

Dissertação de Mestrado

Estudos sobre a Conectividade em Redes de Sensores Sem Fios: Análise de Plataformas e Resultados de Percolação no Plano Contínuo

Marcelo Gabriel Almiron
almiron.marcelo@gmail.com

Orientador:
Prof. Dr. Alejandro C. Frery

Maceió, fevereiro de 2009

A minha amada Lorhena

Marcelo Gabriel Almiron

Estudos sobre a Conectividade em Redes de Sensores Sem Fios: Análise de Plataformas e Resultados de Percolação no Plano Contínuo

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Curso de Mestrado em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. Alejandro C. Frery

Maceió, fevereiro de 2009

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária Responsável: Helena Cristina Pimentel do Vale

A517e Almiron, Marcelo Gabriel.

Estudos sobre a conectividade em redes de sensores sem fios : análise de plataformas e resultados de percolação no plano contínuo / Marcelo Gabriel Almiron. 2009.

84 f. : il.

Orientador: Alejandro C. Frery.

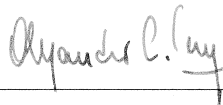
Dissertação (mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2009.

Bibliografia: f. 71-76.

1. Redes de computação. 2. Sistema de comunicação sem fio. 3. Processos pontuais espaciais. 4. Sensores sem fio. 5. Conectividade. I. Título.

CDU: 004.738.5

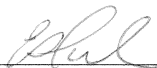
Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional de Conhecimento pelo Programa Multidisciplinar de Pós-Graduação em Modelagem Computacional de Conhecimento, da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina:



Prof. Dr. Alejandro César Frey

UFAL – Instituto de Computação

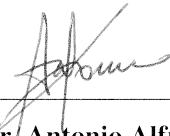
Orientador



Profa. Dra. Eliana Silva de Almeida

UFAL – Instituto de Computação

Examinadora



Prof. Dr. Antonio Alfredo Ferreira Loureiro

UFMG – Departamento de Ciência da Computação

Examinador

Maceió, março de 2009.

Resumo

Apresentamos a caracterização do raio de transmissão mínimo necessário para garantir conectividade, *CTR (Critical Transmission Range)*, num cenário de controle de topologia em RSSF homogêneas e estacionárias. Dada a complexidade de se trabalhar com modelos analíticos, por meio de experiências Monte Carlo obtemos um estimador da distribuição do CTR sobre processos pontuais espaciais que descrevem o posicionamento dos nós sensores no ambiente para diferentes níveis de atratividade. Propomos modelos de otimização práticos que consideram diversos fatores conhecidos *a priori* pelo projetista como, por exemplo, os preços de diversos sensores, o raio de transmissão máximo disponível pelo sensor, os custos de posicionamento no ambiente (função da atratividade), o orçamento total do projeto, a probabilidade de conectividade mínima admissível e o expoente de *path loss* do ambiente. O modelo determina quais e quantos sensores devem ser comprados, com que raio de transmissão devem ser configurados e qual o preço conveniente a pagar pelo posicionamento (função da atratividade), para maximizar o tempo de vida de uma RSSF. Para guiar a escolha da plataforma de simulação e análise de dados, vários resultados a respeito de precisão numérica são apresentados, obtidos aplicando protocolos de avaliação já consolidados. Desta análise, determinamos que R (<http://www.r-project.org>) é a melhor escolha.

Abstract

We study the minimum radius required for connectivity (CTR – Critical Transmission Range) within homogeneous stationary Wireless Sensor Networks (WSN) topology control, considering different levels of attractivity within the sensors. Due to the complexity of dealing with this problem from a theoretical viewpoint, a Monte Carlo experience is devised for estimating the CTR distribution. With this information, we propose optimization procedures that, using as additional input a few known parameters (overall available budget, sensor cost, maximum available transmission radius, minimum probability of connectivity, environmental path loss and deployment cost) leads to the decision of the number and type of sensors to be acquired, their optimal communication radius and the ideal deployment strategy that maximize the WSN lifetime. As a previous result, the accuracy of several computational platforms for statistical computing was assessed, being the main conclusion that R (<http://www.r-project.org>) is the best choice.

Agradecimentos

Gostaria de agradecer o incalculável apoio de Lorhena, das Profas. Elsa Moschetti e Susana Ferrero (do *Departamento de Matemática da Universidad Nacional de Río Cuarto*, Río Cuarto, Córdoba – Argentina) e do Prof. Alejandro C. Frery que, frente ao desafio que representava este empreendimento, sempre me guiaram pelo caminho certo e fizeram tudo o que estava a seu alcance (e mais) para me manter concentrado em um único objetivo: crescer.

Agradeço novamente ao Prof. Alejandro por me dar a oportunidade de trabalhar e apreender, junto a um grão professor e pesquisador, o processo de pesquisa. Além do Prof. Alejandro, durante minha estadia na Universidade Federal de Alagoas, tive o prazer de trabalhar com a Profa. Eliana Silva de Almeida e o Prof. Marcio Nunes de Miranda, e com os colegas Alyson L. Costa Oliveira, Bruno Lopes e Antonio C. Medeiros. Agradeço a todos eles o apoio, o tempo dedicado à interação e a compreensão pelos meus limites, principalmente com o domínio da língua Portuguesa. Também gostaria de agradecer a todos os integrantes do Laboratório de Computação Científica e Análise Numérica pela interação diária. Nas minhas primeiras participações deste grupo de pesquisa, as presenças do Prof. Heitor Soares Ramos Filho, José Alencar-Neto e Alyson, foram muito importantes, e agradeço a eles as horas de conversa dentro e fora do horário de trabalho que sem dúvida foram decisivas no momento de direcionar da melhor maneira possível meus esforços nesta área de pesquisa.

Agradeço também a Branca, Enilson, Morgana, Muda, e a José Tenório Cesar Costa e sua família, pelo acolhimento na maravilhosa capital de Alagoas, e por atenuarem de uma o outra maneira a falta dos seres que amo.

Finalmente, porém não de menor importância, agradeço a todos os Professores que aportaram seu grão de areia na minha formação, ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento, e à Universidade Federal de Alagoas, por terem compartilhado seus conhecimentos, espaços e recursos.

A todos eles, muito obrigado!

Conteúdo

1	Introdução	1
1.1	Breve histórico	1
1.2	RSSFs e características	3
1.3	Espaço de dados nas RSSFs	6
1.4	Aplicações de RSSFs	6
1.5	O estado da arte	7
1.6	Proposta	9
2	Conceitos preliminares	11
2.1	Teoria espectral de grafos	11
2.2	Processos pontuais	15
2.2.1	Definindo processos pontuais	18
2.2.2	O processo pontual C	21
2.3	Modelagem de RSSFs	23
3	Ferramentas de simulação e análise de dados	28
3.1	Ambientes de simulação	29
3.2	Análise de dados	30
3.2.1	Metodologia de avaliação da precisão numérica	30
3.2.2	Resultados de precisão numérica	31
3.3	Comentários	38
4	Estudos de conectividade	39
4.1	Controle de topologia	40
4.2	Conectividade em processos binomiais	43
4.2.1	Estudo Monte Carlo	46
4.2.2	Estudo de caso I: processos pontuais binomiais	50
4.3	Conectividade em casos atrativo e repulsivo	52
4.3.1	Estudos Monte Carlo	53
4.3.2	Estudo de caso II: processos pontuais atrativo e repulsivo	62
5	Conclusões	67
5.1	Resumo dos resultados e conclusões	67
5.2	Trabalhos futuros	68
A	Notação	69
B	Publicações	71

Lista de Figuras

1.1	Publicações sobre RSSFs por ano	2
2.1	Representação gráfica de grafos simples	12
2.2	Comportamento da conectividade algébrica	16
2.3	Exemplos de posicionamento de maneira aleatória uniforme	16
2.4	Exemplos de posicionamento segundo um processo atrativo	17
2.5	Exemplos de posicionamento segundo um processo repulsivo	17
2.6	Tempos de chegada T_i e entre chegadas S_i	18
2.7	Região de contagem $N(B) = 2$	19
2.8	Máxima inibição r na área $W = [0, 1]^2$. Com 16 pontos $r = 1/4$	22
2.9	Área alvo $A = [0, \frac{1}{4}]^2$, e $B = A^c \cap [0, 1]^2 = ([0, \frac{1}{4}]^2)^c \cap [0, 1]^2$	23
2.10	Exemplos de processos pontuais espaciais $C(n, a)$	24
2.11	Representação gráfica do grafo de comunicações $CG = (S, E)$	27
3.1	Níveis de abstração conceituais na modelagem computacional	28
4.1	Escolha do alcance de transmissão dos nós sensores	41
4.2	Relação entre o MST e o CTR	43
4.3	<i>Beanplot</i> de $x_1^{(n,0)}, x_2^{(n,0)}, \dots, x_{1000}^{(n,0)}$, para $n \in \{10, 20, 30, 40, 50\}$	48
4.4	Função empírica das variáveis $X^{(n,0)}$	49
4.5	<i>Beanplot</i> de $x_1^{(n,-30)}, x_2^{(n,-30)}, \dots, x_{1000}^{(n,-30)}$, para $n \in \{10, 20, 30, 40, 50\}$	53
4.6	Função empírica das variáveis $X^{(n,-30)}$	54
4.7	<i>Beanplot</i> de $x_1^{(n,-15)}, x_2^{(n,-15)}, \dots, x_{1000}^{(n,-15)}$, para $n \in \{10, 20, 30, 40, 50\}$	55
4.8	Função empírica das variáveis $X^{(n,-15)}$	56
4.9	<i>Beanplot</i> de $x_1^{(n,5)}, x_2^{(n,5)}, \dots, x_{1000}^{(n,5)}$, para $n \in \{10, 20, 30, 40, 50\}$	57
4.10	Função empírica das variáveis $X^{(n,5)}$	58
4.11	<i>Beanplot</i> de $x_1^{(n,15)}, x_2^{(n,15)}, \dots, x_{1000}^{(n,15)}$, para $n \in \{10, 20, 30, 40, 50\}$	59
4.12	Função empírica das variáveis $X^{(n,15)}$	60
4.13	<i>Beanplot</i> de $x_1^{(n,30)}, x_2^{(n,30)}, \dots, x_{1000}^{(n,30)}$, para $n \in \{10, 20, 30, 40, 50\}$	61
4.14	Função empírica das variáveis $X^{(n,30)}$	62
4.15	Função empírica das variáveis $X^{(10,a)}$	63
4.16	Função empírica das variáveis $X^{(30,a)}$	64
4.17	Função empírica das variáveis $X^{(50,a)}$	65

Lista de Tabelas

3.1	Diversidade da análise de precisão numérica	32
3.2	MD para a média amostral	32
3.3	MD para o desvio padrão amostral	33
3.4	MD para o coeficiente de autocorrelação amostral	34
3.5	MD para o estatístico F de ANOVA	35
3.6	Valores de MD menores para o coeficiente $\hat{\beta}$ e o desvio padrão residual RSD aplicando regressão linear	36
3.7	MD menores para o coeficiente $\hat{\beta}$ e o desvio padrão residual RSD em R aplicando regressão não linear	37
3.8	MD para as funções de quantis de distribuições	37
4.1	Resumo de $x_1^{(n,0)}, \dots, x_{1000}^{(n,0)}$, para $n \in \{10, 20, 30, 40, 50\}$	49
4.2	Resumo de $x_1^{(n,-30)}, \dots, x_{1000}^{(n,-30)}$, para $n \in \{10, 20, 30, 40, 50\}$	54
4.3	Resumo de $x_1^{(n,-15)}, \dots, x_{1000}^{(n,-15)}$, para $n \in \{10, 20, 30, 40, 50\}$	56
4.4	Resumo de $x_1^{(n,5)}, \dots, x_{1000}^{(n,5)}$, para $n \in \{10, 20, 30, 40, 50\}$	57
4.5	Resumo de $x_1^{(n,15)}, \dots, x_{1000}^{(n,15)}$, para $n \in \{10, 20, 30, 40, 50\}$	58
4.6	Resumo de $x_1^{(n,30)}, \dots, x_{1000}^{(n,30)}$, para $n \in \{10, 20, 30, 40, 50\}$	59

Capítulo 1

Introdução

Neste capítulo nos familiarizaremos com as chamadas Redes de Sensores Sem Fios (RSSFs), vendo como surgiram, quais suas características, quais os principais desafios, quais suas aplicações e em que ponto se encontra a comunidade científica em referência aos objetivos principais desta relativamente nova modalidade de rede. É assim que conseguiremos nos situar no contexto deste trabalho de dissertação.

A continuação, na seção 1.1 apresentaremos uma breve análise dos começos da teoria de RSSFs, assim como também a evolução como área de pesquisa. Na seção 1.2 vamos definir as RSSFs e caracterizá-las. Uma apresentação do esquema do processo de colheita de informação, desde o sensoriamento até a tomada de decisões é oferecida na seção 1.3. Para termos maior familiaridade com as diferentes áreas de pesquisa dentro de RSSFs, na seção 1.5 é apresentado um breve resumo dos principais trabalhos de cada área. A partir deste relato, conseguiremos na seção 1.6 apresentar nossa proposta de trabalho e contextualizá-la.

1.1 Breve histórico

Redes de sensores foram fonte de interesse em pesquisas militares durante os anos setenta e oitenta. A missão destas redes foi o monitoramento através de sensores acústicos, de luminosidade, de temperatura e de pressão, dentre outros fenômenos físicos. Estas redes de sensores foram classificadas por Chong & Kumar (2003) em quatro fases, segundo o seu estágio evolutivo. Na primeira fase se encontram as redes de sensores militares da época da guerra fria; o objetivo principal foi monitorar as águas contra possível presença de submarinos inimigos. Uma segunda fase começa nos primeiros anos da década de 1980, e está fortemente marcada por iniciativas em pesquisa da *Defense Advanced Research Projects Agency* (DARPA); o principal projeto foi o *Distributed Sensor Networks* (DSN), envolvendo pesquisadores da *Carnegie Mellon University* e do *Massachusetts Institute of Technology*. A terceira fase ocorreu a finais da década de 1980 com as redes de sensores militares, nesse momento já funcionais, e combinados com tecnologias de interfaces compatíveis com redes e produtos comerciais. Porém o cenário concebido por cientistas da época (sensores pequenos em grandes quantidades) ainda não fosse uma realidade nesse momento, várias redes de sensores foram implementadas nesta terceira fase. Finalmente a quarta fase é a do final do século XX até o presente, onde a tecnologia pretendida já estava começando a estar disponível. Vários grupos de pesquisa de diversas universidades

tinham acesso a sensores cada vez mais pequenos e de baixo custo baseados em tecnologia de *MicroElectroMechanical Systems* (MEMS), de comunicação sem fios e de processadores baratos de baixo consumo energético (ver Chandrakasan et al., 1999).

Dentro desta quarta fase é que surgiram as RSSFs classificadas comercialmente como RSSFs da primeira geração, logo vieram as da segunda geração e finalmente as de terceira geração. Estas gerações são diferenciadas por vários fatores como peso, tamanho, modalidade de distribuição dos nós sensores no ambiente, arquitetura, protocolos e padrões, topologias, autonomia e tempo de vida (ver Sohraby et al., 2007).

Com o começo das gerações comerciais, a quantidade de pesquisadores em RSSFs foi crescendo, e a quantidade de publicações também como consequência. Para ter uma ideia geral, na figura 1.1 vemos o interesse que despertou esta área de pesquisa. O grafico mostra, sobre um total de 2215 publicações, a proporção de publicações por ano.

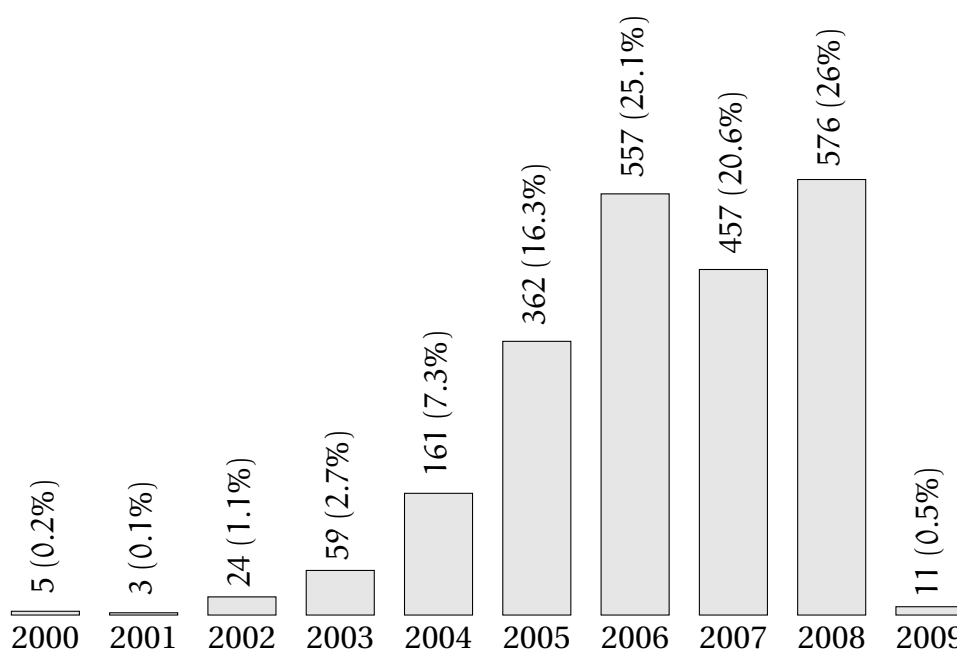


Figura 1.1: Publicações por ano sobre um total de 2215 resultados: pesquisa com a palavra chave: 'wireless sensor' OR 'wireless sensors' OR 'wireless micro-sensor' OR 'wireless microsensors') AND network*

Este gráfico foi gerado a partir dos dados obtidos em pesquisas com a ferramenta *Web of Science* do *ISI Web of Knowledge*. Cada barra indica a quantidade de publicações correspondentes a cada ano e, junto a esse valor, entre parênteses, a percentagem (arredondada a uma casa decimal) sobre o total de publicações. Notemos que estamos utilizando como palavras chaves aquelas que permitem identificar trabalhos publicados já no período de concepção de RSSFs como área específica. Resultados de trabalhos que foram publicados com nomes antigos e variados que deram início às RSSFs não estão presentes.

Na próxima seção vamos definir as RSSFs junto aos diferentes parâmetros que devem se ter em conta ao momento de projetá-las. Desta maneira não somente te-

remos maior conhecimento do que as RSSFs são, mas também o que podem vir a ser.

1.2 RSSFs e características

Em poucas palavras, podemos definir uma RSSF da seguinte maneira (Akyildiz et al., 2002; Nakamura et al., 2007):

Definição 1 (Rede de Sensores Sem Fios) *Uma RSSF é um tipo especial de rede ad hoc composta por muitos nós sensores equipados com direfentes dispositivos sensores. Esta rede tem suporte por meio de avanços tecnológicos em comunicação sem fio de baixo consumo de energia tendo várias funcionalidades tais como sensoriamento, comunicação, e processamento.*

Mas é tão amplo o campo de aplicação e as formas que as RSSFs podem adotar que, para compreender melhor as RSSFs como tais, elencaremos diferentes parâmetros que estão ausentes dentro desta definição e que oferecem informação relevante para conceber um conceito mais concreto.

Römer & Mattern (2004) introduzem algumas características para classificar e diferenciar entre diferentes RSSFs. A continuação vamos expor essas características para ter uma visão mais detalhada e logo assim poder definir o cenário no qual está focalizada esta dissertação:

Distribuição dos sensores no espaço. Os sensores podem ser distribuídos aleatoriamente no espaço do ambiente de interesse (e.g. jogados de um avião) ou instalados em lugares estratégicos com objetivos específicos. Este processo de distribuição de nós no espaço pode ser contínuo no mesmo tempo que a rede encontra-se em funcionamento, ou pode ser feito só uma vez. A distribuição dos nós tem um impacto importante sobre algumas propriedades da rede, como a densidade esperada da rede, a localização de sensores e o grau esperado de dinamismo da rede, isto é, o que se espera quanto ao processo dinâmico que é observado (perda de conectividade ou cobertura, por exemplo) quando sensores deixam de funcionar e quando novos sensores são integrados à rede. Nos referimos à densidade esperada da rede como a quantidade de sensores posicionados fisicamente numa unidade de espaço. É assim que, segundo o posicionamento dos nós no espaço, diferenciamos entre *aleatório* ou *manual*, e segundo as instâncias em *única* ou *iterativa*.

Mobilidade dos nós. Os nós sensores podem mudar de posição durante o tempo de vida da rede, seja por influência do ambiente no qual se acham ou por capacidade própria de movimentação. Como exemplo, suponhamos que dispomos de nós móveis, isto é com capacidade própria de se movimentar. Poderíamos, então, otimizar o consumo de energia movimentando os sensores para setores específicos que tornem a topologia mais eficiente, mas considerando o gasto necessário para gerar esta nova topologia. Por outro lado, redes sem capacidade de movimentação própria podem se diferenciar segundo o grau de mobilidade, isto é, se casual ou constante dependendo novamente do ambiente no qual os sensores se encontram. Resumindo, segundo a mobilidade de seus nós,

diferenciamos entre redes *não móveis*, *parcialmente móveis*, ou *totalmente móveis*; também diferenciamos entre redes com movimento *casual* ou *contínuo* e ainda entre redes com movimentação *ativa* ou *passiva*.

Custo, tamanho, recursos e energia. O tamanho e custo dos nós sensores determinam os recursos e a energia disponíveis, capacidade de comunicação e processamento. Estes parâmetros dependem da aplicação e o orçamento destinado à implementação. É assim que segundo seu tamanho físico pode-se diferenciar entre *muito pequeno*, *pequeno*, *medio* ou *grande*.

Heterogeneidade dos nós sensores. Num principio as RSSFs foram integradas por sensores homogêneos. Atualmente, as RSSFs podem conter uma ampla variedade de dispositivos. Baseados na diversidade de nós sensores na mesma rede, pode-se distinguir entre redes *homogêneas* ou *heterogêneas* do ponto de vista do *hardware*.

Modalidade de comunicação entre nós sensores. Para as comunicações sem fio sobre os nós sensores, já foram utilizadas diferentes tecnologias como *rádio* e *laser*, dentre outras. A tecnologia escolhida é totalmente dependente do ambiente no qual a rede vai operar. Por exemplo, se a rede é aquática com certeza utilizaremos uma tecnologia de comunicação específica para esse ambiente, diferente das que utilizaríamos num ambiente terrestre qualquer. Diferenciando entre características, como linha livre de visão, consumo de energia em comunicação, distância máxima de alcance, pode-se utilizar algumas destas modalidades de acordo com a aplicação.

Infraestrutura da rede. As modalidades de comunicação podem ser de diferentes maneiras. Existem duas formas, uma chamada de redes baseadas em infraestrutura, e outra chamada de redes *ad hoc*. Na rede baseada em infraestrutura, os nós sensores da rede somente podem se comunicar com a estação base enviando a informação requerida, e por meio dela se comunicar com os outros sensores se for necessário. Exemplos deste tipo de rede são as redes da primeira, segunda e terceira fase mencionadas na seção 1.1. Por outro lado, nas redes *ad hoc*, os nós podem se comunicar entre si sem uma infraestrutura. Eles podem fazer de *routers*, de emissores ou receptores, em diferentes momentos durante o tempo de vida da rede. É assim que segundo sua infraestrutura, as redes podem ser *com infra-estrutura* ou *ad hoc*.

Topologia da rede. Uma propriedade importante de uma rede é seu diâmetro (comprimento do máximo caminho de comunicação entre dois nós quaisquer). Isto tem repercussão na latência, e em outros fatores importantes da rede. Existem vários trabalhos em relação ao controle de topologia, área que consiste em modificar a topologia para obter situações eficientes em termos de capacidade e de funcionalidade (Santi, 2005). Para conseguir isto, existem varias metodologias diferentes, que determinam o raio ótimo que devemos utilizar ou, ao inverso, a quantidade de sensores que devemos espalhar para obtermos topologias “convenientes”.

Cobertura da rede. A cobertura de uma rede é uma medida da proporção do ambiente físico que os nós sensores conseguem monitorar em relação ao raio de

sensoriamento. Com cobertura escassa, só partes da região puderam ser observadas pelos nós sensores. Com cobertura densa, a área de interesse está totalmente coberta pelos nós. Com cobertura redundante, vários sensores cobrem o mesmo espaço físico e, conseqüentemente, o reenvio de informação “multiplicada” representa um mal uso dos recursos energéticos, sendo que estes últimos são críticos.

Conectividade da rede. Os intervalos de comunicação e o posicionamento dos nós sensores, individualmente ou coletivamente, determinam a conectividade da rede. É assim que a rede pode ser *conexa*, *intermitente*, quando em alguns momentos a rede fica particionada em dois ou mais componentes conexos, ou *esporádica*, quando a exceção é dispor de conectividade geral.

Tamanho da rede. O número de nós sensores da rede é determinado por requerimentos relacionados à cobertura e conectividade, bem como pelo tamanho da área de interesse. O tamanho também determina requerimentos de escalabilidade nos algoritmos e protocolos da rede.

Tempo de vida da rede. O tempo de vida requerido da rede pode variar de umas horas a vários anos. O tempo de vida necessário tem alto impacto sobre o grau requerido de eficiência no consumo de energia e na qualidade dos nós sensores.

Requerimentos de qualidade de serviço da rede. Segundo a aplicação, uma RSSF pode fornecer suporte a alguns requerimentos de qualidade de serviço (em inglês *QoS Requirements*) tais como restrições de tempo real (isto é, um acontecimento físico deve ser reportado antes de um certo intervalo de tempo), robustez (isto é, a rede deve permanecer operacional incluso se certos problemas bem definidos ocorrerem) ou discrição (isto é, a presença da rede deve ser difícil de descobrir). Uma propriedade necessária na topologia de comunicação que se requer para qualquer tipo de QoS é conectividade, já é a única propriedade que consegue garantir o principal objetivo das RSSFs: colheita de informação para o usuário através de um ou mais nós *sink*.

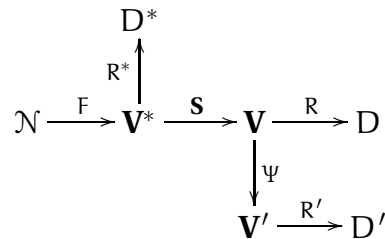
Cada característica, ou conjunto delas, forma um “universo” de pesquisa. Para cada tipo de problema ou necessidade, existem classes de RSSF que podem ser empregadas.

Para esta dissertação nos contextualizamos baseados em algumas destas características. Uma delas é a distribuição dos sensores no espaço. Consideramos variantes com diferentes modelos de aleatoriedade para avaliar o impacto que isto tem sobre a conectividade. Para isso, controlaremos a topologia da RSSF por meio da densidade e do raio de transmissão associado. Também tratamos o problema de escolher entre sensores segundo o custo e a tecnologia de comunicação, considerando, em todos os casos, homogeneidade.

A continuação, na próxima seção, esquematizamos o processo de colheita de informação ao qual toda RSSF está sujeito, através de um diagrama.

1.3 Espaço de dados nas RSSFs

Na área de RSSFs é necessário desenvolver aplicações e soluções que cobrem diferentes etapas de análise no processo de colheita de informação. Cada etapa corresponde-se com um conjunto particular de dados, isto para cada estado da aplicação. A seguir apresentamos um diagrama, proposto por Frery et al. (2008), que é o esquema fundamental de todo o processo de sensoriamento.



\mathcal{N} representa a natureza, \mathbf{V}^* é o conjunto de valores reais que podem ser registrados (quantidades mensuráveis) dado um fenômeno de interesse F . R^* é o conjunto de regras ideais para chegar ao conjunto de decisões ideais D^* . Aqui '*' indica que estamos trabalhando com conjuntos ideais, isto é, valores físicos.

Por o outro lado, temos um conjunto S de sensores por meio dos quais obtemos um conjunto V de dados (recoletados pelos sensores). Agora nesse novo conjunto, através de um conjunto R de regras factíveis, obteremos um conjunto D de decisões factíveis.

Assumindo restrições de tempo, capacidade de processamento e outros fatores, surge a necessidade de uma redução de dados, denominada no diagrama com a letra grega Ψ . Esta função de redução tem o objetivo de reduzir a quantidade de dados com a mínima perda de informação. Aplicando Ψ a V , obtemos um conjunto reduzido V' de dados. Baseados neste conjunto temos um conjunto de regras R' com as quais chegamos a um conjunto D' de decisões.

Nesta dissertação estamos interessados em dar suporte ao projetista de uma RSSF para que a informação presente em V ou em V' possa chegar ao local onde sejam aplicadas as regras para a tomada de decisões. O ponto de saída das informações da rede é denominado "sink", e pode haver mais de um nó sink na rede. Os nós podem enviar as informações recoletadas diretamente ao sink, ou através de outros nós; esta última estratégia é denominada *multi-hop*.

Este problema, denominado "conectividade", pode ser abordado de diferentes maneiras. Muito esforço vem sendo aplicado para a abordagem do ponto de vista de controle de topologia, onde a potência dos sensores é configurada para transmitir à distância mínima necessária para se comunicar com o resto dos sensores e, através de *multi-hop*, chegar até o(s) nó(s) sink.

