)$$

$$\begin{aligned} \frac{1}{2} \int_0^\infty 2p_c(j_c) \int_0^\infty dj Q(j) \Theta(j-j_c) dj_c \delta\left(i - \frac{j+j_c}{2}\right) &= 2 \int_0^\infty dj_c \int_0^\infty dj p_c(j_c) Q(j) \Theta(j-j_c) \delta(2i-j-j_c) \\ &= 2 \int_0^\infty dj_c p_c(j_c) Q(2i-j_c) \Theta(i-j_c) \\ &= 2 \int_0^i dj p_c(j) Q(2i-j). \end{aligned} \quad (3.26)$$

Finalmente a distribuição de fluxos não nulos será dada por:

$$P(t+1; i) = \frac{\mathcal{P}_c(i)}{2} Q(i) + 2 \int_0^\infty dj p_c(j) Q(2i+j) + 2 \int_0^i dj p_c(j) Q(2i-j) \quad (3.27)$$

Para a distribuição de fluxos nos vértices a partir dos fluxos de entrada nos canais, definimos a probabilidade  $B(t)$  de fluxos nulos ao longo de uma linha  $t$ :

$$\mathcal{Q}(t; i) = [B(t)]^2 \delta(i) + 2B(t)P(t; i) + \int_0^i du P(t; u)P(t; i-u) \quad (3.28)$$

A distribuição de correntes num determinado vértice é dada pela soma das probabilidades de dois canais vazios (primeiro termo), um canal vazio e outro não, adicionado à probabilidade de dois canais usados somarem corrente igual à corrente no vértice, por sua vez dada pela convolução da distribuição  $P(t; j)$ , como visualizada na figura 3.8. Observe que a distribuição

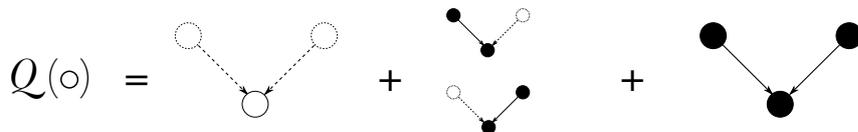


Figura 3.8: Termos para a distribuição de correntes nos vértices, equação 3.28.

$\mathcal{P}$  é escrita em termos da distribuição de correntes finitas nos vértices, i. e., a equação 3.28 sem

o termo de correntes nulas:

$$Q(i;t) = 2B(t)P(t;i) + \int_0^i du P(t;u)P(t;i-u) \quad (3.29)$$

que possui a integral facilmente calculada atribuída à densidade de vértices ocupados com alguma corrente,

$$\int_0^\infty di Q(i;t) = 1 - [B(t)]^2, \quad (3.30)$$

visto que a integral da distribuição em 3.29 é a unidade. Esta observação leva a

$$\int_0^\infty di P(i;t) = 1 - B(t), \quad (3.31)$$

ou seja, a distribuição  $P(j;t)$  possui sua integral associada à densidade de canais usados.

De posse a esses resultados e ao integrarmos a equação 3.27, obtemos:

$$\begin{aligned} 1 - B(t+1) &= \int_0^\infty di P(t+1;i) \\ &= \int_0^\infty di \frac{\mathcal{P}_c(i)}{2} Q(i) + 2 \int_0^\infty di \int_0^\infty dj p_c(j) Q(2i+j) + 2 \int_0^\infty di \int_0^i dj p_c(j) Q(2i-j). \end{aligned} \quad (3.32)$$

Podemos escrever os termos não triviais usando a substituição

$$\begin{cases} x = 2i + j \\ y = j \end{cases} \Rightarrow \begin{cases} i = (x-y)/2 \\ j = y \end{cases} \Rightarrow \mathcal{J} = \frac{1}{2}, \quad (3.33)$$

de Jacobiano

$$\mathcal{J} = \begin{vmatrix} \frac{\partial i}{\partial x} & \frac{\partial i}{\partial y} \\ \frac{\partial j}{\partial x} & \frac{\partial j}{\partial y} \end{vmatrix}$$

a inequação  $0 \leq i$ , do limite inferior da integração em  $i$  do segundo termo na equação 3.33, implica em uma nova relação de desigualdade entre as novas variáveis  $y \leq x$ .

$$\begin{aligned} 2 \int_0^\infty di \int_0^\infty dj p_c(j) Q(2i+j) &= \frac{1}{2} \int_0^\infty dx Q(x) 2 \int_0^x dy p_c(y) \\ &= \frac{1}{2} \int_0^\infty di Q(i) (1 - \mathcal{P}_c(i)). \end{aligned} \quad (3.34)$$

O segundo termo não trivial (terceira parcela da equação 3.33) terá o mesmo jacobiano, porém a transformação  $x = 2i - j$  com a inequação  $j \leq i$  do limite superior da integral em  $j$  desse termo,

implica na condição  $y \leq x$ .

$$\begin{aligned} 2 \int_0^\infty di \int_0^i dj p_c(j) Q(2i-j) &= \frac{1}{2} \int_0^\infty dx Q(x) 2 \int_0^x dy p_c(y) \\ &= \frac{1}{2} \int_0^\infty di Q(i) (1 - \mathcal{P}_c(i)) \end{aligned} \quad (3.35)$$

Assim,

$$\begin{aligned} 1 - B(t+1) &= \int_0^\infty di \frac{\mathcal{P}_c(i)}{2} Q(i) + \int_0^\infty di Q(i) (1 - \mathcal{P}_c(i)) \\ &= \int_0^\infty di Q(i) - \frac{1}{2} \int_0^\infty di \mathcal{P}_c(i) Q(i) \\ \Rightarrow B(t+1) &= B^2(t) + \frac{1}{2} \int_0^\infty di \mathcal{P}_c(i) Q(i) \end{aligned} \quad (3.36)$$

O resultado acima concorda com os termos subtraídos das distribuições  $\mathcal{Q}()$  e  $\mathcal{P}()$ , que aparecem aqui como os termos responsáveis pelo incremento da densidade de canais vazios  $B$ . Em outras palavras, o número de canais vazios aumenta ou por conta dos vértices vizinhos não possuírem fluxo, representado pelo primeiro termo, ou por conta de fluxos muito baixos para serem divididos entre os dois canais, representado pelo segundo termo. Definimos portanto um sistema de equações auto-consistente dado por:

$$\begin{aligned} Q(i;t) &= 2B(t)P(i;t) + \int_0^i dj P(j;t)P(i-j;t) \\ P(i;t+1) &= \frac{\mathcal{P}_c(i)}{2} Q(i;t) + 2 \int_0^\infty dj p_c(j;t) Q(2i+j;t) + 2 \int_0^i dj p_c(j;t) Q(2i-j;t) \\ B(t+1) &= B^2(t) + \frac{1}{2} \int_0^\infty di \mathcal{P}_c(i) Q(i;t) \end{aligned} \quad (3.37)$$

### 3.4.1 Coalescência.

Para melhor entendimento das equações de recorrência desse problema, vamos analisar a forma mais simples do modelo, que é o processo sem divisão de fluxos nos vértices, isto é, o fluxo corre por apenas um dos canais. Nesta formulação temos para a fração de canais vazios:

$$B(t+1) = \frac{1 + B^2(t)}{2}. \quad (3.38)$$

Para o limite de tempo contínuo, teremos

$$\frac{dB}{dt} = \frac{(1-B)^2}{2}. \quad (3.39)$$

Para uma condição inicial  $B(0) = 0$ , todos os canais são usados, a solução será:

$$B(t) = \frac{t}{2+t} \simeq 1 - \frac{2}{t}. \quad (3.40)$$

Esta equação mostra um decaimento  $t^{-1}$  para a fração de canais ativos. De fato essa equação 3.39, corresponde a equação de campo médio para o processo de aniquilação ou coalescência de uma espécie única (KRAPIVSKY; REDNER; BEN-NAIM, 2010), ao procedermos com a substituição  $1 - B \rightarrow c$ :

$$\frac{dc}{dt} = -kc^2, \quad (3.41)$$

sendo  $c$  a fração da população ativa. Esta é a equação de campo médio para dimensão maior que a crítica desse processo,  $d_c = 2$ . Para a dimensão subcrítica, na qual foram realizadas nossas simulações, no limite de correntes pequenas o processo predominante é o de coalescência e temos um expoente para o tempo igual a  $1/2$ , como visto na figura 3.9.

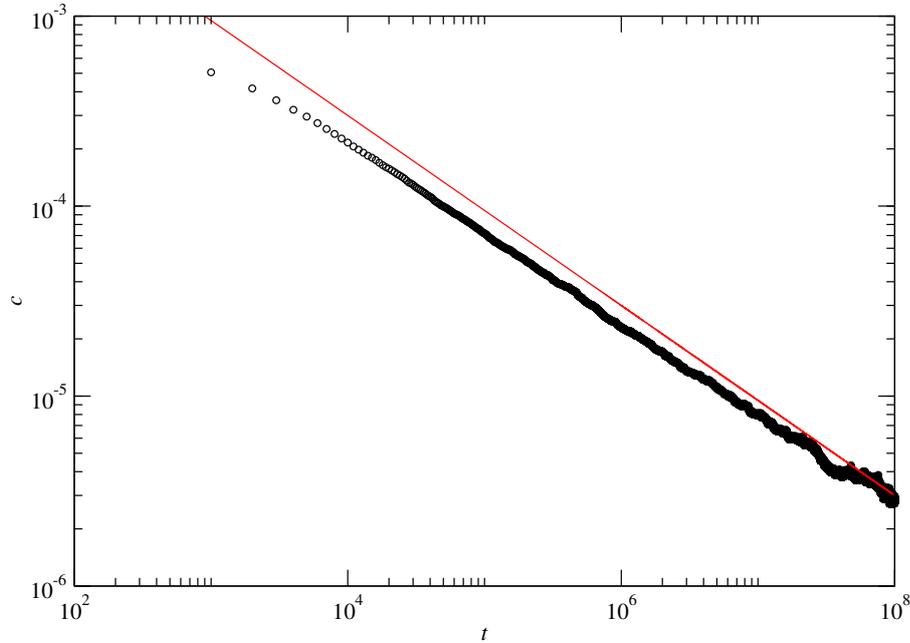


Figura 3.9: Gráfico para o tempo de coalescência. Simulação com  $2L = 20000$  canais,  $\langle j \rangle = 10^{-9}$  e 25 amostras, usamos uma distribuição uniforme de custos para os vértices da rede que define a corrente crítica máxima  $\beta = 1$ . A curva vermelha contínua representa uma lei de potência com expoente  $1/2$ .

### 3.5 Custos uniformemente distribuídos

Para observar o comportamento estacionário desse sistema é necessário obter as soluções das equações de recorrência. Para isso podemos supor uma distribuição conhecida de custos. Um dos casos mais simples é para uma distribuição uniforme de custos dada por:

$$p(c', c'') = p(c')p(c'') = \begin{cases} \frac{1}{(\beta - \alpha)^2}, & \text{if } \alpha \leq c \leq \beta \\ 0, & \text{caso contrário,} \end{cases} \quad (3.42)$$

Os valores de custos de cada canal estão uniformemente distribuído no intervalo  $[\alpha, \beta]$ . Essa distribuição nos fornece uma equivalente para fluxos críticos como uma reta decrescente no intervalo  $[0, \beta - \alpha]$ . Sem perda de generalidade, usaremos a substituição  $\alpha \rightarrow 0$  e  $\beta \rightarrow \beta - \alpha$ , portanto:

$$p_c(j_c) = \begin{cases} \frac{1}{\beta^2}(\beta - j_c), & \text{if } 0 \leq c \leq \beta \\ 0, & \text{caso contrário} \end{cases}, \quad (3.43)$$

que é consistente com o caso uniforme, na equação 2.5, para valores positivos da corrente, nas equações integrais da distribuição  $P(t+1; i)$  3.20. Qualquer que seja a distribuição adotada, dado que a densidade condicional das soluções de fluxos otimizados, equação 3.19, a solução não possui valores críticos negativos.

As equações 3.37 para a distribuição nos canais e a fração de canais vazios serão:

$$\begin{aligned} Q(t; i) &= 2B(t)P(t; i) + \int_0^i dj P(t; j)P(t; i - j) \\ P(t+1; i) &= \begin{cases} \frac{(\beta - i)^2}{2\beta^2} Q(t; i) + \frac{2}{\beta^2} \int_0^i dj (\beta - j) Q(t; 2i - j) + \frac{2}{\beta^2} \int_0^\beta dj (\beta - j) Q(t; 2i + j) & , 0 \leq i \leq \beta \\ \frac{2}{\beta^2} \int_0^\beta dj (\beta - j) Q(t; 2i + j) + \frac{2}{\beta^2} \int_0^\beta dj (\beta - j) Q(t; 2i - j) & , i > \beta \end{cases} \\ B(t+1) &= B^2(t) + \frac{1}{2\beta^2} \int_0^\beta di (\beta - i)^2 Q(t; i). \end{aligned} \quad (3.44)$$

A equação para a distribuição nos vértices permanece inalterada, já que não é influenciada diretamente pelo custo da ligações mas resulta da soma das correntes que chegam aos sítios.

É importante salientar que uma consequência direta da definição da distribuição nos vértices  $Q(j)$  (3.28), é que sua média será o dobro da média da distribuição nos canais  $P(j)$ . Das

equação 3.44 para a distribuição  $P(t = 1; i)$ , podemos extrair também tal conclusão:

$$\begin{aligned} \langle i \rangle &= \int_0^\beta idi \frac{(\beta - i)^2}{2\beta^2} Q(t; i) + \frac{2}{\beta^2} \int_0^\beta idi \int_0^i dj(\beta - j) Q(t; 2i - j) \\ &+ \frac{2}{\beta^2} \int_0^\infty idi \int_0^\beta dj(\beta - j) Q(t; 2i + j) + \frac{2}{\beta^2} \int_\beta^\infty idi \int_0^\beta dj(\beta - j) Q(t; 2i - j). \end{aligned} \quad (3.45)$$

Podemos expandir os termos da equação anterior, omitindo a dependência temporal:

$$\begin{aligned} 2 \int_0^\beta idi \int_0^i dj(\beta - j) Q(2i - j) &= \int_0^\beta \int_0^x dy \left( \frac{x+y}{2} \right) (\beta - y) Q(x) dx + \\ &\int_\beta^{2\beta} \int_0^{2\beta-x} dy \left( \frac{x+y}{2} \right) (\beta - y) Q(x) dx \\ &= + \frac{1}{12} \int_0^\beta (9\beta - 5i) i^2 Q(i) di + \frac{1}{12} \int_\beta^{2\beta} (-4\beta^3 + 12\beta^2 i - 3\beta i^2 - i^3) Q(i) di \end{aligned}$$

$$\begin{aligned} 2 \int_\beta^\infty idi \int_0^\beta dj(\beta - j) Q(2i - j) &= \int_\beta^{2\beta} \int_{2\beta-x}^\beta dy \left( \frac{x+y}{2} \right) (\beta - y) Q(x) dx + \\ &+ \int_{2\beta}^\infty \int_0^\beta dy \left( \frac{x+y}{2} \right) (\beta - y) Q(x) dx \\ &= \frac{1}{12} \int_\beta^{2\beta} (\beta - i)^2 (5\beta + i) Q(i) di \\ &+ \frac{1}{12} \int_{2\beta}^\infty \beta^2 (\beta + 3i) Q(i) di \end{aligned}$$

Para os termos de  $\beta$  até  $2\beta$

$$\int_\beta^{2\beta} (-4\beta^3 + 12\beta^2 i - 3\beta i^2 - i^3 + (\beta - i)^2 (5\beta + i)) Q(i) di = \int_\beta^{2\beta} \beta^2 (\beta + 3i) Q(i) di$$

$$\begin{aligned} 2 \int_0^\infty idi \int_0^\beta dj(\beta - j) Q(2i + j) &= \int_0^\beta \int_0^x dy \left( \frac{x-y}{2} \right) (\beta - y) Q(x) dx \\ &+ \int_\beta^\infty \int_0^\beta dy \left( \frac{x-y}{2} \right) (\beta - y) Q(x) dx \\ &= \frac{1}{12} \int_\beta^\infty \beta^2 (3i - \beta) Q(i) di + \frac{1}{12} \int_0^\beta (3\beta - i) i^2 Q(i) di \end{aligned}$$

Finalmente obtemos:

$$\begin{aligned}
\langle i \rangle &= \int_0^\beta di \frac{i(\beta - i)^2}{2\beta^2} Q(i) + \frac{1}{2\beta^2} \int_0^\beta di (2\beta - i)i^2 Q(i) + \frac{1}{2} \int_\beta^\infty i Q(i) di \\
&= \frac{1}{2\beta^2} \int_0^\beta di [i(\beta - i)^2 + (2\beta - i)i^2] Q(i) + \frac{1}{2} \int_\beta^\infty i Q(i) di \\
\Rightarrow \quad \langle i \rangle_P &= \frac{1}{2} \langle i \rangle_Q
\end{aligned} \tag{3.46}$$

Essa solução é uma consequência direta da lei de conservação adotada nos fluxos que saem dos vértices, representada pelo fluxo médio neste sistema infinito.

A iteração deve ser realizada supondo uma distribuição inicial nos vértices ou ligações e calcular as equações 3.44 seguido de uma interpolação para o armazenamento da distribuição *a posteriori*, que será usada para realizar o cálculo das novas distribuições. Esse procedimento pode ser repetido até que seja obtido um valor da fração estacionária de canais vazios  $B$ .

Podemos verificar o comportamento geral da distribuição de fluxos nos canais na figura 3.10. É possível verificar também o comportamento do parâmetro de ordem, a fração de canais usados  $1 - B$ , em função da razão  $\langle j \rangle / \beta$  que é o único parâmetro livre do modelo, como visto na figura 3.11, que além dos resultados de iteração numérica apresenta outros que serão analisados mais adiante.

## 3.6 Limite para grandes correntes

Para grandes correntes podemos derivar a distribuição final analiticamente, ao usarmos as transformadas de Laplace para as distribuições de fluxos nos vértices e nos canais:

$$\Pi(z) = \int_0^\infty dj P(j) e^{-zj} \tag{3.47}$$

$$K(z) = \int_0^\infty dj Q(j) e^{-zj}. \tag{3.48}$$

As equações de recorrência para as transformadas serão:

$$K(z) = 2B\Pi(z) + \Pi(z)^2, \text{ e} \tag{3.49}$$

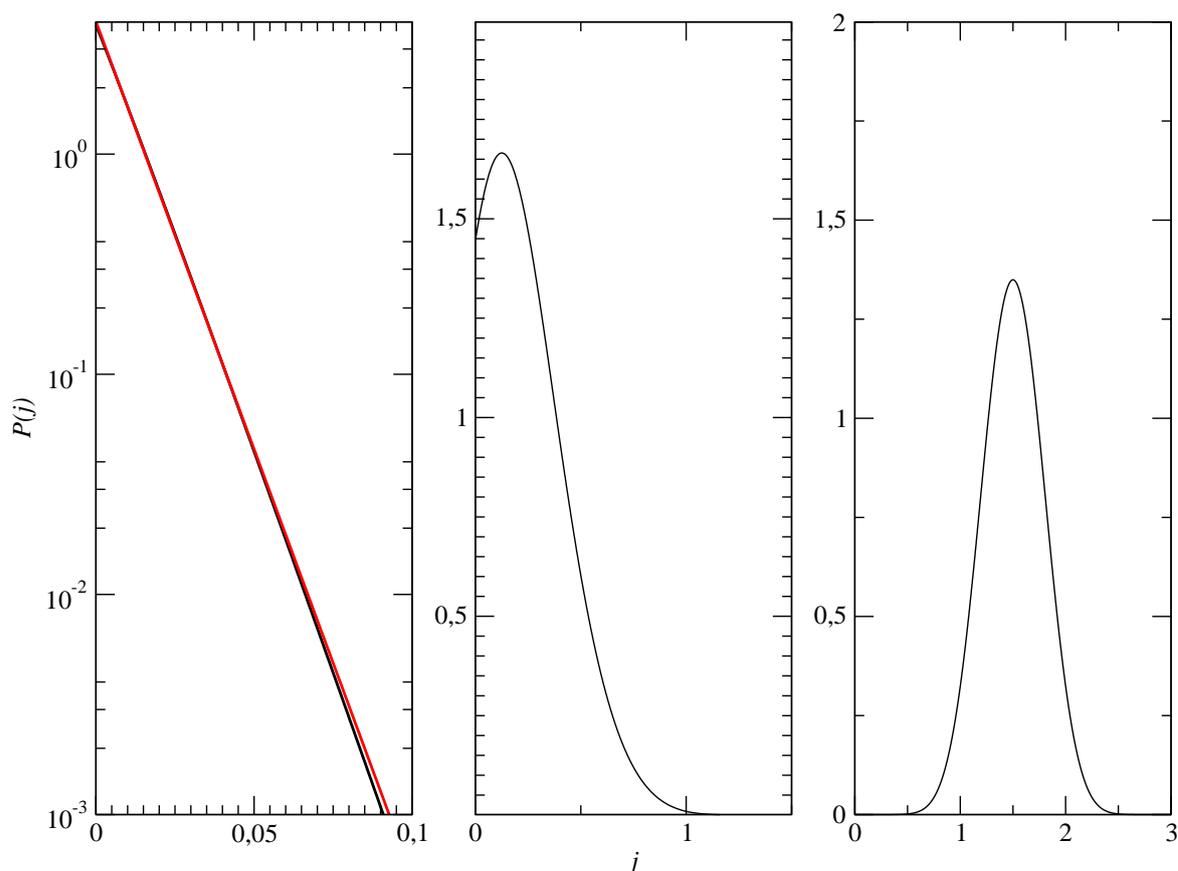


Figura 3.10: Iteração numérica das equações 3.28 e 3.44 para  $t_{\text{máx}} = 250$  passos. Os gráficos correspondem aos valores para corrente média  $\langle j \rangle = 5 \cdot 10^{-4}$ ,  $10^{-2}$ , e  $1,5$  respectivamente. A largura da distribuição uniforme foi escolhida igual a unidade ( $\beta=1$ ), por simplicidade. A linha vermelha para o primeiro gráfico corresponde a um decaimento exponencial do tipo  $\sim \exp\left(-\frac{2}{\sqrt{\langle j \rangle}} j\right)$ .

$$\begin{aligned} \Pi(z) = & \frac{1}{2\beta^2 z^2} \int_0^\beta e^{-xz} \left( 8 - 16e^{\frac{xz}{2}} + 4xz + \beta^2 z^2 + x^2 z^2 + e^{xz} (8 + 4\beta z - 4xz) - 2\beta z (2 + xz) \right) Q(x) dx + \\ & + \frac{1}{\beta^2 z^2} \int_\beta^\infty 4e^{-\frac{1}{2}(\beta+x)z} \left( e^{\frac{\beta z}{2}} - 1 \right)^2 Q(x) dx. \end{aligned} \quad (3.50)$$

No limite para de altas correntes  $\langle j \rangle \gg \beta$  apenas o primeiro termo possui alguma contribuição

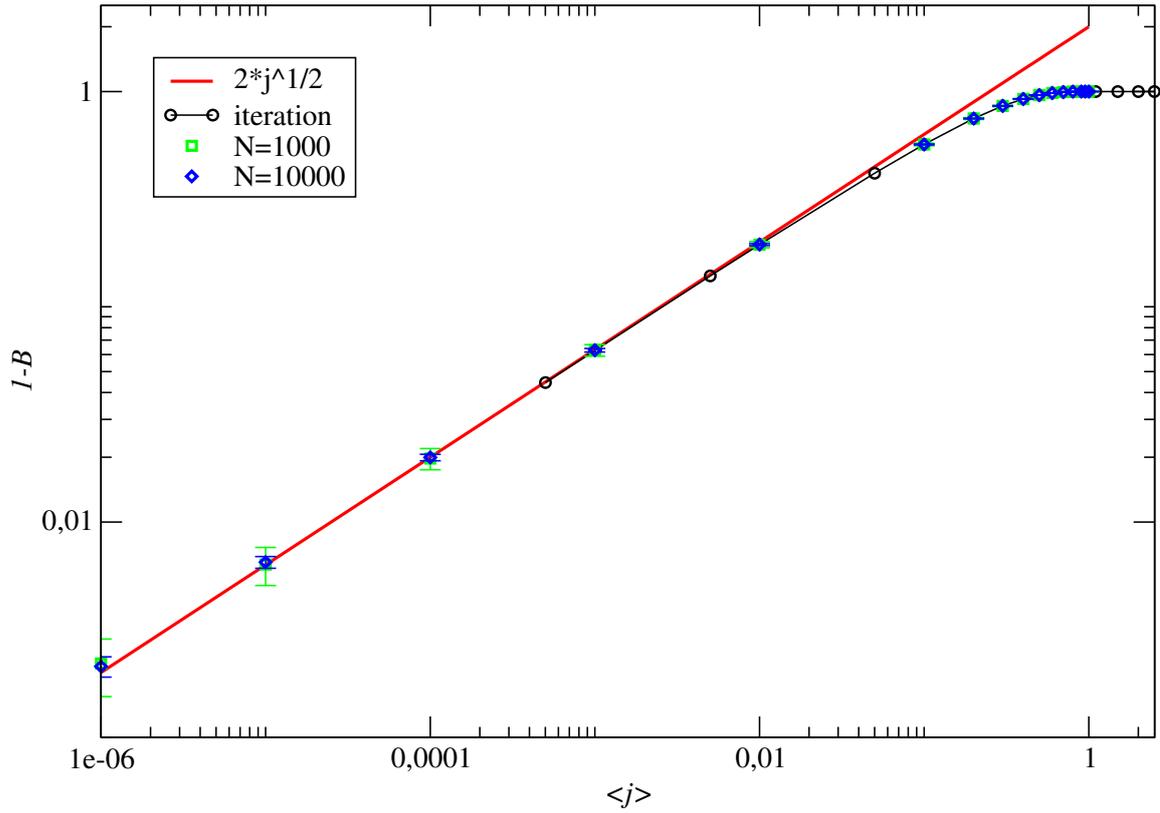


Figura 3.11: Iteração final para a fração de canais usados, foi obtido um comportamento monótono para o parâmetro de ordem com o tempo em torno de 250 iterações partindo de uma distribuição inicial uniforme. Usamos  $\beta = 1$  por simplicidade na implementação.

significativa, e a fração de canais vazios será nula. Desenvolvendo a equação 3.50 neste limite

$$\begin{aligned}
 \Pi(z) &= 4 \frac{\left(e^{\frac{\beta z}{2}} - 1\right)^2 e^{-\frac{\beta z}{2}}}{\beta^2 z^2} \int_0^\infty e^{-\frac{xz}{2}} Q(x) dx \\
 \Pi(z) &= \frac{4}{\beta^2 z^2} e^{-\frac{\beta z}{2}} \left[2e^{\frac{\beta z}{4}} \sinh\left(\frac{\beta z}{4}\right)\right]^2 K\left(\frac{z}{2}\right) \\
 \Pi(z) &= \left[\frac{4}{\beta z} \sinh\left(\frac{\beta z}{4}\right)\right]^2 K\left(\frac{z}{2}\right) \\
 \Rightarrow \Pi(z) &= \left[\frac{4}{\beta z} \sinh\left(\frac{\beta z}{4}\right) \Pi\left(\frac{z}{2}\right)\right]^2. \tag{3.51}
 \end{aligned}$$

Podemos iterar esta equação  $n$  vezes e obteremos:

$$\Pi^{(n)}(z) = \Pi^{(0)}\left(\frac{z}{2^n}\right) \prod_{k=1}^n \left[\frac{2^{k+1}}{\beta z} \sinh\left(\frac{\beta z}{2^{k+1}}\right)\right]^{2^k}. \tag{3.52}$$

Para um número muito grande de iterações ( $n \rightarrow \infty$ ) a seguinte identidade pode ser usada

$$\Pi^{(0)}\left(\frac{z}{2^n}\right)^{2^n} = \left(\Pi^{(0)}(0) + \Pi'^{(0)}(0) \frac{z}{2^n}\right)^{2^n} = e^{-\langle j \rangle z}, \tag{3.53}$$

tal que  $\Pi'(0) = -\langle j \rangle$  e  $\Pi(0) = 1 - B = 1$ .

Esta transformada de Laplace não possui pólos, já que quando  $z = 0$ , verifica-se que

$$\lim_{z \rightarrow 0} \frac{2^{k+1}}{\beta z} \sinh\left(\frac{\beta z}{2^{k+1}}\right) = 1.$$

Podemos proceder com a transformação inversa escolhendo uma linha de integração na variável  $z$ , convenientemente escolhemos a reta em  $Re(z) = 0$ :

$$\begin{aligned} P(j; \langle j \rangle, \beta) &= \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} dz e^{-\langle j \rangle - jz} \prod_{k=1}^{\infty} \left[ \frac{2^{k+1}}{\beta z} \sinh\left(\frac{\beta z}{2^{k+1}}\right) \right]^{2^k}, \quad z \in \Re \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} dz e^{-\langle j \rangle - jiz} \prod_{k=1}^{\infty} \left[ \frac{2^{k+1}}{\beta iz} \sinh\left(\frac{\beta iz}{2^{k+1}}\right) \right]^{2^k}, \quad i^{2^k} = 1 \\ &= \frac{1}{2\pi} \int_0^{\infty} dz \left( e^{\langle j \rangle - jiz} + e^{-\langle j \rangle - jiz} \right) \prod_{k=1}^{\infty} \left[ \frac{2^{k+1}}{\beta z} \sin\left(\frac{\beta z}{2^{k+1}}\right) \right]^{2^k} \\ &= \frac{1}{\pi\beta} \int_0^{\infty} dz \cos\left(\left(\langle j \rangle - j\right) \frac{z}{\beta}\right) \prod_{k=1}^{\infty} \left[ \frac{2^{k+1}}{z} \sin\left(\frac{z}{2^{k+1}}\right) \right]^{2^k}. \end{aligned} \quad (3.54)$$

Podemos esperar uma convergência da expressão, dado que os últimos elementos do produtório se aproximam da unidade e a função no argumento deste decai para seu valor nulo ao fim do intervalo de integração. A figura 3.12 mostra o comportamento dessa função para seis iterações. É possível verificar que esta solução não configura uma dependência com a corrente média para as flutuações. Podemos obter o resultado da variância  $\sigma^2$  independente do primeiro momento ao calcularmos a segundo momento da distribuição 3.54:

$$\langle j^2 \rangle = \lim_{z \rightarrow 0} \Pi''(z) = A\beta^2 + \langle j \rangle^2 \Rightarrow \sigma^2 = A\beta^2 \quad (3.55)$$

onde a constante  $A = 1/12$  é obtida semi-analiticamente por meio de extrapolação do valor das derivadas calculadas com o multiplicatório truncado até o valor 7. A figura 3.13 mostra o valor da segunda derivada de  $Pi(z)$  tomado no limite  $z \rightarrow 0$  (equação 3.55). Tal valor limite utilizado para a escolha da variância da gaussiana da figura 3.12.

Esse mesmo valor da constante na variância foi obtido por citeonlinefabricio para simulação na condição de corrente média igual a maior corrente crítica da rede, como evidenciado na figura 3.14.

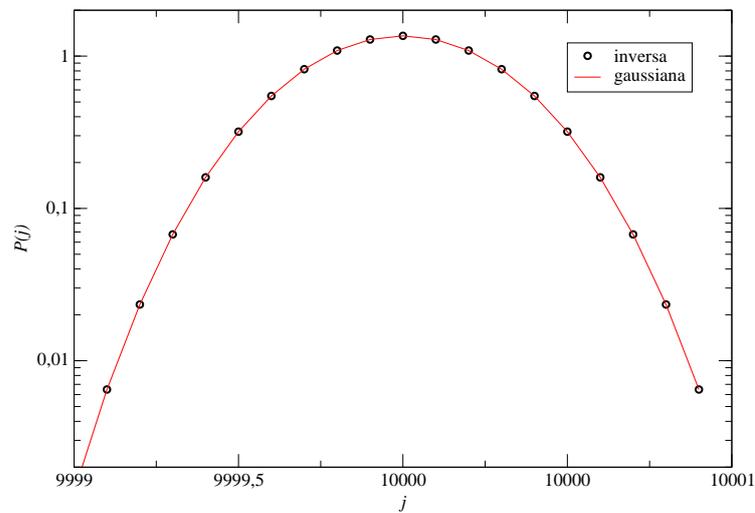


Figura 3.12: Solução numérica para distribuição estacionária de fluxos nos canais, eq. 3.54, no limite de grandes fluxos,  $B = 0$  (pontos). Parâmetros utilizados:  $\langle j \rangle = 10^4$ ,  $\beta = 1$ . A fim de comparação a curva contínua corresponde a uma gaussiana de variância  $1/12$ .

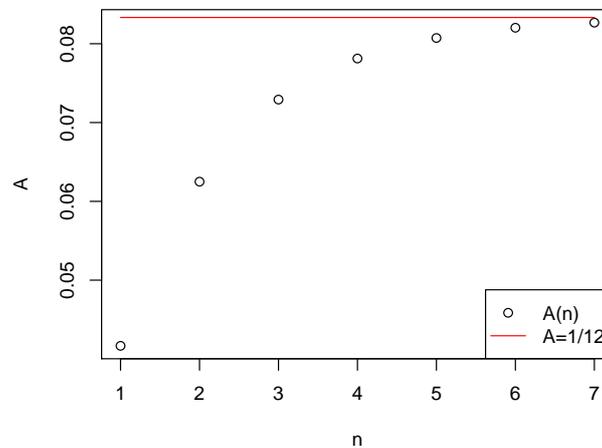


Figura 3.13: Determinação do coeficiente da variância para a distribuição estacionária para altas correntes.