Para mensurar a importância das RSSFs como ferramenta de colheita de informação, na seção seguinte se apresentam os diferentes campos de aplicações das RSSFs.

1.4 Aplicações de RSSFs

As aplicações das RSSFs se encontram cada dia em maior proporção e diversidade. Conforme Yick et al. (2008) apresentam, estas podem ser classificadas em duas gran-

des modalidades: monitoramento e acompanhamento. Estas duas modalidades vêm sendo aplicadas em diferentes áreas, dentre elas:

- militares, para detecção de perigos ou acompanhamento do inimigo;
- ambientais, para monitoração de fatores biológicos, físicos e meteorológicos, assim com também para atividade vulcânica;
- habitat, na monitoração e rastreamento de animais;
- industriais, no acompanhamento de carros, ônibus e caminhões, no monitoramento de estruturas civis, fábricas, fatores químicos e robótica, dentre outras;
- saúde, no monitoramento de pacientes.

Esta diversidade de aplicações é uma consequência do grau de dinamismo e do vasto espaço de características apresentados na seção 1.2.

Para mais informações a respeito de projetos concretos, com repostas a problemas empíricos, recomendamos a leitura de Simon et al. (2004), Zhang et al. (2004), Tolle et al. (2005), Krishnamurthy et al. (2005), Vasilescu et al. (2005), Johnstone et al. (2007), Werner-Allen et al. (2006) e Baker et al. (2007).

Para todas as aplicações citadas aqui há um conjunto de problemas que devem ser tratados adequadamente. Um desses problemas comum a todas estas aplicações é a conectividade. Como deixamos claro anteriormente, a conectividade é a que permite que as informações recoletadas pelos sensores estejam disponíveis para quem necessita delas.

A continuação apresentamos um breve resumo dos trabalhos relacionados e aqueles que apresentaram-se como fundamento para nosso estudo.

1.5 O estado da arte

Do começo, foi claro que os principais problemas a estudar em RSSF estão relacionados à largura de banda reduzida na comunicação e à quantidade limitada de energia disponível nos nós sensores. Alguns dos primeiros trabalhos importantes que consideram o estudo do consumo de energia como fator principal são os de Wang et al. (2001), Heinzelman et al. (2000) e Heinzelman et al. (2002). Wang et al. (2001) implementam um protocolo de *clustering*, utilizando algoritmos de *beamforming* para agregar dados, em particular o *Maximum Power* e o LMS (*Least Mean Squares*). Por outra parte, Heinzelman et al. (2000) e Heinzelman et al. (2002) propõem o protocolo LEACH como uma alternativa eficiente do ponto de vista do consumo de energia. Este último protocolo foi amplamente utilizado como referência de comparação por outros trabalhos que propõem novas formas de economizar energia, dentre eles Tan & Korpeoglu (2003), Chang & Jan (2005), Noh et al. (2005), Qian et al. (2006) e Kim & Youn (2006).

Embora a administração de recursos seja de fundamental interesse, especialmente a eficiência em consumo de energia, existem outros objetivos de uma RSSF que devem ser atingidos. Ghosh & Das (2006) apresenta dois conceitos que formam parte do planejamento, ou deveriam, de toda RSSF: cobertura e conectividade. Estes problemas foram estudados massivamente em RSSF e também em outras áreas

como grafos aleatórios (Erdős & Rényi, 1960; Harary & Gupta, 1997; Watts & Strogatz, 1998; Penrose, 1999) e percolação (Gilbert, 1961; Meester & Roy, 1996; Diaz et al., 2000).

Em particular, o trabalho de Santi (2005), um *survey* sobre controle de topologia, foi uma das principais motivações para esta dissertação. Santi apresenta formalmente o conceito de controle de topologia em RSSFs, e os principais resultados que foram obtidos até esse momento. A ideia por trás desta área é manipular com algum critério o raio de transmissão dos nós sensores para obter uma topologia adequada, isto é, que garanta conectividade com a menor quantidade de energia possível. Para isso, diferentes técnicas foram desenvolvidas, e em particular nesta dissertação utilizamos uma versão da teoria de percolação no contínuo adaptada para uma região compacta.

Muitos dos trabalhos mais importantes em controle de topologia estão totalmente relacionados ao estudo probabilístico, já que é por meio de ferramentas de probabilidade que conseguimos modelar o dinamismo de uma RSSF (Santi, 2005).

Um trabalho interessante é o de Bettstetter & Hartmann (2005), onde os autores estudam o fenômeno de conectividade de redes sem fios, analisando o impacto que o expoente de *path loss*, α , causa na funcionalidade da rede. Este trabalho expõe também a ideia de, a partir do conhecimento do raio de comunicação necessário, analisar a quantidade de sensores em relação ao raio e às dimensões do ambiente, para obter uma solução ótima. Entretanto, não achamos na literatura uma combinação entre resultados de conectividade e a tomada de decisões de vários fatores para dar suporte ao projetista de uma RSSF.

Notemos que todo problema de controle de topologia resulta não trivial somente quando o posicionamento dos nós sensores no ambiente não é regular em forma de grade.

Já no caso de posicionamentos aleatórios uniformes e independentes (posteriormente formalizaremos estas situações), vários resultados têm sido obtidos. Panchapakesan & Munjunath (2001), por meio do modelo de grafos aleatórios geométrico que definiremos posteriormente, obtêm uma expressão funcional para o raio de transmissão mínimo necessário para manter funcionalidade em uma rede sem fios, isto num ambiente de duas dimensões. Narayanaswamy et al. (2002) propõem um algoritmo capaz de deduzir o raio de transmissão necessário para manter conectividade com a máxima capacidade de tráfego. A partir do trabalho de Gupta & Kumar (2000), conhecemos que este resultado implica que a rede é mínima em termos de custos energéticos, isto é, em redes sem fios, minimização de custos energéticos é equivalente a maximização da capacidade de tráfego da rede.

Os trabalhos de Glauche et al. (2003) e Booth et al. (2003) apresentam, dentre outros resultados, como a teoria de percolação no contínuo pode ser utilizada para caracterizar o raio de transmissão mínimo necessário para garantir conectividade em uma rede sem fios. No mesmo ano, Franceschetti et al. (2003) mostram que o raio de transmissão em comunicação isotrópica representa o caso mais exigente, isto é, conectividade em modelos isotrópicos implicam conectividade em modelos anisotrópicos usuais na prática.

Vários resultados de geometria estocástica foram aplicados na área de conectividade. Consideramos que o principal é o de Penrose (1997) e Penrose (1999), onde o autor encontra a equivalência entre a aresta máxima da árvore geradora mínima das distâncias euclidianas dos nós, com o raio de transmissão mínimo necessário para ga-

rantir conectividade na rede. Este resultado, se combinado com o de Gupta & Kumar (2000), resulta que a aresta máxima da árvore geradora mínima não somente garante conectividade mas que também maximiza a capacidade de uma rede sem fios.

Outros modelos foram empregados neste problema além do de percolação no contínuo como, por exemplo, o de percolação discreto onde o posicionamento aleatório é modelado com uma função que traduz posições em probabilidades de conexão (ver Ding & Guan, 2008).

Finalmente, a última referência importante, no que tange a conectividade, é o trabalho de Ammari & Das (2008), onde os autores propõem o denominado “modelo de discos correlacionados”. Este novo modelo de percolação permitiu dar respostas analíticas sobre a caracterização da transição de fase em conectividade e cobertura. Entende-se como transição de fase o que acontece com a conectividade como função do raio de transmissão ou quantidade de sensores. Com poucos sensores, ou raio de transmissão pequeno, os sensores estão isolados e com o aumento, de vários componentes conexos passamos a ter, de forma abrupta, uma RSSF funcional.

1.6 Proposta

Por causa da complexidade de se trabalhar analiticamente com RSSFs quando o posicionamento dos nós sensores não é regular ou uniforme independente, há poucos trabalhos que considerem situações mais “reais” quanto ao posicionamento dos sensores no ambiente.

Por outro lado, há vários resultados na caracterização do raio de transmissão mínimo necessário para garantir conectividade que consideram o posicionamento dos sensores no ambiente como uniforme e independente. Entanto, o trabalho de Panchapakesan & Munjunath (2001) foi uma das motivações para o estudo em situações mais gerais. Como comentamos na seção anterior, Panchapakesan & Munjunath (2001) propõem uma expressão funcional para caracterizar o raio “adequado” para manter conectividade em cenários onde os nós são posicionados uniforme e independentemente. Nesta expressão há uma constante desconhecida, a qual consideramos em nosso estudo. A proposta inicial deste trabalho foi no sentido de caracterizar esta constante, não somente no caso para o qual a expressão foi proposta, mas também avaliar o seu comportamento para processos de posicionamento mais gerais.

Quanto à modelagem de posicionamentos mais gerais, o trabalho de Frery et al. (2008) resultou num modelo que achamos adequado, em quanto à dose de diversidade de situações práticas. É assim que a proposta deste trabalho é a seguinte:

Caracterizar o raio de transmissão mínimo necessário para garantir conectividade em RSSF utilizando modelos de posicionamento dos nós sensores no ambiente que sejam mais gerais e “realistas”.

Pelo comentado anteriormente, modelaremos o posicionamento com o processo pontual espacial proposto por Frery et al. (2008), e estudaremos a caracterização do raio de transmissão “ótimo”, isto é, que por uma lado seja mínimo, garantindo assim o uso eficiente de energia, e que por outro lado garanta conectividade para manter a RSSF funcional.

Dentro deste contexto, um dos primeiros passos a seguir foi a escolha da plataforma de simulação adequada para este trabalho. Análises prévias, como os traba-

lhos de Cavin et al. (2002) e de Andel & Yasinsac (2006), mostram que escolher uma plataforma de simulação em redes sem fios não é uma tarefa fácil.

Dada a complexidade de se avaliar uma plataforma de simulação de redes, um dos primeiros objetivos deste trabalho, foi avaliar plataformas de computação estatística para logo escolher a mais adequada e empregá-la em nossos estudos.

Esta dissertação está estruturada em quatro capítulos, além deste. Eles são:

Capítulo 2: apresentamos os principais conceitos empregados neste trabalho a respeito de grafos, processos estocásticos pontuais espaciais e o modelo de RSSF adotado.

Capítulo 3: realizamos a análise da precisão numérica de plataformas usuais para análise de dados, e comentamos critérios para a escolha de plataformas de simulação de RSSF.

Capítulo 4: estudamos o fenômeno de conectividade em RSSF para a situação dos nós serem espalhados no quadrado unitário. Estendemos aqui resultados prévios de percolação no plano para o caso de suporte compacto e diversos modelos de distribuição espacial de sensores.

Capítulo 5: analisamos os resultados alcançados nos capítulos anteriores, e traçamos possíveis direções para pesquisas futuras.

Capítulo 2

Conceitos preliminares

Este capítulo está dedicado ao estudo de conceitos prévios necessários para a leitura e compreensão dos tópicos tratados nesta dissertação. Além de fornecer suporte a esses conceitos que podem ser achados na literatura, a contribuição principal é a unificação da notação de teorias que proliferaram sob diferentes áreas de pesquisa.

Na seção 2.1 começamos com definições básicas da teoria de grafos, para logo apresentar conceitos de interesse para este trabalho que provêm da teoria espectral de grafos. Já na seção 2.2 nos concentramos na teoria de processos pontuais com o objetivo de descrever os processos utilizados para modelar o posicionamento dos nós sensores em nossos estudos. Finalmente na seção 2.3 introduzimos o modelo de RSSFs que consideramos mais apropriado, unificando a notação para termos uma linguagem adequada sem ambigüidades, e reutilizando a teoria apresentada nas seções 2.1 e 2.2.

2.1 Teoria espectral de grafos

Um grafo é um formalismo matemático que serve para representar objetos e relações entre eles. Esta simples estrutura encontra-se em uma grande diversidade de aplicações como circuitos elétricos, estradas, vários tipos de redes, ecossistemas, relações sociais, interação molecular, bases de dados, estruturas de controle de programas e sistemas de permissões de sistemas de segurança, só para citar algumas.

Nesta dissertação utilizamos alguns resultados específicos da teoria espectral de grafos. Esta área da teoria de grafos tem seu foco nas relações de propriedades topológicas de grafos com aquelas derivadas de matrizes de representação ou caracterizações dos mesmos.

Na continuação vamos definir formalmente um grafo, para logo descrever suas características, e finalmente definir grupos de grafos com propriedades que são de especial interesse para este trabalho.

Um dos grafos mais gerais, isto é, com menor quantidade de restrições possíveis, aplicável em RSSFs, é o grafo direcionado simples, definido a seguir.

Definição 2 (Grafo direcionado simples) *Um grafo direcionado simples é uma dupla $\vec{G} = (V, \vec{E})$ onde V é o conjunto finito de vértices, e \vec{E} é o conjunto de arestas. Cada aresta é um par (u, v) de vértices em V , com $u \neq v$.*

O termo “simples” indica que não é permitido expressar que um vértice mantenha relacionamento com ele mesmo. Em RSSFs isto faz muito sentido, já que não estamos

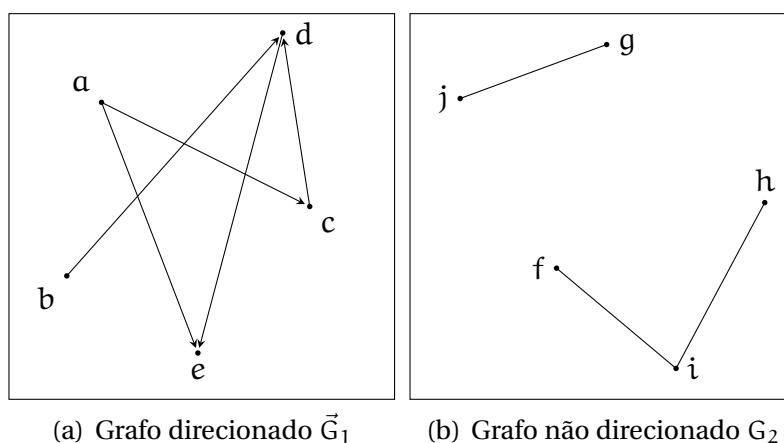


Figura 2.1: Representação gráfica de grafos simples

interessados em modelar autocomunicação pelo meio. Um caso particular deste tipo de grafos, é o grafo não direcionado simples. A única diferença com o grafo definido na definição 2, é que as arestas são representadas como um conjunto, e não como um par ordenado de vértices.

Definição 3 (Grafo não direcionado simples) *Um grafo não direcionado simples é uma dupla $G = (V, E)$ onde V é o conjunto finito de vértices, e E é o conjunto de arestas. Cada aresta é um conjunto $\{u, v\}$ de vértices em V , com $u \neq v$.*

Neste grafo não direcionado simples, cada aresta $\{u, v\}$ não direcionada, representa a presença das arestas direcionadas (u, v) e (v, u) no grafo direcionado simples. A figura 2.1 apresenta estes tipos de grafos. Na figura 2.1(a), o gráfico corresponde ao grafo direcionado $\vec{G}_1 = (V, \vec{E})$ com $V = \{a, b, c, d, e\}$ e $\vec{E} = \{(a, c), (a, e), (b, d), (c, d), (d, e)\}$. Já na figura 2.1(b), o gráfico representa o grafo não direcionado $G_2 = (V, E)$ com $V = \{f, g, h, i, j\}$ e $E = \{\{f, i\}, \{g, j\}, \{h, i\}\}$.

Nesta dissertação vamos trabalhar sempre com grafos simples não direcionados. Quando não for o caso, vamos deixar explícito o tipo de grafo utilizado.

Uma das representações mais utilizadas para grafos, tanto direcionados quanto não direcionados, é a matriz de adjacência. Este conceito é fundamental para continuar construindo definições e propriedades sobre grafos.

Definição 4 (Matriz de adjacência) *A matriz de adjacência é uma representação de grafos. Para um grafo $\vec{G} = (V, \vec{E})$, é uma matriz $A(\vec{G})$ de dimensão $|V| \times |V|$, definida por*

$$A(\vec{G})[i, j] = \begin{cases} 1 & \text{se } (i, j) \in \vec{E}, \\ 0 & \text{caso contrario.} \end{cases}$$

Nesta definição, e no resto do texto, expressões do tipo “ $|V|$ ” representam a cardinalidade do conjunto V . É importante esclarecer que quando dizemos que a matriz de adjacência é uma representação, estamos afirmando que o grafo \vec{G} e sua matriz de adjacência A contêm a mesma informação, isto é, são equivalentes em conteúdo. Baseados no grafo da figura 2.1(a), construímos a matriz de adjacência correspondente

e obtemos

$$A(\vec{G}_1) = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Neste caso, para indexar, utilizamos como exemplo a função de ordem lexicográfico $I_1(x): V \rightarrow \mathbb{N} \cup 0$, isto é $I_1(a) = 0$, $I_1(b) = 1$, $I_1(c) = 2$, $I_1(d) = 3$ e $I_1(e) = 4$. Agora criamos uma nova função de indexação $I_2(x)$, definida por extensão da seguinte maneira: $I_2(f) = 0$, $I_2(g) = 1$, $I_2(h) = 2$, $I_2(i) = 3$ e $I_2(j) = 4$. O matriz de adjacência associada ao grafo da figura 2.1(b) é a seguinte:

$$A(G_2) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Para denotar as relações diretas que um vértice mantém com outros, vamos introduzir os conceitos de adjacência e grau de um vértice.

Definição 5 (Adjacência e grau) Num grafo direcionado $\vec{G} = (V, E)$, o vértice v é adjacente a u se $A[u, v] = 1$. O número de vértices adjacentes a u chama-se grau de u , e diferenciamos entre grau de saída (output)

$$k_u^{\text{out}}(\vec{G}) = \sum_{v \in V: v \neq u} A[u, v]$$

e grau de entrada (input)

$$k_u^{\text{in}}(\vec{G}) = \sum_{v \in V: v \neq u} A[v, u].$$

Denotamos o grau de u simplesmente por $k_u(\vec{G}) = k_u^{\text{out}}(\vec{G}) + k_u^{\text{in}}(\vec{G})$.

É de particular interesse neste trabalho de dissertação, um tipo de grafos chamados de “conexos”. Para definir este tipo de grafos é necessário, primeiro, introduzir o conceito de caminho simples. Notemos que até agora, definindo conceitos para grafos direcionados simples, estamos também incluindo a definição para grafos não direcionados, já que os grafos direcionados são mais gerais.

Definição 6 (Caminho simples) O caminho simples entre v_i e v_j , vértices do grafo direcionado $\vec{G} = (V, \vec{E})$, é qualquer seqüência de arestas

$$(v_i, v_{ij_1}), (v_{ij_1}, v_{ij_2}), \dots, (v_{ij_{n-2}}, v_{ij_{n-1}}), (v_{ij_{n-1}}, v_j),$$

que denotaremos $v_i \rightsquigarrow v_j$. Cada v_{ij} denota um vértice intermediário do caminho desde v_i até v_j . Aqui, o comprimento do caminho é n .

Note que os nós intermediários v_{ij} são irrelevantes para a definição de caminho simples. No sentido mais específico, vamos definir um grafo conexo para grafos não direcionados como segue.

Definição 7 (Grafo conexo) *Um grafo não direcionado simples $G = (V, E)$ é dito conexo se existe um caminho simples $u \rightsquigarrow v$ de qualquer vértice $u \in V$ a qualquer vértice $v \in V$, com $v \neq u$.*

Esta definição pode se aplicar para grafos direcionados simples, mas temos que fazer primeiro uma transformação. Vamos dizer que um grafo direcionado simples \vec{G} é conexo se sua “bi-orientação” é um grafo não direcionado simples G conexo. A bi-orientação é uma função sobre a matriz de adjacência que para cada elemento $A[i, j] = 1$, atribui adjacência para o seu par simétrico, isto é, $A[j, i] = 1$. Outro conceito de interesse é a conectividade forte em grafos direcionados simples, que definimos a seguir:

Definição 8 (Grafo fortemente conexo) *Um grafo direcionado simples $G = (V, E)$ é dito fortemente conexo se existe um caminho simples $u \rightsquigarrow v$ de qualquer vértice $u \in V$ a qualquer vértice $v \in V$, com $v \neq u$.*

Este tipo de conectividade é a desejada em estudos de RSSFs com raio de comunicação anisotrópicos, onde a comunicação, que é descrita pelas arestas, pode ser unilateral.

Dentro da teoria espectral de grafos uma definição básica é a de característica polinomial. A partir deste conceito é que vamos definir os autovalores de um grafo, definição necessária para introduzir a definição principal desta seção: a conectividade algébrica.

Definição 9 (Característica polinomial) *A característica polinomial de um grafo \vec{G} é o determinante $\det(xI - A(\vec{G}))$, onde I é a matriz identidade de dimensão $|A| \times |A|$, e x uma variável real.*

Quando expandimos o determinante da definição 9, vemos que a característica polinomial é um sistema de equações lineares. Resolvendo este sistema, podemos achar os autovalores do grafo, definidos a continuação.

Definição 10 (Autovalores) *Os autovalores de um grafo $\vec{G} = (V, \vec{E})$ são as raízes da característica polinomial.*

Notemos que para um grafo $\vec{G} = (V, \vec{E})$, existem exatamente $|V|$ autovalores, como veremos em breve. Estes autovalores determinam o espectro do grafo definido da seguinte maneira:

Definição 11 (Espectro) *O espectro de um grafo $\vec{G} = (V, \vec{E})$ é o multiconjunto de autovalores associados. Seja $|V| = n$, então existem n autovalores associados a \vec{G} .*

O espectro de um grafo fornece caracterizações do mesmo. Mas é por meio do espectro do laplaciano do grafo que obtemos a caracterização de maior interesse neste trabalho: a conectividade algébrica. Com o laplaciano de um grafo é possível determinar se um grafo é conexo apenas com operações matriciais. Porém não se conhece ainda muito sobre as propriedades espectrais do laplaciano de um grafo, trabalhos relativamente novos nesta área sugerem que há muito mais por descobrir (ver Gross & Yellen, 2003, seção 6.5.6). A definição do laplaciano de um grafo é introduzida da seguinte maneira:

Definição 12 (Laplaciano de um grafo) O laplaciano L de um grafo \vec{G} define-se como $L = D - A(\vec{G})$, onde D é a matriz diagonal de graus dos vértices de \vec{G} .

O segundo menor autovalor do laplaciano de um grafo fornece informação útil a respeito da conectividade do grafo. Antes de conhecer estas propriedades, vamos definir formalmente a conectividade algébrica introduzida por Fiedler (1973).

Definição 13 (Conectividade algébrica) A conectividade algébrica do grafo $\vec{G} = (V, \vec{E})$ com laplaciano L de autovalores $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ é λ_2 .

Como antecipamos, a conectividade algébrica é o segundo menor autovalor do laplaciano associado ao grafo.

O principal resultado a respeito da conectividade algébrica é apresentado no seguinte teorema.

Teorema 1 A conectividade algébrica é positiva ($\lambda_2(L_{\vec{G}}) > 0$) se e somente se o grafo \vec{G} é fortemente conexo.

Uma característica interessante deste teorema (a prova encontra-se em Fiedler, 1973) é que permite determinar se um grafo é conexo ou não apenas com operações matriciais. Para ferramentas como R e Oct, isto representa uma vantagem, já que as operações matriciais estão implementadas com eficiência. Por outro lado, deveríamos analisar o impacto da precisão numérica nos resultados obtidos. Por exemplo, para um grafo conexo, a conectividade algébrica poderia ser 10^{-20} . Um valor negativo muito próximo de zero, implicaria que não teremos classificado o grafo corretamente segundo a propriedade de ser conexo ou não. Mas a conectividade algébrica tem um significado maior. Quanto menor o seu valor, “mais longe” está o grafo de ter conectividade, e quando o valor cresce passando a ser positivo, quanto maior for, “mais forte” é sua conectividade, isto é, a eliminação de algumas arestas do grafo não tira a propriedade do mesmo ser conexo. A figura 2.2 mostra o comportamento da conectividade algébrica segundo o grafo.

A continuação vamos apresentar a teoria de processos pontuais focalizando-nos especificamente em processos pontuais espaciais de duas dimensões. O objetivo é modelar o posicionamento dos nós sensores de maneira aleatória procurando maior identificação em situações reais.

2.2 Processos pontuais

Os primeiros estudos sobre RSSFs basearam-se no posicionamento regular dos nós sensores no espaço. Este modelo de distribuição de sensores no espaço têm suas limitações. Suponhamos que, por exemplo, estejamos interessados em utilizar RSSFs para monitorar grandes áreas onde não seja possível implantar com exatidão sensor a sensor na área. Nesse caso, poderíamos “jogar” os sensores desde uma determinada altura com um avião, ou poderíamos contratar mão de obra e várias pessoas terem o trabalho de posicionar com certa distância variável uns dos outros tentando ocupar toda a área.

Surgiram, assim, novos modelos onde o posicionamento dos sensores já não formam uma grade regular. Um exemplo típico é considerar as coordenadas do posicionamento de cada sensor como um par de variáveis aleatórias uniformes indepen-

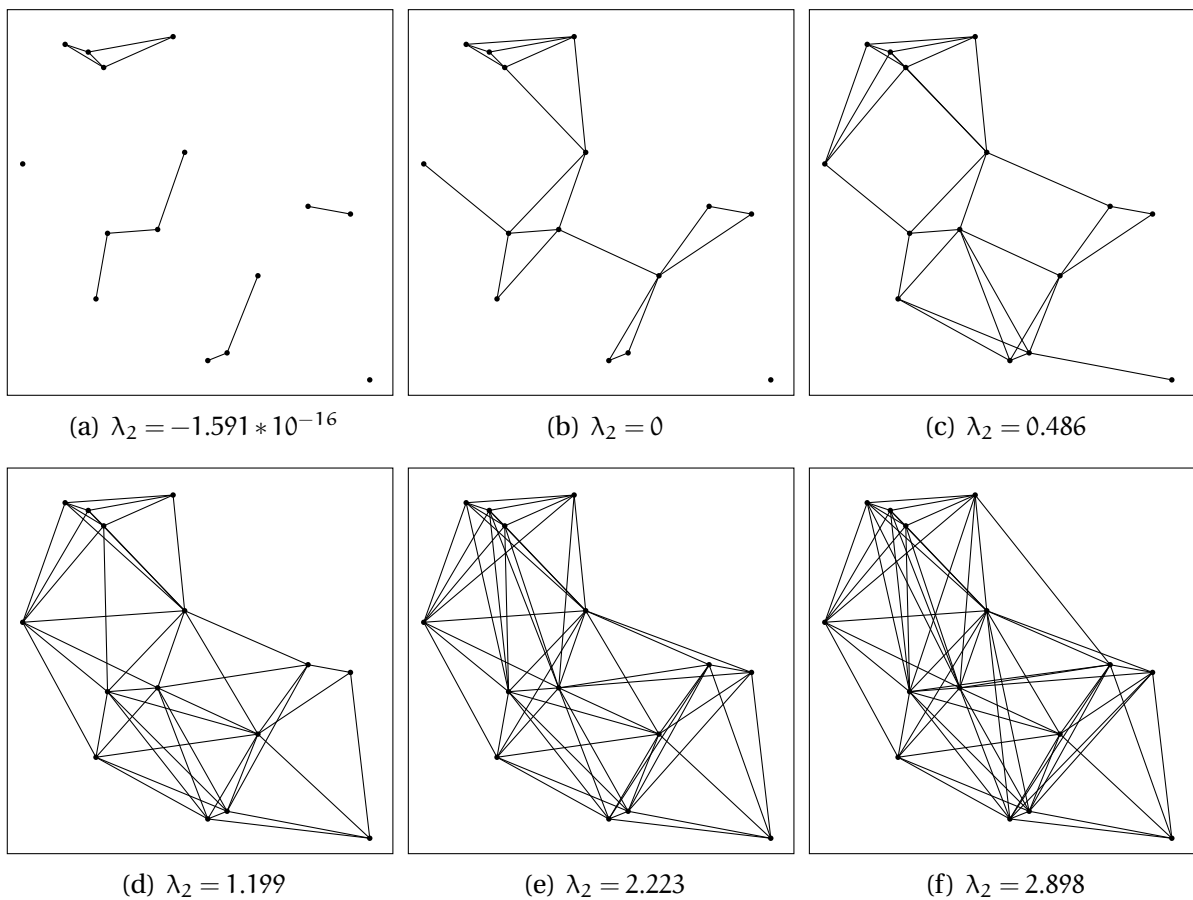


Figura 2.2: Comportamento da conectividade algébrica

dentadas umas de outras. A figura 2.3 ilustra esta situação com três redes, A_p , B_p e C_p , com 15 sensores cada.