Isso mostra que o formalismo está de acordo com o teorema do limite central (REIF, 1965), pois a corrente é uma variável que é acumulada a cada instante de tempo nos canais. Portanto a distribuição resultante de uma variável que é a soma de variáveis aleatórias com distribuições com primeiro e segundo momento bem definidos deve tender a uma gaussiana quando a quantidade de variáveis acumuladas é muito grande. Esse limite é válido quando temos a fração

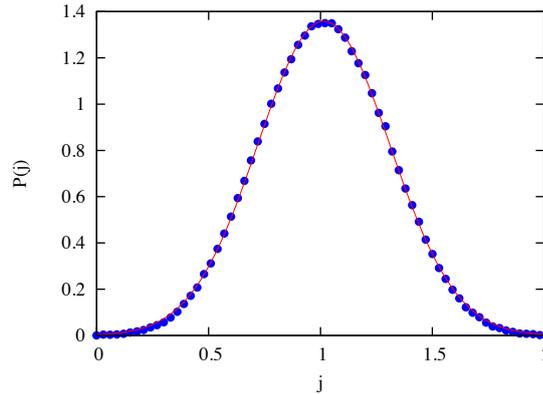


Figura 3.14: Simulação para uma rede com custos uniformemente distribuído no intervalo  $(0, 1]$ ,  $\langle j \rangle = 1$ ,  $L = 1000$  e 100 amostras. A gaussiana ajustada possui média  $\mu = 1,01$  e variância  $\sigma^2 = 1/2$ . Figura retirada de Forgerini (2013, página 89).

$\langle j \rangle / \beta = 1$ , fazendo com que a probabilidade de canais não utilizados seja nula. As figuras 3.12 e 3.14 sugerem portanto:

$$P_{\sigma}(j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(j-\langle j \rangle)^2}{2\sigma^2}}, \quad (3.56)$$

onde  $\sigma$  é o desvio padrão, dada a equação 3.55 possui valor  $\sigma = 1/\sqrt{12}$ . Conclui-se de tal análise que o único fator determinante para as flutuações em grandes correntes deve ser a largura da distribuição  $\sigma \rightarrow f(\beta)$ .

### 3.7 Limite para baixas correntes

Neste limite em que a largura da distribuição de fluxos críticos é muito maior que a corrente média ( $\langle j \rangle \ll \beta$ ) as equações de recorrência para a transformada de Laplace – primeiro termo da equação 3.50 – levam a uma relação pouco prática para executar alguma recorrência:

$$\Pi(z) = \frac{K(z)}{2} + \frac{K'(z)}{\beta} + \frac{K''(z)}{2\beta^2} + 4\frac{K(0)}{\beta^2 z^2} + \frac{4K(z)}{\beta^2 z^2} - \frac{8K\left(\frac{z}{2}\right)}{\beta^2 z^2} + \frac{2K(0)}{\beta z} - \frac{2K(z)}{\beta z} + \frac{2K'(0)}{\beta^2 z} - \frac{2K'(z)}{\beta^2 z}.$$

Em contrapartida podemos usar as iterações numéricas a fim de verificar a forma assintótica para a distribuição estacionária. Tendo em mãos a distribuição exponencial na figura 3.10, po-

demostremos obter facilmente a constante de normalização a partir das equações de recorrência:

$$\begin{aligned}
P(0) &= \frac{\beta^2}{2\beta^2} Q(0) + \frac{2}{\beta^2} \int_0^\beta dj(\beta - j)Q(j) \\
&= \frac{1}{2} Q(0) + \frac{2}{\beta}(1 - B^2) - \frac{4}{\beta^2} \langle j \rangle - \frac{2}{\beta^2} \int_\beta^\infty dj(\beta - j)Q(j) \\
&= BP(0) + \frac{2}{\beta}(1 - B^2) - \frac{4}{\beta^2} \langle j \rangle \\
P(0) &= \frac{2}{\beta}(1 + B) - \frac{4}{\beta^2} \frac{\langle j \rangle}{1 - B}. \tag{3.57}
\end{aligned}$$

Aqui usamos o fato que a integração sobre valores maiores ou iguais ao máximo fluxo crítico  $\beta$  das distribuições em questão pouco contribuem para a distribuição final nesse limite. Portanto, para  $\langle j \rangle \rightarrow 0$ , o valor assintótico de  $P(0)$  será 4, faltando constatar que o denominador  $1 - B$  vai pra zero de forma mais lenta que a corrente média. Para isso a equação de  $B$  pode ser desenvolvida de forma semelhante:

$$\begin{aligned}
B &= B^2 + \frac{1}{2\tilde{\beta}^2} \int_0^{\tilde{\beta}} dj(\tilde{\beta} - j)^2 Q(j) \\
2B^2 - 2B + \frac{1}{\tilde{\beta}^2} \int_0^{\tilde{\beta}} dj(\tilde{\beta}^2 - 2\tilde{\beta}j + j^2) Q(j) &= 0 \\
2B^2 - 2B + (1 - B^2) - 2 \left( \frac{2\langle j \rangle}{\tilde{\beta}} \right) + \frac{\langle j^2 \rangle_Q}{\tilde{\beta}^2} &= 0 \\
B^2 - 2B + 1 - 4 \frac{\langle j \rangle}{\tilde{\beta}} + \frac{\langle j^2 \rangle_Q}{\tilde{\beta}^2} &= 0 \\
\Rightarrow B &= 1 - \frac{1}{2} \sqrt{4 - 4 \left( 1 - 4 \frac{\langle j \rangle}{\tilde{\beta}} + \frac{\langle j^2 \rangle_Q}{\tilde{\beta}^2} \right)} \tag{3.58}
\end{aligned}$$

sendo  $\langle j^2 \rangle_Q$  o segundo momento da distribuição  $Q(j)$  e supondo  $\langle j^2 \rangle_Q / \beta^2$  indo a zero mais rápido que  $\langle j \rangle / \beta$ , obtemos:

$$B(\langle j \rangle) = 1 - 2\sqrt{\frac{\langle j \rangle}{\beta}}, \tag{3.59}$$

confirmando assim a curva com expoente  $1/2$  para a corrente média na figura 3.11 como expoente de campo médio.

Com isso obtemos a condição para a normalização da distribuição  $P(j)$  que unido ao valor do coeficiente de normalização obtido por meio de 3.57 e o *ansatz* fornecido pela iteração

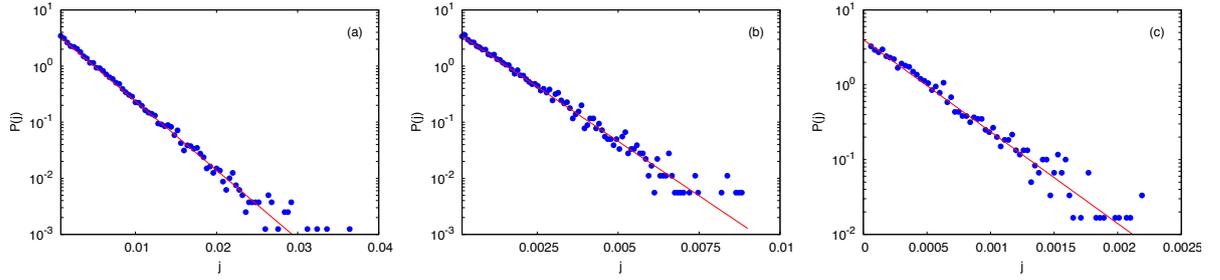


Figura 3.15: Distribuição de correntes para o limite de baixa correntes. Parâmetros utilizados:  $\beta = 1$ ,  $\langle j \rangle = 5 \cdot 10^{-5}$ ,  $5 \cdot 10^{-6}$ ,  $5 \cdot 10^{-7}$ , respectivamente. Os ajustes seguem a equação 3.60. Figura retirada de Forgerini (2013, página 89).

numérica de (3.10), a suposição para nossa distribuição estacionária será:

$$P(j) = 4e^{-2/\sqrt{\langle j \rangle}/\beta j}. \quad (3.60)$$

Essa forma funcional foi observada também por ?? ao simular o limite de baixas corrente médias, o histograma das distribuições resultantes nessas simulações, figura 3.15, concordam bem com a expressão aqui obtida com as iterações numéricas.

Uma forma possível para verificar se esta distribuição é atrator desse sistema para pequenas correntes seria: escrever as equações de recorrência para os momentos da distribuição neste limite; escolher a expressão 3.60 como distribuição inicial e verificar se os momentos seguintes são mantidos. A equação 3.44 resulta numa recorrência para os momentos:

$$\begin{aligned} \langle j^2 \rangle_P &= \frac{\langle j^2 \rangle_Q}{2} - \frac{\langle j^3 \rangle_Q}{3\beta} + \frac{\langle j^4 \rangle_Q}{8\beta^2} \\ \langle j^3 \rangle_P &= \frac{\langle j^3 \rangle_Q}{2} - \frac{\langle j^4 \rangle_Q}{2\beta} + \frac{3\langle j^5 \rangle_Q}{16\beta^2} \\ &\vdots \\ \langle j^n \rangle_P &= \frac{\langle j^n \rangle_Q}{2} - \frac{1}{\beta} \left( \frac{n-1}{n+1} \right) \langle j^{n+1} \rangle_Q + \frac{1}{\beta^2} \frac{(2^{n+1} - 2^n n + 2^n n^2 - 4)}{2^{n+1}(1+n)(2+n)} \langle j^{n+2} \rangle_Q, \end{aligned} \quad (3.61)$$

Para os momentos da distribuição de fluxos nos vértices em termos dos fluxos nos canais, teremos:

$$\langle j^n \rangle_Q = 2 \langle j^n \rangle_P + \sum_{k=1}^{n-1} \binom{n}{k} \langle j^k \rangle_P \langle j^{n-k} \rangle_P \quad (3.62)$$

Em posse dessas relações é possível verificar que para a distribuição 3.60 os momentos poste-

riores ganham apenas termos de ordem superior após cada iteração, permanecendo inalterado o termo de menor potência na corrente média, validando assim o limite:

$$\lim_{\langle j \rangle \rightarrow 0} \frac{\langle j^n \rangle(t+1)}{\langle j^n \rangle(t)} = 1 \quad (3.63)$$

### 3.7.1 Tempo de relaxação

Podemos verificar o tempo no qual o sistema retorna ao seu valor estacionário após uma perturbação, que ocorre naturalmente na escolha aleatória dos canais a serem utilizados no processo de otimização local. Procedemos com uma perturbação na fração de canais vazios, equação 3.44:

$$B + \delta B(t+1) = B^2 + 2B\delta B(t) + (\delta B(t))^2 + \frac{1}{2} \left\{ 1 - (B(t))^2 - 2\langle j \rangle + \langle j^2 \rangle(t) \right\}, \quad (3.64)$$

subtraindo  $\delta B(t)$  de ambos os lados,

$$\begin{aligned} B + \frac{d(\delta B)}{dt} &= B^2 + 2B\delta B(t) + (\delta B(t))^2 - \delta B(t) + \\ &\quad + \frac{1}{2} \left\{ 1 - B^2 - 2B\delta B(t) - (\delta B(t))^2 - 2\langle j \rangle + \langle j^2 \rangle(t) \right\} \\ \frac{d(\delta B)}{dt} &= \left\{ \frac{(1-B)^2}{2} - 2\langle j \rangle \right\} - (1-B)\delta B(t). \end{aligned} \quad (3.65)$$

Para esta equação usamos apenas termos lineares em  $\delta B$  e  $\langle j \rangle$ , o termo entre chaves já vimos que é nulo, dada a equação 3.59. A forma assintótica para a perturbação no limite de tempo contínuo será portanto:

$$\frac{d(\delta B)}{dt} = -(1-B)\delta B(t) \quad (3.66)$$

esta equação tem por solução um decaimento exponencial, portanto o tempo de relaxação é dado por:

$$\tau = \frac{1}{1-B} = \frac{1}{2\sqrt{\langle j \rangle/\beta}} \quad (3.67)$$

### 3.7.2 Número arbitrário de canais de entrada e saída

O limite para baixas correntes revela uma peculiaridade desse tipo de problema que torna-se bastante geral com respeito à forma funcional da distribuição de canais utilizados e a corrente média. Primeiramente podemos analisar a equação 3.37 para a fração de canais vazios neste limite. Ao compararmos a distribuição de correntes críticas  $p_c(j_c)$  com a distribuição de correntes nos vértices  $Q(j)$  podemos usar uma suposição bastante forte, com base nas iterações numéricas, e tendo em vista que usamos neste problema apenas distribuições com primeiro e segundo momento bem definidos, que a integração do produto  $p_c(j)Q(j)$  possuem como contribuição apenas os primeiros fatores da expansão de  $p_c(j)$  em torno de zero. A equação fica, então:

$$1 - B = 2\sqrt{p_0 \langle j \rangle} \quad (3.68)$$

onde  $p_0 = p_c(0)$  é o valor da distribuição de fluxos críticos no ponto zero, desde que essa função tenha as características descritas anteriormente, possui uma forma de decaimento simétrico centrada em zero, caso a distribuição de custos seja uniforme esse valor é igual à  $\beta^{-1}$ .

A equação 3.68 mostra-se bastante geral independente da forma como correntes maiores que a corrente crítica se dividem, e também para um número arbitrário de canais, desde que o número de canais de entrada e saída sejam iguais. Essa generalização é imediata e decorre de:

$$B(t+1) = B(t)^m + \frac{(m-1)}{m} \int_0^\infty dj \mathcal{P}_c Q(t; j) \quad (3.69)$$

onde  $m$  é o número de canais de saída. Aqui utilizamos a regra da escolha do canal de menor custo caso a corrente de entrada seja inferior à corrente crítica. No limite de pequenas correntes recuperamos a equação 3.68 ao usarmos uma expansão também no valor de  $B \rightarrow 1 - b$ , onde  $b$

é uma pequena fração de canais usados, visto que  $B(t)$  é muito próximo a unidade neste limite:

$$\begin{aligned}
B &= B^m + \frac{m-1}{m} [1 - B^m - 2mp_0 \langle j \rangle] \\
&= \frac{1}{m} B^m + \frac{m-1}{m} - 2(m-1)p_0 \langle j \rangle \\
1-b &= \frac{1}{m} - b + \frac{(m-1)}{2} b^2 + \frac{m-1}{m} - 2(m-1)p_0 \langle j \rangle + \mathcal{O}(b^3, \langle j^2 \rangle) \\
\Rightarrow b &= 1 - B = 2\sqrt{p_0 \langle j \rangle}. \tag{3.70}
\end{aligned}$$

Usamos apenas termos até a segunda potência em  $b$  para a expansão binomial da penúltima linha da equação 3.70.

### 3.7.3 Correlação entre dois pontos

As flutuações no números de canais utilizados ao longo do escoamento dos fluxos podem ser determinadas a partir dos processos onde ocorrem alteração no número de sítios. Neste modelo podemos enumerar os processos de coalescência, o qual um sítio é reduzido, e ramificação, em que um filete de corrente gera dois ramos.

Para o processos de coalescência podemos verificar a probabilidade de ocorrência do processo de aglutinação de correntes  $\mathcal{P}_{merge}(j' + j'' \geq j_c)$ , ou seja, a coalescência ocorrerá quando a soma de duas correntes de entrada  $j'$  e  $j''$  for maior que a corrente crítica definida nos canais. Sem perda de generalidade usaremos uma distribuição exponencial de correntes críticas com parâmetro de escala  $\gamma = 1$ , o que leva à:

$$\begin{aligned}
\mathcal{P}_{merge}(j' + j'' \leq j_c) &= \int_0^\infty dj_c e^{-j_c} \int_0^{j_c} dj \int_0^j du P(u) P(j-u) \\
&\simeq \frac{\langle j \rangle}{(1 + \sqrt{\langle j \rangle})^2} \sim \langle j \rangle. \tag{3.71}
\end{aligned}$$

onde a convolução representa a soma das correntes de entrada distribuídas segundo a equação 3.60 para  $P(j)$ . Por simplicidade, omitimos a constante de normalização que, por sua vez, possui dependência com a distribuição de correntes críticas escolhida.

Usando o mesmo procedimento para a ramificação  $\mathcal{P}_{bran.}(j > j_c)$ , se a corrente proveniente de um único canal é alta o suficiente para se dividir em dois ramos, a fração de canais que

participam desse processo será:

$$\begin{aligned} \mathcal{P}_{bran.}(j \geq j_c) &= \int_0^\infty dj_c e^{-j_c} \int_{j_c}^\infty dj A e^{-\frac{\alpha j}{\sqrt{j}}} \\ &\simeq \frac{\langle j \rangle}{1 + \sqrt{\langle j \rangle}} \sim \langle j \rangle \end{aligned} \quad (3.72)$$

Os demais processos que ocorrem, difusão e aglutinação seguida de ramificação, podem ser computados usando o mesmo protocolo, porém nestes não há flutuação do número de partículas. Assim, o número de canais utilizados se mantém devido ao processo de difusão da corrente na rede, que relaciona-se à integral da função estacionária de um canal  $P(j)$  nos valores de corrente menores que as correntes críticas possíveis. No limite de pequenas correntes isso coincide com a integral de  $P(j)$  ponderada pela distribuição de correntes críticas que, como vimos na equação 3.59, é o fator que dá origem ao termo de expoente  $1/2$ , na fração de canais usados  $1 - B$ .