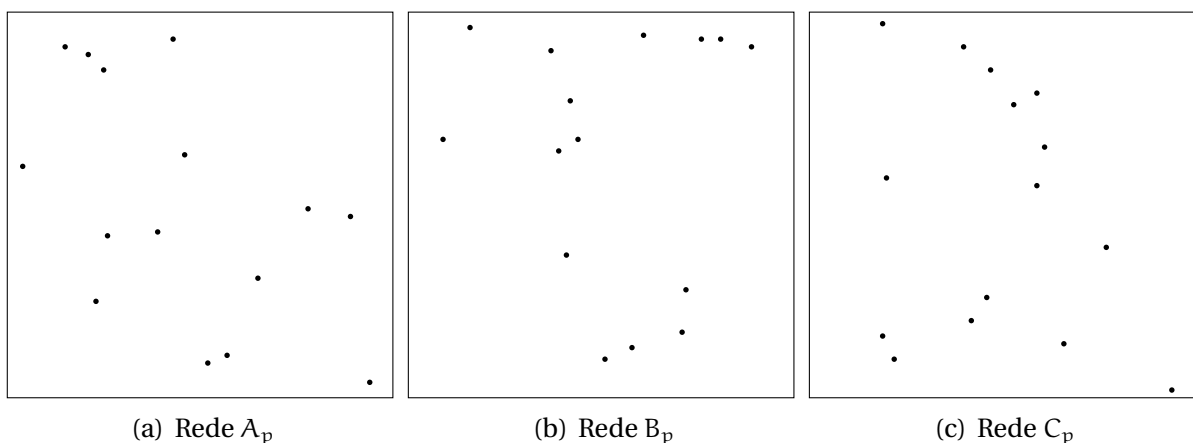


Figura 2.3: Exemplos de posicionamento de sensores de maneira aleatória uniforme

Porém este modelo parece ter maior realismo, existem situações onde o posicionamento de nós sensores de maneira aleatória uniforme não são necessariamente representativas. Um exemplo pode ser o da figura 2.4, onde a superfície na qual

os sensores são posicionados apresenta uma inclinação fortemente marcada, mantendo os sensores principalmente em um centro de baixo nível (lateral esquerda inferior dos gráficos), o que denominamos processo atrativo.

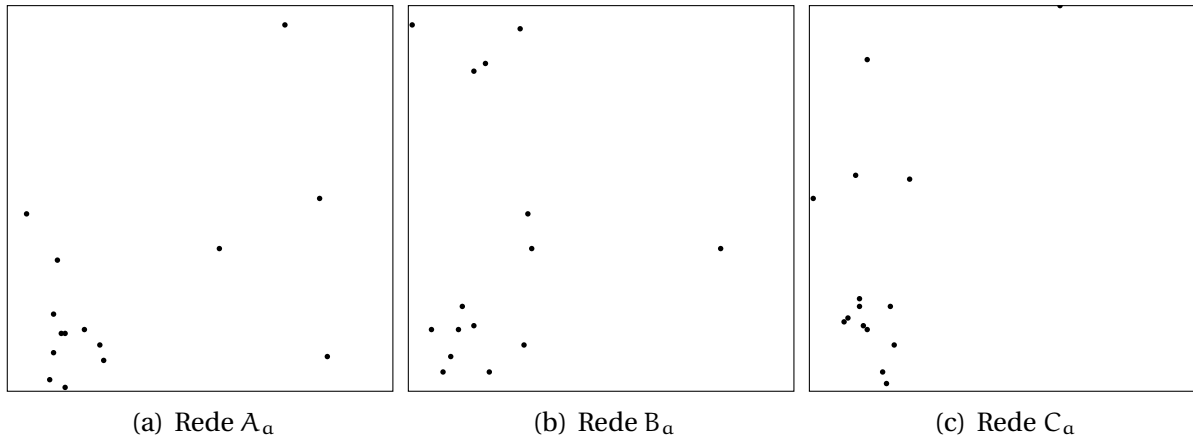


Figura 2.4: Exemplos de posicionamento de sensores segundo um processo atrativo

Por outro lado, voltando à situação de contratar mão de obra para o posicionamento, poderíamos exigir aos contratados manter o mais longe possível os sensores uns dos outros como mostra a figura 2.5, evitando ter dois ou mais sensores em uma região reduzida da área total a cobrir. Este tipo de posicionamento é conhecido como processo repulsivo.

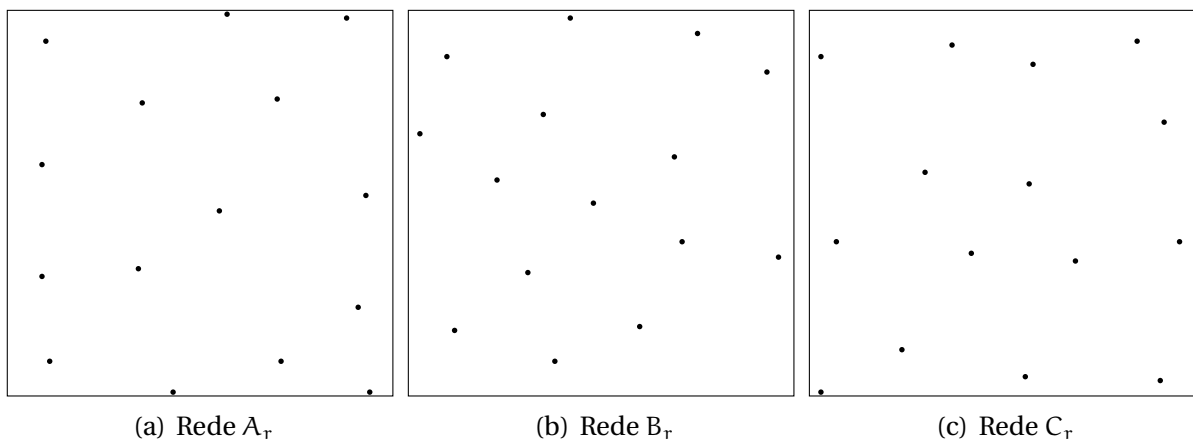


Figura 2.5: Exemplos de posicionamento de sensores segundo um processo repulsivo

Estes três tipos de processos foram identificados como situações de interesse para RSSFs por Frery et al. (2008). Os autores compõem estes processos em um processo chamado $C(n, \alpha)$, onde n determina a quantidade de sensores na área, e α é o nível de atratividade do processo. Mas para definir o processo C formalmente, primeiro vamos introduzir alguns conceitos da teoria de processos pontuais espaciais, especificamente em duas dimensões.

A continuação, na seção 2.2.1 vamos formular os processos pontuais como objetos matemáticos, e também vamos definir os processos pontuais de nosso interesse.

Logo a partir dos processos pontuais básicos, na seção 2.2.2 vamos definir e nos familiarizar com o processo pontual C introduzido por Frery et al. (2008).

2.2.1 Definindo processos pontuais

O processo aleatório de Poisson em uma dimensão (tempo), pode ser utilizado para modelar a seqüência de instantes aleatórios que determinam o acontecimento de eventos. O mesmo processo pode ser utilizado para modelar o padrão aleatório de pontos no espaço d -dimensional. Baddeley (2007) define de maneira conveniente os processos pontuais espaciais a partir do processo pontual unidimensional, e neste trabalho adotamos essa metodologia de formulação, devido à facilidade de compreensão que apresenta.

Em um processo pontual unidimensional, existem várias alternativas teóricas de modelagem. Consideremos por exemplo, os tempos de chegada $T_1 < T_2 < \dots$ que determinam a posição em \mathbb{R}_+ . Estatisticamente, estas variáveis aleatórias são fortemente dependentes e como alternativa útil, poderíamos nos concentrar nos tempos entre chegadas $S_i = T_{i+1} - T_i$ (ver figura 2.6). Uma outra alternativa é estudar o pro-

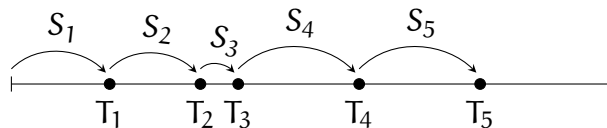


Figura 2.6: Tempos de chegada T_i e entre chegadas S_i

cesso de contagem acumulado, definido a seguir.

Definição 14 (Processo de contagem acumulado) *Sejam $\{T_i\}_{i \in \mathbb{N}}$ variáveis aleatórias que indicam posições em \mathbb{R}_+ , definimos o processo de contagem acumulado N_t como*

$$N_t = \sum_{i=1}^{\infty} \mathbb{1}[T_i \leq t],$$

para todo $t > 0$, sendo $\mathbb{1}[B]$ a função indicadora, definida por

$$\mathbb{1}[B] = \begin{cases} 1 & \text{se } B \equiv \text{TRUE}, \\ 0 & \text{caso contrario.} \end{cases}$$

Analogamente, é útil utilizar o processo de contagem por intervalos (ver definição 15). O intervalo de contagem $N[a, b)$ registra o número de pontos no intervalo $[a, b)$.

Definição 15 (Intervalos de contagem) *Seja N_t o processo de contagem acumulado, definimos o processo de contagem no intervalo $[a, b)$ como*

$$N[a, b) = N_b - N_a,$$

para $0 \leq a < b$.

Para dimensões maiores que um não existe uma maneira natural de ordenar os pontos. Fica, assim, evidente a dificuldade de se trabalhar com os tempos entre chegadas S_i ou os processos de contagem N_t nesse caso. A maneira mais fácil para definir processos pontuais espaciais (d -dimensionais, com $d > 1$), é generalizar o conceito de intervalos de contagem $N[a, b)$ para região de contagem $N(B)$ definida como o número de pontos dentro da região fechada e limitada $B \subset \mathbb{R}^d$ (ver figura 2.7). As va-

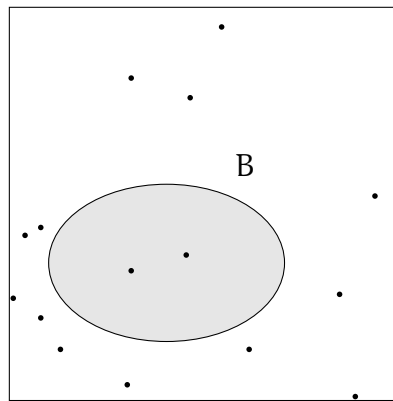


Figura 2.7: Região de contagem $N(B) = 2$

riáveis $N(B)$ têm propriedades interessantes. Os valores de $N(B)$, de todos os B possíveis, fornecem informação suficiente para reconstruir o processo ponto por ponto. Mais ainda, os pontos do processo são aqueles x tal que $N(x) > 0$. Portanto, definimos um processo pontual da seguinte maneira:

Definição 16 (Processo pontual) *Um processo pontual P é uma coleção de variáveis aleatórias $N(B)$ indexadas pelos subconjuntos B da região $R(P)$, onde $N(\cdot)$ satisfaz as seguintes propriedades:*

1. $A_1 \cap A_2 = \emptyset \Rightarrow N(A_1 \cup A_2) = N(A_1) + N(A_2)$,
2. $N(\emptyset) = 0$,
3. se A_n é uma seqüência decrescente de regiões fechadas limitadas ($A_n \supseteq A_{n+1}$) com $\bigcap_n A_n = A$ então $N(A_n) \rightarrow N(A)$.

Estas propriedades levam a dizer que $N(\cdot)$ é uma medida sob $R(P)$.

Generalmente consideramos que o processo pontual é localmente finito, e simples. Estes conceitos são apresentados nas definições 17 e 18, respectivamente.

Definição 17 (Processo pontual localmente finito) *Um processo pontual P definido por $N(\cdot)$ é dito localmente finito, se*

$$\Pr(N(B) < \infty) = 1,$$

para toda região limitada $B \subset \mathbb{R}^d$.

Esta definição diz que com probabilidade 1 qualquer região limitada somente contém um número finito de pontos.

Definição 18 (Processo pontual simples) *Um processo pontual definido por $N(\cdot)$ é dito simples, se*

$$\Pr(N(\{x\}) \leq 1) = 1,$$

para todo $x \in \mathbb{R}^d$.

Num processo pontual simples, com probabilidade 1 não existem no processo dois pontos coincidentes.

Com estes conceitos já estamos em condições de definir o processo pontual de Poisson em duas dimensões.

Definição 19 (Processo pontual espacial de Poisson) *O processo pontual espacial de Poisson com intensidade $\eta > 0$, é um processo pontual em \mathbb{R}^2 tal que:*

1. *para cada conjunto fechado limitado B , a variável aleatória $N(B)$ segue uma distribuição Poisson com média $\eta\xi(B)$, onde $\xi(B)$ denota a área da região B ,*
2. *se B_1, \dots, B_m são regiões disjuntas, então, $N(B_1), \dots, N(B_m)$ são independentes.*

Nesta definição, a constante η é o número de pontos esperados por área unitária. Se estivéssemos interessados em obter uma instanciação do processo de Poisson com exatamente n pontos, então a seguinte propriedade é importante (ver prova em Baddeley, 2007, página 11):

Teorema 2 (Propriedade condicional: Processo Binomial) *Consideremos um processo pontual espacial de Poisson em \mathbb{R}^2 com intensidade uniforme $\eta > 0$. Seja $W \subset \mathbb{R}^2$ uma região onde $0 < \xi(W) < \infty$. Dado o caso que $N(W) = n$, a distribuição condicional de $N(B)$ para $B \subseteq W$ segue uma distribuição binomial, isto é,*

$$\Pr(N(B) = k \mid N(W) = n) = \binom{n}{k} p^k (1-p)^{n-k},$$

onde $p = \xi(B)/\xi(W)$. *Mais ainda, a distribuição condicional conjunta de $N(B_1), \dots, N(B_m)$ para qualquer $B_1, \dots, B_m \subseteq W$ é a mesma que a distribuição conjunta destas variáveis em um processo binomial.*

Esta maneira de definir um processo binomial resulta muito conveniente quando a região de interesse W é um retângulo $[0, \ell]^2$, com $\ell > 0$. Em outras palavras, dado o caso de ter n pontos de um processo Poisson em W , estes n pontos são condicionalmente independentes e distribuídos uniformemente em W , isto é, cada par das coordenadas dos n pontos, são variáveis aleatórias uniformes.

Para estudar outros tipos de processos, nos quais a intensidade não é a mesma em qualquer região, introduzimos na definição 20 o conceito de função de intensidade.

Definição 20 (Intensidade) *Uma intensidade é uma função $\eta: W \rightarrow \mathbb{R}_+$ tal que $\int_A \eta < \infty$ para qualquer região $A \subset W$.*

Esta definição é necessária para introduzir o processo pontual geral de Poisson. Através de instanciações deste processo geral podemos obter outros processos de interesse.

Definição 21 (Processo pontual geral de Poisson) *Seja S um espaço, e Λ uma medida sobre S . O processo pontual geral de Poisson sobre S com intensidade Λ é um processo pontual tal que:*

1. *para cada conjunto limitado e fechado $B \subset S$, $N(B)$ segue uma distribuição Poisson com média $\Lambda(B)$,*
2. *se B_1, \dots, B_m são conjuntos compactos disjuntos, então $N(B_1), \dots, N(B_m)$ são independentes.*

O processo pontual não-homogêneo de Poisson é um caso particular da definição 21, quando η é constante sobre W .

Definição 22 (Processo pontual de Poisson não-homogêneo) *O processo pontual de Poisson não-homogêneo em \mathbb{R}^2 com intensidade $\eta(u)$, $u \in \mathbb{R}^2$ é definido utilizando $S = \mathbb{R}^2$ e $\Lambda(B) = \int_B \eta(u) du$ na definição 21.*

Um outro processo de interesse é o processo pontual *Simple Sequential Inhibition* (SSI) de Matérn definido a contunuação.

Definição 23 (Processo pontual SSI de Matérn) *Denotado $M(n, r)$, é definido iterativamente pelo algoritmo que tenta colocar sequencialmente n pontos sobre a região W tendo como restrição a impossibilidade de dois pontos estarem a uma distância inferior a r . Recursivamente, enquanto os n pontos não estiveram posicionados, o algoritmo posiciona um novo ponto seguindo uma distribuição uniforme com o seguinte critério: se a distância euclideana do novo ponto a qualquer ponto já posicionado for menor que r , então o ponto é rejeitado e o processo inicia uma nova iteração para gerar o proximo ponto candidato. Se o número de tentativas no posicionamento de um ponto passar o limite de iterações t_0 predefinido, então o processo de posicionamento termina.*

O critério de exclusão do processo SSI pode ser generalizado para qualquer forma geométrica, não se limitando a círculos de raio r .

Com os processo definidos nesta seção, já estamos em condições de introduzir processos compostos. A continuação na seção 2.2.2 definimos o processo composto C proposto por Frery et al. (2008).

2.2.2 O processo pontual C

Neste trabalho vamos utilizar o processo $C(n, a)$ proposto por Frery et al. (2008) para modelar o posicionamento dos nós sensores no espaço, onde n é a quantidade de nós que desejamos posicionar, e a é o parâmetro que determina a atratividade. Sem perda de generalidade, vamos definir a região de interesse como $W = [0, 1]^2$. O processo $C(n, a)$ é uma composição de três processos pontuais espaciais:

1. o processo pontual espacial Binomial $B(n)$, introduzido no teorema 2,
2. o processo pontual espacial SSI de Matérn $M(n, r(p))$ com parâmetro de exclusão $r(p) = (1/\sqrt{n} - 0.05)p$ e o número de iterações $t_0 = \infty$ (pois a convergência é garantida), onde $p \in (0, 1)$ é um parâmetro que vamos utilizar para variar a exclusão,

3. o processo pontual espacial $S(n, a, \eta)$ composto por dois processos Poisson, um aplicado com intensidade $a\eta$ sobre a região alvo $[0, 1/4]^2$, e outro com intensidade η na região $([0, 1/4]^2)^c \cap [0, 1]^2$ (ver figura 2.9).

Para compreender melhor a parametrização do processo $M(n, r(p))$, na figura 2.8 ilustramos por que a máxima inibição é $r = 1/\sqrt{n}$. Simplificando, na figura, fixamos $n = 16$ dentro de $W = [0, 1]^2$. É importante considerar que W encontra-se num contexto maior que é \mathbb{R}^2 . Desta maneira, percebemos que a maior inibição possível é $1/\sqrt{16} = 1/4$ (ver figura 2.8). Com esta inibição, o processo geraria uma grade regular. A partir desta máxima inibição, para convergir rapidamente, adicionamos uma margem de tolerância 0.05 acelerando também os tempos de simulação. Embora a margem introduzida não permita a geração de uma grade regular, esta situação pode ser estudada analiticamente, portanto não vamos nos preocupar com a geração desta cenário; sempre que desejarmos sensores regularmente espaçados, essas posições serão geradas de forma exata sem lançar mão do algoritmo de simulação.

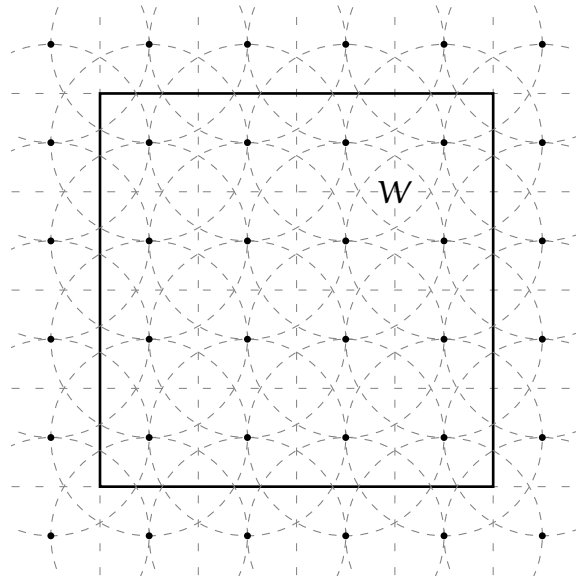


Figura 2.8: Máxima inibição r na área $W = [0, 1]^2$. Com 16 pontos $r = 1/4$

É assim que com a composição destes três processos pontuais espaciais definimos o processo pontual espacial $C(n, a)$ como segue:

Definição 24 (Processo pontual espacial $C(n, a)$) *Sejam $n \in \mathbb{N}$ e $a \in \mathbb{R}$; o processo pontual espacial $C(n, a)$ é*

$$C(n, a) = \begin{cases} M(n, r(e^a)) & \text{se } a < 0, \\ B(n) & \text{se } 0 \leq a \leq 1, \\ S(n, a) & \text{se } a > 1. \end{cases}$$

Os autores implementaram este processo em R com as funções do pacote `spatstat`. Devido ao fato do processo ser implementado originalmente para $[0, 100]^2$, tivemos que fazer alterações minúsculas para dimensioná-lo para $[0, 1]^2$. As figuras 2.10(a),

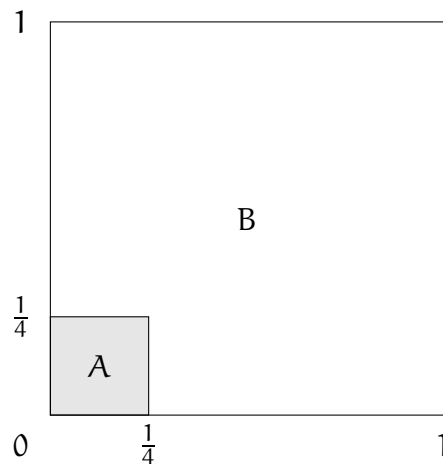


Figura 2.9: Área alvo $A = [0, \frac{1}{4}]^2$, e $B = A^c \cap [0, 1]^2 = ([0, \frac{1}{4}]^2)^c \cap [0, 1]^2$

2.10(b) e 2.10(c) apresentam alguns exemplos de instanciações de processos repulsivos, com $\alpha = -30$. Exemplos para processos independentes podem se apreciar nas figuras 2.10(d), 2.10(e) e 2.10(f). Finalmente as figuras 2.10(g), 2.10(h) e 2.10(i), apresentam implantações com atratividade $\alpha = 30$.

Na seção seguinte, vamos introduzir o modelo de RSSF utilizado ao longo desta dissertação. Os conceitos que foram apresentados até aqui serão de grande importância, especificamente na modelagem do posicionamento dos nós sensores no ambiente.

2.3 Modelagem de RSSFs

Nesta seção, vamos apresentar o formalismo para modelagem de RSSFs mais utilizado na literatura (ver Santi, 2005). Este modelo foi adotado nesta dissertação, e vai ser muito útil principalmente para termos uma linguagem única e livre de ambigüidades.

Destacamos que a falta de ambigüidade que este modelo fornece está sujeita ao tipo de estudo abordado nesta dissertação; outros modelos, se descritos nesta linguagem, não necessariamente estarão livres de ambigüidades. Em particular, essa linguagem não está apta a descrever mobilidade com toda generalidade.

Isto é, nada dizemos respeito de outras áreas de RSSFs, ou inclusive outros estudos relacionados, onde este modelo pode não ser conveniente.

No sentido mais geral, vamos introduzir as RSSFs como redes potencialmente móveis. O primeiro conceito que vamos introduzir é o de configuração de uma RSSF móvel.

Definição 25 (Configuração de uma RSSF móvel) *A configuração de uma RSSF móvel em um espaço d -dimensional é um par $M_d = (S, P)$, onde S é um conjunto de nós sensores, com $|S| = N_s$, e $P: S \times T \rightarrow [0, \ell]^d$ é a função de posicionamento, para algum $\ell > 0$.*

Nesta definição, a função de posicionamento atribui a cada sensor $s \in S$ para cada tempo $t \in T$ coordenadas no cubo d -dimensional de lado ℓ .

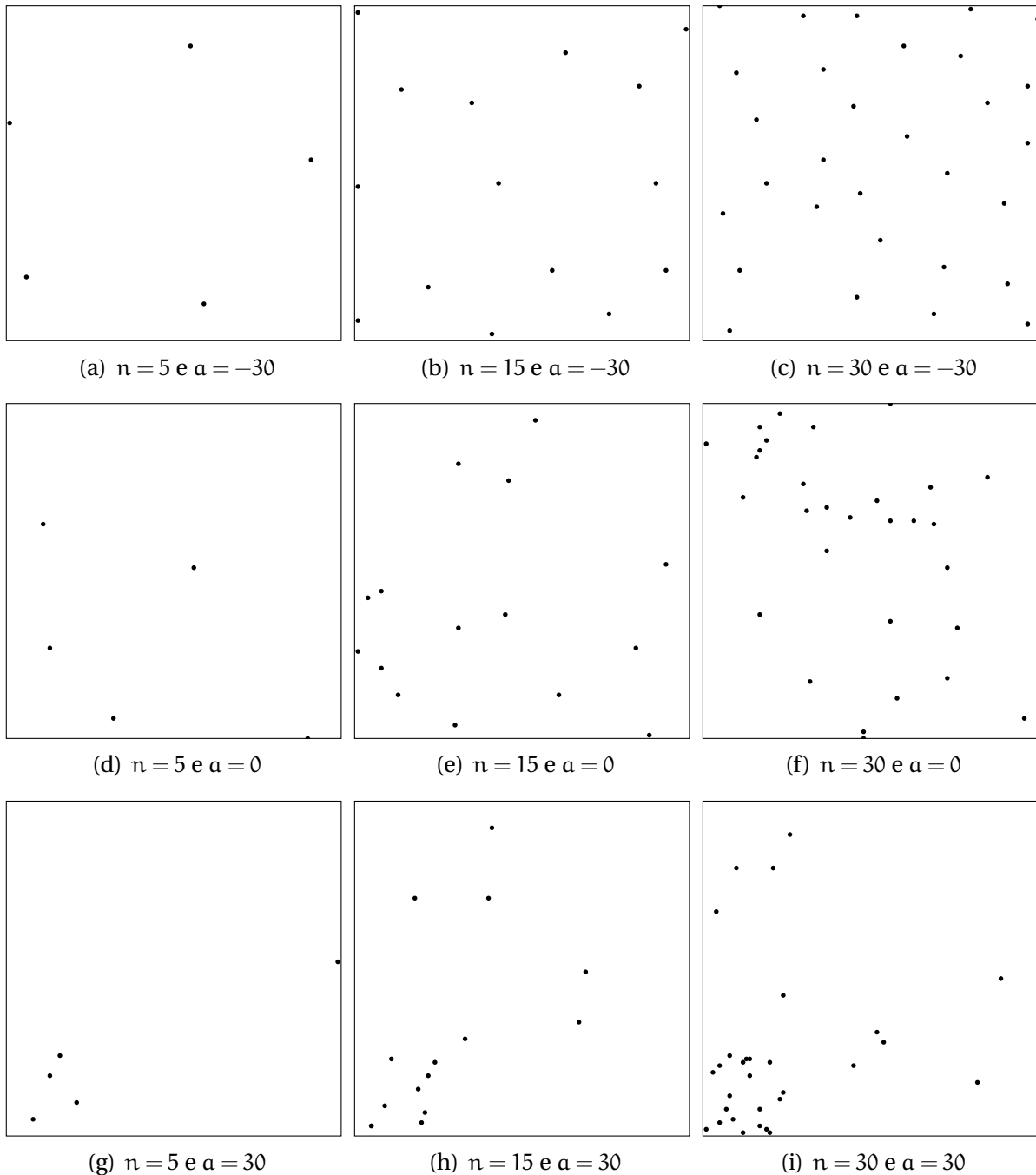


Figura 2.10: Exemplos de processos pontuais espaciais $C(n, \alpha)$

Porém há RSSFs onde os nós sensores têm capacidade de se movimentar, a definição 25 considera também situações realistas onde os nós sensores são movimentados de maneira passiva por causa de fenômenos do ambiente. Por exemplo, numa rede aquática, os nós sensores podem estar sujeitos a movimentos casuais não voluntários, devido a correntes d'água, que poderiam ser tidos em conta no modelo. Por outro lado, quando a mobilidade dos nós sensores não é de interesse existem os conceitos de nó sensor e rede estacionários.

Definição 26 (Sensor e rede estacionários) *Um nó sensor $s \in S$ é dito estacionário se*

o seu posicionamento físico não muda com o tempo. Analogamente, uma RSSF é dita estacionária se cada nó sensor da rede é estacionário. Nesta situação a função de posicionamento pode ser representada por $P: S \rightarrow [0, \ell]^d$.

Para a descrição do posicionamento dos nós sensores podemos utilizar qualquer dos processos pontuais descritos na seção 2.2.

Um outro componente a ser caracterizado é o raio de transmissão dos nós sensores. Neste modelo utilizamos o conceito de função de atribuição de raio de comunicação definido como segue.

Definição 27 (Função de atribuição do raio de comunicação) *Para uma configuração de uma RSSF $M_d = (S, P)$, a função de atribuição do raio de comunicação $RA(S) \rightarrow (0, r_{\max})$ determina o alcance de transmissão de cada sensor da rede. O parâmetro r_{\max} é o raio de transmissão máximo.*

Notemos que esta definição supõe que todos os sensores da rede estão equipados com os mesmos transmissores com alcance não nulo e no máximo r_{\max} .

Há cenários em controle de topologia nos quais este modelo apresentaria problemas. Uma situação onde a definição 27 é muito restritiva é quando a rede é formada por sensores heterogêneos, onde cada um deles pode, por limitações de hardware, possuir um raio máximo de comunicação diferente dos outros. Aqui a atribuição dos raios de comunicação dos sensores pode ser feita ao longo da vida útil dos nós. Nesta situação não estariamos tratando corretamente o problema, ficando expostos a resolver atribuir um raio de comunicação maior do que é possível a um sensor da rede, prejudicando potencialmente a validade da solução.