## 3.8 Conclusões

Estudamos um modelo de otimização de fluxos que proporciona interações entre sítios vizinhos muito semelhantes a um modelo de reações entre partículas de espécie única, envolvendo os processos de difusão na rede, coalescência de sítios e ramificação. Ao estudarmos a solução exata proposta por Alimohammadi, Khorrami e Aghamohammadi (2001) para o caso de pequenas taxas de criação, ou ramificação ( $\gamma \rightarrow 0$ ), observamos uma correspondência direta entre essa taxa e a razão da corrente média pela largura da distribuição de custos definida em nosso modelo de otimização

$$\frac{\langle j \rangle}{\beta} \rightarrow \gamma^2$$

dados os comportamentos limites estudados nos dois modelos.

O expoente  $1/2$  da densidade de canais usados com a corrente média é esperado, uma vez que utilizamos um conjunto de recorrência apenas para todos os sítios da rede, como um campo médio das interações. Tal expoente é confirmado em nossas simulações. A justificativa é que como as correntes podem ser misturadas no processo de aglutinação num vértice, ao final da

simulação todos os sítios terão contribuído, em média, com as mesmas interações para chegar ao estado final.

O formalismo de campo médio só permite uma solução analítica para o limite de altas correntes. Neste limite ainda observamos que a distribuição de correntes nos canais se parece muito com uma gaussiana e as flutuações de corrente dependem apenas da largura da distribuição de custos nos canais. Já para o limite de baixas correntes, a razão da corrente pela largura da distribuição de custos governa as flutuações do número de canais usados. Temos que na condição de uso total dos canais, apenas os custos são responsáveis pela desordem no sistema, observada nas flutuações de corrente. Já no uso escasso dos canais, essa desordem introduzida influencia na flutuação de correntes juntamente com a corrente média, fazendo com que a flutuação do número de canais usados seja definida pela razão dessas últimas.

O modelo aqui desenvolvido redefine as taxas de ramificação e difusão de partículas de espécie única sujeitas a esse tipo de interação entre primeiros vizinhos numa única dimensão. Dessa forma a taxa de ramificação  $R$  será uma função dos recursos injetados na rede, representados pela corrente média nos canais ( $\langle j \rangle$ ) ou sítios ( $2\langle j \rangle$ ), i. e.  $R = f(\langle j \rangle)$ , e a taxa de difusão  $D$  dada pela estrutura intrínseca ao meio, aqui representada pela largura da distribuição de custos nos vértices  $\beta$ , logo  $D = f(\beta)$ . No caso de grandes fluxos todos os processos de reação se misturam restando apenas flutuações de corrente e não de sítios, muito próximas a flutuações definidas por uma distribuição normal. No problema real de tráfego, a condição de altos fluxos é a mais usual, uma interpretação para esse modelo é que durante o fluxo normal (sem interrupções) as flutuações do mesmo se dará pelo custo intrínseco a cada via, visto que a variância na distribuição de fluxos é uma função apenas da distribuição de custos.

## ***4 Formação de estruturas modulares na rede de dependência do Debian GNU/Linux***

Neste capítulo apresentaremos uma análise por meio do conceito de rede para o sistema operacional de código-livre Debian GNU/Linux. Iremos verificar a relação entre os elementos constituintes desse sistema, i. e., os programas responsáveis pelo acesso às diversas funções da máquina que são conjuntos de algoritmos empacotados num único arquivo a fim de facilitar a instalação dos mesmos.

Tais “pacotes” possuem relações de interdependência, o que permite a representação em forma de grafo desse sistema. Dessa forma observamos propriedades gerais como a distribuição de graus desse sistema, bem como os agrupamentos que emergem dessa rede e sua relação com os agrupamentos pré-estabelecidos pela comunidade de desenvolvimento do Debian.

### **4.1 Rede de dependências entre pacotes Debian**

Podemos olhar para a rede de dependências como sendo um agrupamento de sítios, definidos como um conjunto de linhas de código empacotadas para, após serem compiladas, exercer alguma funcionalidade no sistema, os chamados pacotes. Estes estão conectados por meio de ligações direcionadas conforme a dependência de bibliotecas ou funções específicas entre pacotes: caso seja necessária para instalação do pacote  $j$  a instalação prévia de  $i$ , então temos uma ligação  $e_{ji}$  (sentido de  $j$  para  $i$ ). Um exemplo é que diversos pacotes podem depender do pacote *iceweasel*, navegador de páginas *Web*, porém não há necessidade dessa dependência ser direta,

um ou mais pacotes intermediários podem existir no caminho, veja figura 4.1<sup>1</sup>.

A rede pode ser montada por meio da leitura da saída do comando no *shell*<sup>2</sup>:

```
apt-cache show $(apt-cache pkgnames).
```

No presente trabalho usamos o arquivo que é lido por esse comando unindo os tipos *main*, *contrib* e *non-free*, isto é, a rede foi formada supondo uma instalação completa do Debian GNU/Linux 6.0, codenome *squeeze*. Tal arquivo pode ser encontrado num endereço de espelho de repositório do Debian, a exemplo do endereço `http://ftp.br.debian.org/debian/dists/squeeze/main/binary386/Packages.bz2`. O diretório *main* pode ser trocado por *contrib* ou *non-free* a fim de obter os dados que designam os pacotes de acordo com a licença que rege o direito de uso dos mesmos: *main* é constituído por pacotes licenciados conforme a definição de *software livre* para o projeto Debian, *contrib* por pacotes livres que trazem dependências com códigos-fonte sem uma licença de livre uso, e *non-free* os pacotes de uso e distribuição restritas.

Dada a definição acima temos analisado dados obtidos em julho de 2011, referente à versão 6.0.0, que corresponde ao primeiro lançamento da então versão estável do Debian – aos quatro dias de maio de 2013 foi lançado a versão estável Debian 7.0 *wheezy*. São portanto  $n = 26391$  vértices (pacotes) e  $m = 121992$  ligações direcionadas (relações de dependência entre pacotes). O que resulta uma componente gigante<sup>3</sup> com  $n_g = 26147$  sítios e  $m_g = 121833$  ligações.

Uma vez feita a leitura da rede, podemos verificar grandezas de interesse, como a conectividade, o coeficiente de agrupamento (*clustering*), a centralidade que depende do menor caminho como a centralidade de entroncamento (*betweenness*), além de verificar a estrutura modular dessa rede, uma detecção de agrupamentos por afinidade de funções, isto é, comunidades.

---

<sup>1</sup> Ao longo desse capítulo, utilizaremos ilustrações de grafos, para isso o *software* empregado é o de código-livre escrito por Bastian, Heymann e Jacomy (2009), o *Gephi*.

<sup>2</sup> Terminal que oferece uma interface de entrada de dados para o usuário do sistema operacional.

<sup>3</sup> Componente gigante é o maior conjunto de sítios em que existe algum caminho passando por eles (DO-ROGVTSEV, 2010, página13).



### 4.2.1 Distribuição de conectividade

As principais definições de grau de conexões, ou conectividade, numa rede direcionada é o número de sítios primeiros vizinhos  $j$  de um vértice  $i$  que enviam uma ligação (grau de entrada) e a quantidade de primeiros vizinhos que recebem ligação de um dado sítio  $i$  (grau de saída). Portanto, dada a matriz adjacência  $\mathbf{A}$  cujos elementos  $a_{ij}$  são unitários caso exista ligação de  $j$  para  $i$  e nulos caso contrário, a conectividade de saída  $k_i^{out}$  será dada por:

$$k_i^{out} = \sum_{j \neq i} a_{ji} = (a_{1i} \ a_{2i} \dots \ a_{ni}) \mathbf{1}, \quad (4.1)$$

onde  $\mathbf{1}$  é o vetor coluna com  $n$  componentes  $(1 \ 1 \dots 1)^T$ . A conectividade de entrada  $k_i^{in}$  é definida portanto,

$$k_i^{in} = \sum_{j \neq i} a_{ij} = (a_{i1} \ a_{i2} \dots \ a_{in}) \mathbf{1}. \quad (4.2)$$

Definimos assim a distribuição de conectividade  $p(k)$  como sendo a frequência em que uma dada conectividade de valor  $k$  ocorre na rede, e sua respectiva distribuição acumulada  $P(k)$  na qual o grau de um sítio na rede não pode ser menor que  $k$ :

$$P(k) = \sum_{j=0}^k p(j) \quad (4.3)$$

A rede estudada apresenta distribuição de graus de saída com um comportamento em lei de potência,  $p(k_{out}) \sim k_{out}^{-(\alpha+1)}$ , isto é, neste tipo de sistema existe uma pequena quantidade de pacotes que são fundamentais para praticamente todos os pacotes finais do sistema. Em contrapartida a conectividade de entrada segue um decaimento exponencial,  $p(k_{in}) \sim e^{-\beta k_{in}}$ , que fornece uma escala característica para o número médio de pacotes que recebem dependências de pacotes mais fundamentais dada pelo inverso do coeficiente de decaimento  $\beta \simeq 0,15$ . Um exemplo concreto é o pacote *libc6* que tem um grau de saída da ordem do número de pacotes no sistema,  $k(\text{libc6}) \sim 10^4$ , enquanto que para os pacotes finais, nunca recebem mais que  $10^2$  dependências. A figura 4.2 mostra tal comportamento, usando a distribuição cumulativa de conectividade.

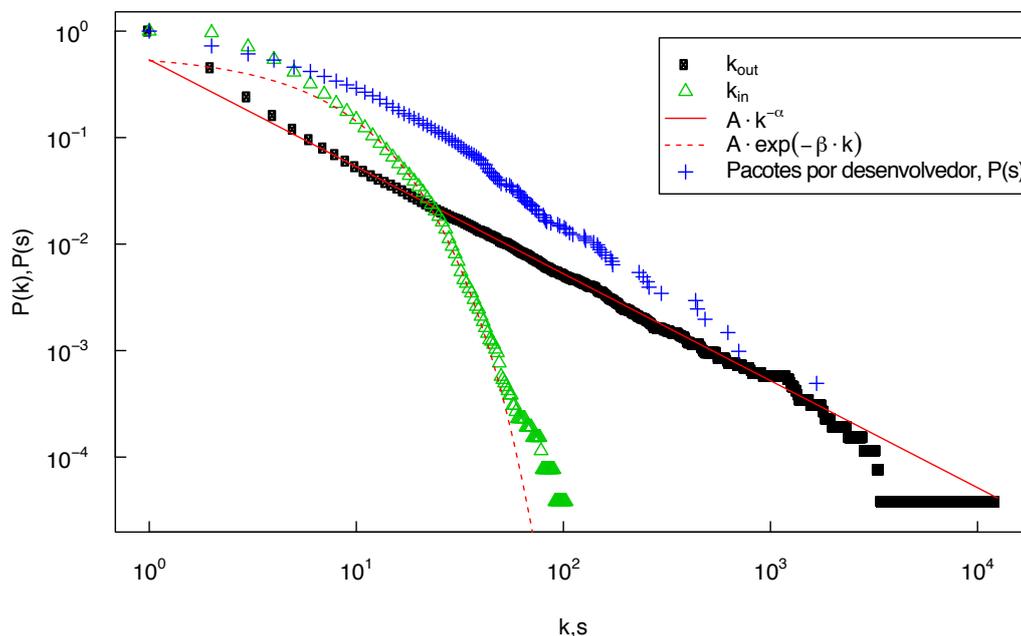


Figura 4.2: Distribuição cumulativa de conectividade  $P(k)$  da rede de dependências entre pacotes Debian *squeeze*. A linha contínua corresponde a lei de potência com  $\alpha = 1$  e a linha tracejada ao coeficiente de decaimento  $\beta = 0,15$ ; os erros para cada coeficiente estão estimados em  $6,8 \cdot 10^{-4}$  e  $1,4 \cdot 10^{-3}$  respectivamente, conforme ajuste. A distribuição de tamanhos de *clusters* entre pacotes de mesmo colaborador  $P(s)$  são os pontos em cruz.

Podemos observar também que a distribuição de pacotes por colaboradores se aproxima mais da distribuição de graus de saída, tendo em vista a tendência do desenvolvimento ocorrer dos pacotes mais fundamentais aos pacotes para usuários finais.

#### 4.2.2 Menor caminho médio e centralidade de entroncamento, *betweenness*

Definimos o menor caminho  $\ell_{ij}$  entre dois vértices  $i$  e  $j$  como sendo o menor número de passos entre ligações que chegam de  $i$  até  $j$ . Assim o menor caminho médio,  $\bar{\ell}$ , é a média sobre todos os pares  $(i, j)$  da rede os quais existe ao menos um caminho conectando os mesmos (DOROGVTSEV, 2010).

A distância entre vértices da rede pode ser caracterizada tanto pelo menor caminho médio quanto pelo diâmetro da rede  $\ell_D$  que é a maior distância geodésica da rede, i.e., o maior dos

menores caminhos  $\ell_{ij}$ . A componente gigante da rede de dependências entre pacotes Debian extraída neste trabalho possui respectivamente  $\bar{\ell} = 3,6$  e  $\ell_D = 13$ .

O menor caminho – ou caminho geodésico – também é usado para medir o potencial comunicativo de um dado sítio, em outras palavras, a razão do número caminhos entre dois sítios  $i$  e  $j$  que passam por um dado sítio  $k$ ,  $s(i, k, j)$ , e o total de caminhos geodésicos entre os dois primeiros sítios,  $s(i, j)$ . Este potencial comunicativo acumulado nos sítios da rede define a centralidade de entroncamento, ou *betweenness*. Esse tipo de centralidade mostra a frequência com que menores caminhos passam por um dado nó (FREEMAN, 1977), ou seja:

$$C_B(k) = \sum_{i,j \neq k} \frac{s(i, k, j)}{s(i, j)} \quad (4.4)$$

É perceptível para a rede em questão que ela apresenta também a distribuição dessa centralidade como uma lei de potência  $p(C_B) \sim C_B^{-(1+\gamma)}$ , a figura 4.3 mostra bem esse comportamento com a distribuição acumulada, com  $\gamma = 0,74$ .

Esse comportamento sugere uma relação de escala do tipo

$$C_B \sim k^\beta,$$

i.e., ao assumirmos que a quantidade de menores caminhos possui uma relação de escala com a conectividade do sítio, facilmente obtemos a relação para o expoente  $\beta$  (VÁZQUEZ; SATORRAS; VESPIGNANI, 2002). Dado que a relação de escala permanece válida para as distribuições acumuladas

$$P(C_B) \sim C_B^\gamma$$

e

$$P(k) \sim k^\alpha.$$

Portanto, com a relação de escala do *betweenness*, temos  $P(k) \sim k^{\gamma\beta} \sim k^\alpha$ , o que nos fornece,

$$\beta = \frac{\alpha}{\gamma}. \quad (4.5)$$

De fato o sistema em questão apresenta tal comportamento geral, como visto na figura 4.4.

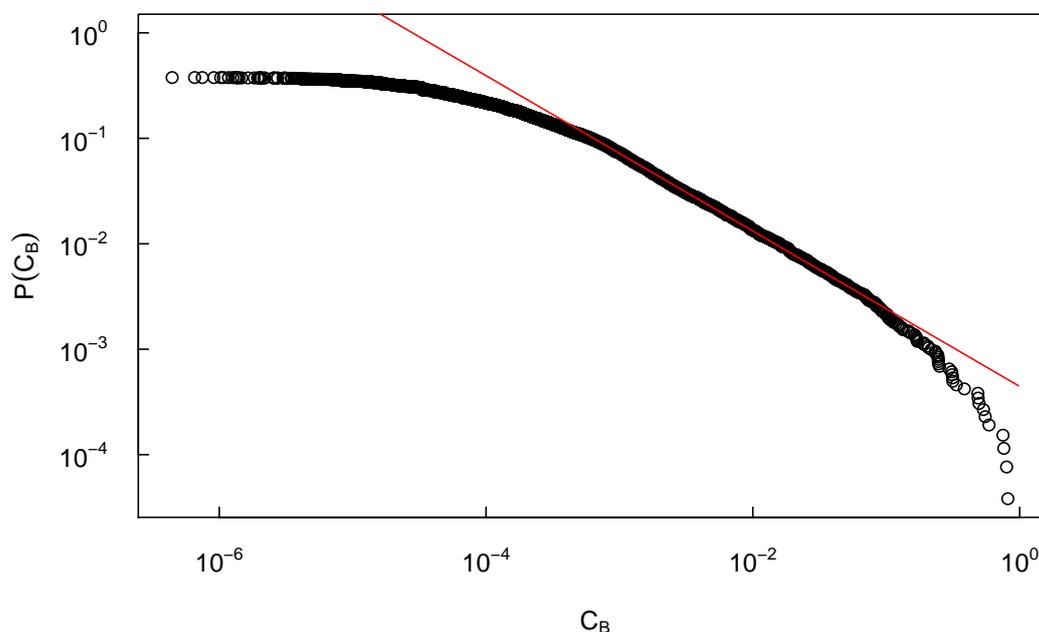


Figura 4.3: Distribuição cumulativa de *betweenness*  $P(C_B)$  da rede de dependências entre pacotes Debian *squeeze*. A linha contínua corresponde a lei de potência com  $\gamma = 0,74$ . O erro para o valor do expoente foi estimado em  $5,6 \cdot 10^{-4}$ , conforme ajuste. O *betweenness* foi normalizado pela quantidade de ligações na rede.