O caso mais simples de rede homogênea é descrito pela seguinte definição.

Definição 28 (Atribuição r -homogênea do raio de comunicação) *Uma função de atribuição do raio de comunicação constante $RA(i) = r$, para todo $i \in S$ com $0 < r \leq r_{\max}$, é dita atribuição r -homogênea do raio de comunicação.*

É importante em qualquer estudo sobre RSSFs ter uma relação entre o raio de comunicação e o consumo de energia necessário para enviar dados corretamente. Um modelo largamente aceito na comunidade é o da equação (2.1), que relaciona a potência p_i que o nó sensor $i \in S$ necessita para transmitir dados corretamente até o nó sensor $j \in S$, sendo $\delta_{(P(i), P(j))}$ a distância entre eles (Rappaport, 2002):

$$\frac{p_i}{\delta_{(P(i), P(j))}^\alpha} \geq \beta. \quad (2.1)$$

Esta desigualdade indica uma condição que deve ser satisfeita. Aqui, $2 \leq \alpha \leq 6$ é o gradiente distância-potência e $\beta \geq 1$ é a qualidade de transmissão. Baseados na equação (2.1), podemos definir o custo energético da função de atribuição do raio de comunicação RA como:

Definição 29 (Custo energético da função RA) *O custo energético da função de atribuição do raio de comunicação RA é definida por $c_e(RA) = \sum_{i \in S} (RA(i))^\alpha$.*

Notemos que para redes r -homogêneas, o custo energético da função RA é $c_e = |S|r^\alpha$. Em geral, nesta dissertação trabalhamos com redes r -homogêneas, que chamaremos

simplesmente de “redes homogêneas” quando o conhecimento do valor de r não for relevante ou for evidente no contexto.

Uma RSSF é definida baseados na configuração da rede (definição 25) e na função de atribuição do raio de comunicação (definição 27), através do grafo de comunicações induzido.

Definição 30 (Grafo de comunicações) *O grafo de comunicações CG induzido pela configuração $M_d = (S, P)$ e pela função de atribuição do raio de comunicação RA no tempo t é denotado $CG = (S, \vec{E}(t))$, onde $(i, j) \in \vec{E}(t) \equiv RA(i) \geq \delta_{(P(i,t), P(j,t))}$.*

Em situações onde a função de atribuição do raio de comunicação seja r -homogênea (ver definição 28), o grafo de comunicações, em cenários isotrópicos, é um grafo não direcionado (ver definição 3).

A figura 2.11 apresenta a representação do grafo de comunicações $CG = (S, E)$ induzido pela configuração $M_2 = (S, P)$, com $S = \{s_0, s_1, s_2, \dots, s_{14}\}$ e $P = C(15, 0)$. A função de atribuição do raio de comunicação utilizada é r -homogênea, com $r = 0.33$. Nesta figura, vemos que $\{s_0, s_3\} \in E$ como consequência de $\delta_{(P(s_0), P(s_3))} \leq r$. O mesmo acontece para $\{s_0, s_5\}$, $\{s_0, s_{11}\}$ e $\{s_0, s_{12}\}$. Notemos que não estamos indexando o tempo, isto é, estamos considerando uma rede estacionária (definição 26). Os rótulos correspondentes aos nós sensores não serão utilizados no resto da dissertação, já que o que nos interessa é estudar o comportamento da rede como um todo, analisando propriedades gerais da rede como conectividade, e não propriedades locais.

Neste capítulo foram apresentados os principais conceitos utilizados nesta dissertação. Em particular, a teoria espectral de grafos fornece o conceito de conectividade algébrica λ_2 (Fiedler, 1973), que é uma caracterização muito importante que será utilizada no estudo de conectividade. O processo pontual espacial $C(n, \alpha)$ (Frery et al., 2008) foi definido e vai ser utilizado posteriormente como função de posicionamento na configuração de RSSFs. Foi definido também um modelo de RSSFs unificado e livre de ambigüidades que facilitará a leitura e compreensão deste trabalho baseado no trabalho de Santi (2005).

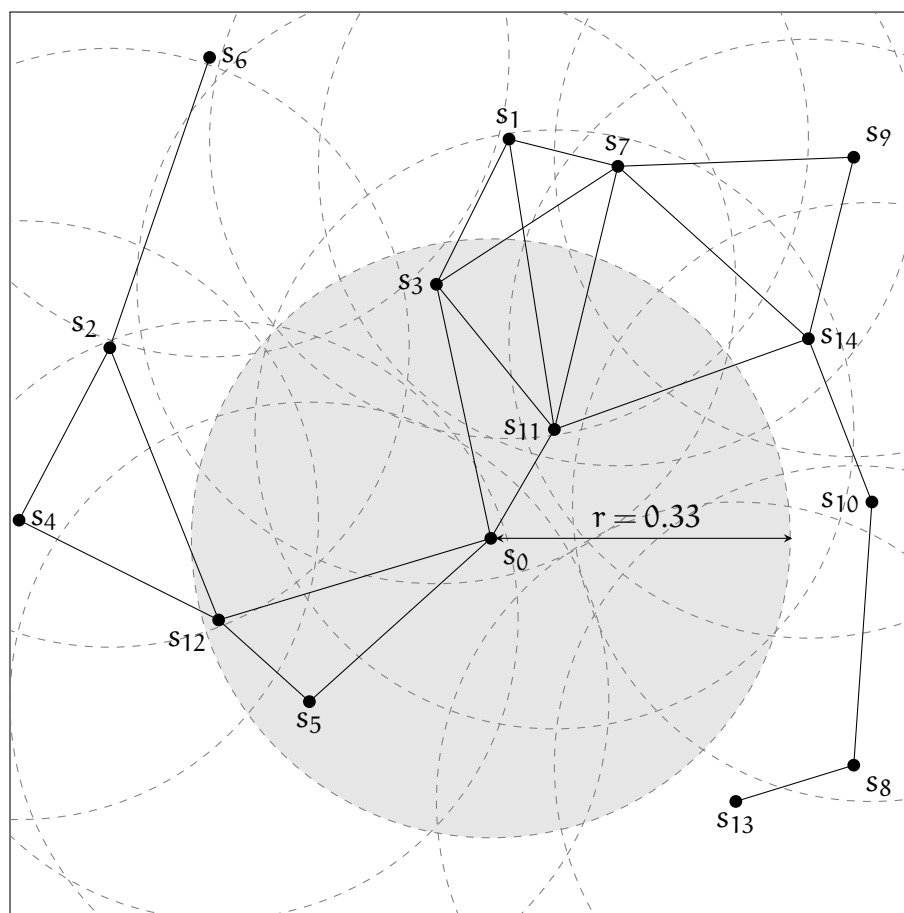


Figura 2.11: Representação gráfica do grafo de comunicações $CG = (S, E)$

Capítulo 3

Ferramentas de simulação e análise de dados

No capítulo anterior vimos os conceitos e notações básicos de RSSF. Este capítulo apresenta resultados a respeito da qualidade de ferramentas de simulação e análise de dados frequentemente empregadas nos estudos de RSSFs.

É comum, no estudo de RSSFs, o uso de ingredientes estocásticos com o objetivo de dotar os modelos de realismo; a natureza raramente se comporta de forma puramente determinística. É possível trabalhar de forma analítica com alguns modelos simples que incorporam ingredientes estocásticos mas, por via de regra, o uso de simulação estocástica é a ferramenta mais empregada. Ela não apenas permite calcular as quantidades de interesse, mas também ilustra o funcionamento do modelo através de geração de cenários típicos e extremos. Mediante simulações estocásticas, utilizando os modelos que contêm fatores estocásticos explícitos, obtemos respostas às nossas perguntas em situações específicas consideradas.

Segundo Velho et al. (2008), o processo de modelagem computacional de muitos sistemas, e em particular do nosso, pode ser representado esquematicamente como na figura 3.1. No “Universo Físico” estão os elementos e propriedades físicas

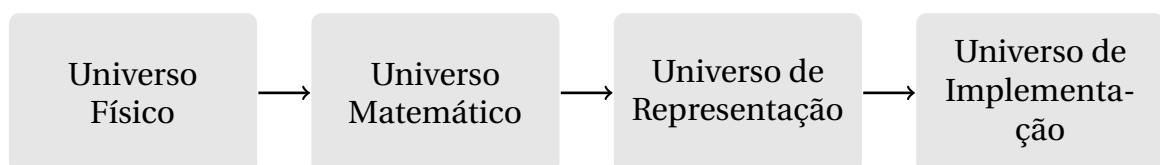


Figura 3.1: Níveis de abstração conceituais na modelagem computacional

que são de interesse no nosso estudo. Logo, apenas algumas propriedades físicas do fenômeno serão escolhidas e modeladas matematicamente com um certo nível de abstração. No “Universo Matemático” determinamos o nível de abstração possível, segundo as propriedades dos objetos e as relações de interesse. A estrutura de representação escolhida encontra-se no “Universo de Representação”, e aqui é onde, dentre outras simplificações, iremos discretizar os processos para deixá-los passíveis de tratamento em computadores digitais. Finalmente, é por meio da implementação em uma linguagem de programação que concluímos o processo de representação. Este último passo é o que no paradigma de abstração chamamos de “Universo de Implementação”.

Muitas ferramentas de simulação e análise de dados fornecem suporte para simulações com elementos estocásticos. Algumas destas ferramentas são mais úteis do que outras, dependendo da situação que desejamos estudar, oferecendo assistência em componentes específicos para a simulação. Este é, geralmente, o critério considerado de maior ponderação no momento de escolher a plataforma de trabalho para simulações em RSSFs, no entanto, pouca importância se dá à precisão numérica destas ferramentas.

Identificamos dois grandes tipos de problemas no “Universo de Implementação” que podem ter impacto negativo nas conclusões tiradas a partir de estudos de simulação: erros devidos ao usuário (falta de conhecimento sobre as funções e configurações do simulador) e erros oriundos da plataforma (arredondamento, erro de acumulação, truncamento e cancelamento).

Neste capítulo discutimos a dificuldade de escolher dentre ferramentas de simulação para RSSFs e apresentamos alguns resultados de precisão numérica de ferramentas para análise de dados. Na seção 3.1 discutimos brevemente quais os problemas que tornam difícil a escolha de uma plataforma adequada para simulações de sistemas distribuídos. As ferramentas de análise de dados serão avaliadas em detalhe na seção 3.2.

3.1 Ambientes de simulação

Atualmente existem vários simuladores utilizados em RSSFs, sendo que a maioria deles é guiada por eventos discretos. Este paradigma reflete fielmente a natureza dos sistemas distribuídos clássicos, onde cada entidade reage, basicamente, a eventos internos ou externos. O comportamento de cada entidade pode ser representado por um ponto do espaço produto cartesiano de eventos e estados das próprias entidades (Santoro, 2007).

Como Andel & Yasinsac (2006) observam, é impossível simular todos os aspectos e situações possíveis de uma rede móvel *ad hoc*, e o mesmo vale para RSSFs. A complexidade de escolher um nível de abstração “correto” na modelagem, faz com que os resultados obtidos possam perder sua validade. Cavin et al. (2002) mostram que, do ponto de vista do usuário, existem inconsistências nos modelos implementados em vários simuladores amplamente utilizados. Os autores implementaram um protocolo simples em três simuladores, a saber, Opnet, NS-2, e GloMoSim, obtendo resultados significativamente diferentes (e conflitantes) em cada um deles.

Uma das fontes de erro comprovadas foi a simplificação extrema do modelo de comunicação disponível nos simuladores. Outros simuladores, como Castalia, tentam se concentrar mais no nível de abstração “mínimo”, e forçam o usuário a lidar com vários parâmetros necessários para configurar o modelo de comunicação, evitando, desta maneira, simplificação extrema.

Por outro lado, supondo resultados válidos nas simulações, inclusive tão válidos que não conseguiríamos diferenciá-los de resultados obtidos em experiências reais, ainda não estamos livres de fontes de erro em nossas conclusões. Isto vai depender da precisão numérica da ferramenta utilizada na análise de dados feita para chegar a essas conclusões.

Na continuação vamos nos concentrar na análise de dados. A avaliação que apresentamos na seção seguinte, vai permitir a escolha de uma ferramenta dentre as di-

ferentes opções, ponderando principalmente a sua precisão numérica.

3.2 Análise de dados

Depois da simulação concluir, imediatamente vem a análise dos dados obtidos. O objetivo deste passo é se familiarizar com os mesmos e aplicar técnicas estatísticas que forneçam respostas às nossas perguntas sobre o fenômeno sob estudo. A aplicação destas técnicas estatísticas implica calcular reiteradamente funções estatísticas simples além de operações numericamente mais complexas. Erros numéricos de pequena magnitude nas funções estatísticas simples podem levar a conclusões totalmente erradas. A precisão numérica, portanto, deve ser um fator decisivo no momento de escolher uma ferramenta para análise de dados. Nesta seção apresentaremos um resumo das análises feitas em Almiron et al. (in press) e Almiron et al. (2008).

Para aferir a precisão numérica de ferramentas estatísticas, McCullough (1998) propõe um protocolo de avaliação tomando como referências conjuntos de dados fornecidos pelo *National Institute of Standards and Technology*. Este protocolo foi aplicado às ferramentas mais conhecidas e amplamente utilizadas na análise de dados. Esta metodologia de avaliação da precisão numérica encontra-se já consolidada e foi empregada e adotada em vários trabalhos. Alguns dos mais interessantes são os de McCullough (1999), McCullough & Wilson (1999), Altman (2002), Bustos & Frery (2005), Altman et al. (2007), Keeling & Pavur (2007) e McCullough & Heiser (2008).

No que segue, mostraremos os resultados de aplicar esse protocolo a várias plataformas de uso frequente. Dividimos a análise segundo a função estatística avaliada.

Para as ferramentas de planilhas eletrônicas como Open Office Calc, Microsoft Excel, Gnumeric, NeoOffice e GNU Oleo, foram avaliadas as funções estatísticas simples como a média, o desvio padrão, o coeficiente de autocorrelação e o estatístico F do procedimento de análise de variância ANOVA.

Oleo é a planilha eletrônica do projeto GNU. A característica principal é a não dependência de sistemas gráficos, isto é, Oleo pode ser utilizado em modo de caracteres. Algumas das suas funções são carregadas do pacote GNU Scientific Library (GLS—Galassi et al., 2003).

No grupo das ferramentas mais completas, onde encontram-se Octave, Ox, Python e R, foram avaliados, além das funções anteriormente mencionadas, procedimentos estatísticos mais complexos como regressão linear e não linear, funções para o cálculo de quantis nas caudas das distribuições mais importantes e geração de números pseudoaleatórios.

3.2.1 Metodologia de avaliação da precisão numérica

À primeira vista, existem dois fatos que dificultam o estudo de qualidade da implementação de um algoritmo numérico em uma ferramenta específica. O primeiro é a indisponibilidade do código fonte do algoritmo em ferramentas que não são de código aberto. Mas, ainda tendo a nossa disposição o código fonte, a complexidade das implementações frequentemente fará com que a tarefa de análise do código seja muito árdua. Por esta razão avaliamos a precisão numérica do ponto de vista do

usuário, isto é, independente da maneira como foi implementada a função: o que nos interessa são os resultados.

O procedimento de avaliação é o seguinte. Para cada ferramenta, submetemos as funções a diferentes amostras, para as quais conhecemos o resultado exato. Logo comparamos o resultado obtido (estimado) com o resultado exato, e obtemos uma métrica que indica a quantidade de dígitos significativos estimados corretamente.

Para calcular esta métrica, utilizamos a função MD (*Matching Digits*) definida como:

Definição 31 (Matching Digits) *A quantidade aproximada de dígitos significativos que coincidem entre dois valores é determinada por:*

$$MD(x, c) = \begin{cases} -\log_{10} \frac{|x-c|}{|c|} & \text{se } c \neq 0, \\ -\log_{10} |x| & \text{caso contrario.} \end{cases} \quad (3.1)$$

Em nossa situação, $MD(x, c)$ indica a quantidade aproximada de dígitos significativos que foram estimados corretamente, onde x é o valor estimado e c o correto. A equação 3.1 é uma generalização da medida proposta por McCullough (1998) que permite analisar a situação $c = 0$.

Como suporte, utilizamos as amostras do *Statistical Reference Datasets* coletadas pelo *National Institute of Standards and Technology*. Estas amostras foram agrupadas, segundo o tipo de procedimento estatístico que deseja se avaliar, em quatro classes: um grupo para funções de resumo univariados, outro grupo para ANOVA, outro para regressão linear, e finalmente outro para avaliar regressão não linear. Todos os valores exatos fornecidos têm quinze dígitos significativos corretos. No caso da avaliação das funções de quantis nas caudas de distribuições, foram considerados como valores exatos, com seis dígitos significativos, os obtidos pela função ELV de Knüsel (1989) (ver também Knüsel, 1998).

3.2.2 Resultados de precisão numérica

Os resultados apresentados a seguir foram produzidos em duas plataformas:

- um *notebook* com processador “i386” (32 *bits*) com os seguintes sistemas operacionais: GNU/Linux Ubuntu Intrepid com o *kernel* “2.6.27-7-generic” e Microsoft Windows Vista SP1.
- uma *desktop* “iMac” com processador “amd64” (64 *bits*) e sistema operacional MacOS X Leopard.

A tabela 3.1 descreve a diversidade da análise indicando, para cada ferramenta, a plataforma de *hardware* utilizada, além do sistema operacional e a versão da ferramenta. Em todos os casos, as versões avaliadas foram as últimas estáveis disponíveis na data de execução deste trabalho.

Na continuação apresentamos os resultados da avaliação de precisão numérica. Algumas das ferramentas podem, eventualmente, estar ausentes em alguma ou algumas das tabelas apresentadas. Isto acontece quando não há uma função específica

Tabela 3.1: Diversidade da análise de precisão numérica

Plataforma		Calc		Excel		Gnumeric		NeoOffice	Octave	Oleo	Ox	Python	R
Hardware	OS	2.4.1	3.0.0	2007	2008	1.8.3	1.9.1	2.2.5	2.9.12	1.99.16	4.10a	2.5.2	2.6.2
i386	Windows	✓	✓	✓			✓						
	Ubuntu	✓	✓			✓			✓	✓	✓	✓	✓
amd64	Mac OS		✓		✓			✓					

implementada para realizar esse cálculo nessa ferramenta, e desta maneira mantemos a fidelidade da análise à plataforma, evitando introduzir dependência de código estranho à mesma.

Nas amostras de referência existem diferentes níveis de complexidade: baixa, média, e alta. Em cada amostra nas tabelas é indicado entre parêntesis o nível de dificuldade. Para facilitar a visualização, quando há diferença dentre as ferramentas, ressaltamos com **negrito** os melhores resultados e com sublinhado os piores valores de precisão, isto para todas as tabelas.

Estatísticos de resumo univariados

As tabelas 3.2, 3.3 e 3.4, apresentam os resultados da quantidade de dígitos significativos estimados corretamente para o cálculo da média amostral, o desvio padrão amostral e o coeficiente de autocorrelação amostral, respectivamente.

Tabela 3.2: MD para a média amostral

Amostra	Calc	Excel	Gnumeric	NeoOffice	Octave	Oleo	Ox	Python	R
Lew (baixa)	15	15	15	15	15	15	15	15	15
Lottery (baixa)	15	15	15	15	15	15	15	15	15
Mavro (baixa)	15	15	15	15	15	15	15	15	15
Michelso (baixa)	15	15	15	15	15	15	15	15	15
NumAcc1 (baixa)	15	15	15	15	15	<u>7</u>	15	15	15
PiDigits (baixa)	15	15	15	15	15	15	15	15	15
NumAcc2 (média)	14	14	15	14	15	14	15	14	15
NumAcc3 (média)	15	15	15	15	15	<u>7</u>	15	14	15
NumAcc4 (alta)	14	14	15	14	14	<u>8</u>	15	15	15

Para o cálculo da média amostral, Calc, Excel, Gnumeric e NeoOffice fornecem a função chamada *average*, e Oleo provê a função *avg*. No caso de Octave, Ox, Python e R, a função para a média amostral é *mean*. Em Python, esta função encontra-se em *stats* dentro do pacote *scipy*. A respeito do desvio padrão, *stdev* é a função utilizada em Calc, Excel, Gnumeric e NeoOffice. A função correspondente em Octave e Oleo é *std*. Em Ox utilizamos a função *moments*, em Python a função *stdev* (do pacote *scipy.stats*) e em R a função *sd* (do pacote *stats*). No caso do coeficiente de autocorrelação de primeira ordem em Calc, Excel, Gnumeric e NeoOffice, utilizamos a função *correl*. Destacamos que Oleo não fornece nenhuma função específica para este procedimento. Ox e R utilizam a função *acf*. Octave utiliza a função *corrcoef* e Python provê a função *pearsonr* (do pacote *scipy.stats*).

No cálculo da média (ver tabela 3.2), todas as funções apresentam resultados bons. Notemos que Oleo teve alguns problemas com as amostras *NumAcc1* e *NumAcc3* e por consequência foi a menos precisa. Por outro lado, Gnumeric, Ox, e R, apresentaram os melhores resultados no cálculo da média obtendo a precisão máxima em todas as amostras.

Tabela 3.3: MD para o desvio padrão amostral

Amostra	Calc	Excel 2007	Excel 2008	Gnumeric	NeoOffice	Octave	Oleo	Ox	Python	R
Lew (baixa)	15	15	15	15	15	15	15	<u>2</u>	15	15
Lottery (baixa)	15	15	15	15	15	15	15	<u>2</u>	15	15
Mavro (baixa)	13	13	9	13	13	15	9	<u>1</u>	13	15
Michelso (baixa)	14	14	8	14	14	15	8	<u>2</u>	13	15
NumAcc1 (baixa)	15	15	15	15	15	15	–	<u>0</u>	15	15
PiDigits (baixa)	15	15	15	15	15	14	15	<u>4</u>	14	15
NumAcc2 (média)	15	11	11	15	15	15	12	<u>3</u>	14	15
NumAcc3 (média)	9	9	<u>1</u>	9	9	9	–	<u>3</u>	9	15
NumAcc4 (alta)	8	8	<u>0</u>	8	8	8	–	<u>3</u>	8	15

No que diz respeito ao cálculo do desvio padrão, a tabela 3.3 mostra que os piores resultados foram aqueles obtidos com Excel 2008, Oleo, e Ox. O símbolo ‘–’, na coluna correspondente a Oleo, indica que não somente não foi estimado nenhum dígito significativo corretamente, mas também que o resultado foi muito longe do valor correto, resultando negativo o valor obtido ao aplicar a função MD. Também observamos que para as amostras que oferecem maior dificuldade numérica, a saber, *NumAcc3* e *NumAcc4*, a única ferramenta que não teve problemas foi R, sendo desta vez a única que obteve a máxima precisão em todas as amostras qualquer seja a sua dificuldade.

Concluindo com as funções para o cálculo de estatísticos de resumo univariados, na tabela 3.4 encontram-se os resultados para o cálculo do coeficiente de autocorrelação de primeira ordem. Claramente percebe-se grande diferença por parte de Ox e R, em relação ao resto das ferramentas, em todas as amostras. Particularmente na

Tabela 3.4: MD para o coeficiente de autocorrelação amostral

Amostra	Calc	Excel	Gnumeric	NeoOffice	Octave	Ox	Python	R
Lew (baixa)	3	3	3	3	5	14	2	15
Lottery (baixa)	2	2	2	2	4	14	2	15
Mavro (baixa)	2	2	2	2	4	13	1	13
Michelso (baixa)	3	3	3	3	8	13	3	13
NumAcc1 (baixa)	0	0	0	0	0	15	-	15
PiDigits (baixa)	4	4	4	4	3	15	3	13
NumAcc2 (média)	3	3	3	3	7	15	3	14
NumAcc3 (média)	3	3	3	-	3	12	-	14
NumAcc4 (alta)	3	3	3	2	3	11	2	14

amostra NumAcc1, salvo Ox e R, nenhuma ferramenta obteve um dígito significativo estimado de maneira correta. Em todos estes casos, o valor correto é $c = -0.5$ sendo os valores estimados $x \approx -1$, a exceção de Python onde o resultado foi mais distante ainda.

Análise de variância

Para ANOVA, *National Institute of Standards and Technology* oferece duas amostras ‘reais’ (isto é, oriundas da prática) obtidas em pesquisas não relacionadas à análise de precisão numérica, e nove geradas a partir da metodologia apresentada por Simon & Lesage (1989). Na tabela 3.5, observamos que os melhores resultados foram atingidos com Gnumeric e R. Por outro lado, Python teve muita dificuldade em todas as amostras de média e alta complexidade.

Para obter o estatístico F em Excel 2007 e Gnumeric, utilizamos a ferramenta disponível para essa finalidade embutida no pacote de “análise de dados” que pode ser acionada através da interface. Por outro lado, com Octave, Python e R, utilizamos as funções `anova`, `f_oneway` (do pacote `scipy.stats`) e `aov` (do pacote `stats`), respectivamente.

Regressão linear

Para regressão linear, *National Institute of Standards and Technology* oferece amostras com ajustes correspondentes a modelo lineares, quadráticos, polinomiais e multilíneares. As funções utilizadas na avaliação foram as seguintes: em Octave a função `regress`, em Ox a classe `PcFiml` e em R a função `lm` do pacote `stats`. Python não foi avaliado já que não provê uma função predefinida para este procedimento.

A tabela 3.6 resume os resultados obtidos. Podemos observar que tanto Octave quanto Ox não conseguem calcular os residuais com boa precisão. A respeito dos coeficientes, as três ferramentas avaliadas obtêm resultados bons, destacando-se R sobre o resto.

Tabela 3.5: MD para o estatístico F de ANOVA

Amostra	Excel 2007	Gnumeric	Octave	Python	R
SiRstv (baixa)	13	13	12	8	13
SmLs01 (baixa)	15	15	14	13	15
SmLs02 (baixa)	14	15	13	11	15
SmLs03 (baixa)	13	15	12	11	15
AtmWtAg (média)	10	10	8	<u>0</u>	9
SmLs04 (média)	10	10	8	<u>0</u>	10
SmLs05 (média)	10	10	8	–	10
SmLs06 (média)	10	10	6	–	10
SmLs07 (alta)	4	4	<u>2</u>	–	4
SmLs08 (alta)	<u>1</u>	4	<u>2</u>	–	4
SmLs09 (alta)	–	4	–	–	4

As expressões ‘NA’ nesta tabela indicam que não foram obtidos resultados, isto é, o algoritmo não retornou nenhuma resposta numérica. Esta situação, embora pareça absurda, é um bom sinal. Quando um algoritmo não consegue satisfazer as necessidades para retornar uma boa resposta, é desejável que se abstenha de responder de maneira incorreta.

Regressão não linear

Octave, Ox, e Python estão ausentes nesta avaliação, já que nenhum deles dispõe de funções predefinidas para este tipo de procedimento. Para avaliar a função de regressão não linear de R são utilizados modelo exponenciais e racionais, entre outros.

Cada conjunto de dados inclui dois pontos iniciais para os algoritmos, o primeiro deles é um valor distante da solução, e o segundo é mais próximo. Somente quando a função não chegou a uma solução a partir do primeiro ponto inicial fornecido, executamos o algoritmo com o segundo ponto inicial. Denotamos estes casos, na tabela, com ‘*’.

Em todas as experiências fizemos uso da função `nls` do pacote `stats`. Na tabela 3.7 vemos valores de $MD \geq 4$ com exceção do residual para a amostra `Lanczos2`, sendo que, segundo McCullough (1998), valores de quatro a cinco dígitos significativos estimados corretamente representam uma precisão razoável.

Novamente, como nos resultados de regressão linear, encontramos expressões ‘NA’, neste caso para a amostra `Lanczos1`. Reforçamos dizendo que esta situação é melhor que o caso da amostra `Lanczos2`. Para compreender, vamos trazer um exemplo real apresentado por McCullough & Heiser (2008). Consideremos a amostra `Misra1a` para ajustar com o modelo $y = \beta_1(1 - \exp(-\beta_2x)) + \epsilon$, com um tamanho de amostra $n = 14$ e valores iniciais para $(\beta_1, \beta_2) = (500, 0.0001)$, os valores corretos finais são $\hat{\beta}_1 = 238.94212918$ e $\hat{\beta}_2 = 0.00055015643181$. De maneira predefinida, Excel

Tabela 3.6: Valores de MD menores para o coeficiente $\hat{\beta}$ e o desvio padrão residual RSD aplicando regressão linear

Amostra	Octave		Ox		R	
	$\hat{\beta}$	RSD	$\hat{\beta}$	RSD	$\hat{\beta}$	RSD
Norris (baixa)	12	<u>1</u>	13	5	12	13
Pontius (baixa)	11	<u>1</u>	<u>1</u>	<u>0</u>	12	12
NoInt1 (média)	14	<u>2</u>	14	5	14	14
NoInt2 (média)	15	<u>2</u>	15	<u>4</u>	15	15
Filip (alta)	7	<u>1</u>	NA	NA	NA	NA
Longley (alta)	7	<u>0</u>	12	<u>4</u>	12	14
Wampler1 (alta)	8	8	NA	NA	9	9
Wampler2 (alta)	12	12	NA	NA	12	14
Wampler3 (alta)	8	<u>0</u>	9	<u>4</u>	9	14
Wampler4 (alta)	7	<u>0</u>	9	<u>4</u>	8	14
Wampler5 (alta)	5	<u>0</u>	9	<u>4</u>	6	14

2007 utiliza um algoritmo com o método “*Newton*”, mas com opção para utilizar o método “*Conjugate*”. Também de maneira predefinida, a tolerancia de convergência é 0.001. Da maneira predefinida, com o método *Newton*, obtemos a mensagem “*Solver has converged to the current solution. All constraints are satisfied*”, e o resultado foi $\hat{\beta}_1 = 454.12$ e $\hat{\beta}_2 = 0.0002676$. Se aplicamos o outro método, o de *Conjugate*, aparece na tela a mensagem “*Solver has converged to the current solution. All constraints and optimality conditions are satisfied*”. Desta vez, parece ser que o resultado da aplicação do algoritmo foi melhor, já que as *condições de otimalidade* estão agora presentes, diferente da primeira experiência onde tais condições não foram satisfeitas, mas os resultados desta vez foram $\hat{\beta}_1 = 500$ e $\hat{\beta}_2 = 0.0002422$!

Claramente, as mensagens fazem com que nós, os usuários, tenhamos certeza de haver achado a solução ótima, e conseqüentemente estaremos concluindo afirmações erradas a respeito do problema, o que é pior, sem dúvida, do que não ter resolvido nada.

Quantis em caudas de distribuições

Lembraremos a seguir a definição de quantil.

Definição 32 (Quantil) *Seja X uma variável aleatória com função de distribuição acumulada F_X , e $p \in [0, 1]$. O quantil de ordem p para F_X é o menor q_p tal que*

$$F_X(q_p) = \Pr(X \leq q_p) = p.$$

Para variáveis aleatórias contínuas, geralmente F_X é estritamente crescente, e logo podemos determinar o quantil de ordem p da seguinte maneira:

$$F_X^{-1}(p) = q_p.$$

Tabela 3.7: MD menores para o coeficiente $\hat{\beta}$ e o desvio padrão residual RSD em R aplicando regressão não linear

Amostra	$\hat{\beta}$	RSD	Amostra	$\hat{\beta}$	RSD
Chwirut1 (baixa)	5	6	Mgh17* (média)	5	5
Chwirut2 (baixa)	4	5	Misra1c (média)	8	6
Danwood (baixa)	8	7	Misra1d (média)	6	6
Gauss1 (baixa)	6	6	Rozzman1 (média)	5	6
Gauss2 (baixa)	6	6	Nelson* (média)	5	5
Lanczos3 (baixa)	6	4	Bennett5 (alta)	5	4
Misra1a (baixa)	6	6	Boxbod* (alta)	5	5
Misra1b (baixa)	6	6	Eckerle4* (alta)	7	7
Enso (média)	4	5	Mgh09* (alta)	4	4
Gauss3 (média)	6	5	Mgh10* (alta)	6	6
Hahn1 (média)	6	6	Rat42* (alta)	7	6
Kirby2 (média)	6	6	Rat43* (alta)	5	5
Lanczos1 (média)	NA	NA	Thurber (alta)	5	4
Lanczos2 (média)	7	-			

A precisão numérica das funções de quantis de ordem $p = 2 \cdot 10^{-7}$ das distribuições gaussianas padrão, χ^2 com um grau de liberdade, e t-student com um grau de liberdade foram aferidas. Também avaliamos a função de quantis de ordem $p = 10^{-5}$ para a distribuição F com (1, 1) graus de liberdade. Na tabela 3.8 apresentamos os resultados obtidos de aplicar a função MD para os valores estimados pelas diferentes ferramentas analisadas quando comparadas com os valores corretos.

Tabela 3.8: MD para as funções de quantis de distribuições

Distribuição	Octave		Ox		R	
	função	MD	função	MD	função	MD
gaussiana	norminv	6	quann	6	qnorm	6
F	probit	6	quanf	6	qf	6
t-student	tinv	<u>0</u>	quant	4	qt	6
χ^2	chi2inv	<u>0</u>	quanchi	3	qchisq	3

Tendo em conta que os valores máximos de precisão possíveis nesta análise é seis, percebemos que, a exceção de χ^2 , R obteve ótimos resultados. Por outro lado, observamos que Octave tem serios problemas com as distribuições t-student e χ^2 .

Geração de números pseudoaleatórios

Para geração de números pseudoaleatórios, Octave, Ox, Python e R utilizam como predefinido o algoritmo Mersenne-Twister (ver Matsumoto & Nishimura, 1998). Este algoritmo é um gerador com dimensionalidade de 623, um período de 2^{19937} e um

conjunto 624-dimensional de sementes inteiras de 32 *bits*. Este gerador passa o teste *Diehard* de Marsaglia (1998), igual a outros geradores que podem ser utilizados nestas ferramentas.

Para maior informação das alternativas de algoritmos de geração de números pseudoaleatórios nestas plataformas, recomendamos ver a seção pertinente em Al-miron et al. (in press).

3.3 Comentários

O procedimento de avaliação apresentado neste capítulo pode ser utilizado futuramente como ponto inicial para determinar qual ou quais das plataformas de simulação de sistemas distribuídos existentes são adequadas para estudar o desempenho de RSSFs sob diferentes cenários.

Os resultados apresentados e a diversidade de pacotes numéricos e gráficos disponíveis para a ferramenta R nos levaram a adotá-la como plataforma de análise de dados para este trabalho. Já que não precisamos simular ambientes distribuídos nesta dissertação, também foi adotada como ferramenta de simulação.

Este capítulo apresentou resultados de precisão numérica de ferramentas estatísticas na análise de dados. Como fruto da avaliação, surge R como a melhor opção. Também ficou em evidência a falta de validade de conclusões tiradas a partir de ferramentas amplamente utilizadas na comunidade científica.

Capítulo 4

Estudos de conectividade

Uma RSSF, como apresentamos no capítulo 1, tem por objetivo principal recoletar informação do ambiente, por meio de diversos sensores e enviar alguma função desses dados, por nós intermediários, até um ou mais nós *sink*. Para este objetivo ser atingido, precisamos garantir que exista ao menos um caminho simples (ver definição 6 na página 13) de cada nó sensor a algum nó *sink*.

Mas existem outras necessidades em termos de conectividade que devem ser consideradas. Por exemplo, no processo de sensoriamento, é sabido que os nós sensores estão sujeitos, freqüentemente, a erros de colheita de dados devido a diferentes fatores (erros de *hardware* ou aplicação de técnicas de *clustering* baseadas em distância geográfica, por exemplo). Conseqüentemente, o trabalho colaborativo tanto de processamento de informação quanto de sinais resulta uma tarefa crítica (Zhao et al., 2003; Ghosh & Das, 2006; Nakamura et al., 2007). Desta maneira, fica evidente a necessidade de garantir que o grafo que representa a rede seja fortemente conexo (ver definição 8 na página 14).

Uma atividade de grande importância na implementação de uma RSSF é o processo de posicionamento dos nós sensores no ambiente. Este processo influi diretamente sobre a eficiência no funcionamento. Os nós sensores podem ser distribuídos regularmente formando uma grade (ver Reis et al., 2007), ou posicionando individualmente cada um deles em pontos estratégicos de interesse (Ghosh & Das, 2006); mas na prática, devido a custos e a limitações físicas, tipicamente os nós são posicionados aleatoriamente, seja jogando os mesmos desde alguma altura com algum critério (ou trajetória) ou simplesmente posicionando-os rapidamente da maneira mais espalhada possível respeitando uma distância máxima tal que os nós consigam se comunicar entre si, no entanto, sem estar preocupados com as coordenadas exatas do posicionamento.

Os posicionamentos com diferentes graus de aleatoriedade apresentam duas vantagens substanciais:

1. redução de custos na implementação da RSSF, especificamente no processo de posicionamento dos nós sensores no ambiente,
2. custo de manutenção da RSSF funcional, do ponto de vista da conectividade, relativamente baixo.

O item 1 é facilmente comprovável. Na afirmação do item 2 consideramos a situação típica onde os nós da RSSF vão “morrendo” e ao mesmo tempo é requerido

que novos nós sensores sejam posicionados em substituição dos que já não estão em condições de continuar suas tarefas, seja por falta de energia ou falha de hardware. Desta maneira podemos considerar o problema de determinar a proporção de sensores em funcionamento necessária para manter uma rede conexa, dado que estamos frente a um processo de posicionamento de sensores aleatório segundo alguma lei de probabilidade. Em particular, vamos considerar o processo pontual espacial $C(n, \alpha)$ definido na seção 2.2.2 para descrever tal lei.

Nesta dissertação estamos interessados em estudar o “comportamento” da conectividade como um fenômeno de percolação no plano contínuo. Sem perda de generalidade para os propósitos de interesse desta dissertação, estudaremos detalhadamente este fenômeno em $[0, 1]^2$. A formulação deste problema teve início na literatura com o estudo de problemas de *controle de topologia* no ponto de vista *homogêneo*, especificamente em redes densas com nós distribuídos aleatoriamente de maneira independente.

Os resultados apresentados baseam-se no estudo do ponto crítico do raio de transmissão que faz a rede conexa. Além da distribuição de nós aleatória e independente, como antecipamos anteriormente, analisamos o fenômeno em cenários onde os nós são distribuídos com diferentes graus de atratividade modelados com o processo $C(n, \alpha)$.

4.1 Controle de topologia

Em RSSFs, como vimos anteriormente, a topologia não é fixa devido a vários fatores. Em particular, temos a possibilidade de reconfigurá-la por varios meios: variando o raio de transmissão dos sensores, controlando a densidade (quantidade de sensores por unidade de espaço), ou por meio da capacidade de movimentação nos sensores, no caso de sensores móveis.

O controle de topologia é uma área que procura administrar o raio de transmissão dos nós sensores com os seguintes fins:

1. manter alguma propriedade no grafo de comunicações associado, sendo tipicamente a conectividade o objetivo,
2. reduzir o custo energético c_e da rede, que está fortemente relacionado com o raio de transmissão utilizado pelos nós sensores.

Há varios enfoques que tratam este problema sob diferentes pontos de vista, dependendo do cenário em estudo.

Seguindo a classificação de Santi (2005), a primeira grande distinção é entre controle de topologia homogêneo e não homogêneo. No primeiro caso, que representa o tipo de controle de topologia mais simples, consideramos que os nós sensores possuem o mesmo raio de comunicação. Já no controle de topologia não homogêneo é permitido que cada nó tenha um raio de transmissão diferente, sempre que dentro do intervalo $(0, r_{\max}]$ ditado pelas características operacionais dos nós. Neste último tipo de controle de topologia, existem três categorias de soluções: as baseadas na posição, as baseadas na direção, e as baseadas na vizinhança.

Nesta dissertação estudaremos o caso mais simples de controle de topologia, isto é, o homogêneo; para quem tiver interesse em algumas das técnicas aplicadas para

problemas de controle de topologia não homogêneo, um bom ponto de partida é o trabalho de Santi (2005), que apresenta uma visão clara e geral da área.

Quando o alcance de transmissão dos sensores é muito grande em relação a densidade numa RSSF, a comunicação perde eficiência em termos de custo energético e capacidade, como ilustramos na figura 4.1(a). Lembremos que a energia gasta para transmitir mensagens é de ordem quadrática em ambientes ideais, e de quarta ordem em média em situações práticas, chegando a ser em casos extremos até de sexta ordem.

Quanto à limitação da capacidade de tráfego, especificamente, quando um nó sensor está transmitindo dados, o resto dos sensores que estão dentro do raio de comunicação do transmissor não podem utilizar o meio de comunicação; caso contrário aconteceriam colisões e a integridade dos dados estaria comprometida. Notemos que na figura 4.1(a) alguns nós têm grau muito grande, $k \geq 8$. Isto implica que a largura de banda dos *links* de comunicação deve ser dividida por 8 ou mais fragmentos, qualquera seja a técnica de divisão utilizada.

Por outro lado, se o alcance de transmissão for muito pequeno, a conectividade se perde e conseqüentemente também a funcionalidade. A figura 4.1(b) ilustra esta situação, onde a rede encontra-se fragmentada.

Finalmente, consideremos a figura 4.1(c), onde o alcance de transmissão dos nós sensores é o mínimo que garante conectividade. Esta situação não somente é eficiente no gasto de energia, mas também representa a situação ótima absoluta (com sensores homogêneos), considerando como função objetivo a maximização da capacidade de tráfego e como restrição a preservação da conectividade (Gupta & Kumar, 2000).

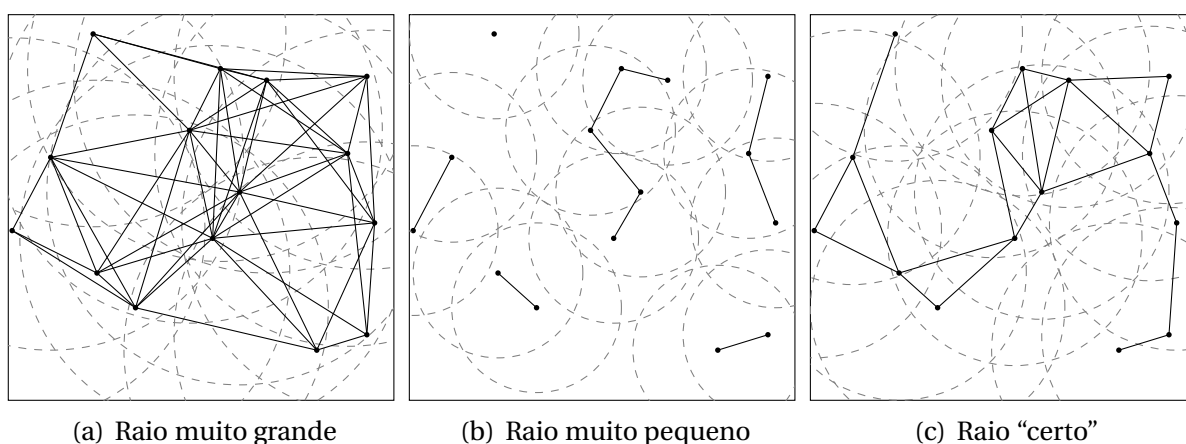


Figura 4.1: Escolha do alcance de transmissão dos nós sensores

Num cenário com nós homogêneos, o problema principal é conhecido como *alcance de transmissão crítico* definido como segue:

Definição 33 (Alcance de transmissão crítico) *Sejam n nós sensores na região $\mathcal{R} = [0, \ell]^d$, com $d \in \{1, 2, 3\}$. Definimos o raio de transmissão crítico como o mínimo necessário que os sensores devem ter para garantir conectividade em uma rede de sensores homogêneos.*

Outra maneira possível de interpretar este problema é perguntando-nos o seguinte:

qual é a densidade mínima necessária que a rede deve possuir para garantir conectividade em uma RSSF r -homogênea?

Numa primeira leitura, o problema tal como formulado na definição 33 pode parecer impreciso pela falta de definição da regra de localização dos sensores. Há, porém, uma resposta trivial para o problema em qualquer situação: há conectividade em \mathcal{R} quando o raio é de pelo menos $\ell\sqrt{d}$. Esta situação extrema, contudo, não é de interesse prático.

A motivação principal para estudar o problema do alcance de transmissão crítico, em inglês *Critical Transmission Range* (CTR), surge do fato de sensores de baixo custo de aquisição não terem a capacidade de ajustar dinamicamente o alcance de transmissão dos seus dispositivos de comunicação. Assim, uma vez fixado o raio r de transmissão e posicionados no ambiente, o custo de mudar o raio de comunicação resulta muito alto, ou simplesmente não é possível fazê-lo. Nesta situação, resulta conveniente determinar o mesmo raio de transmissão r para todos os sensores da rede, e desta maneira manter *links* de comunicação bilaterais. Por definição, para reduzir o consumo de energia e aumentar a capacidade da rede sob estes supostos, a melhor escolha é o CTR.

No caso de posicionamento aleatório, há duas abordagens típicas para modelar redes sem fios *ad hoc*, ambas baseadas na teoria de grafos aleatórios. Uma delas é o grafo aleatório binomial (ou Bernoulli), denotado por $\mathcal{G}_p(n)$, e a outra é o grafo aleatório uniforme (ou Erdős-Rényi) $\mathcal{G}_m(n)$. O grafo aleatório binomial pode ser definido da seguinte maneira:

Definição 34 (Grafo completamente aleatório ou bernoulli) *Sejam $n \geq 2$ nós e a probabilidade $0 \leq p \leq 1$. O grafo aleatório binomial, denotado por $\mathcal{G}_p(n)$, é o resultado de observar $b_1, \dots, b_{\binom{n}{2}}$ ocorrências independentes da variável aleatória A com distribuição Bernoulli e probabilidade p de sucesso e de colocar o arco respectivo no grafo.*

Desta maneira, a probabilidade de qualquer grafo $G \in \mathcal{G}_p(n)$ com m arestas é $p^m q^{N-m}$, sendo $N = \binom{n}{2}$ a quantidade de arestas possíveis.

O grafo aleatório uniforme é definido como:

Definição 35 (Grafo aleatório uniforme ou Erdős-Rényi) *Para $0 \leq m \leq N$, um grafo aleatório uniforme, denotado por $\mathcal{G}_m(n)$, é o resultado de escolher uniformemente um grafo no conjunto de todos os grafos que contêm exatamente n vértices e m arestas.*

Notemos que, nesta classe de grafos, a probabilidade de qualquer grafo $G \in \mathcal{G}_m(n)$ com n vértices e m arestas é

$$\binom{N}{m}^{-1}.$$

Nenhum destes modelos, para os quais já foram feitas análises de conectividade em diferentes aplicações, são adequados para descrever situações reais. O maior problema é a falta de consideração do alcance de transmissão dos nós no modelo (Santi, 2005).

Para estudar o fenômeno de conectividade, a teoria de *grafos aleatórios geométricos*, em inglês, *Geometric Random Graphs* (GRG), representa um modelo adequado. Formalmente, definimos GRG da seguinte maneira:

Definição 36 (Grafo aleatório geométrico) É um grafo $\mathcal{G}_r(n, \ell) = (V, E)$ onde os n vértices se distribuem no espaço d -dimensional $[0, \ell]^d$ segundo a distribuição induzida pela função de densidade de probabilidade f . Uma vez atribuídas as posições dos vértices, a aresta $(u, v) \in E$ estará presente se e somente se $\delta_{(P(u), P(v))} < r$.

Esta classe de grafos, largamente utilizada em problemas de RSSFs, coincide em expresividade com a noção de grafo de comunicações apresentado na definição 30 da página 26.

Na próxima seção estudaremos o fenômeno de conectividade no caso de posicionamento dos nós sensores de maneira aleatória independente, isto é, consideraremos um processo pontual espacial Binomial.

4.2 Conectividade em processos binomiais

A caracterização do CTR permite, ao momento de projetar uma RSSF, responder perguntas básicas tais como a quantidade de nós sensores necessária em uma região ou o valor mínimo de atribuição do raio de transmissão aos sensores que deve ser utilizada para garantir conectividade, ou inversamente, dado que dispomos de uma tecnologia de transmissão específica, com um raio de transmissão r , podemos decidir quantos nós sensores precisamos utilizar para cobrir uma região determinada, mantendo conectividade na rede.

Na literatura foram propostas diferentes soluções ao problema do CTR dependendo da informação da qual dispomos ao momento de implementar a rede, especificamente sobre o posicionamento dos sensores no ambiente. Se a posição de cada sensor é totalmente conhecida, então o CTR é o peso da maior aresta da árvore de cobertura mínima (MST – *Minimum Spanning Tree*) euclideana contruída a partir do grafo de comunicações (Penrose, 1997). Esta situação é ilustrada na figura 4.2.

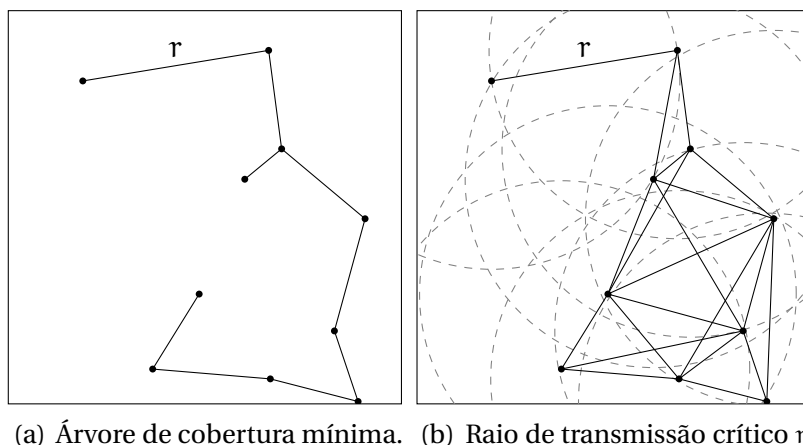


Figura 4.2: Relação entre o MST e o CTR

Este resultado seria de grande utilidade em cenários onde os nós sensores possuam tecnologia GPS (*Global Positioning System*), ou algum outro sistema de referência espacial de cada sensor. Ainda assim, o custo computacional, em termos de mensagens, para uma implementação distribuída, resultaria numa opção proibitiva tratando-se de RSSFs.

Num cenário inverso, onde não temos informação alguma sobre o posicionamento dos nós sensores, o valor mínimo do raio de transmissão que garante conectividade em todos os casos é $r \approx \ell\sqrt{d}$, já que poderíamos ter nós sensores posicionados nos extremos da região de observação. Esta solução é pouco realista e, conseqüentemente, não é relevante. Conforme Santi (2005) afirma, o propósito no estudo do CTR é supor que os nós sensores são posicionados em uma certa região seguindo um processo pontual espacial, para logo caracterizar o CTR com alta probabilidade.

Existem diferentes modelos probabilísticos que podem ser utilizados para caracterizar as relações entre sensores. O modelo maiormente empregado é o de grafos aleatórios geométricos (ver definição 36) em redes densas. A conexão introduzida por Penrose (1997, 1999) deu início a uma serie de trabalhos nesta linha de pesquisa. Em redes 1-dimensionais foi demonstrado que para n nós distribuídos aleatoriamente de maneira uniforme e independente em $[0, 1]$, o CTR é $r = \ln n/n$ com alta probabilidade. Em redes 3-dimensionais, se conhece que se n nós são distribuídos aleatoriamente de maneira uniforme e independente em $[0, 1]^3$, então o CTR é

$$r = \sqrt{\frac{\ln n - \ln \ln n}{n\pi} + \frac{3 \cdot 1.41 + g(n)}{2\pi n}}$$

com alta probabilidade, onde $g(n)$ é uma função qualquer tal que $\lim_{n \rightarrow \infty} g(n) = +\infty$.

A teoria de grafos aleatórios geométricos apresenta algumas desvantagens. Em particular, ela supõe que a região \mathcal{R} onde os nós sensores são posicionados é fixa, observando o comportamento a medida que a densidade da rede cresce ao infinito. Assim, este modelo não deve ser utilizado para tirar conclusões analíticas em redes esparsas.

Modelos similares ao de grafos aleatórios geométricos foram propostos. O de maior êxito e aceitação considera como parâmetro principal o lado ℓ da região de interesse, sendo o raio r e a quantidade n de sensores funções arbitrárias de ℓ . Nesta situação o CTR é analisado assintoticamente quando $\ell \rightarrow \infty$.

Santi et al. (2001) e Santi & Blough (2002, 2003) provaram que supondo que os n nós são posicionados aleatoriamente de maneira uniforme e independente em $\mathcal{R} = [0, \ell]^d$, o CTR é $r = \ell\sqrt{(c \ln n)/n}$ com alta probabilidade, para alguma constante $c > 0$.

Também existem antecedentes de caracterização do CTR por meios mais práticos. Um exemplo deste enfoque é o trabalho de Narayanaswamy et al. (2002), onde os autores apresentam um protocolo chamado COMPOW, que, por meio de heurísticas, tenta determinar o raio mínimo de transmissão que garanta conectividade na rede.

Nesta dissertação estudamos o resultado apresentado por Panchapakesan & Munjunath (2001), onde os autores, empregando grafos aleatórios geométricos sobre um processo pontual espacial binomial, propõem uma função de limiar (*threshold*) para o alcance de transmissão crítico em $[0, 1]^2$:

Teorema 3 (CTR em $[0, 1]^2$) *Seja o grafo $G \in \mathcal{G}_r(n, 1)$ induzido pelo processo pontual $C(n, 0)$ em $[0, 1]^2$. A função de threshold para o grafo G ser conexo é $r(n) = \varphi\sqrt{\frac{\ln n}{n}}$, para algum valor $\varphi > 0$.*

A prova deste teorema pode ser achada em Panchapakesan & Munjunath (2001), tendo em conta que $C(n, 0)$ é um processo pontual binomial com n pontos.

Embora este resultado seja muito forte, é de especial interesse conhecer o valor da constante φ que faz a rede se tornar conexa. Desta maneira, poderíamos fornecer uma função para escolher o CTR para os sensores de uma RSSF distribuída segundo um processo pontual binomial no ambiente. No caso de não dispor da tecnologia para configurar o alcance de transmissão, uma função que retorne a quantidade de sensores necessário para tornar o alcance predefinido igual ao CTR seria muito útil.

Para estudar o fenômeno de conectividade sobre processos pontuais espaciais, o natural é considerar este processo junto com o raio de comunicação como um modelo de percolação no contínuo (Meester & Roy, 1996). Os principais resultados de percolação no contínuo são, no entanto, assintóticos e requerem, dentre outras condições, um número infinito de nós. Os modelos mais simples de percolação no contínuo são os modelos booleanos. Em duas dimensões, estes consideram processos espaciais pontuais em \mathbb{R}^2 , onde são considerados círculos de raio r com centro em cada ponto. Observando as interseções entre círculos, \mathbb{R}^2 é particionado em duas partes: a união das regiões que contêm ao menos um círculo, conhecidas como “regiões ocupadas”, e o seu complemento, conhecida como “regiões de vacância”. Para uma região $A \subset \mathbb{R}^2$, denotamos com $W(A)$ à união de todas as regiões ocupadas que têm interseção não vazia com A . Quando $A = \mathbf{0} = (0, 0)$ é a origem do espaço bidimensional, $W(A) = W(\{\mathbf{0}\}) = W$ representa o conjunto de todas as regiões (componentes) que tem interseção não vazia com a origem. Em percolação no contínuo definimos densidade crítica τ_c como:

$$\tau_c = \inf\{\tau \geq 0 : \Theta(\tau) > 0\},$$

onde τ é a intensidade do processo pontual espacial e $\Theta(\tau)$ é a função de percolação que, por definição, indica a probabilidade de termos uma componente de tamanho infinito que passe pela origem, isto é, $\Theta(\tau) = \mathbb{P}r(d(W) = \infty)$, onde $d(W) = \sup_{x,y \in W} \{\delta_{(x,y)}\}$ é o diâmetro de W .

O estudo da densidade crítica é um problema onde procura-se determinar a menor densidade que garanta a existência de ao menos uma região ocupada de tamanho infinito. Da teoria de percolação no contínuo, se conhece que τ_c é estritamente positivo e finito. Isto implica que para valores $\tau < \tau_c$ a origem tem probabilidade zero de estar contida em uma componente ocupada de tamanho infinito e, no caso de processos estacionários, o mesmo acontece para qualquer ponto em \mathbb{R}^2 . Para valores $\tau > \tau_c$, componentes ocupadas não limitadas, isto é, com diâmetro infinito, existem com probabilidade positiva. Quando $\tau < \tau_c$, dizemos que o sistema está na fase sub-crítica. No caso $\tau > \tau_c$, dizemos que o sistema está na fase super-crítica. Para τ_c dizemos que o sistema está na fase crítica. O sistema se torna interessante fora da região sub-crítica.

Contudo, estudar percolação em sistemas finitos, requer redefinir intuitivamente os conceitos, os quais, infelizmente, não têm um equivalente concreto. Em nossos estudos, para não depender diretamente da potência necessária para garantir transmissão de dados sem perda ou o raio de transmissão, utilizaremos a alternativa descrita a continuação.

Consideremos a RSSF denotada por $M_2 = (S, C(n, a))$, para n e a fixos. Lembremos, $C(n, a)$ é um processo em $[0, 1]^2$ e n a quantidade de sensores posicionados segundo o parâmetro de atratividade a . Sobre este cenário estamos interessados em achar o valor mínimo de φ no teorema 3 que garanta conectividade. À medida que o

valor de φ crece, para n e a fixos, variando no intervalo contínuo

$$0 < \varphi \leq \frac{\sqrt{2}\sqrt{n}}{\sqrt{\ln n}},$$

a RSSF começa tendo todos os seus nós isolados, passando cada vez a ter menor quantidade de componentes conexas e finalmente, para um valor determinado de φ , teremos uma única componente conexa que inclui todos os sensores da rede, isto é, o grafo de comunicações associado à rede se torna fortemente conexo.