Indicando que a rede de dependência possui diferentes partes conectadas por menores caminhos que não passam pelo vértice central (BARTHELEMY, 2003) – vértice de maior conectividade, o pacote *libc6* no sistema em questão. Para esses vértices a quantidade de menores caminhos é da ordem do número de ligações existentes nessa região

$$C_B(k) \sim k(k-1) \sim k^2,$$

o que não é refletido no Debian Squeeze,  $\beta = 1,3 < 2$ .

### 4.2.3 Coeficiente de agrupamento

Uma característica não-local dos sítios da rede é o coeficiente de agrupamento, uma medida do quão conectados são os vizinhos de um sítio entre si. Ao contrário do menor caminho médio ou do *betweenness*, associado a extensão das cadeias de sítios na rede, essa grandeza

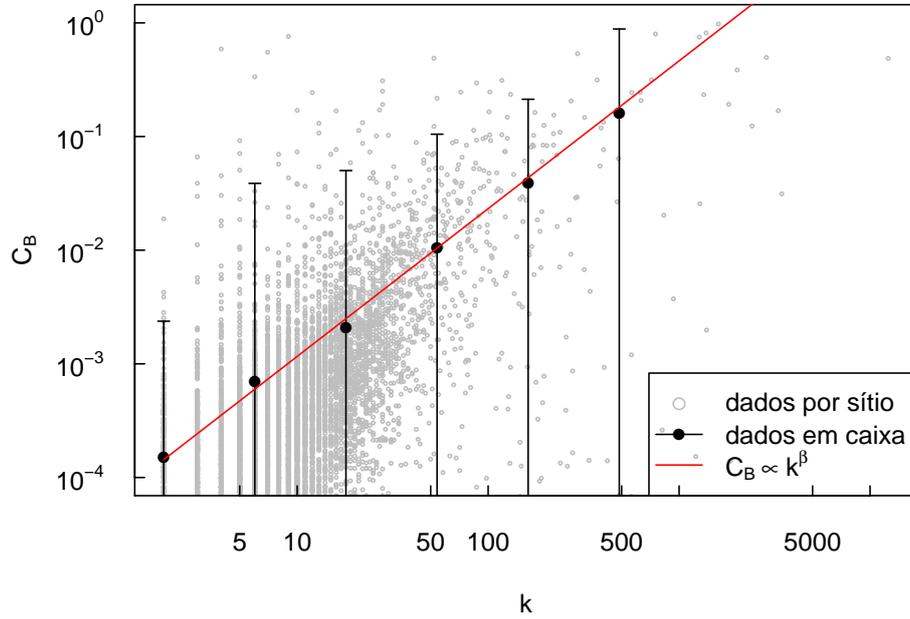


Figura 4.4: Centralidade de entrocamento acumulada nos sítios em função da conectividade. A linha contínua apresenta a lei de potência com expoente  $\beta_{fit} = 1,3$  e erro estimado de  $9,6 \cdot 10^{-3}$ . Esse apresenta uma boa concordância com a equação 4.5 usando os expoentes obtidos nas figuras 4.2 e 4.3,  $\beta = 1/0.74 \simeq 1,4$ .

esta associada aos pequenos grupos que podem ser formados ao longo dessas cadeias. Numa rede não direcionada podemos definir o coeficiente de agrupamento de um sítio  $i$  por:

$$c_i(A) = \frac{\frac{1}{2} \sum_{j \neq i} \sum_{k \neq (i,j)} A_{ij} A_{ik} A_{jk}}{\frac{1}{2} k_i (k_i - 1)} = \frac{(\mathbf{A})_{ii}}{k_i (k_i - 1)}, \quad (4.6)$$

que é a razão de triângulos presentes no subgrafo formado pelo sítio  $i$  e seus  $k_i$  vizinhos pela quantidade de triângulos possíveis deste subgrafo, i.e., a quantidade de agrupamentos possíveis entre dois vizinhos  $\binom{k_i}{2} = \frac{k_i(k_i-1)}{2}$ .

Para redes direcionadas podemos definir ainda agrupamentos com base no sentido das ligações que constituem triângulos entre o sítio  $i$  e seus dois vizinhos. Esses agrupamentos são chamados de: (i) ciclo, agrupamentos que definem uma relação cíclica entre os sítios do triângulo ( $i \rightarrow j \rightarrow k \rightarrow i$ ); (ii) entrada, na qual o sítio  $i$  recebe ligação de seus vizinhos; (iii) saída, os vizinhos do sítio  $i$  recebem uma ligação deste cada; (iv) *middleman*, o sítio  $i$  é o ponto médio no caminho que sai de  $j$  até  $k$  por exemplo (FAGIOLO, 2007). A figura 4.5 mostra tais

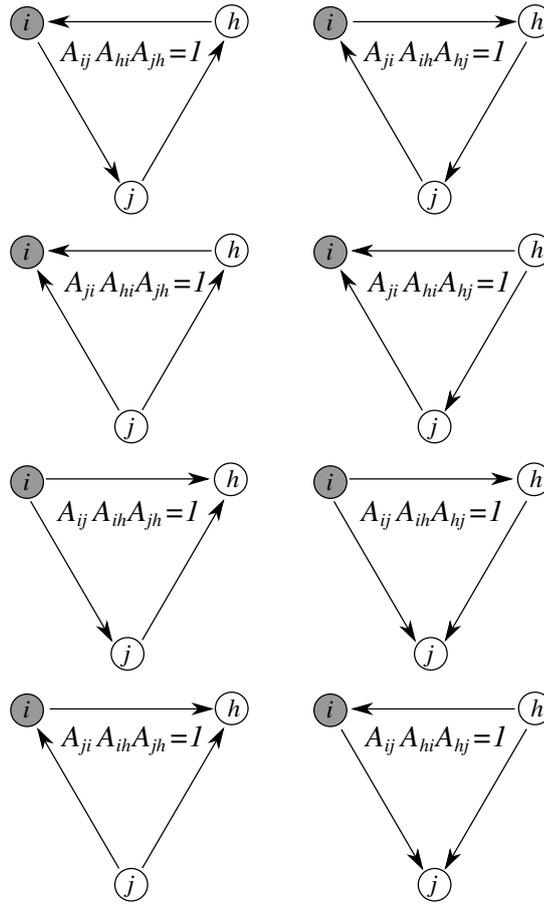


Figura 4.5: Tipos de agrupamento em torno de um sítio  $i$ , de cima para baixo as duas possibilidades de cada um: ciclo, entrada, saída e *middleman*. Ilustração construída com base no trabalho do Fagiolo (2007).

definições explicitamente. Ou para uma rede direcionada definida pela matriz de adjacências  $\mathbf{A}$  temos:

$$\begin{aligned}
 c_i^{cyc} &= \frac{(\mathbf{A}^3)_{ii}}{k_i^{\leftarrow} k_i^{\rightarrow} - k_i^{\leftrightarrow}}, \\
 c_i^{mid} &= \frac{(\mathbf{A}\mathbf{A}^T\mathbf{A})_{ii}}{k_i^{\leftarrow} k_i^{\rightarrow} - k_i^{\leftrightarrow}}, \\
 c_i^{in} &= \frac{(\mathbf{A}^T\mathbf{A}^2)_{ii}}{k_i^{\leftarrow} (k_i^{\leftarrow} - 1)} \mathbf{e} \\
 c_i^{out} &= \frac{(\mathbf{A}^2\mathbf{A}^T)_{ii}}{k_i^{\rightarrow} (k_i^{\rightarrow} - 1)},
 \end{aligned} \tag{4.7}$$

onde a seta sobrescrita defini a conectividade de entrada ( $\leftarrow$ ), saída ( $\rightarrow$ ) e os dois tipos sobrepostos ( $\leftrightarrow$ ).

O coeficiente pode ser acumulado para a vizinhança de todos os sítios da rede e um valor

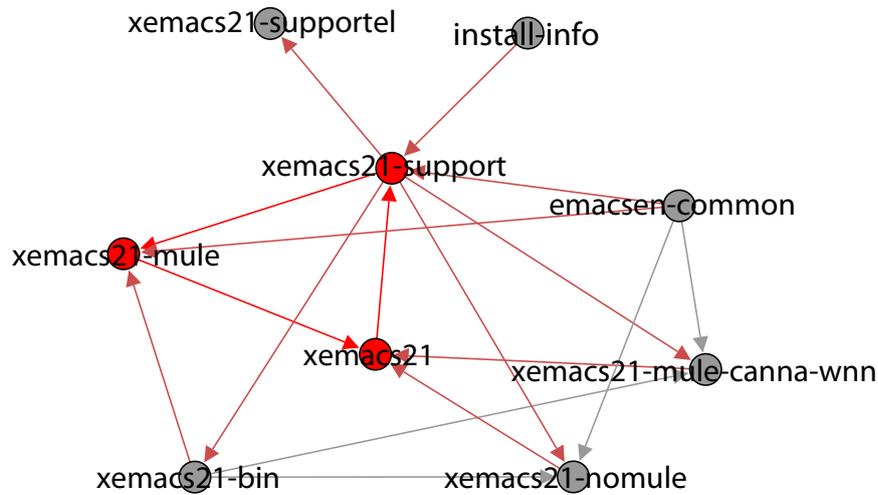


Figura 4.6: Exemplo de ciclo. Os pacotes em vermelhos somente poderão ser instalados sem levar em consideração a relação de dependências entre o ciclo.

médio é definido para toda a rede:

$$c^{tipo} = \frac{1}{n} \sum_i^n c_i^{tipo}. \quad (4.8)$$

Esse valor pode ser comparado com o valor desse coeficiente caso a rede tivesse suas ligações distribuídas ao acaso, i. e., uma rede que está contida no *ensemble* dos grafos aleatórios. O coeficiente de agrupamento no caso de uma rede aleatória,  $c^{ER}$  em homenagem ao modelo para grafos aleatório definido em (ERDÖS; RÉNYI, 1959), coincide com a densidade da rede:

$$c^{ER} = \rho = \frac{m}{n(n-1)}, \quad (4.9)$$

onde  $m$  e  $n$  são os números de ligações e sítios respectivamente.

Para a rede de pacotes Debian podemos perceber que a formação de agrupamentos do tipo ciclo,  $c^{cyc} = 4,58 \cdot 10^{-5}$ , constitui um defeito na relação de dependências, dada a impossibilidade de instalação de um pacote do ciclo se for exigida a instalação prévia de sua dependência. A figura 4.6 mostra um exemplo de ciclo para a instalação dos pacotes *xemacs21*, *xemacs21-support* e *xemacs21-mule*. O valor pequeno desse coeficiente mostra que esse não é um erro muito recorrente, apenas 0,07% dos pacotes mostram esse tipo de relação de dependência.

Os tipos de agrupamento *out* e *middleman* são os agrupamentos predominantes,  $c^{out} = 0,193$  e  $c^{mid} = 0,148$  respectivamente, sendo que agrupamentos de entrada um são pouco mais escassos  $c^{in} = 1,40 \cdot 10^{-2}$ , i.e., se um pacote depende de dois outros para sua instalação é menos provável que estes possuam alguma relação de dependência entre si, o contrário porém ocorre com maior frequência, se dois pacotes dependem de um outro certamente eles poderão compartilhar alguma relação de dependência com mais facilidade. Isso mostra um viés na construção desse sistema, pois neste é verificado uma densidade de aglomerados superior a sua equivalente aleatória,  $c^{ER} = 1,78 \cdot 10^{-4}$ , podemos assim proceder com uma detecção mais aprofundada de padrões de ligações nessa rede, que será discutido nas seções que seguem.

### 4.3 Estrutura de comunidades ou modularidade

As estruturas em comunidades – agrupamentos com alta densidade de ligações entre sítios da mesma comunidade e baixa densidade entre sítios de agrupamentos distintos – podem ser classificadas por diversas técnicas: retirada iterativa de ligações ou sítios que possuem grande centralidade como a de entroncamento até que formem ilhas de grande densidade de ligações entre si (GIRVAN; NEWMAN, 2002); construção de matriz de vizinhança reescalando a matriz de adjacências para que tenha elementos ligados na ordem de primeiros vizinhos (ANDRADE; PINHO; LOBÃO, 2009), dentre outros. Usaremos aqui um dos mais importantes conceitos para essas estruturas, a modularidade, introduzida por Newman (2006), que discutiremos nas duas próximas subseções.

#### 4.3.1 Modularidade em redes não-direcionadas

Os agrupamentos que são formados ao longo da estrutura de uma rede complexa a princípio não são conhecidos, porém é razoável a suposição da existência dos mesmos: redes sociais, por exemplo, são compostas por grupos de pessoas que possuem alguma afinidade; também notamos na rede de pacotes que o coeficiente de agrupamento é três ordens maior que a densidade do grafo. Uma possível definição para tais agrupamentos é dada pelo conceito de modulari-

dade, i. e., conjunto de sítios muito conectados entre os sítios do mesmo módulo e com menos conexões entre módulos diferentes.

Esse conceito foi introduzido por Newman (2006), como sendo a diferença entre a estrutura dada – a rede a ser analisada definida pela matriz adjacência  $\mathbf{A}$  – e sua contraparte aleatória, i.e., dada a distribuição de conectividade da rede, podemos calcular a probabilidade de dois sítios escolhidos ao acaso ( $i$  e  $j$ ) estarem conectados

$$p(e_{ij}) = \frac{k_i k_j}{2m}.$$

Essa contra-parte que serve de base ao modelo configuracional (DOROGOVTSSEV; MENDES, 2003), permite a escrita da modularidade como sendo:

$$Q = \frac{1}{2m} \sum_{ij} \left\{ A_{ij} - \frac{k_i k_j}{2m} \right\} \delta_{c_i c_j}. \quad (4.10)$$

Aqui o  $\delta_{c_i c_j}$  faz com que a soma ocorra apenas nas partições existentes na rede, quando dois sítios possuem rótulos de partições  $c_i$  e  $c_j$  iguais. Essa é chamada de modularidade configuracional (dado o segundo termo advindo do modelo de mesmo nome) pois podemos ter uma modularidade menos utilizada medindo a distância entre a rede em questão e sua contraparte aleatória de acordo com uma distribuição de conectividade uniforme de média  $\langle k \rangle$ , i.e.,  $(\langle k \rangle / 2m)^2$ .

A idéia de modularidade surge naturalmente ao observarmos a estabilidade de uma caminhada aleatório dentro de um conjunto de partições  $\mathcal{P}$  existentes na rede (LAMBIOTTE; DELVENNE; BARAHONA, 2009). A estabilidade de uma divisão de comunidades é uma medida da autocovariância desse particionamento sob um processo de Markov, i.e., podemos quantificar a covariância de um particionamento dentro de uma escala temporal  $\tau$  por

$$\text{cov}(X_t, X_{t+\tau}) = \mathbb{E}[X_t X_{t+\tau}] - \mathbb{E}[X_t]^2,$$

onde  $X_t$  é uma variável aleatória associada ao processo, essa variável é estacionária mas não necessariamente markoviana, e  $\mathbb{E}$  denota o valor esperado para o particionamento. Para a divisão de um grafo em que as conexões entre comunidades são pouco frequentes  $X_t$  e  $X_{t+\tau}$  serão correlacionadas por um longo tempo (DELVENNE; YALIRAKI; BARAHONA, 2010). Essa medida

pode ser escrita mais explicitamente ao considerarmos uma caminhada aleatória no grafo, a equação dinâmica para a densidade  $p_{i,n}$  de caminhantes num num sítio  $i$  de um grafo ao final de  $n$  passos é dada por:

$$p_{i,n+1} = \sum_j \frac{A_{ij}}{k_j} p_{j,n}, \quad (4.11)$$

que possui solução estacionária dada por

$$p_i^* = \frac{k_i}{2m},$$

com a conectividade  $k_i = k_i^{in} = k_i^{out}$ , dada de forma análoga às equações 4.1 e 4.2. Dessa forma a probabilidade de um caminhante estar num estado estacionário na comunidade  $C$  é  $\mathbb{E}[X_t]_C = \sum_{j \in C} k_j / 2m$ , e a probabilidade de um caminhante da comunidade  $C$  após um passo saindo do estado estacionário é dado por  $\mathbb{E}[X_t X_{t+1}]_C = \sum_{i,j \in C} (A_{ij}/k_j) (k_j/2m)$ . Logo a covariância para uma particionamento  $\mathcal{P}$  do grafo é dado por:

$$\text{cov}(X_t, X_{t+1}) = \sum_{C \in \mathcal{P}} \sum_{i,j \in C} \left[ \frac{A_{ij}}{k_j} \frac{k_j}{2m} - \frac{k_i}{2m} \frac{k_j}{2m} \right] = Q. \quad (4.12)$$

Assim a modularidade dada pelo modelo configuracional surge naturalmente de uma propriedade dinâmica do processo de caminhada aleatória na rede, em outras palavras as regiões onde o caminhante ficou preso numa escala temporal de um passo ( $\tau = 1$ ) são as regiões definidas pela maximização da modularidade na equação 4.10.

### 4.3.2 Modularidade em redes direcionadas

No caso de redes direcionadas a modularidade pode ser generalizada, porém o conceito de modularidade por meio de propriedades dinâmicas do passeio aleatório na rede e definições estatísticas não são mais coincidentes, i.e., para redes não-direcionadas a ideia de fluxo de probabilidade e densidade entre *links* nas comunidades não são coincidentes. Para verificar isso podemos ver que a densidade estacionária de caminhantes nos sítios da rede é dada pelo autovetor dominante  $\pi$  da matriz de transferência de elementos  $A_{ij}/k_j^{out}$  e a estabilidade na escala

$\tau = 1$  é calculada de forma análoga à equação 4.12:

$$\text{cov}(X_t, X_{t+1}) = \sum_{C \in \mathcal{P}} \sum_{i, j \in C} \left[ \frac{A_{ij}}{k_j^{\text{out}}} \pi_j - \pi_i \pi_j \right]. \quad (4.13)$$

A existência de um estado estacionário  $\pi_i$  é garantida ao introduzir um teletransporte aleatório, i. e., um processo de destruição de um caminhante num sítio sem saída seguido da criação do caminhante num outro sítio escolhido ao acaso. Isso deixa a dinâmica novamente ergódica (DEL-VENNE; YALIRAKI; BARAHONA, 2010).

Outra maneira de definir a modularidade em redes direcionadas foi proposta primeiramente por Leicht e Newman (2008), de forma análoga à equação 4.10, dessa vez o termo do modelo configuracional é dado pelo produto das conectividades de saída e entrada:

$$Q = \frac{1}{m} \sum_{ij} \left\{ A_{ij} - \frac{k_i^{\text{in}} k_j^{\text{out}}}{m} \right\} \delta_{c_i c_j}, \quad (4.14)$$

a matriz adjacência não é mais simétrica, portanto o fator 2 no denominador não faz-se necessário dado que  $m = \sum_{i,j} A_{ij}$ .

A diferença entre as modularidades definidas pela equação 4.13 e pela 4.14 é que a primeira favorece comunidades em que existe um grande fluxos de probabilidade da dinâmica e a segunda favorece comunidades em que a densidade de ligações é maior entre sítios de um mesmo aglomerado do que entre sítios de aglomerados diferentes.

Neste trabalho iremos nos ater à definição de Newman para modularidade em redes direcionadas por conta das características da rede Debian na qual grupos de desenvolvedores constroem essa rede, não existindo uma motivação para uma interpretação dinâmica da mesma.

## 4.4 Comunidades na rede Debian

Nos projetos de *software* livre a idéia de comunidade surge naturalmente com a comunidade de colaboradores, neste trabalho, os colaboradores do projeto Debian GNU/Linux. Tais pessoas trabalham voluntariamente no desenvolvimento do sistema – entende-se por desenvolvimento tanto a criação do *software* quanto a manutenção do mesmo no Debian, adotando-se

apenas pacotes já criados por terceiros – dando para cada pacote uma assinatura do grupo de desenvolvimento ou do indivíduo, caso este não faça parte de um time, ou tal pacote não seja da responsabilidade do time, caso em que o desenvolvedor contribui com um determinado time Debian mas adota pacotes fora do escopo do referido grupo. O foco deste capítulo é a comparação entre a existência de comunidades dada uma organização humana definida pelos desenvolvedores Debian e as estruturas surgidas espontaneamente nesse processo que podem ser detectadas usando algoritmos de maximização da função de modularidade definida nas seções anteriores.

Agrupamentos de pacotes também são previamente definidos neste sistema como as categorias as quais cada pacote é caracterizado por sua funcionalidade, e.g. *Administration Utilities, Databases, Perl*, etc. que são categorias com pacotes responsáveis pelo gerenciamento do sistema, banco de dados e linguagem de interpretação *perl* respectivamente, bem como outras diversas categorias encontradas em <http://packages.debian.org/stable/>. Pela definição da equação 4.14, com a matriz de adjacência  $\mathbf{A}$  definida pela componente gigante da rede de dependências, obtemos um valor de modularidade para as categorias pré-definidas de  $Q_{pré} = 0,24$  referentes a 51 categorias, quando colorimos os rótulos  $c_i$  dos sítios conforme os dados da seção de cada pacote. Ao colorir os rótulos usando as assinaturas dos times (ou colaboradores) de desenvolvimento, por sua vez, temos um valor de modularidade muito próximo,  $Q_{dev} = 0,23$  referente a 2024 desenvolvedores ou times de desenvolvimento, o que nos fornece uma base de comparação dos valores calculados no decorrer desse trabalho.

## 4.5 Maximização da modularidade

O cálculo inicial entre seções de pacotes pré-definidas e por desenvolvedores mostra uma primeira dificuldade ao usar a modularidade para definir comunidades, dado que valores muito próximos de modularidade são definidos por números muito diferentes de partições. De fato o problema de busca pela partição que maximize a função de modularidade 4.10 (ou 4.14) é

conhecido ser um problema NP-completo<sup>4</sup>, o que implica que nenhum algoritmo pode resolver o problema num tempo polinomial quanto a entrada do problema, no caso de grafos o número de conexões, a menos que tal tipo de problema tenha um tempo polinomial determinístico  $P$ , logo  $NP = P$  (BRANDES *et al.*, 2008). Este é um problema em aberto da matemática, se  $NP \neq P$  ou  $NP = P$ . Assumindo a primeira assertiva, dado o não conhecimento de um algoritmo  $P$  para a maximização da modularidade, o que nos resta são algoritmos “espertos” de busca por uma solução ótima, a exemplo da busca gulosa no sentido do aumento da modularidade, usando a direção das ligações (BLONDEL *et al.*, 2008; BRANDES *et al.*, 2008) ou arrefecimento simulado (LIU; LIU, 2010) para determinar o crescimento da modularidade, ou métodos em que é calculado um gradiente para maximização da modularidade (LEICHT; NEWMAN, 2008), dentre outros.

### 4.5.1 Método espectral

Esse método para redes direcionadas foi proposto por Leicht e Newman (2008) juntamente com a definição da equação 4.14. Consiste em divisões sucessivas da rede em duas comunidades, após a primeira divisão basta considerar os subgrafos resultantes como problemas isolados de divisão em duas comunidades, tomando o cuidado de redefinir a função de modularidade para cada subgrafo.

O problema de duas comunidades torna-se razoavelmente fácil de ser resolvido dada a natureza da equação 4.14 em podemos proceder com a substituição  $\delta_{c_i c_j} \rightarrow \frac{1}{2}(s_i s_j + 1)$  em que os rótulos de comunidades  $c_i$ 's são representados por variáveis binárias  $s_i = \pm 1$  que distingue cada agrupamento pelo seu sinal. Definindo um vetor de estado  $\mathbf{s}$ , cujas componentes associam os sítios da rede a sua respectiva comunidade reduzimos a equação 4.14 a:

$$Q = \frac{1}{2m} \left[ \sum_{ij} B_{ij} s_i s_j + \sum_{ij} B_{ij} \right] = \frac{1}{2m} \mathbf{s}^T \mathbf{B} \mathbf{s}, \quad (4.15)$$

<sup>4</sup>NP é a sigla para *nondeterministic polynomial time* ou tempo polinomial não determinístico, que define problemas cuja solução pode ser encontrada e verificada em tempo polinomial por meio de algoritmos não determinísticos, i.e., um algoritmo que pode fornecer palpites corretos para a solução a cada passo (DASGUPTA; PAPADIMITRIOU; VAZIRANI, 2006), o termo “completo” vem do fato de que todo problema de busca do tipo NP pode ser reduzido a um só problema.

onde  $B_{ij}$  é tão chamada matriz de modularidade definida por

$$B_{ij} = A_{ij} - \frac{k_i^{in} k_j^{out}}{m},$$

que tem a propriedade  $\sum_{ij} B_{ij} = 0$ , e o número de sítios  $n$  e o número de ligações  $m$  da rede definem os vínculos existentes no problema:

$$\sum_{ij} s_i^2 = n \text{ e } \sum_{ij} A_{ij} = \sum_i k_i^{in} = \sum_j k_j^{out} = m. \quad (4.16)$$

Para facilitar a obtenção da máxima modularidade na primeira divisão podemos recuperar a simetria da matriz de modularidade ao somarmos a transposta e dividirmos por dois a equação 4.15:

$$Q = \frac{1}{4m} \mathbf{s}^T (\mathbf{B}^T + \mathbf{B}) \mathbf{s}. \quad (4.17)$$

Usando a mesma metodologia para obtenção da solução 3.4, no problema de otimização de fluxos, a fim de achar o extremo da função  $Q$ , um multiplicador de Lagrange  $\lambda$  para o vínculo do número de sítios, equação 4.16, pode ser empregado:

$$\Lambda(\mathbf{s}, \lambda) = \frac{1}{4m} \sum_{ij} (B_{ij} + B_{ji}) s_i s_j + \lambda \left( \sum_i s_i^2 - n \right), \quad (4.18)$$

e temos que o gradiente de  $\Lambda$  deve ser nulo num estado extremo de modularidade (máximos, mínimos ou pontos de sela em  $Q$ ), logo para uma componente  $s_k$  de  $\nabla \Lambda$  neste ponto,

$$\frac{\partial}{\partial s_k} \Lambda(\mathbf{s}, \lambda) = \sum_i (B_{ik} + B_{ki}) s_i + 8m\lambda s_k = 0,$$

que descreve toda componente de  $\mathbf{s}$ . Esse sistema pode ser identificado como uma equação de autovalor

$$(\mathbf{B}^T + \mathbf{B}) \mathbf{v} = \beta \mathbf{v}, \quad (4.19)$$

identificando  $\beta \equiv -4m\lambda$ . Portanto precisamos obter o autovetor  $\mathbf{v}$  referente ao maior autovalor positivo, porém nesse tipo de problema as componentes de  $\mathbf{v}$  são reais pertencentes ao intervalo  $[0, 1]$ . Ao fim desse processo de obtenção deste autovetor, podemos obter o vetor  $\mathbf{s}$  equivalente, de coordenadas  $s_k = \pm 1$ , mais paralelo possível; escolhendo  $s_k = 1$  se  $v_k > 0$  e  $s_k = -1$  se  $v_k < 0$ , caso a componente seja  $v_k = 0$  pode-se usar o critério de sorteio do valor de  $s_k$  ou um

critério “guloso” que escolha o valor de  $s_k$  que tenha maior variação de modularidade.

Ao contrário do que sugere o nome do método, não é necessária a tarefa de conhecer o espectro da matriz de modularidade para obter as divisões de comunidades, dado o gasto excessivo de uma diagonalização da matriz de modularidade. Para isso podemos usar um método iterativo, aplicando sucessivas vezes a matriz de modularidade a um vetor inicial e dividindo-o por sua norma esse tipo de iteração converge para o autovetor dominante:

$$\mathbf{v}_{t+1} = \frac{\mathbf{B}'\mathbf{v}_t}{\|\mathbf{B}'\mathbf{v}_t\|}.$$

Rigorosamente temos  $\lim_{t \rightarrow \infty} \mathbf{v}_t = \mathbf{v}$ , com  $\mathbf{v}$  solução da equação 4.19 para o autoutovalor de maior módulo, mas na prática não precisamos em geral um pouco mais de alguns milhares de iterações pra obter um resultado convergido.

Após a primeira divisão da rede podemos tratar cada módulo encontrado como outro problema, apenas generalizando a matriz de modularidade por uma nova matriz constituída por aquele subagrupamento ( $B_{ij} \rightarrow B^{(g)}_{ij}$ ). A variação da modularidade representada pela nova divisão daquele subgrafo  $g$ , em face a modularidade total da rede:

$$\begin{aligned} \Delta Q(g) &= Q_1(g) - Q_0(g) = \frac{1}{2m} \left[ \sum_{i,j \in g} (B_{ij} + B_{ji}) \frac{s_i s_j + 1}{2} - \sum_{i,j \in g} (B_{ij} + B_{ji}) \right] \\ &= \frac{1}{4m} \sum_{i,j \in g} [(B_{ij} + B_{ji}) s_i s_j - (B_{ij} + B_{ji}) s_i^2] \\ &= \frac{1}{4m} \sum_{i,j \in g} \left[ (B_{ij} + B_{ji}) - \delta_{ij} \sum_{k \in g} (B_{ik} + B_{ki}) \right] s_i s_j \\ &= \frac{1}{4m} \mathbf{s}^T \left( \mathbf{B}^{(g)T} + \mathbf{B}^{(g)} \right) \mathbf{s}, \end{aligned}$$

agora temos que obter o autovetor dominante para a matriz

$$B^{(g)}_{ij} = B_{ij} - \frac{1}{2} \delta_{ij} \sum_{k \in g} (B_{ik} + B_{ki}).$$

A vantagem em usar um método iterativo para encontrar o autovetor dominante pode converter-se numa desvantagem caso o autovalor dominante seja negativo, visto que o método iterativo escolhe o autovalor que tem a maior norma. Neste caso, um refinamento é exigido como uma

busca gulosa fornecendo como entrada o vetor mais paralelo ao autovetor dominante. As divisões acabam quando não for mais possível uma variação positiva da modularidade.

### 4.5.2 Arrefecimento simulado

A equação 4.14 é semelhante ao hamiltoniano do modelo de Potts em vidro de spin (ELDERFIELD; SHERRINGTON, 1983) sem campo, quando escrita na forma:

$$Q = \frac{1}{m} \sum_{ij} \left\{ A_{ij} - \frac{k_i^{in} k_j^{out}}{m} \right\} \delta_{c_i c_j} = -\frac{1}{2} \sum_{ij} J_{ij} \delta_{c_i c_j} = -H(\mathbf{c}), \quad (4.20)$$

onde identificamos a matriz de modularidade como a matriz de troca do vidro de spin, e o vetor  $\mathbf{c} = (c_1, c_2, \dots, c_n)^T$  é o vetor  $q$  estados, i.e, comunidades.

Podemos resolver o problema de maximização da modularidade (ou minimização da função  $H = -Q$ ) com uma dinâmica de Metropolis, que consiste na transição microscópica entre estados vizinhos com probabilidade nula de transição caso tais estados não pertençam a vizinhança e taxa finita para transições vizinhas (TOMÉ; OLIVEIRA, 2001), asseguramos a existência de um estado estacionário definindo a matriz estocástica entre dois estados vizinhos  $\mathbf{c}$  e  $\mathbf{c}'$ :

$$T(\mathbf{c}', \mathbf{c}) = \begin{cases} \frac{1}{q} \exp \{-\beta [H(\mathbf{c}') - H(\mathbf{c})]\}, & \text{se } H(\mathbf{c}') > H(\mathbf{c}) \\ \frac{1}{q}, & \text{se } H(\mathbf{c}') < H(\mathbf{c}) \end{cases}, \quad (4.21)$$

onde  $\beta$  é um parâmetro que depende do inverso da temperatura, o conceito de temperatura num problema de maximização da modularidade implicará na regulação da liberdade dos sítios mudarem seu estado de comunidade conforme o aumento ou diminuição desse parâmetro, limitando as mudanças para temperatura pequenas e permitindo muitas mudanças para altas temperaturas. De fato ao observarmos a definição acima temos para temperaturas infinitas,  $\beta \rightarrow 0$ , aceitação igual para qualquer que seja o estado de maior ou menor modularidade e para temperaturas nulas,  $\beta \rightarrow \infty$ , apenas estados de mínima energia são aceitos. Para verificar que tal escolha obedece ao balanço detalhado, condição de reversibilidade microscópica do estado estacionário do sistema, podemos observar dois estados vizinhos  $c_1$  e  $c_1$  tal que  $H(\mathbf{c}_1) < H(\mathbf{c}_2)$ .

Assim a matriz estocástica 4.21 nos fornece para as transições entre esses estados

$$T(\mathbf{c}_1, \mathbf{c}_2) = \frac{1}{q} \text{ e } T(\mathbf{c}_2, \mathbf{c}_1) = \frac{1}{q} e^{-\beta[H(\mathbf{c}_2) - H(\mathbf{c}_1)]}$$

temos que a probabilidade estacionária do sistema estar em  $\mathbf{c}_1$  é dada por  $P(\mathbf{c}_1) = \frac{1}{Z} e^{-\beta H(\mathbf{c}_1)}$  e analogamente em  $\mathbf{c}_2$  será  $P(\mathbf{c}_2) = \frac{1}{Z} e^{-\beta H(\mathbf{c}_2)}$ . Para a transição  $\mathbf{c}'$  a partir do estado  $\mathbf{c}$  teremos

$$T(\mathbf{c}_2, \mathbf{c}_1)P(\mathbf{c}_1) = \frac{1}{q} e^{-\beta[H(\mathbf{c}_2) - H(\mathbf{c}_1)]} \frac{1}{Z} e^{-\beta H(\mathbf{c}_1)} = T(\mathbf{c}_1, \mathbf{c}_2)P(\mathbf{c}_2),$$

onde  $Z = \sum_{\mathbf{c}} e^{-\beta H(\mathbf{c})}$  é a função de partição do sistema.