Em particular, para a RSSF ser eficiente nos aspectos que citamos anteriormente, isto é, capacidade e consumo de energia, queremos determinar o menor dos valores de φ que com alta probabilidade mantenha a RSSF conexa. Neste contexto, definimos o valor crítico φ_c da RSSF M_2 como:

Definição 37 (Valor crítico) *Seja $M_2 = (S, C(n, a))$ com $|S| = n$, o valor crítico φ_c é caracterizado por*

$$\varphi_c = \inf\{\varphi \geq 0: \Theta^*(\varphi) > 0\},$$

onde Θ^* é a função de percolação no finito, definida por

$$\Theta^*(\varphi) = \mathbb{P}r(\text{GC é conexo}),$$

sendo GC o grafo de comunicações associado com $RA = \varphi \sqrt{\frac{\ln n}{n}}$.

Destacamos que, por definição, utilizando $RA = \varphi$ para valores de φ menores que φ_c , a probabilidade da RSSF ser conexa para uma instância do processo C é nula. Por outro lado, quando utilizamos valores de φ maiores que φ_c , a probabilidade de a RSSF ser conexa é positiva, e aumenta quando φ cresce. Este fenômeno, como mencionamos anteriormente, é conhecido como *transição de fase*, e pode se distinguir entre fase sub-crítica, quando $\varphi < \varphi_c$, fase crítica, quando $\varphi = \varphi_c$ e, fase super-crítica, quando $\varphi > \varphi_c$.

Na continuação apresentamos uma abordagem Monte Carlo para determinar o valor de φ_c , e analisamos a percolação continua sobre o processo $C(n, a)$.

4.2.1 Estudo Monte Carlo

Consideremos a rede $M_2 = (S, C(n, a))$, homogênea e estacionária. Em particular, estamos interessados em determinar valores de φ que nos levem para a fase super-crítica do sistema e, para isso, introduzimos a família de variáveis aleatórias $X^{(n,a)}$ definidas da seguinte maneira:

Definição 38 (Variável aleatória $X^{(n,a)}$) *Seja $M_2 = (S, C(n, a))$, definimos a variável aleatória $X^{(n,a)}: \Omega \rightarrow \left[0, \frac{\sqrt{2}\sqrt{n}}{\sqrt{\ln n}}\right]$ como:*

$$X^{(n,a)} = \min \left\{ \varphi \in \left[0, \frac{\sqrt{2}\sqrt{n}}{\sqrt{\ln n}}\right] : r = \varphi \sqrt{\frac{\ln n}{n}} \wedge r = RA \wedge CG(M_2, RA) \text{ é conexo} \right\}.$$

A variável aleatória $X^{(n,a)}$ modela o mínimo valor de φ que induz o raio de comunicação $r = \varphi\sqrt{n^{-1} \ln n}$ que garante a conectividade na rede instanciada e formada pelos n pontos distribuídos espacialmente de acordo com o processo pontual espacial $C(n, a)$, onde $a \in \mathbb{R}$ descreve a atratividade do modelo. Convém notar que a aleatoriedade de $X^{(n,a)}$ é fruto da aleatoriedade do processo pontual espacial, isto é, dada a distribuição espacial de nós, o valor mínimo de φ que induz o raio de comunicação $r = \varphi\sqrt{n^{-1} \ln n}$ que garante a conectividade na rede é fixo. A variável aleatória $X^{(n,a)}$, em princípio, forma uma família pelo fato de esperarmos ter comportamentos diferentes para valores diferentes de n (o número de nós) e de a (a atratividade com que estão distribuídos espacialmente).

O estudo desta família de variáveis aleatórias vai fornecer uma ideia do comportamento da constante φ que garante conectividade quando a expressão $\varphi\sqrt{n^{-1} \ln n}$ é o CTR associado (ver definição 33) para a um processo com atratividade a e com uma densidade específica. Esse estudo irá empregar o conceito de função empírica, definido a seguir.

Definição 39 (Função empírica) *Dadas as k observações reais x_1, \dots, x_k , eventos das variáveis aleatórias X_1, \dots, X_k independentes e identicamente distribuídas segundo a lei de probabilidade caracterizada pela função de distribuição acumulada $F: \mathbb{R} \rightarrow [0, 1]$, a função empírica*

$$\widehat{F}(x) = k^{-1} |\{0 \leq j \leq k : x_j \leq x\}|$$

é um estimador de F (Huber, 1981).

Para simplificar a leitura da apresentação de resultados, no resto do texto, vamos fazer as seguintes considerações:

1. $X_1^{(n,a)}, X_2^{(n,a)}, \dots, X_k^{(n,a)}$ são variáveis aleatórias independentes e identicamente distribuídas, e especificamente vamos discretizar a análise variando $n \in \{10, 20, 30, 40, 50\}$ e $a \in \{-30, -15, 0, 5, 15, 30\}$, obtendo uma família para cada n e a fixos. Assim sendo, n e a são os “fatores” do nosso estudo, e cada par de valores é uma “situação”.
2. As famílias de variáveis aleatórias $X_1^{(n,a)}, X_2^{(n,a)}, \dots, X_k^{(n,a)}$ serão instanciadas pelas amostras $x_1^{(n,a)}, x_2^{(n,a)}, \dots, x_k^{(n,a)}$;
3. $\widehat{F}_k^{(n,a)}$ denota a função de distribuição empírica construída com uma amostra de tamanho k de ocorrências de variáveis aleatórias independentes e identicamente distribuídas da variável aleatória $X^{(n,a)}$, isto é,

$$\widehat{F}_k^{(n,a)}(x_1^{(n,a)}, x_2^{(n,a)}, \dots, x_k^{(n,a)})$$

é um estimador da função de distribuição acumulada $F_k^{(n,a)}$ que caracteriza a distribuição de $X^{(n,a)}$.

Os conjuntos de valores $x_1^{(n,a)}, x_2^{(n,a)}, \dots, x_k^{(n,a)}$, para cada $n \in \{10, 20, 30, 40, 50\}$ e $a \in \{-30, -15, 0, 5, 15, 30\}$ com $k = 1000$, foram obtidos com simulações no ambiente R na sua versão 2.8.0. A plataforma utilizada foi um *notebook* com processador i386 (32 bits) e GNU/Linux Ubuntu Intrepid como sistema operacional. O processo

pontual $C(n, a)$, utilizado no posicionamento dos sensores, foi implementado com funções do pacote `spatstat` como adaptação para o espaço $[0, 1]^2$ do apresentado por Frery et al. (2008), onde espaço que representava o ambiente era $[0, 100]^2$. Todos os valores $x_i^{(n,a)}$ foram calculados com cinco dígitos decimais de precisão. Para cada n e a , o tamanho de amostra utilizado foi $k = 1000$; este número de replicações foi considerado satisfatório em uma análise exploratória inicial.

Para nos familiarizarmos com os resultados das simulações, vamos fazer uma análise exploratória gráfica e numérica.

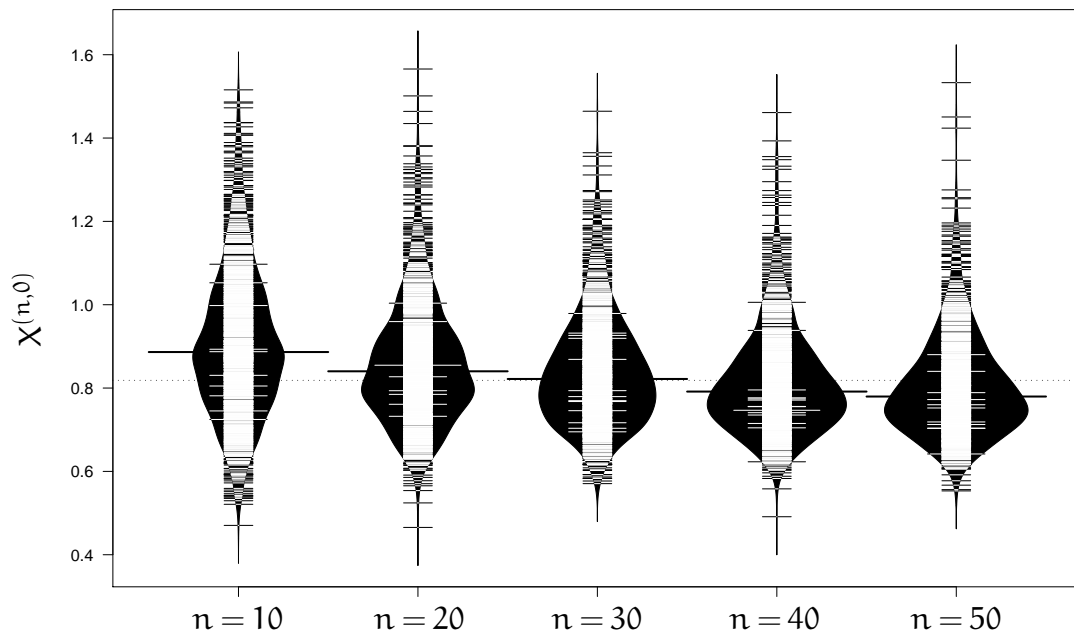


Figura 4.3: *Beanplot* de $x_1^{(n,0)}, x_2^{(n,0)}, \dots, x_{1000}^{(n,0)}$, para $n \in \{10, 20, 30, 40, 50\}$

A figura 4.3 resume os resultados obtidos nas simulações para a família $X^{(n,0)}$ com $n = \{10, 20, 30, 40, 50\}$, como fixamos anteriormente. Neste gráfico chamado *beanplot*, cada *bean* (gráfico corresponde à amostra de uma situação) contém um estimador da densidade (baseado em *kernel* gaussiano) espelhado, formando uma figura verticalmente simétrica (ver Kampstra, 2008, para detalhes sobre o pacote `beampplot` e suas funções). Os dados *per se*, são plotados unidimensionalmente com traços que fazem contraste com o kernel. Traços maiores indicam empates nos dados; quanto maior o traço, mais elementos da amostra têm o mesmo valor. O traço maior em cada *bean* indica a média da amostra, e o traço pontilhado indica a média geral, calculada com todos os dados de todas as amostras.

Notemos neste gráfico que quanto maior a quantidade de sensores, maior é a concentração de dados. Também observamos que as médias variam entre $\varphi = 0.9$, para $n = 10$, e $\varphi = 0.8$, para $n = 50$. Complementando o gráfico 4.3 com informação numérica, a tabela 4.1 apresenta as medidas mais interessantes: valor mínimo (\min), quartil inferior empírico ($q_{1/4}$), mediana amostral (Med), média amostral (\bar{x}), quartil supe-

rior empírico ($q_{\frac{3}{4}}$), valor máximo (min), distância interquartil (IQR – *Inter-Quartile Range*) e desvio padrão amostral (s).

Tabela 4.1: Estatísticos de resumo de $x_1^{(n,0)}, x_2^{(n,0)}, \dots, x_{1000}^{(n,0)}$, para $n \in \{10, 20, 30, 40, 50\}$

n	min	$q_{\frac{1}{4}}$	Med	\bar{x}	$q_{\frac{3}{4}}$	max	IQR	s
10	0.470	0.783	0.886	0.905	1.011	1.516	0.227	0.176
20	0.465	0.761	0.840	0.859	0.936	1.566	0.175	0.148
30	0.570	0.743	0.821	0.839	0.914	1.464	0.170	0.132
40	0.491	0.731	0.791	0.814	0.875	1.461	0.143	0.125
50	0.553	0.721	0.779	0.804	0.864	1.533	0.143	0.122

Por meio da distância interquartil (IQR), confirmamos o que já tínhamos antecipado com o *beanplot* da figura 4.3: a probabilidade de termos conectividade aumenta sua inclinação quando n cresce. Em geral, com a exceção dos valores mínimos e máximos, com o aumento de n os estatísticos considerados na tabela 4.1 diminuem, o que indica que além de termos uma redução na variabilidade, isto é, no espalhamento, também há um deslocamento negativo nos valores.

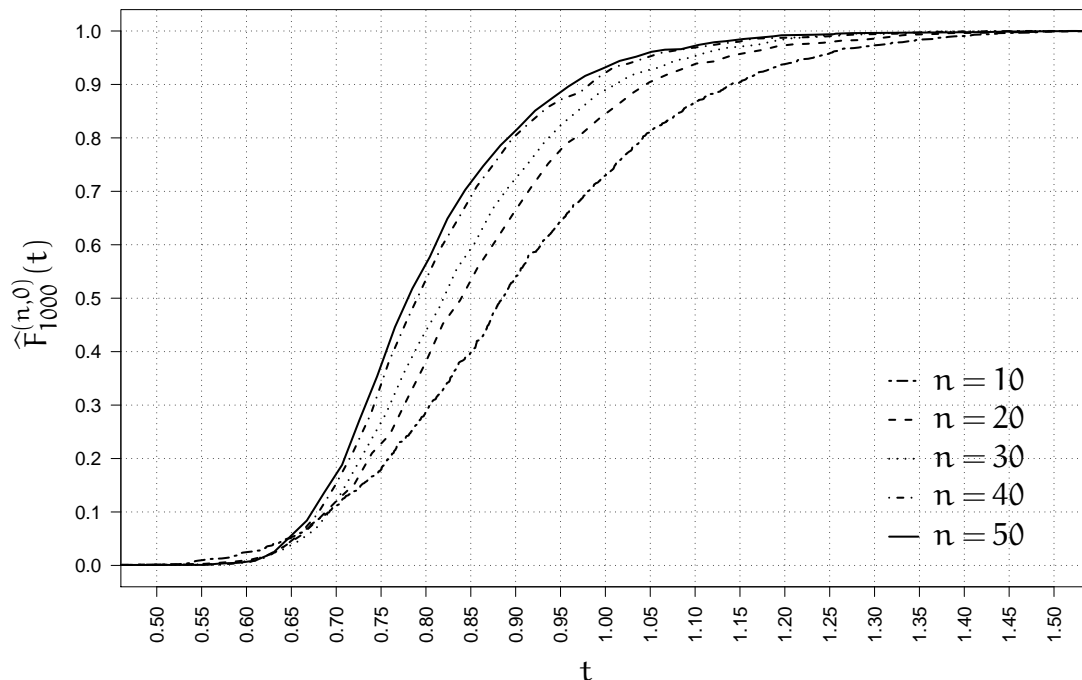


Figura 4.4: Função empírica das variáveis $X^{(n,0)}$

Na figura 4.4 apresentamos as funções empíricas das variáveis $X^{(10,0)}$, $X^{(20,0)}$, $X^{(30,0)}$, $X^{(40,0)}$ e $X^{(50,0)}$. Vemos em este gráfico que quanto maior a quantidade de sen-

sores no ambiente, mais marcada é a transição de fase, isto é, a inclinação da função é maior ao mesmo tempo que o intervalo de transição é menor.

Notemos que os resultados obtidos correspondem aos valores de φ que garantem que o raio de transmissão $r = \varphi \sqrt{n^{-1} \ln n}$ seja o CTR, e não diretamente os valores do raio de transmissão. A conversão de φ para valores (práticos) de r é imediata pelo uso da relação entre ambos.

Notemos que a a função empírica cumpre um papel principal do ponto de vista prático, já que com ela conseguiremos estabelecer de antemão o raio de transmissão necessário para que uma rede com n sensores tenha conectividade com a probabilidade p desejada em relação aos custos de utilizar uma tecnologia de maior alcance sacrificando energia e os benefícios funcionais. Estes custos podem ser tanto de valor monetário para adquirir a tecnologia necessária, quanto de energia, o que implicaria menor tempo de vida da RSSF. Como vimos anteriormente, também podemos pensar no problema inverso, e responder qual a quantidade de sensores necessária para manter uma rede funcional com uma tecnologia de transmissão predeterminada, isto é, com um raio r predeterminado. Novamente, a decisão se reduz a uma análise de custos e benefícios que vão depender tanto dos recursos disponíveis quanto dos objetivos da RSSF.

4.2.2 Estudo de caso I: processos pontuais binomiais

Ao momento de implementar uma RSSF, varios fatores devem ser levados em conta. Sem perda de generalidade, vamos estudar estes parâmetros restritos ao espaço $[0, 1]^2$, sendo que este estudo pode se estender para $[0, \ell]^2$, com $\ell \in \mathbb{R}_+$. Algumas das primeiras considerações que devem se analisar, são as seguintes:

1. Custo de cada nó sensor que vai ser utilizado: estes se acham em diversidades de preços segundo as características de processamento, comunicação e dispositivos sensores. Assumindo que vamos escolher dentre sensores que somente variem na sua tecnologia de comunicação, escolheremos posteriormente, com algum critério, dentre alguns preços que listaremos para ilustrar o processo de escolha dos diferentes fatores na fase de implementação da RSSF.
2. Alcance de transmissão máximo r_{\max} : como comentado no item 1, o sensor escolhido vai ter associada uma tecnologia de comunicação específica, a qual restringe o possível ou os possíveis valores de raio de transmissão disponíveis, e em particular, o valor máximo de transmissão r_{\max} .
3. Custo de posicionar cada sensor no ambiente: em nosso caso, o posicionamento é feito de forma aleatória independente e uniforme e deve ter um valor fixo.
4. Importância do tempo de vida da rede: segundo o propósito da RSSF, existe um tempo de vida mínimo admissível, que pode variar de aplicação em aplicação.
5. Probabilidade mínima admissível de conectividade: semelhante ao item 4, isto varia segundo o propósito para o qual a RSSF deve ser implementada.
6. Custo máximo: como em todo projeto, existe um limite de capital disponível para implementá-lo, e não podemos exceder esse valor.

7. Hostilidade do meio de comunicação: o parâmetro α , conhecido como *path loss* apresentado no modelo de comunicação da equação (2.1) da página 25, pode ser muito diferente de um cenário para outro. Isto pode, conseqüentemente, causar variações nos custos relacionados à potencia necessária para enviar dados corretamente de um sensor a outro.

A partir destes itens elencados, podemos formular um problema de otimização para cada necessidade.

Para resolver qualquer problema de otimização em nossa situação, vamos necessitar de maior resolução para o parâmetro n . Em <http://sites.google.com/site/almironmarcelo/dissertacao-ufal> disponibilizamos os valores das funções empíricas para vários valores de n .

Baseados nestes resultados, e na informação complementar necessária dos itens acima elencados, poderíamos resolver problemas de otimização de interesse. Para exemplificar vamos introduzir um problema específico, mas poderíamos ter escolhido qualquer outro exemplo para resolver.

Suponhamos que na implementação de uma RSSF, dispomos de cinco tipos diferentes de sensores, digamos, A, B, C, D e E, cada um deles com alcance máximo de transmissão r_A , r_B , r_C , r_D e r_E respectivamente. Representamos com R_t o preço do sensor de tipo t , para cada tipo t identificado no mercado. Para a compra dos sensores e o posicionamento no ambiente dispomos de um capital T . Posicionar cada sensor no ambiente tem um custo m . Considerando o propósito da RSSF, achamos aceitável uma chance de 95% de possibilidade da rede ser conexas. Por experiências em comunicação sem fio no ambiente estipulamos o expoente de *path loss* $\alpha = 4$. Com estas informações, desejamos saber o seguinte:

quais e quantos sensores devem ser comprados, e com que raio de transmissão vamos utilizá-los de maneira de maximizar o tempo de vida da RSSF?

Para responder esta pergunta, devemos contruir o modelo de otimização. Assim, na continuação vamos descrever os dados do modelo, seguido das variáveis que vão ser consideradas. Desta maneira teremos o necessário para construir as restrições e a função objetivo. Os dados do modelo são os seguintes:

- r_t : denota o raio máximo de transmissão do sensor de tipo t ,
- R_t : é o preço de aquisição por sensor de tipo t ,
- m : é o custo de posicionar um sensor no ambiente, independentemente do tipo,
- T : denota o capital disponível,
- p_{sr} : é a probabilidade de conectividade da RSSF para s sensores com raio de transmissão r ,
- $\alpha = 4$: é o parâmetro de *path loss*.

Desta maneira as variáveis a determinar são:

- ϕ : valor da constante ϕ do teorema 3,

- s : quantidade de sensores que deve ser comprada,
- t : tipo de sensor que devemos utilizar,
- r : raio de transmissão empregado na implementação.

A primeira restrição que vamos considerar é a respeito do capital disponível que não pode ser excedido, isto é:

$$s(R_t + m) \leq T.$$

Também devemos certificar que não consideraremos raios de transmissão maiores que o que a tecnologia do sensor que compraremos disponibiliza:

$$r \leq r_t.$$

Para respeitar o 95% de possibilidade de conectividade requerida, utilizamos:

$$r = \phi \sqrt{\frac{\ln s}{s}}, \text{ e}$$

$$p_{s\phi} \geq 0.95.$$

Finalmente, a função objetivo para maximizar o tempo de vida, pode se reformular como uma minimização do custo c_e da RSSF (ver definição 29),

$$\min\{c_e : c_e = sr^\alpha\}.$$

Implementando este sistema de otimização vamos obter a melhor solução à pergunta formulada. Notemos a flexibilidade que permitiu trabalhar com a constante ϕ em lugar de ter calculado diretamente o raio de transmissão, ou ainda a potência necessária. Desta maneira, os resultados obtidos nas simulações admitem trabalhar com diferentes expoentes de *path loss*. Inclusive, se novos modelos de comunicação que relacionem o raio de transmissão com a potência necessária ou energia gasta forem propostos na literatura, os resultados aqui apresentados poderiam ser utilizados com a mesma validade.

Observemos também que com os parâmetros definidos é relativamente simples formular novos modelos de otimização segundo nossas necessidades e objetivos.

4.3 Conectividade em processos não independentes (casos atrativo e repulsivo)

Como vimos na seção 2.2, na prática podemos ter cenários diferentes segundo o ambiente e a modalidade na qual são posicionados os nós sensores. Um dos trabalhos que considera esta diversidade é o de Panchapakesan & Munjunath (2001). Os autores analisam o raio mínimo de transmissão necessário para garantir conectividade, por meio de simulações, em diferentes situações com um processo pontual controlado por um parâmetro $-1 \leq \beta \leq 1$. Quando $\beta = -1$, os nós sensores são posicionados com maior intensidade perto das bordas, e quando $\beta = 1$ os sensores são posicionados com maior intensidade no centro da região de interesse.

Por outro lado, Frery et al. (2008) analisam, entre outros fatores, o impacto do posicionamento dos sensores no processo de reconstrução de informação utilizando

RSSF com dois protocolos diferentes: LEACH e SKATER. Para modelar as diferentes situações de posicionamento de sensores no ambiente, os autores propõem um novo processo pontual espacial $C(n, \alpha)$ introduzido formalmente na seção 2.2.2.

Em particular, nesta seção utilizamos o processo pontual espacial $C(n, \alpha)$ (ver definição 24 na página 22) para analisar o impacto que os diferentes níveis de atratividade, sobre o posicionamento dos nós sensores, têm sobre o valor φ do teorema 3 (página 44) que resolve o problema CTR.

4.3.1 Estudos Monte Carlo

Como acabamos de antecipar, utilizaremos também neste estudo a expressão proposta por Panchapakesan & Munjunath (2001). Porém este resultado foi proposto originalmente para processos pontuais independentes Poisson, percebemos com estudos de simulação que o fenômeno de transição de fase também acontece quando aplicamos esta expressão no processo $C(n, \alpha)$ para vários valores de n e de α . Para facilitar a análise, vamos distinguir entre diferentes níveis de atratividade, com $\alpha \in \{-30, -15, 0, 5, 15, 30\}$, onde o caso $\alpha = 0$ corresponde ao estudado na seção anterior (seção 4.2). Da mesma forma, utilizaremos $n = \{10, 20, 30, 40, 50\}$ e começaremos analisando a situação para processos repulsivos, diminuindo a repulsividade até chegar a processos atrativos.

Para tomarmos familiaridade com os dados obtidos para $X^{(n, -30)}$, a figura 4.5 mostra o *beanplot* correspondente.

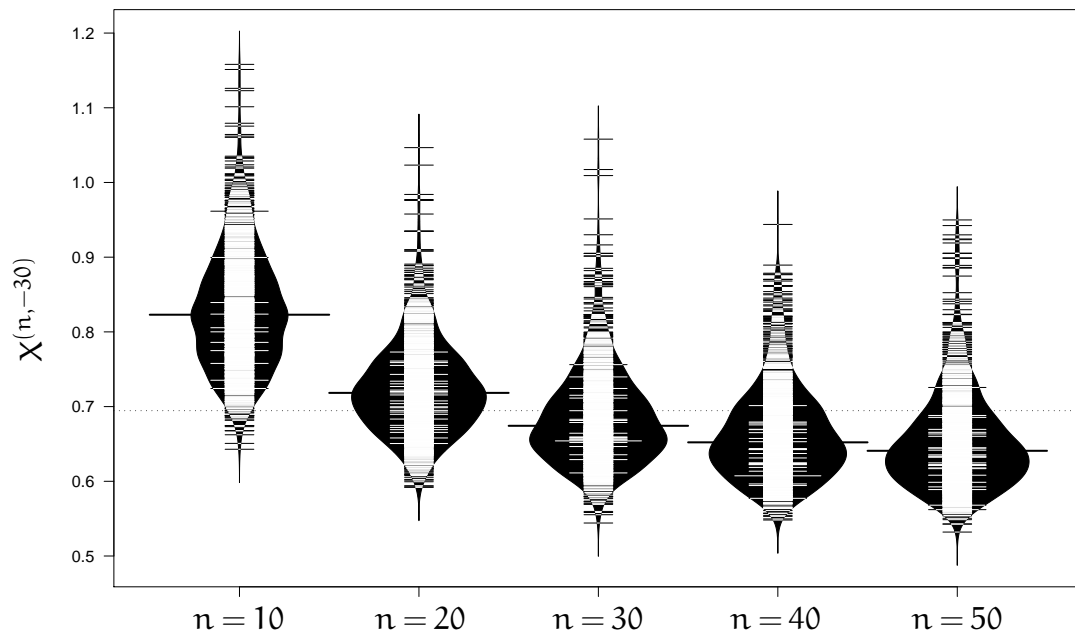


Figura 4.5: *Beanplot* de $x_1^{(n, -30)}, x_2^{(n, -30)}, \dots, x_{1000}^{(n, -30)}$, para $n \in \{10, 20, 30, 40, 50\}$

Com execução do caso $n = 10$, em geral tirando os valores extremos temos redução

da variabilidade dos dados de forma similar ao já observado, mas com um deslocamento mais marcado que diminui quando aumentamos a quantidade de sensores.

Para dar suporte a essas conclusões, a tabela 4.2 mostra os valores dos estatísticos de resumo. Observemos que a distância interquartil se mantém estável com exceção do caso $n = 10$. Isto indica que a diminuição no espalhamento dos dados ocorre para $n \geq 20$, mas há um deslocamento dentro dos valores.

Tabela 4.2: Estatísticos de resumo de $x_1^{(n,-30)}, x_2^{(n,-30)}, \dots, x_{1000}^{(n,-30)}$, para $n \in \{10, 20, 30, 40, 50\}$

n	min	$q_{\frac{1}{4}}$	Med	\bar{x}	$q_{\frac{3}{4}}$	max	IQR	s
10	0.643	0.770	0.823	0.830	0.878	1.158	0.108	0.080
20	0.592	0.683	0.718	0.725	0.760	1.047	0.076	0.061
30	0.544	0.641	0.674	0.686	0.717	1.058	0.076	0.064
40	0.548	0.619	0.652	0.664	0.698	0.944	0.078	0.062
50	0.532	0.609	0.641	0.653	0.684	0.950	0.075	0.064

A figura 4.6 apresenta a distribuição empírica amostral para as variáveis $\chi^{(n,-30)}$. Notemos que em relação ao problema estudado na seção anterior, isto é, para as variáveis $\chi^{(n,0)}$, percebemos que realmente existe uma transição de fase marcada.

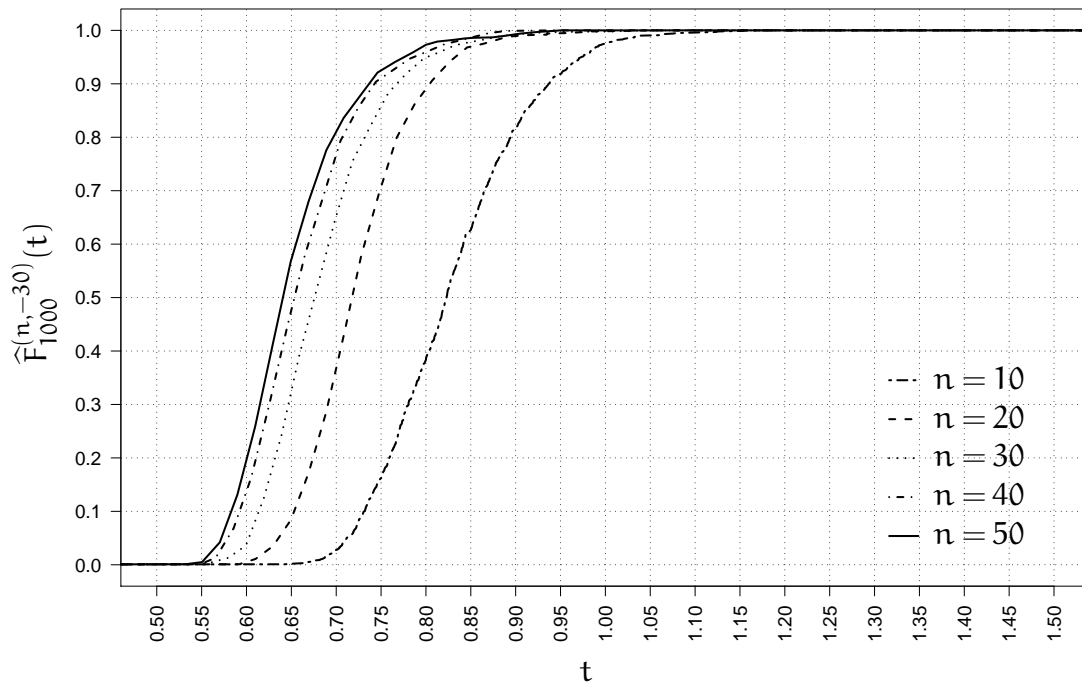


Figura 4.6: Função empírica das variáveis $\chi^{(n,-30)}$

Notemos no caso $n = 50$, por exemplo, que fixando $\varphi = 0.6$ o raio de transmissão $r_1 = 0.6\sqrt{(\ln 50)/50} \approx 0.168$ induz 19.4% de possibilidade da RSSF ser conexas,

enquanto que um valor um pouco maior, digamos $\varphi = 0.75$ o raio de transmissão $r_2 = 0.75\sqrt{(\ln 50)/50} \approx 0.2$ leva a 92.6% de possibilidade da RSSF ser conexas. Este tipo de processo demonstra ser mais estável, no sentido que não há muita variabilidade se comparado com o caso de processo pontuais espaciais binomiais.