Para obtenção do valor inicial de temperatura executamos a dinâmica de Metropolis com temperatura nula, i.e., os estados de maior e menor modularidade são igualmente permitidos, deixando o sistema convergir por um tempo num estado de modularidade maior que zero. Neste ponto é que o sistema poderá arrefecer com o controle de temperatura, a escolha neste trabalho é para um decaimento em lei de potência:

$$\beta(t) = \beta_0 \cdot t^\gamma \Rightarrow T(t) = T_0 t^{-\gamma}, \quad (4.22)$$

com  $T_0 = Q_0$  dada pelo valor da modularidade após uma breve aplicação da dinâmica de Metropolis a temperatura nula. A escolha do expoente de decaimento está ligada diretamente à quantidade de estados visitados a cada passo, visto que um decaimento mais rápido limitará rapidamente a mudança de estados desfavoráveis ao aumento da modularidade.

A vantagem no uso desse método está no cálculo da diferença de modularidade entre estados vizinhos que é feita a cada passo, dado que podemos fazer a mudança de estados de forma sequencial com um sítio sendo modificado a cada passo, a escolha da taxa de transição  $1/q$  é por conta do sorteio de cada um dos  $q$  estados que cada sítio pode assumir. Seja  $c'_i$  a nova comunidade sorteada para o sítio  $i$  e  $c_i$  o estado anterior deste sítio, o incremento da modularidade a cada mudança pode ser calculada com auxílio da equação 4.15:

$$\Delta Q_{c_i \rightarrow c'_i} = \frac{1}{2m} \left[ \sum_j (B_{ij} + B_{ji}) \delta_{c'_i c_j} - (B_{ij} + B_{ji}) \delta_{c_i c_j} \right], \quad (4.23)$$

em outras palavras, somam-se todos os termos da função de modularidade do novo sítio na

comunidade  $c'$  e subtraem-se os termos que contribuíam com esse sítio na comunidade antiga.

### 4.5.3 Algoritmo guloso

Originalmente proposto por Blondel *et al.* (2008) para redes não-direcionadas, esse algoritmo consiste em seguir um caminho pré-estabelecido para o crescimento da modularidade. Isso o torna muito mais rápido que um algoritmo determinístico como o método espectral que possui uma etapa de cálculo do maior autovetor, e não-determinístico com o estado de maior modularidade sendo visitado após longos passos de Monte Carlo. A receita proposta por Blondel *et al.* (2008) escolhe como caminho pré-estabelecido, os caminhos entre os vértices da rede: (i) iniciando-se com cada sítio pertencente a sua própria comunidade é verificado uma aglutinação dos sítios por meio do cálculo da variação da modularidade ao retirar o sítio de sua comunidade e acrescentá-lo à comunidade de seu vizinho; (ii) após percorrer toda a rede unindo sítios que contribuem positivamente para a modularidade, caso aglomerados, é criada a rede seguinte com os vértices sendo as comunidades e as ligações ponderadas com a frequência de *links* entre comunidades distintas e as ligações entre comunidades consistem em autoconexões (*loops*) desses sítios derivados, dessa forma o passo (i) pode ser aplicado novamente até que nenhum incremento possa ser calculado.

A equação 4.23 é de grande valia nesse método uma vez que o incremento da modularidade é calculado sítio a sítio. Neste trabalho uma alteração desse método foi usada para que pudéssemos avaliar o crescimento da modularidade por meio da junção de times de desenvolvedores em novos grupos, ou indivíduos colaboradores aglutinando-se em novos grupos ou times existentes. Assim o estado inicial do algoritmo não é cada sítio definindo sua própria comunidade, mas sim as comunidades estabelecidas no projeto Debian.

A heurística escolhida é unir vértices na mesma comunidade que seu primeiro vizinho que envia ligações para os mesmos, para isso ordenando as ligações de forma decrescente quanto ao valor do peso destas. Apesar de extremamente rápido esse algoritmo não possui prova alguma de que o estado de maior modularidade será alcançado, mas é bastante útil para aglutinar



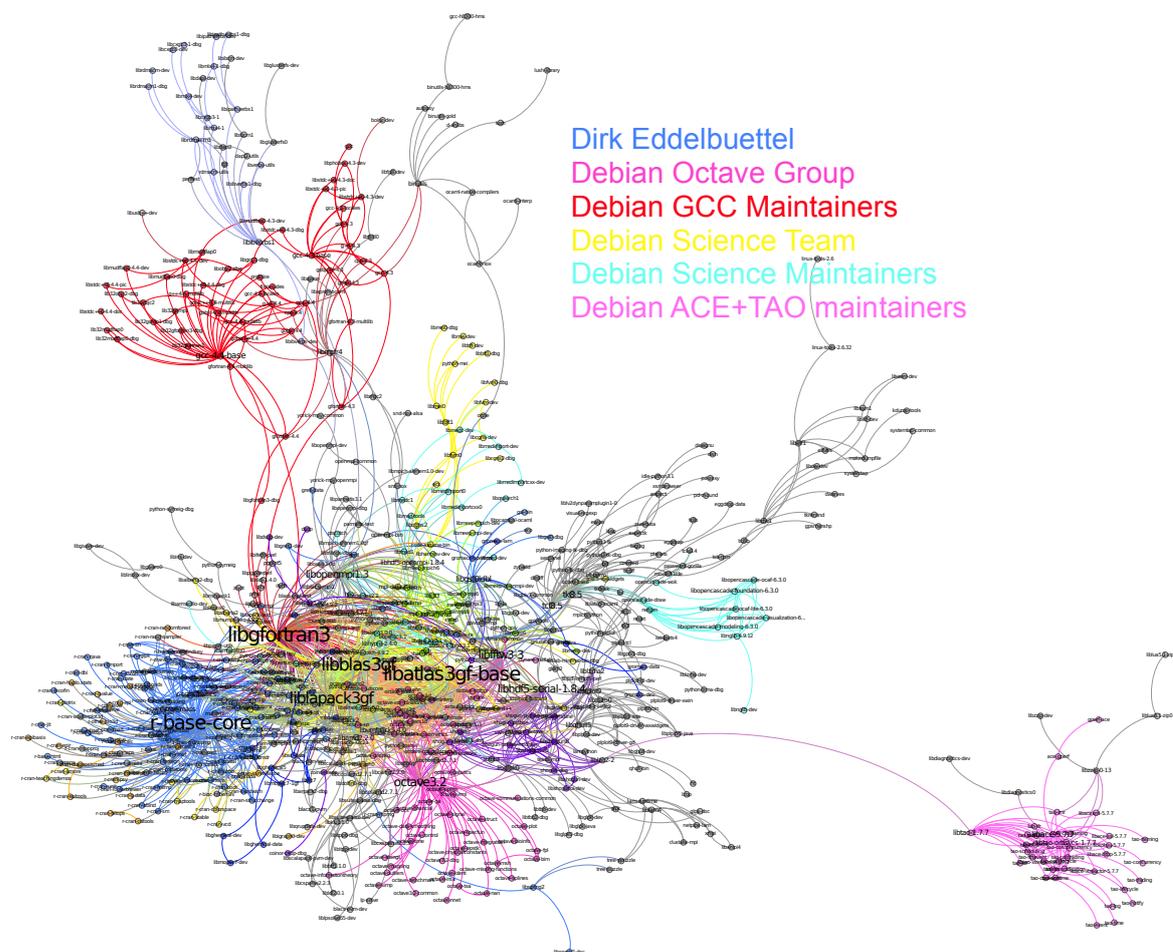


Figura 4.8: Subgrafo obtido com o método espectral, contém pacotes como o compilador de linguagem C (o *gcc*) e *software* de análise estatística (o *R*). As cores indicam os times de desenvolvedores contidos neste subgrafo. Na legenda constam os times ou colaborador que responsáveis por mais pacotes dentro desse módulo.

para as principais comunidades obtidas com o método espectral por meio de uma matriz de coincidência entre a distribuição de comunidades obtidas por meio do método espectral e a distribuição definida pelos times de colaboradores Debian, essa matriz é montada ordenando o índice dos módulos pela quantidade de pacotes dentro deles, de forma decrescente, e calculando a frequência de pacotes coincidentes em relação às comunidades definidas pelos times de colaboradores, a figura 4.9.

A figura 4.10 dá uma idéia ilustrativa do quanto a distribuição de pacotes por times de colaboradores não é feita por estruturas modulares, apesar de constituir subgrupos dentro dessas

estruturas como mostrado acima. Numericamente temos para as estruturas por colaboradores e pelo método espectral respectivamente,  $Q_{dev} \simeq 0,23$  e  $Q_{spec} \simeq 0,58$ . Isso se reflete numa coloração mais difusa entre os agrupamentos de organização humana e cores melhores distinguidas nas comunidades do método espectral.

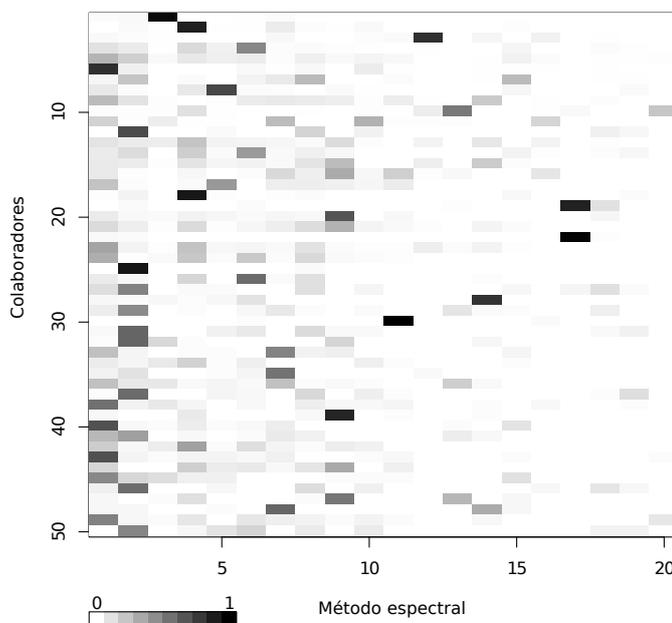


Figura 4.9: Matriz de frequências entre pacotes tanto nos módulos obtidos pelo método espectral (abcissas) quanto nos agrupamentos definidos pelos times de desenvolvedores (ordenadas), normalizadas pelo tamanho dos agrupamentos por colaboradores. Blocos mais próximos do preto indicam agrupamentos de pacotes dos colaboradores Debian praticamente inseridos numa única comunidade pelo método espectral.

A distribuição de tamanhos de comunidades  $p(s)$  mostra uma lei de potência para o estado de máxima modularidade para o método espectral com expoente menor que dois e para a distribuição de pacotes por colaboradores esse comportamento é mais acentuado, mostrando um expoente  $\gamma$  entre 2 e 3. Estudos anteriores sugerem que o estado de máxima modularidade é caracterizado por uma lei de potência na distribuição de tamanho das mesmas (ARENAS *et al.*, 2003; CLAUSET; NEWMAN; MOORE, 2004). De fato ao calcularmos melhor o estado de máxima modularidade usando o arrefecimento simulado (*simulated annealing*), obtemos  $Q_{sim} \simeq 0,59$  com 395 módulos, obtemos uma lei de potência mais acentuada. A figura 4.11 com as distribuições cumulativas de tamanhos nos sugerem para as distribuições de comunida-

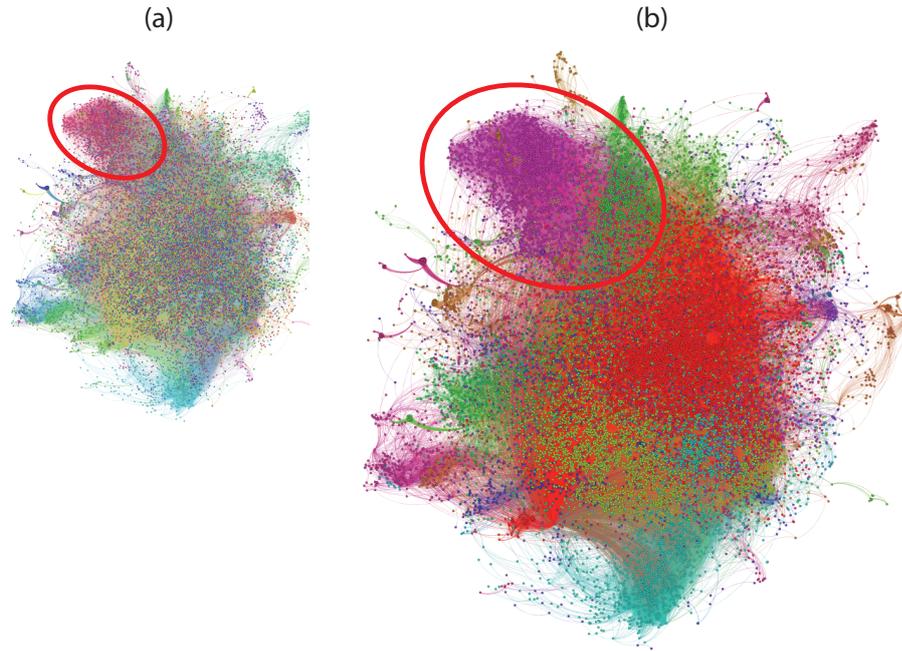


Figura 4.10: Componente gigante para a rede de dependência entre pacotes Debian 6.0,  $n = 26149$  pacotes e  $m = 121833$  ligações direcionadas. A componente em (a) é colorida pelos agrupamentos de pacotes por colaboradores,  $Q_{dev} \simeq 0,23$  e a mesma componente em (b) é colorida conforme o método espectral,  $Q_{spec} \simeq 0,58$ . A região em destaque é a comunidade dos pacotes mantidos pelo time *Perl* e seu módulo correspondente obtido com o método espectral.

des de pacotes por colaboradores, método espectral e por arrefecimento são respectivamente:

$$p_{dev}(s) \sim s^{-2,43}, p_{spec}(s) \sim s^{-1,29}, p_{sim}(s) \sim s^{-1,58} \quad (4.24)$$

Temos que para a distribuição de seções de pacotes de mesmo tipo, uma cauda exponencial é melhor ajustada com tamanho médio igual a 461 pacotes, o que implica que distribuições em lei de potência não são as únicas possíveis num estado de modularidade positiva, veja figura 4.11. Apesar do número de módulos no método espectral estar na mesma ordem do número de tipos de pacotes, a cauda da distribuição se estende por mais valores de números de pacotes, se afastando de uma distribuição com escala definida.

Para o processo de aglutinação de times e colaboradores Debian usando o algoritmo guloso, foi possível obter um acréscimo de  $\Delta Q \simeq 0,17$  na modularidade, partindo de um estado de modularidade  $Q_{dev} = 0,23$  com 2024 grupos ou colaboradores para um estado de  $Q'_{dev} = 0,40$  com

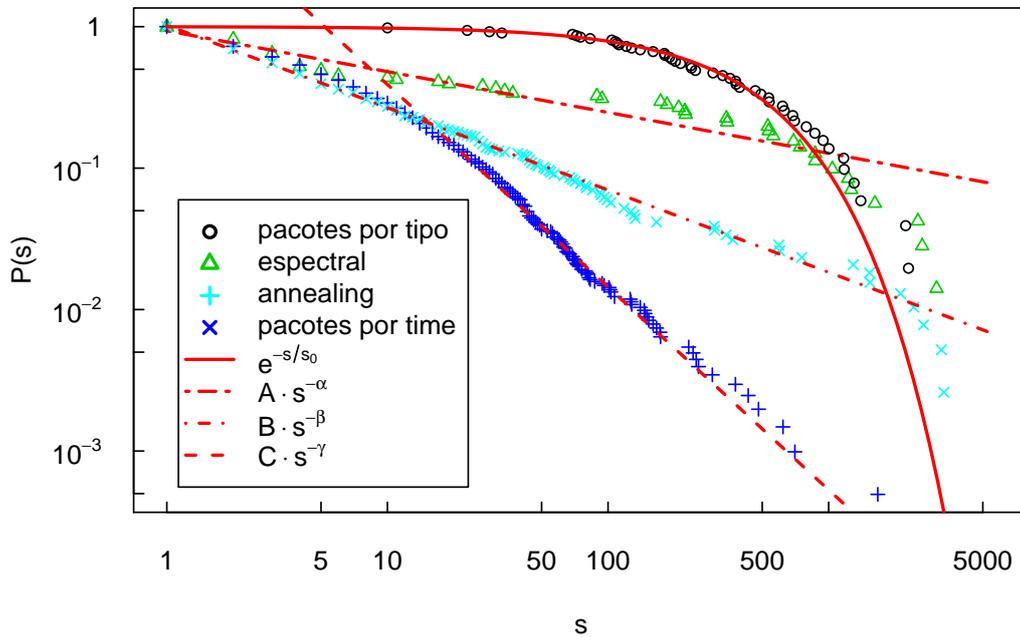


Figura 4.11: Distribuição de tamanhos de comunidades, os valores dos expoentes são:  $\alpha \simeq 0,289 \pm 0,013$ ;  $\beta \simeq 0,5821 \pm 0,0028$ ; e  $\gamma \simeq 1,433 \pm 0,014$ . Temos o coeficiente de decaimento exponencial dado por  $s_0 = \langle s \rangle = 461$ .