A transição de fase ocorre quase com a mesma intensidade para φ entre 0.55 e 1.05 para qualquer valor de n .

No caso de $\alpha = -15$, na figura 4.7 não percebemos muita diferença respeito do cenário $\alpha = -30$.

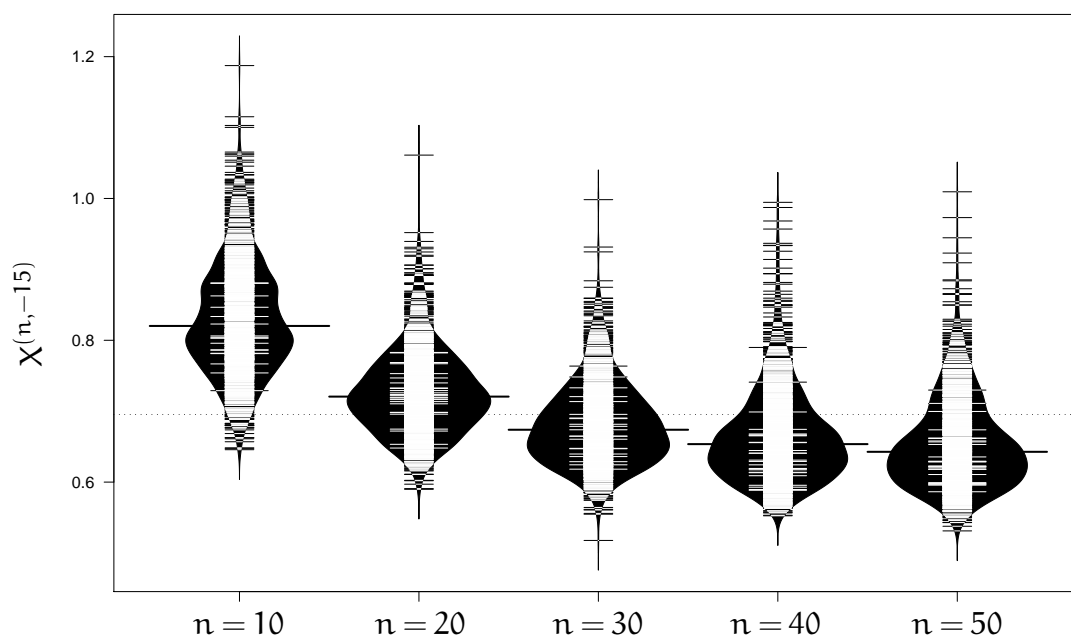


Figura 4.7: *Beanplot* de $x_1^{(n,-15)}, x_2^{(n,-15)}, \dots, x_{1000}^{(n,-15)}$, para $n \in \{10, 20, 30, 40, 50\}$

A tabela 4.3 apresenta os valores dos mesmos estatísticos escolhidos anteriormente. Notamos que a redução da variabilidade dos dados se mantém igual que no caso $\alpha = -30$.

Quando a repulsividade é um pouco menor, notemos que para os casos de menor quantidade de nós começamos ter uma leve diferença entre os extremos de n , quanto à inclinação da função empírica mostrada na figura 4.8.

Também percebemos que, para o caso $n = 20$, a função empírica começa se afastar das distribuições correspondentes a n maiores em comparação aos resultados para $\alpha = -30$.

Agora vamos analisar um caso com um pouco de atratividade. Na figura 4.9 apresentamos o *beanplot* para $\alpha = 5$. Em esta figura observamos que não existe um deslocamento tão marcado como nas situações $\alpha = -30$ e $\alpha = -15$. Para o caso $n = 40$ e $n = 50$ as distribuições são muito similares na massa principal de dados.

Numericamente, os desvios e as distâncias interquartis indicam que quando passamos de processos repulsivos para processos atrativos, a variabilidade dentro de

Tabela 4.3: Estatísticos de resumo de $x_1^{(n,-15)}, x_2^{(n,-15)}, \dots, x_{1000}^{(n,-15)}$, para $n \in \{10, 20, 30, 40, 50\}$

n	min	$q_{\frac{1}{4}}$	Med	\bar{x}	$q_{\frac{3}{4}}$	max	IQR	s
10	0.646	0.775	0.820	0.831	0.881	1.187	0.106	0.081
20	0.590	0.686	0.721	0.726	0.758	1.061	0.072	0.059
30	0.518	0.640	0.674	0.682	0.714	0.998	0.074	0.058
40	0.552	0.618	0.654	0.665	0.696	0.995	0.078	0.066
50	0.531	0.607	0.643	0.654	0.687	1.009	0.08	0.065

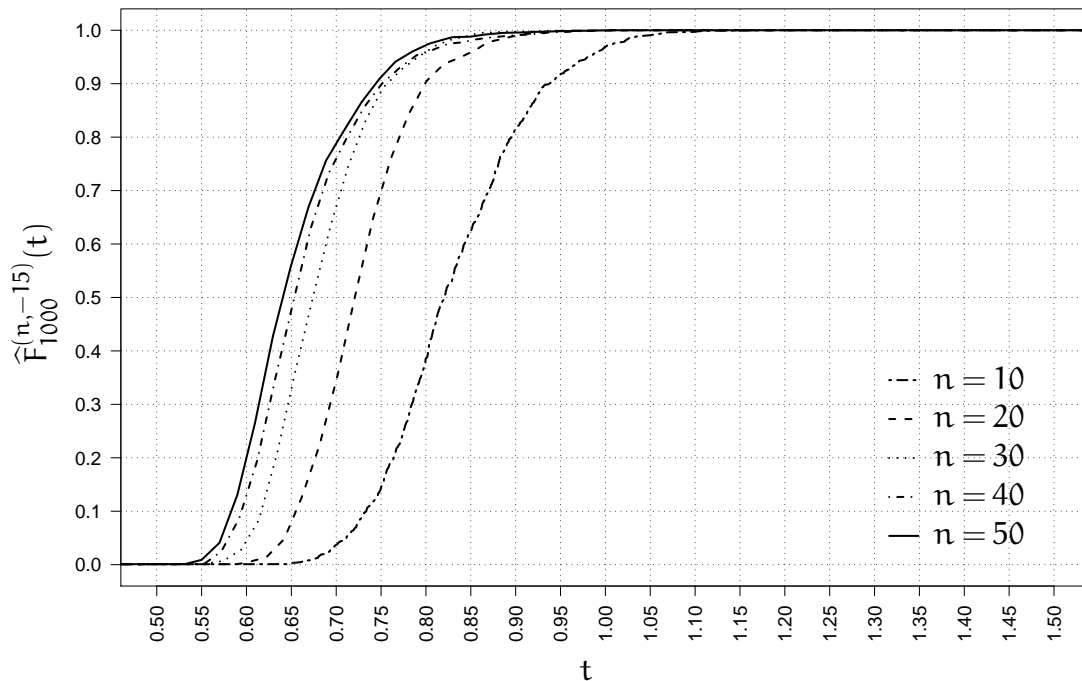


Figura 4.8: Função empírica das variáveis $X^{(n,-15)}$

cada amostra aumenta (ver tabela 4.4).

Por outro lado, vemos que as diferenças nas amostras para cada n vão diminuindo, o que pode se conferir tanto no *beanplot* da figura 4.9 quanto na tabela 4.4.

Finalmente, na figura 4.10, apresentamos as funções empíricas amostrais que correspondem a $\alpha = 5$. O primeiro que observamos nesta figura é a semelhança entre $n = 40$ e $n = 50$, como antecipamos com os estatísticos de resumo e o gráfico de *beanplot* correspondente. Também percebemos neste gráfico que para valores entre 0.55 e 0.68, aproximadamente, com menor quantidade de sensores obtemos, com os mesmos valores de φ , probabilidade maior de conectividade.

Em relação aos casos repulsivos, para situações de pouca atratividade a transição de fase não é marcada, isto é, não existe um valor de φ_c para o qual valores inferior-

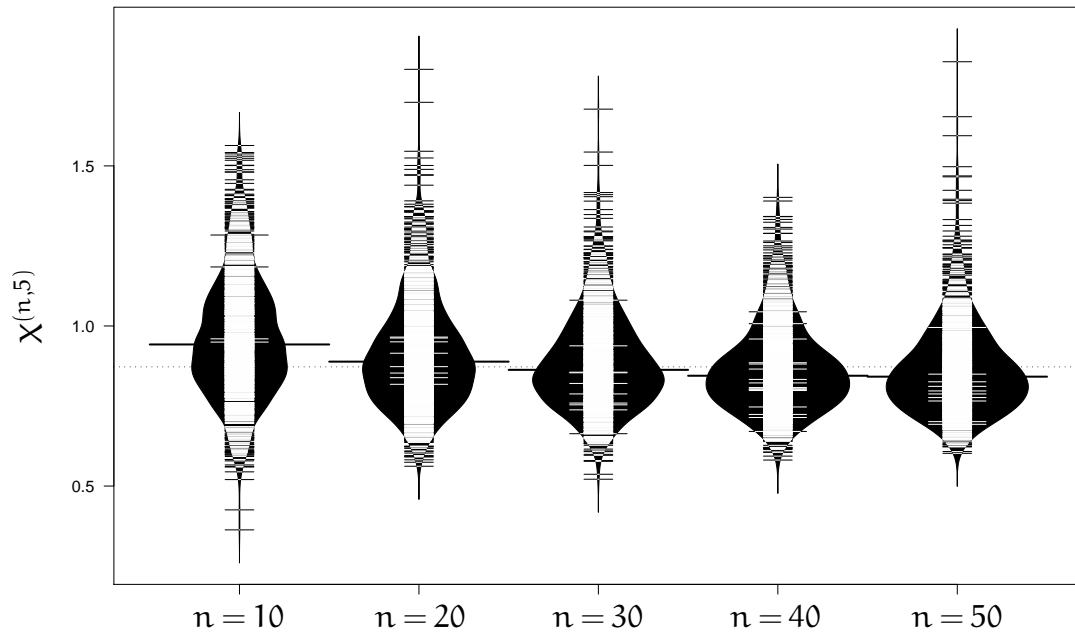


Figura 4.9: *Beanplot* de $x_1^{(n,5)}, x_2^{(n,5)}, \dots, x_{1000}^{(n,5)}$, para $n \in \{10, 20, 30, 40, 50\}$

Tabela 4.4: Estatísticos de resumo de $x_1^{(n,5)}, x_2^{(n,5)}, \dots, x_{1000}^{(n,5)}$, para $n \in \{10, 20, 30, 40, 50\}$

n	min	$q_{\frac{1}{4}}$	Med	\bar{x}	$q_{\frac{3}{4}}$	max	IQR	s
10	0.363	0.830	0.942	0.962	1.080	1.564	0.250	0.191
20	0.562	0.795	0.889	0.917	1.010	1.802	0.215	0.172
30	0.521	0.781	0.863	0.888	0.973	1.677	0.193	0.150
40	0.581	0.772	0.845	0.867	0.935	1.402	0.162	0.134
50	0.602	0.768	0.842	0.866	0.937	1.825	0.169	0.144

res muito próximos tenham associados uma probabilidade baixa de conectividade e para valores superiores muito próximos a probabilidade de conectividade seja muito alta. Deixamos claro que o termo “muito próximo” é subjetivo, na frase anterior; baseamos-nos na comparação com processos repulsivos, que consideramos ideais.

Agora continuamos a nossa análise com a figura 4.11 (página 59), que apresenta o *beanplot* para $\alpha = 15$. Nesta figura vemos que o deslocamento que existia em situações anteriores não está mais presente. Notamos que os casos $n = 40$ e $n = 50$ são muito similares, com exceção de valores extremos. Observemos também que em situações com menor quantidade de sensores, a conectividade está presente com valores menores de φ quando comparados com situações com maior quantidade de sensores no ambiente.

A tabela 4.5 apresenta os valores numéricos dos estatísticos usuais neste trabalho,

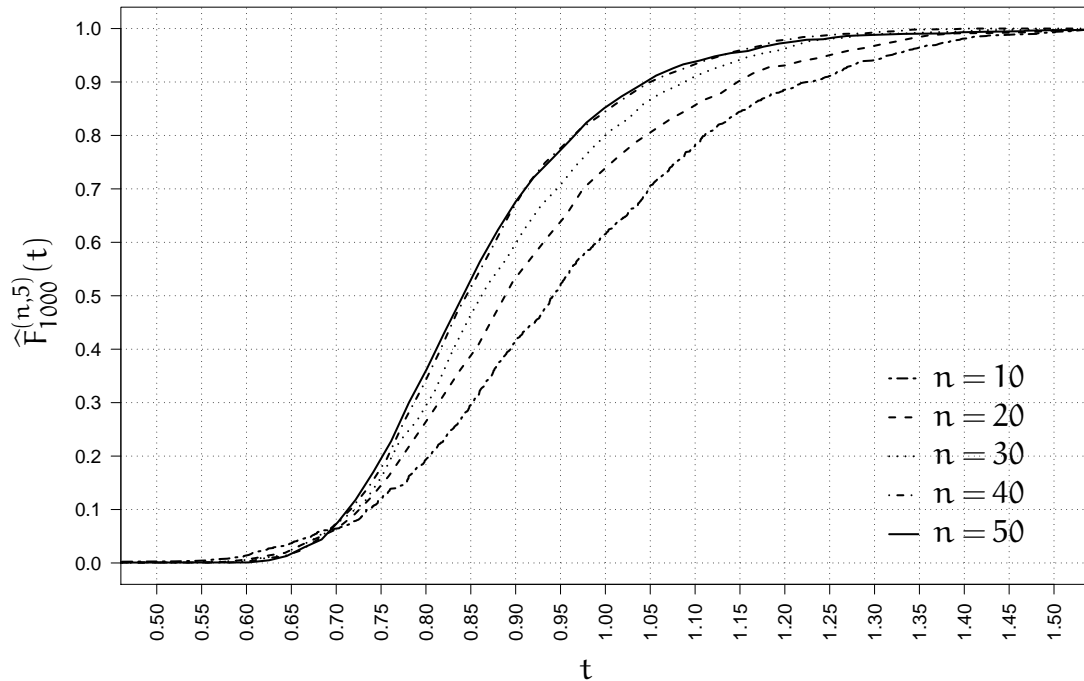


Figura 4.10: Função empírica das variáveis $X^{(n,5)}$

para os casos com $\alpha = 15$.

Tabela 4.5: Estatísticos de resumo de $x_1^{(n,15)}, x_2^{(n,15)}, \dots, x_{1000}^{(n,15)}$, para $n \in \{10, 20, 30, 40, 50\}$

n	min	$q_{\frac{1}{4}}$	Med	\bar{x}	$q_{\frac{3}{4}}$	max	IQR	s
10	0.267	0.875	1.034	1.052	1.217	1.893	0.341	0.257
20	0.499	0.903	1.035	1.060	1.189	1.978	0.285	0.223
30	0.605	0.899	1.025	1.043	1.154	1.945	0.255	0.201
40	0.645	0.905	1.010	1.033	1.140	1.888	0.235	0.184
50	0.638	0.891	0.994	1.022	1.122	1.755	0.231	0.179

Como vínhamos acompanhando, a variabilidade aumenta e também o intervalo onde a transição de fase acontece em casos com maior atratividade, tornando a primeira menos pronunciada.

A respeito da função empírica, apresentada na figura 4.12 (página 60), notamos pouca diferença quando variamos n .

Também percebemos que para valores de φ entre 0.4 e 0.9, quanto menor a quantidade de nós sensores, maior é a possibilidade de ter conectividade na rede com esses valores associados. Mas temos que ter precaução, já que não devemos confundir o raio de transmissão com o valor de φ , isto é, a função $\sqrt{n^{-1} \ln n}$ é decrescente. Ob-

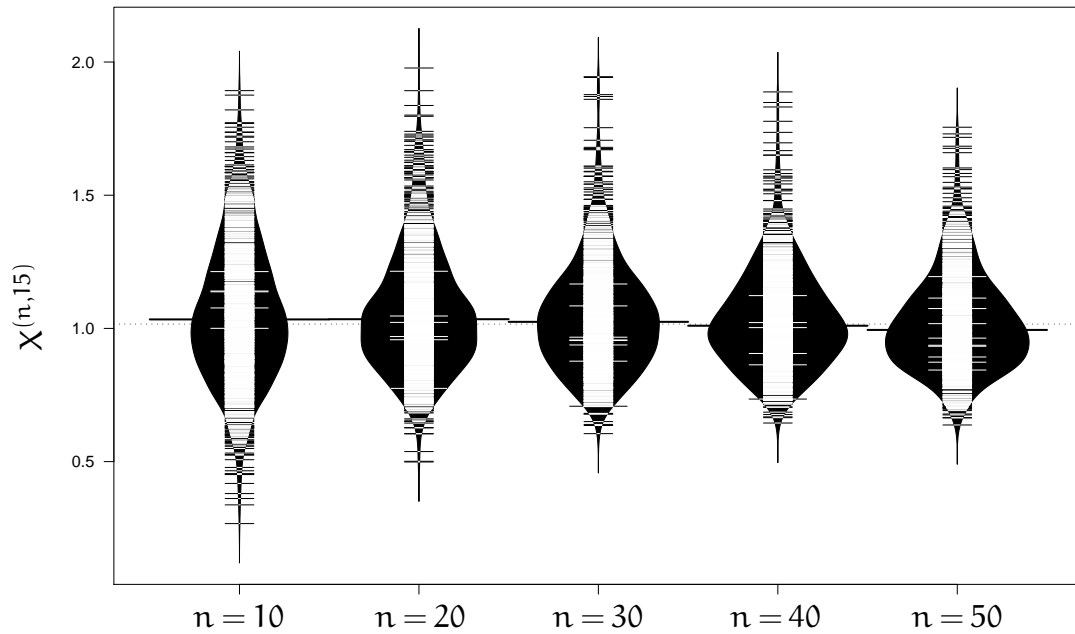


Figura 4.11: *Beanplot* de $x_1^{(n,15)}, x_2^{(n,15)}, \dots, x_{1000}^{(n,15)}$, para $n \in \{10, 20, 30, 40, 50\}$

servemos que a escala foi trocada no gráfico 4.12, já que ao aumentar a variabilidade, uma janela maior foi necessária para visualizar completamente os resultados.

Para este caso, achamos mais proveitoso aprimorar a utilização da menor quantidade de sensores possível, já que para garantir conectividade com muitos sensores, a capacidade da rede resulta seriamente prejudicada, como veremos posteriormente.

Para concluir a apresentação dos dados, a figura 4.13 (página 61) apresenta o caso de maior atratividade considerado, isto é, para processos com $\alpha = 30$.

Observamos neste gráfico, que as amostras de valores de n maiores contêm valores de φ maiores também. Com ajuda da tabela 4.6 vemos que para $n = 10$ o valor de φ em média que garante conectividade é 1.119, no entanto, para $n = 50$, o valor de φ em média que garante conectividade é 1.212.

Tabela 4.6: Estatísticos de resumo de $x_1^{(n,30)}, x_2^{(n,30)}, \dots, x_{1000}^{(n,30)}$, para $n \in \{10, 20, 30, 40, 50\}$

n	min	$q_{\frac{1}{4}}$	Med	\bar{x}	$q_{\frac{3}{4}}$	max	IQR	s
10	0.173	0.910	1.104	1.119	1.333	2.106	0.422	0.319
20	0.212	1.000	1.168	1.203	1.380	2.102	0.380	0.281
30	0.642	1.018	1.179	1.219	1.381	2.295	0.363	0.266
40	0.74	1.05	1.18	1.222	1.36	2.38	0.316	0.254
50	0.722	1.041	1.167	1.212	1.333	2.494	0.292	0.239

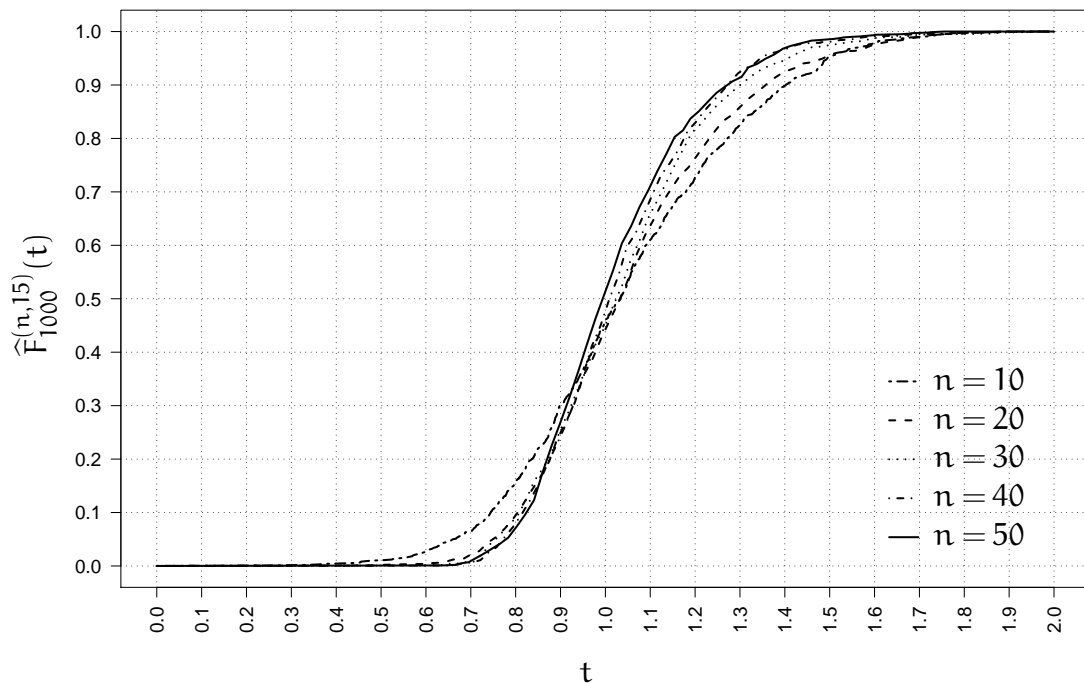


Figura 4.12: Função empírica das variáveis $X^{(n,15)}$

Como comentamos em várias oportunidades, isto não diz muito a respeito do raio de transmissão. Especificamente, para a situação média em $n = 10$, o raio de transmissão associado é $r_1 0 = 1.119 \sqrt{(\ln 10)/10} \approx 0.53$, em quanto que para $n = 50$, o raio de transmissão em média que garante conectividade é $r_3 0 = 1.212 \sqrt{(\ln 50)/50} \approx 0.33$.

A figura 4.14 (página 62) apresenta a função empírica associada a esta última situação analisada.

Observamos que as funções empíricas correspondentes aos casos com n maiores possuem ainda maior inclinação. Por outro lado, o deslocamento nos valores de φ são cada vez maiores quando aumentamos a quantidade de sensores, levando os casos mais atrativos a demorar muito para começar a transição de fase.

A análise feita até agora esteve focada em cada nível de atratividade, comparando principalmente o comportamento da transição de fase que segue o fenômeno de conectividade. Isto é, para um processo com uma atratividade a estivemos interessados no impacto que pode causar agregar mais sensores ao ambiente, para diversos valores de φ , concluindo o seguinte:

1. para casos onde o processo pontual espacial que descreve o posicionamento dos nós sensores é repulsivo, não percebemos muita diferença na intensidade do fenômeno de transição de fase, mas sim no ponto no qual começa, isto é, para maior quantidade de nós sensores, mais cedo começa a se manifestar a transição de fase;
2. para casos onde o processo pontual espacial que descreve o posicionamento

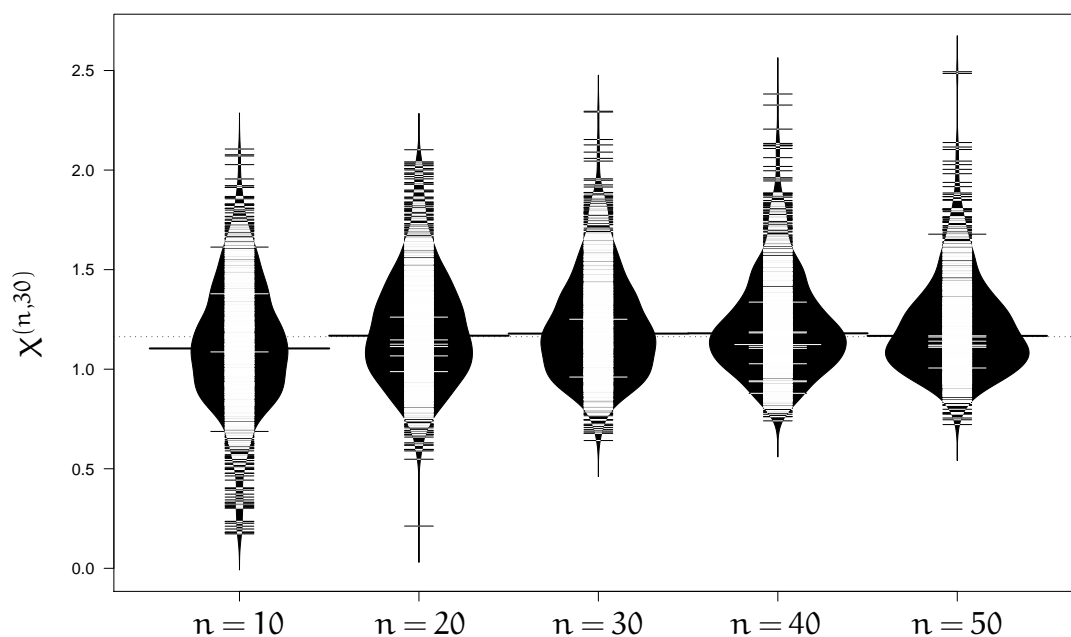


Figura 4.13: *Beanplot* de $x_1^{(n,30)}, x_2^{(n,30)}, \dots, x_{1000}^{(n,30)}$, para $n \in \{10, 20, 30, 40, 50\}$

dos nós sensore é atrativo, a conectividade em média tende a ser a mesma para qualquer processo, sendo que, ao termos maior variabilidade para n pequeno, os primeiros casos de conectividade acontecem para valores de φ relativamente pequenos, tornando a transição de fase mais suave.

Para termos uma melhor compreensão do comportamento do menor valor de φ que garante conectividade nos diferentes níveis de atratividade, a continuação comparamos as funções empíricas fixando n . Para facilitar a análise, vamos confrontar para $n \in \{10, 30, 50\}$ os níveis de atratividade $\alpha \in \{-30, 0, 30\}$.

A figura 4.15 (página 63) apresenta a confrontação entre diferentes níveis de atratividade com $n = 10$.

Para o caso atrativo, percebemos que a transição de fase é muito devagar respeito do caso repulsivo.

Na figura 4.16 apresentam-se os resultados variando a atratividade, mas desta vez, para $n = 30$.

Comparado com o caso $n = 10$ apresentado, aqui observamos que com atratividade $\alpha = 30$ ao aumentar a quantidade de sensores, os valores de interesse se mantêm, isto é, não percebemos melhora. Contudo, a situação para atratividade nula ou negativa, percebemos que com o aumento de sensores, existe uma mudança nestes valores. O processo de transição começa com antecedencia, isto é, o valor φ_c é menor. Também para os casos de atratividade nula ou negativa, vemos que a transição de fase acontece mais rapidamente, num intervalo menor.

Finalmente, para concluir com nossa análise, a figura 4.17 (página 65) mostra as diferentes situações de atratividade para $n = 50$.