24 agrupamentos de pacotes obtidos pela união dos primeiros aglomerados. É possível notar que essa união ocorre de formas distintas para as duas maiores comunidades: a segunda maior constituída pelo *Debian Perl Team* que na figura 4.10 já é possível perceber que é uma comunidade que possui suas relações de dependência bastante densa nos pacotes que a constituem, dada que ela se mantém em quase toda sua totalidade ao maximizarmos a modularidade. No processo de aglutinação dos times de colaboradores ela atrai as demais comunidades formando o segundo maior módulo. Para o maior módulo temos como maior subgrupo o *Debian Python Modules Team* que corresponde apenas a pouco menos de 6% do novo subgrafo, i.e., o maior agrupamento não é formado em torno do maior grupo original, mas a partir de agrupamentos periféricos. A figura 4.6 fornece uma descrição quantitativa dessa junção.

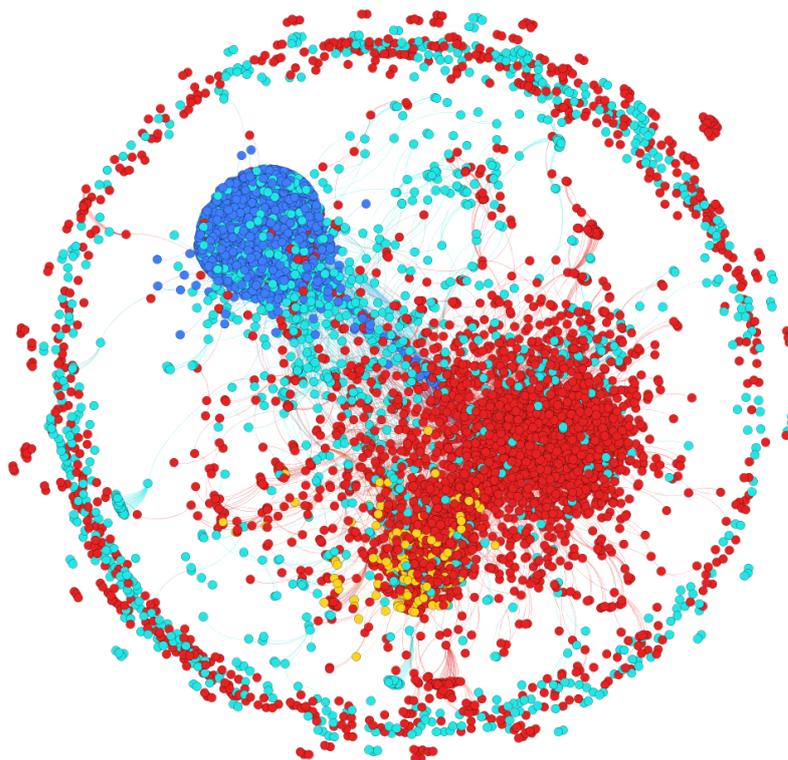


Figura 4.12: Subgrafo correspondente aos dois maiores aglomerados de time de colaboradores. Sítios de cores azuis correspondem ao segundo maior aglomerado de 213 grupos Debian com 3449 pacotes, em tons mais escuros os pacotes mantidos pelo *Debian Perl Group* (48,4% dos pacotes no grafo). Sítios de cor vermelha e amarela constituem o maior aglomerado de 341 times responsáveis por 4606 pacotes, o *Debian Python Modules Team* corresponde aos sítios amarelos (5,6% dos pacotes no grafo).

## 4.7 Conclusão

A otimização da modularidade não possui uma solução fechada nos dias atuais por constituir um problema *NP*-completo, porém podemos utilizar características existentes nos sistemas reais para atingir uma modularidade mais otimizada, dentro das limitações impostas pelo sistema. O Debian por exemplo é um sistema operacional livre em que uma comunidade de voluntários trabalha na manutenção e desenvolvimento do mesmo, logo podemos utilizar o conceito de modularidade para guiar os desenvolvedores independentes a agruparem-se para que a maior quantidade de informação entre dependências de pacotes fiquem no mesmo grupo.

Outras características estruturais, bem como o estudo da modularidade independentemente

da rede de desenvolvedores, foi realizada neste trabalho. A distribuição de tamanhos de comunidades apresenta uma concordância maior entre as distribuições de pacotes por tipo e pacotes divididos pelo método espectral (curvas das distribuições mais próximas), o que indica uma distribuição de dependências entre pacotes pertencentes a um determinado fim, porém uma característica do desenvolvimento fica evidente quando a distribuição pelo método de arrefecimento se aproxima mais da distribuição de tamanhos por desenvolvedor, evidenciando os indivíduos não ligados a um grupo em específico que adotam um simples pacote, introduzindo um caráter de escolha do pacote não correlacionada aos tipos de pacotes.

As grandezas fundamentais da rede de pacotes Debian mostram um comportamento livre de escala para a distribuição de conectividade de saída, isso reflete na distribuição de centralidade de entroncamento que também segue uma lei de potência antes de um decaimento exponencial, dessa forma pode-se inferir que pacotes fundamentais também são responsáveis pela interdependência entre agrupamentos distintos. A existência de ciclos de dependência ainda é presente na atual versão estável do Debian, porém está bem diluída dentro da rede, representando apenas 0,07% do total de pacotes, este corresponde a um *bug* que permite a instalação por meio de um comando para ignorar as dependências, numa instalação típica o usuário não irá passar por esse problema, o que facilita a ocultação desse defeito.

## 5 *Conclusão*

Duas abordagens foram realizadas nesse trabalho de tese. Primeiramente o estudo de um modelo aplicado a redes regulares, usando uma aproximação de campo médio, permitindo uma solução semi-analítica do modelo unidimensional estudado. A segunda abordagem uma análise de sistema real utilizando técnica de otimização da função de modularidade para obter as estruturas de interesse.

Nessas duas abordagens utilizamos o conceito de otimização, ora aplicado ao nosso modelo de fluxos por meio da noção de equilíbrio de tráfego (HELBING, 2001) ora aplicado a definição da função de modularidade (NEWMAN, 2006), tendo em vista que os sistemas reais tendem a obedecer alguma regra de otimização, normalmente uma função que representa a energia do sistema, nesse trabalho é estendido para o caso de otimização de um custo diretamente ligado ao tempo gasto entre diferentes rotas de tráfego. E no caso da análise da rede de pacotes, esse conceito de otimização foi usado para identificar agrupamentos na estrutura, propondo um caminho para a formação de novos agrupamentos, dados os grupos pré-estabelecidos.

O nosso modelo de tráfego apresenta uma particularidade ao coincidir no nível de simples partículas participando de reações ao longo de um espaço unidimensional com um modelo com solução analítica por meio da técnica da evolução de espaços vazios, equação 3.14. Ao visualizarmos tais reações ocorrendo por conta de fluxos otimizados numa rede com uma distribuição de custos por ligações a taxa de criação de partículas passa a ser a razão entre a corrente média pela largura da distribuição de custos. Isso fornece uma outra interpretação para os processos

de difusão e ramificação, o primeiro terá sua taxa definida pela distribuição de custos na rede,

$$D \propto \beta^{\frac{1}{2}},$$

e o segundo será definido pela distribuição de recursos no sistema, a corrente média

$$R \propto \langle j \rangle^{\frac{1}{2}}.$$

Para a rede de dependências é verificada uma abordagem para auxiliar na formação de grupos de desenvolvimento dentro de um grande projeto como o Debian GNU/Linux, visto que podemos utilizar o conceito de modularidade para aglutinar desenvolvedores em grupos a fim de maximizar o número de dependências entre pacotes de um mesmo grupo. O algoritmo guloso pode ser empregado a fim de oferecer uma rápida estimativa do acréscimo na modularidade caso haja fusões entre indivíduos de um projeto.

## *Referências Bibliográficas*

- ALIMOHAMMADI, M.; KHORRAMI, M.; AGHAMOHAMMADI, A. Exactly solvable models through the empty-interval method. *Physical Review E*, v. 64, p. 056116, 2001.
- ANDRADE, R. F. S.; PINHO, S. T. R.; LOBÃO, T. P. Identification of community structure in networks using higher order neighborhood concepts. *International Journal of Bifurcation and Chaos*, v. 19, n. 08, p. 2677–2685, 2009.
- ARENAS, A. *et al.* Community analysis in social networks. *The European Physical Journal B - Condensed Matter*, v. 38, n. 2, p. 373–380, 2003. ISSN 1434-6028.
- ASHCROFT, N.; MERMIN, D. *Solid state physics*. 1. ed. [S.l.]: Holt, Rinehart and Winston, 1976. ISBN 0030839939.
- BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 509–512, 1999. ISSN 1095-9203.
- BARTHELEMY, M. Comment on: 'universal behavior of load distribution in scale-free networks'. *Physical Review Letters*, v. 91, n. 18, p. 189803, 2003. ISSN 0031-9007.
- BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In: . [s.n.], 2009. Disponível em: <<http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>>.
- BLONDEL, V. *et al.* Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2008, n. 10, p. P10008, 2008. ISSN 1742-5468.
- BRANDES, U. *et al.* On modularity clustering. *Knowledge and Data Engineering, IEEE Transactions on*, v. 20, n. 2, p. 172–188, 2008. ISSN 1041-4347.
- CLAUSET, A.; NEWMAN, M.; MOORE, C. Finding community structure in very large networks. *Physical Review E*, v. 70, n. 6, p. 066111, 2004.
- DASGUPTA, S.; PAPADIMITRIOU, C.; VAZIRANI, U. *Algorithms*. 1. ed. [S.l.]: McGraw-Hill Science/Engineering/Math, 2006. ISBN 0073523402.
- DELVENNE, J.; YALIRAKI, S.; BARAHONA, M. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, v. 107, n. 29, p. 12755–12760, 2010. ISSN 1091-6490.
- DOROGVTSEV, S. N.; MENDES, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and Www (Physics)*. [S.l.]: Oxford University Press, 2003. Hardcover. ISBN 0198515901.
- DOROGVTSEV, S. N. *Lectures on Complex Networks*. [S.l.]: Oxford University Press, 2010.

- ELDERFIELD, D.; SHERRINGTON, D. The curious case of the potts spin glass. *Journal of Physics C: Solid State Physics*, v. 16, n. 15, p. L497, 1983. ISSN 0022-3719.
- ERDÖS, P.; RÉNYI, A. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, v. 6, p. 290–297, 1959.
- FAGIOLO, G. Clustering in complex directed networks. *Physical Review E*, v. 76, n. 2, p. 026107, 2007.
- FORGERINI, F. L. *Interação de agentes e processos estocásticos em redes complexas*. Tese (Doutorado) — Universidade de Aveiro, 2013.
- FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry*, v. 40, n. 1, p. 35–41, 1977. Disponível em: <<http://dx.doi.org/10.2307/3033543>>.
- GIRVAN, M.; NEWMAN, M. E. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA. girvan@santafe.edu, v. 99, n. 12, p. 7821–7826, junho 2002. ISSN 0027-8424. Disponível em: <<http://dx.doi.org/10.1073/pnas.122653799>>.
- GRIFFITHS, D. *Introduction to Quantum Mechanics (2nd Edition)*. 2nd. ed. [S.l.]: Pearson Prentice Hall, 2004. ISBN 0131118927.
- HELBING, D. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, American Physical Society, v. 73, n. 4, p. 1067–1141, dez. 2001. Disponível em: <<http://dx.doi.org/10.1103/RevModPhys.73.1067>>.
- KAPLAN, W. *Advanced Calculus (4th Edition)*. 4 sub. ed. [S.l.]: Addison Wesley Publishing Company, 1991. ISBN 0201578883.
- KRAPIVSKY, P. L.; REDNER, S.; BEN-NAIM, E. *A Kinetic View of Statistical Physics*. [S.l.]: Cambridge University Press, 2010. ISBN 9780521851039.
- LAMBIOTTE, R.; DELVENNE, J.; BARAHONA, M. Laplacian dynamics and multiscale modular structure in networks. 2009.
- LEICHT, E. A.; NEWMAN, M. E. J. Community structure in directed networks. *Physical Review Letters*, APS, v. 100, n. 11, 2008. Disponível em: <<http://dx.doi.org/10.1103/PhysRevLett.100.118703>>.
- LIGGETT, T. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes (Grundlehren der mathematischen Wissenschaften)*. 1. ed. [S.l.]: Springer, 1999. ISBN 3540659951.
- LILJEROS, F. *et al.* The web of human sexual contacts. *Nature*, v. 411, p. 907–908, junho 2001.
- LIU, J.; LIU, T. Detecting community structure in complex networks using simulated annealing with k-means algorithms. *Physica A: Statistical Mechanics and its Applications*, v. 389, n. 11, p. 2300–2309, 2010. ISSN 03784371.
- MAERIVOET, S.; MOOR, B. D. Cellular automata models of road traffic. *Physics Reports*, v. 419, n. 1, p. 1–64, nov. 2005. ISSN 03701573. Disponível em: <<http://dx.doi.org/10.1016/j.physrep.2005.08.005>>.

MARTINO, D. D. *et al.* A minimal model for congestion phenomena on complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2009, n. 08, p. P08023+, ago. 2009. ISSN 1742-5468. Disponível em: <<http://dx.doi.org/10.1088/1742-5468/2009/08/P08023>>.

MARTINO, D. D. *et al.* Congestion phenomena on complex networks. *Physical Review E*, American Physical Society, v. 79, n. 1, p. 015101+, jan. 2009. ISSN 1539-3755. Disponível em: <<http://dx.doi.org/10.1103/PhysRevE.79.015101>>.

MOBILIA, M.; BARES, P. A. Generalized empty-interval method applied to a class of one-dimensional stochastic models. *Physical Review E*, v. 64, n. 066123, p. 18, julho 2001.

NAGURNEY, A. *Sustainable Transportation Networks (Elgar Monographs)*. 1st. ed. Edward Elgar Pub, 2000. Hardcover. ISBN 1840643579. Disponível em: <<http://www.worldcat.org/isbn/1840643579>>.

NEWMAN, M. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, v. 98, n. 2, p. 404–409, 2001. ISSN 1091-6490.

NEWMAN, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, v. 103, n. 23, p. 8577–8582, 2006. Disponível em: <<http://www.pnas.org/content/103/23/8577.abstract>>.

PATHRIA, R. *Statistical Mechanics, Second Edition*. 2. ed. [S.l.]: Butterworth-Heinemann, 1996. ISBN 0750624698.

PEREGRINO, N. M. M. C. *Topologia e Dinâmica de Sistemas Complexos: Transições de Fase e Fenômenos Críticos*. Tese (Doutorado) — Instituto de Física - Universidade Federal Fluminense, 2011.

REIF, F. *Fundamentals of Statistical and Thermal Physics*. 1. ed. [S.l.]: McGraw-Hill Science/Engineering/Math, 1965. Hardcover. ISBN 0070518009.

ROBINSON, K.; COHEN, T.; COLIJN, C. The dynamics of sexual contact networks: Effects on disease spread and control. *Theoretical Population Biology*, v. 81, n. 2, p. 89–96, 2012. ISSN 00405809.

ROSS, S. *Stochastic Processes (Wiley Series in Probability and Statistics)*. 2. ed. [S.l.]: Wiley, 1995. ISBN 0471120626.

ROUSSAS, G. *A Course in Mathematical Statistics*. Third edition. [S.l.]: Academic Press, 1997. ISBN 0120884461.

SCHIFF, J. *The laplace transform: Theory and applications*. 1999.

SERIDONIO, A. *et al.* Spin-resolved local density of states for an anderson adatom in a ferromagnetic island. *Journal of Superconductivity and Novel Magnetism*, v. 26, n. 6, p. 2197–2200, junho 2013.

SERIDÔNIO, A. C. *et al.* Non-zeeman splitting for a spin-resolved stm with a kondo adatom in a spin-polarized two-dimensional electron gas. *Physical Review B*, v. 85, p. 165109, 2012.

SOUSA, O. F. de; PENNA, T. J. P. Community analysis of software maintenance in the debian gnu/linux project. Submetido a *Physica A*. Janeiro 2013.

TOMÉ, T.; OLIVEIRA, M. J. de. *Dinâmica Estocástica e Irreversibilidade*. [S.l.]: Edusp, 2001. 124–125 p.

VÁZQUEZ, A.; SATORRAS, R.; VESPIGNANI, A. Large-scale topological and dynamical properties of the internet. *Physical Review E*, v. 65, n. 6, p. 066130, 2002.