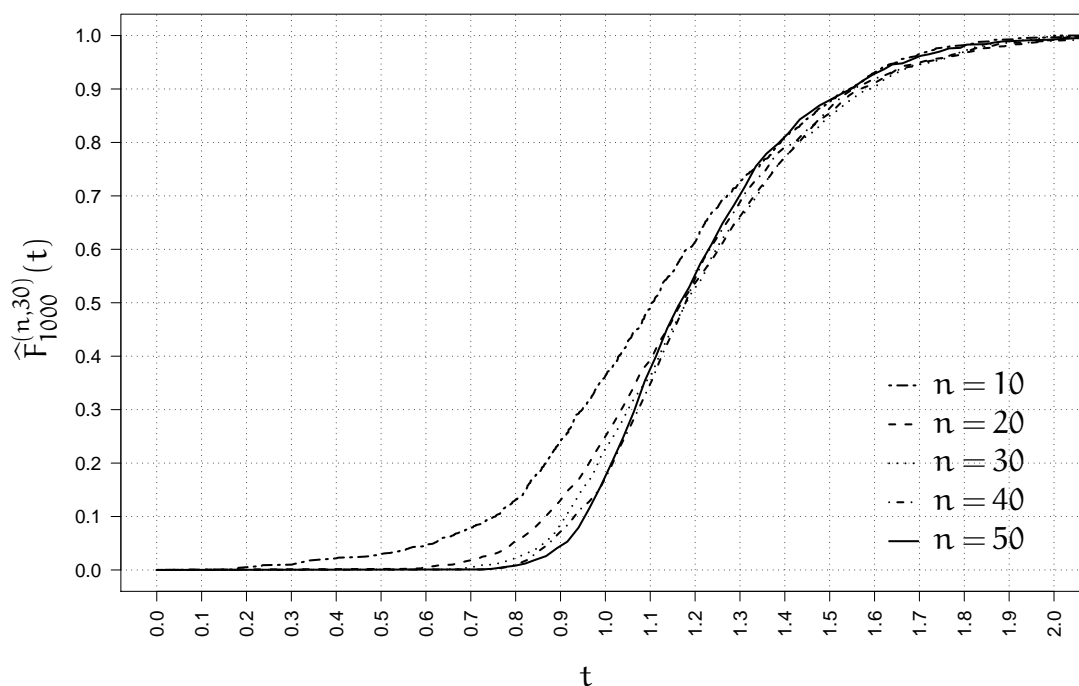


Figura 4.14: Função empírica das variáveis $X^{(n,30)}$

Aqui percebemos que os casos com atratividade nula ou negativa continuam melhorando consideravelmente, no entanto, não há uma melhora significativa para atratividade positiva $\alpha = 30$.

Notemos que para poucos sensores, com atratividade $\alpha = 0$, isto é, para o caso binomial, a situação é razoavelmente boa.

4.3.2 Estudo de caso II: processos pontuais atrativo e repulsivo

Como vimos na seção 4.2.2, existem diferentes fatores que devem ser levados em conta ao momento de projetar uma RSSF, dentre eles o custo dos sensores, o alcance de transmissão, o tempo de vida da rede, a probabilidade pretendida de conectividade, o capital disponível para a implementação, o meio no qual vão ser utilizados os sensores e o custo de posicionar os mesmos no ambiente. Este último item é de especial interesse para o problema estudado em esta seção.

O posicionamento dos sensores de maneira regular, ou seguindo uma metodologia manual de espalhamento, posicionando sensor a sensor, mantendo uma distância máxima e cobrindo uma área desejada, pode redundar em um custo maior do que simplesmente jogá-los de um avião. A última situação, aparentemente mais prática, pode não ser favorável não somente do ponto de vista da conectividade, mas também sob a óptica da qualidade da informação reconstruída.

Particularmente, os estudos sobre conectividade fornecem informação sobre o nível de funcionalidade que uma RSSF pode oferecer. A informação deve ser enviada corretamente aos nós *sink* e a topologia de comunicação deve fornecer suporte

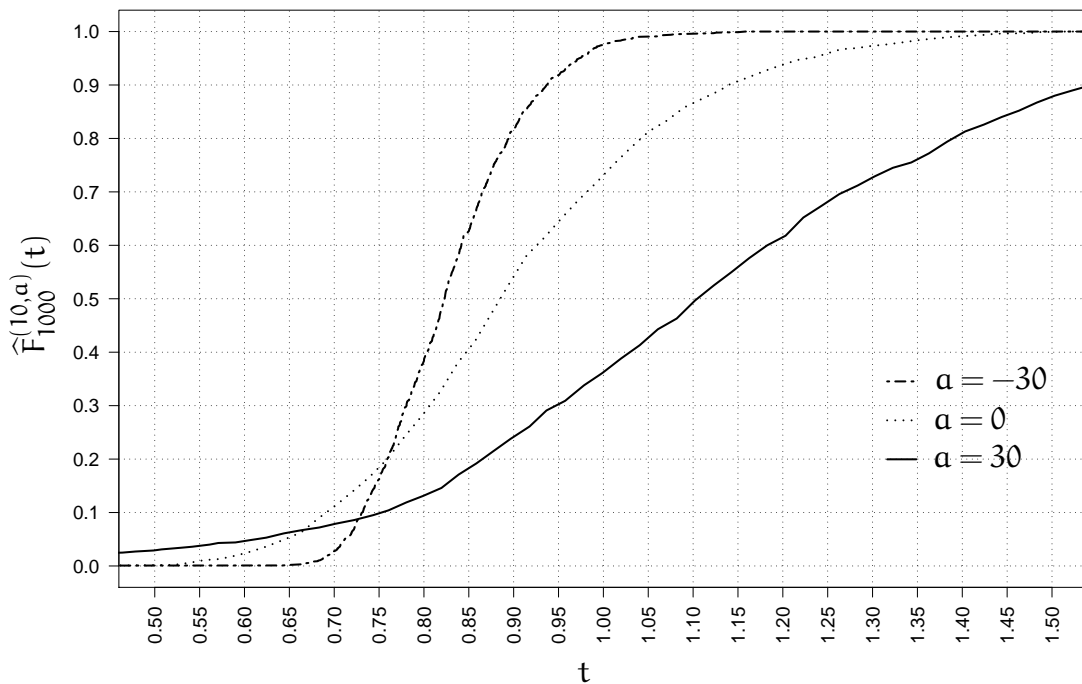


Figura 4.15: Função empírica das variáveis $\chi^{(10,a)}$

a este procedimento. Em ambiente hostis, segundo a aplicação, às vezes estamos interessados em garantir k -conectividade. Esta propriedade garante que na topologia existem de qualquer nó a qualquer outro nó, k caminhos independentes. Isto garante que podemos aleatoriamente tirar $k - 1$ nós sensores e a rede ainda continuará funcional, isto é, haverá pelo menos 1-conectividade, ou simplesmente a rede será conexa após essas retiradas. Isto pode ser requerido para garantir funcionalidade em ambientes onde os nós sensores estão propensos a falhas de *hardware*.

Por outro lado, funcionalidade não implica que a informação recoletada seja de qualidade. Cobertura trata sobre o estudo de topologias que tornam o vetor que representa os dados mensurados \mathbf{V} o mais perto o possível do vetor de dados físicos \mathbf{V}^* (ver o diagrama da página 6).

O estudo de cobertura, contudo, vai além do propósito desta dissertação. Destacamos que procesos atrativos irão exibir, tipicamente, deficiências do ponto de vista da cobertura. Isto é, caso os sensores estejam juntos, ocupando uma região reduzida, não conseguiremos extrair informação do resto do ambiente.

Suponhamos que desejamos implementar uma RSSF num ambiente. Dispomos de diferentes maneiras de posicionar os sensores no ambiente, e cada uma delas tem um custo associado. Em particular, vamos considerar que existe uma função $m: a \rightarrow \mathbb{R}_+$, que modela o custo de posicionar um nó sensor de maneira planejada segundo um processo pontual C com atratividade a . Espera-se que esta função seja decrescente, isto é, que se paga um preço cada vez maior quanto mais regularmente se deseje espalhar os sensores.

Poderíamos, desta maneira, considerar os mesmos parâmetros escolhidos na for-

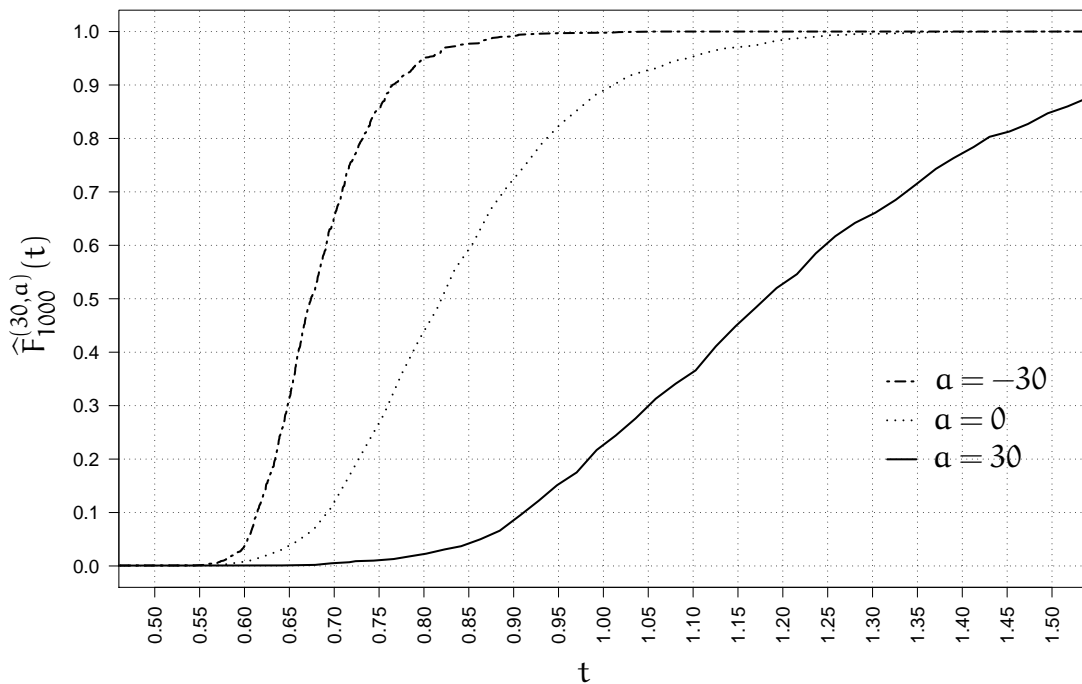


Figura 4.16: Função empírica das variáveis $\chi^{(30,a)}$

mulação do caso de estudo da seção 4.2.2, mas agora acrescentando uma variável a mais: a atratividade do processo.

Consideremos novamente que na implementação de uma RSSF, dispomos de vários tipos diferentes de sensores t , cada um deles com alcance máximo de transmissão r_t . O custo unitário de cada tipo de sensor é R_t . Para a compra dos sensores e para o posicionamento no ambiente dispomos de um capital T . Posicionar cada sensor no ambiente segundo um processo com atratividade a tem um custo m_a . Considerando o propósito da RSSF, achamos aceitável uma chance de $p100\%$ de possibilidade da rede ser conexas. Consideramos a comunicação sem fio no ambiente com expoente de *path loss* α . Estamos interessados em responder a seguinte pergunta:

quais e quantos sensores devem ser comprados, com que raio de transmissão vamos utilizá-los, e qual a atratividade com a qual vamos posicioná-los no ambiente de maneira a maximizar o tempo de vida da RSSF?

Para responder esta pergunta, novamente vamos formular um problema de otimização muito similar ao formulado na seção 4.2.2. Os dados que devemos fornecer ao modelo são os seguintes:

- r_t : denota o raio máximo de transmissão do sensor de tipo t ,
- R_t : é o preço de aquisição unitário do tipo de sensor t ,
- m_a : é o custo de posicionar um sensor no ambiente, independentemente do tipo, em função da regularidade a ,

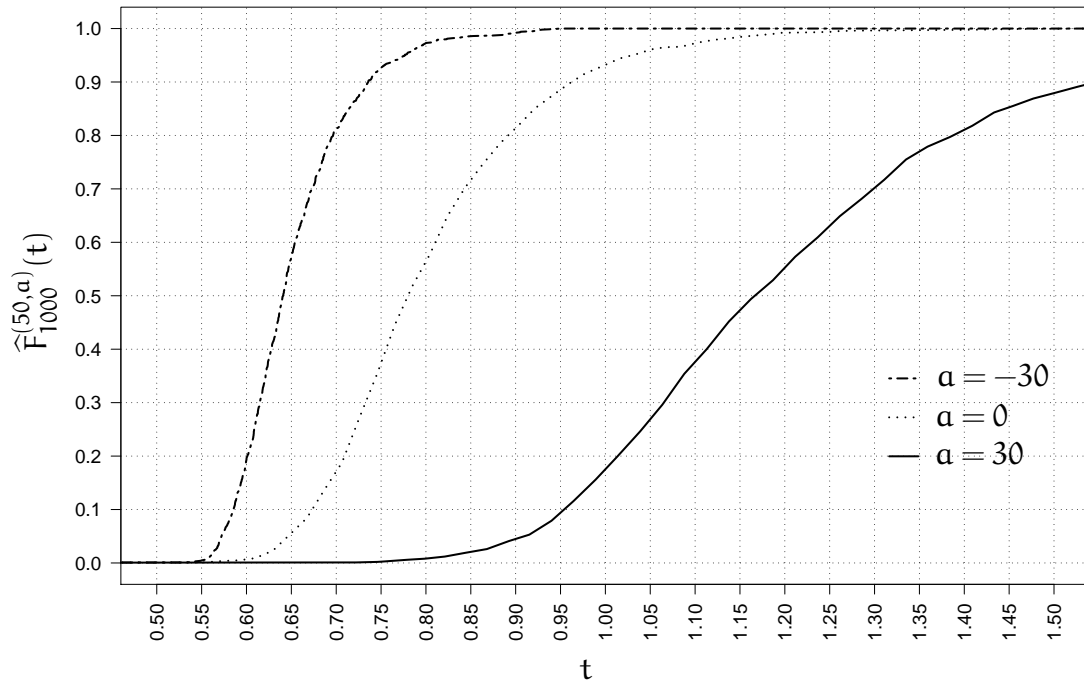


Figura 4.17: Função empírica das variáveis $\chi^{(50,\alpha)}$

- T : denota o capital disponível,
- p_{sr}^a : é a probabilidade de conectividade da RSSF para s sensores com raio de transmissão r e atratividade a ,
- α : é o parâmetro de *path loss*.

Desta maneira as variáveis a determinar são:

- ϕ : valor da constante φ do teorema 3,
- s : quantidade de sensores que deve ser comprada,
- t : tipo de sensor que devemos utilizar,
- r : raio de transmissão empregado na implementação,
- a : a atratividade com a qual vamos posicionar os nós sensores no ambiente.

As restrições do modelo são as seguintes:

$$\begin{aligned} s(R_t + m_a) &\leq T, \\ r &\leq r_t, \\ r &= \phi \sqrt{\frac{\ln s}{s}}, \text{ e} \\ p_{s\phi}^a &\geq p. \end{aligned}$$

Finalmente, a função objetivo para minimizar do custo c_e da RSSF é, novamente,

$$\min\{c_e : c_e = sr^\alpha\}.$$

Este capítulo apresentou a caracterização da conectividade por meio da constante desconhecida φ presente na expressão do CTR, para processos pontuais binomiais, proposta por Panchapakesan & Munjunath (2001). Também foi caracterizada a conectividade para o processo pontual $C(n, \alpha)$, proposto por Frery et al. (2008), para diferentes valores de atratividade α utilizando a mesma constante e a mesma expressão. Em ambos os casos, foi apresentado um modelo de otimização simples para resolver problemas práticos através dos resultados numéricos obtidos na caracterização.

Capítulo 5

Conclusões

Concluimos este trabalho de dissertação apresentando na seção 5.1 o resumo dos principais resultados obtidos, e na seção 5.2 comentamos potenciais trabalhos futuros e novas direções de pesquisa.

5.1 Resumo dos resultados e conclusões

Um dos primeiros resultados deste trabalho foi a respeito da avaliação da precisão numérica de ferramentas de análise de dados. Diante dos resultados de precisão numérica obtidos, recomendamos fortemente trabalhar com a plataforma R já que com esta ferramenta obtivemos os melhores resultados em todos os casos de estudo. Concluimos que resultados obtidos com algumas ferramentas usualmente empregadas em análise de dados carecem de respaldo, e deveriam estar, no mínimo, sujeitos a revisões. Dentro das ferramentas que avaliamos neste trabalho, as que devem se evitar variam segundo os procedimentos em que estejamos interessados em utilizar. Em particular, para a aplicação das funções para o cálculo da média e desvio padrão, sugerimos não utilizar Oleeo, Ox nem Excel 2008 para Mac OS. Já para o cálculo do coeficiente de autocorrelação, somente podemos confiar nos resultados obtidos com Ox e R. Se desejamos aplicar análise da variância, recomendamos utilizar R e Gnumeric, evitando utilizar Excel, Octave e principalmente Phyton. No que tange a regressões, tanto linear quanto não linear, a única ferramenta que deve se empregar é R, evitando ao máximo utilizar Excel (ver McCullough & Heiser, 2008). Quanto às implementações de funções de quantis de distribuições, tanto R, quanto Ox e Octave podem se empregar nos casos das distribuições gaussiana e F. Já no caso das distribuições t-student e χ^2 somente é recomendável utilizar as funções da ferramenta R. Finalmente, concluindo com a análise da precisão numérica das ferramentas para análise de dados, observamos que tanto Octave, Ox, Phyton, quando R, estão equipados com geradores de números pseudoaleatórios excelentes. Esses resultados redundaram nas publicações Almiron et al. (2008, in press) e no artigo Almiron et al. (submitted), em avaliação no *Journal of Statistical Software*.

Nos estudos de conectividade, vários resultados foram obtidos. Numa primeira instância, conseguimos caracterizar um estimador da distribuição do CTR (*Critical Transmission Range*) em cenários onde n nós sensores se encontram posicionados segundo um processo pontual espacial binomial. Nossa contribuição foi a caracterização da constante livre do teorema 3 (apresentado na página 44) derivado por

Panchapakesan & Munjunath (2001).

A expressão que modela o comportamento do valor do CTR para o posicionamento segundo processos pontuais binomiais foi estendido para processos atrativos e repulsivos. Para estes cenários foi utilizado o processo pontual espacial $C(n, \alpha)$ de Frery et al. (2008). Aqui também foi caracterizada, via um estimador, a distribuição do CTR em processos pontuais com atratividade α e n sensores.

Com estes resultados, foram modelados problemas de otimização simples que, empregando os estimadores obtidos, dão suporte ao projeto de RSSFs. Este modelo é formulado como um esquema geral, e a mesma metodologia pode ser utilizada para personalizá-lo segundo os objetivos da RSSF. A partir de dados como preços dos sensores, raio de transmissão máximo, custo de posicionamento como função do parâmetro de atratividade α , orçamento para a implementação da RSSF, probabilidade mínima admissível de conectividade e o parâmetro de *path loss*, conseguimos responder a perguntas tais como:

quais e quantos sensores devem ser comprados, com que raio de transmissão vamos utilizá-los, e qual o preço que é conveniente pagar pelo posicionamento de cada sensor (função da atratividade α) no ambiente de maneira a maximizar o tempo de vida da RSSF?

Na próxima seção apresentamos as direções futuras para pesquisa que surgiram a partir do trabalho aqui apresentado.

5.2 Trabalhos futuros

Diante da relevância dos resultados obtidos analisando a precisão numérica de plataformas computacionais quando aplicadas ao cálculo de funções estatísticas simples, e diante da necessidade de fazer medições críticas para analisar a conectividade de redes de sensores sem fios, deixamos o registro da importância de dar continuidade a esses estudos. Em particular, pretendemos propor e aplicar um protocolo, similar ao proposto por McCullough (1998), para a análise da precisão do cálculo de autovalores em estudos de conectividade algébrica. Resultados exploratórios preliminares mostraram que há situações onde a rede é conexa, porém λ_2 calculada na plataforma R é um número negativo muito próximo de zero. Esse tipo de erro levaria a concluir que a rede não possui conectividade quando, na verdade, ela é conexa.

Com a conectividade algébrica poderíamos estender o problema da conectividade, que é binário, a um problema contínuo. Com isso, estaríamos aptos a analisar a possibilidade de sacrificar um pouco de recursos permitindo uma pequena proporção de isolamento de nós sensores no ambiente, para obtermos um ganho significativo no consumo de energia e na capacidade de tráfego da RSSF. No momento de conclusão da escrita desta dissertação estávamos escrevendo um artigo com estes resultados de conectividade.

Apêndice A

Notação

A continuação apresentamos a notação utilizada nesta dissertação. A ordem adotada foi a mesma escolhida no texto, isto é, a ordem em que foram apresentadas nos diferentes capítulos.

Expressão	Significado
\vec{G}	Grafo direcionado simples (pág. 11)
G	Grafo não direcionado simples (pág. 12)
$A(\vec{G})$	Matriz de adjacência associada ao grafo \vec{G} (pág. 12)
$k_u^{\text{out}}(\vec{G})$	Grau de saída do vértice u do grafo \vec{G} (pág. 13)
$k_u^{\text{in}}(\vec{G})$	Grau de entrada de vértice u do grafo \vec{G} (pág. 13)
$k_u(\vec{G})$	Grau do vértice u do grafo \vec{G} (pág. 13)
$v_i \rightsquigarrow v_j$	Caminho simples entre v_i e v_j (pág. 13)
$L(\vec{G})$ ou L	Laplaciano do grafo \vec{G} (pág. 15)
$\lambda_2(\vec{G})$ ou λ_2	Conectividade algébrica do grafo \vec{G} (pág. 15)
N_t	Processo de contagem acumulado (pág. 18)
$\mathbb{1}[B]$	Função indicadora no conjunto B (pág. 18)
$N[a, b]$	Processo de contagem no intervalo $[a, b]$ (pág. 18)
$N(B)$	Região de contagem de B (pág. 19)
$\xi(B)$	Área da região B (pág. 20)
$B(n)$	Processo pontual binomial com n pontos (pág. 21)
$M(n, r(p))$	Processo pontual SSI de Matérn de n pontos com inibição r (pág. 21)
$S(n, a, \eta)$ ou $S(n, a)$	Processo pontual composto por dois processos Poisson com n pontos, intensidade $a\eta$ numa região alvo e intensidade η no complemento (pág. 21)
$C(n, a)$	Processo pontual de Frery et al. (2008) com n pontos e atratividade a (pág. 22)
$M_d = (S, P)$	Configuração de uma RSSF no espaço d -dimensional (pág. 23)
$P(\cdot)$	Função de posicionamento dos sensores no espaço (pág. 23)
$RA(\cdot)$	Função de atribuição do raio de comunicação (pág. 25)
p_i	Potência gasta pelo nó sensor i (pág. 25)

$\delta_{(P(i),P(j))}$	Distância euclidiana entre as posições dos sensores i e j (pág. 25)
$c_e(RA)$	Custo energético da função RA (pág. 25)
$CG = (S, \vec{E})$	Grafo de comunicações associado à RSSF (pág. 26)
$MD(x, c)$	Função de <i>Matching Digits</i> (pág. 31)
$F_X(q_p)$	Quantil de ordem p (pág. 36)

Apêndice B

Publicações

Este apêndice contém as publicações realizadas e (uma publicada, uma aceita e uma submetida) durante a realização do mestrado. São elas: Almiron et al. (2008), Almiron et al. (in press) e Almiron et al. (submitted).

Bibliografia

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. & Cayirci, E. (2002), 'Wireless sensor networks: a survey', *Computer Networks-The International Journal of Computer and Telecommunications Networking* **38**(4), 393–422.
- Almiron, M. G., Almeida, E. L. & Miranda, M. N. (in press), 'The reliability of statistical functions in four software packages freely used in numerical computation', *Brazilian Journal of Probability and Statistics*. URL <http://www.imstat.org/bjps>.
- Almiron, M. G., Lopes, B., Oliveira, A. L. C. & Medeiros, A. C. (2008), Precisão das funções estatísticas de ferramentas spreadsheet: uma análise comparativa, in 'XXXI Congresso Nacional de Matemática Aplicada e Computacional'. URL http://www.sbmac.org.br/eventos/cnmac/xxxi_cnmac/PDF/208.pdf.
- Almiron, M. G., Lopes, B., Oliveira, A. L. C., Medeiros, A. C. & Frery, A. C. (submitted), 'On the numerical accuracy of spreadsheets', *Journal of Statistical Software*.
- Altman, M. (2002), 'A review of JMP 4.03 with special attention to its numerical accuracy', *American Statistician* **56**(1), 72–75.
- Altman, M., Gill, J. & McDonald, M. P. (2007), 'Accuracy: Tools for accurate and reliable statistical computing', *Journal of Statistical Software* **21**(1), 1–30.
- Ammari, H. M. & Das, S. K. (2008), 'Integrated coverage and connectivity in wireless sensor networks: A two-dimensional percolation problem', *IEEE Transactions on Computers* **57**(10), 1423–1434.
- Andel, T. R. & Yasinsac, A. (2006), 'On the credibility of MANet simulations', *Computer* **39**(7), 48–54.
- Baddeley, A. (2007), Spatial point processes and their applications, in A. Baddeley, I. Barany, R. Schneider & W. Weil, eds, 'Stochastic Geometry', Vol. 1892 of *Lecture Notes in Mathematics*, Springer, Berlin / Heidelberg, pp. 1–75.
- Baker, C., Armijo, K., Belka, S., Benhabib, M., Bhargava, V., Burkhart, N., Minassians, A., Dervisoglu, G., Gutnik, L., Haick, M., Ho, C., Koplow, M., Mangold, J., Robinson, S., Rosa, M., Schwartz, M., Sims, C., Stoffregen, H., Waterbury, A., Leland, E., Perring, T. & Wright, P. (2007), Wireless sensor networks for home health care, in '21th International Conference on Advanced Information Networking and Applications Workshops, Ontario, Canada', IEEE Computer Press, pp. 832–837.
- Bettstetter, C. & Hartmann, C. (2005), 'Connectivity of wireless multihop networks in a shadow fading environment', *Wireless Networks* **11**(5), 571–579.

- Booth, L., Bruck, J., Franceschetti, M. & Meester, R. (2003), 'Converging algorithms, continuum percolation and the geometry of wireless networks', *The Annals of Applied Probability* **13**(2), 722–741.
- Bustos, O. H. & Frery, A. C. (2005), 'Statistical functions and procedures in IDL 5.6 and 6.0', *Computational Statistics and Data Analysis* **50**(2), 301–310.
- Cavin, D., Sasson, Y. & Schiper, A. (2002), On the accuracy of MANet simulators, in 'Proceedings of the 2nd ACM International Workshop on Principles of Mobile Computing', ACM, New York, USA, pp. 38–43.
- Chandrakasan, A., Amirtharajah, R., Cho, S., Goodman, J., Konduri, G., Kulik, J., Rabiner, W. & Wang, A. (1999), Design considerations for distributed microsensor systems, in 'IEEE Custom Integrated Circuits Conference', IEEE Computer Press, San Diego, CA, USA, pp. 279–286.
- Chang, J. H. & Jan, R. H. (2005), An energy aware, cluster-based routing algorithm for wireless sensor networks, in 'Embedded and Ubiquitous Computing - EUC 2005', Vol. 3824, pp. 255–266.
- Chong, C.-Y. & Kumar, S. (2003), 'Sensor networks: evolution, opportunities, and challenges', *Proceedings of the IEEE* **91**(8), 1247–1256.
- Diaz, J., Penrose, M. D., Petit, J. & Serna, M. (2000), 'Convergence theorems for some layout measures on random lattice and random geometric graphs', *Combinatorics, Probability and Computing* **9**, 489–511.
- Dijkstra, E. W. (1982), 'Why numbering should start at zero'. URL <http://www.cs.utexas.edu/users/EWD/ewd08xx/EWD831.PDF>, circulado de maneira privada.
- Ding, L. & Guan, Z. H. (2008), 'Modeling wireless sensor networks using random graph theory', *Physica A-Statistical Mechanics and its Applications* **387**(12), 3008–3016.
- Erdős, P. & Rényi, A. (1960), 'On the evolution of random graphs', *Bulletin of the International Statistical Institute* **38**(4), 343–347.
- Fiedler, M. (1973), 'Algebraic connectivity of graphs', *Czechoslovak Mathematical Journal* **23**(2), 298–305.
- Franceschetti, M., Booth, L., Cook, M., Meester, R. & Bruck, J. (2003), Percolation in multi-hop wireless networks, Technical Report CaltechPARADISE:2003.ETR055, California Institute of Technology. URL <http://www.paradise.caltech.edu/papers/etr055.pdf>.
- Frery, A. C., Ramos, H., Alencar-Neto, J. & Nakamura, E. (2008), Error estimation in wireless sensor networks, in 'Proceedings of the 23rd Annual ACM Symposium on Applied Computing', pp. 1923–1927.
- Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M. & Rossi, F. (2003), *GNU Scientific Library Reference Manual*, second ed., Network Theory Ltd.

- Ghosh, A. & Das, S. K. (2006), Coverage and connectivity issues in wireless sensor networks, *in* R. Shorey, A. L. Ananda & M. C. C. W. T. Ooi, eds, 'Mobile, Wireless, and Sensor Networks: Technology, Applications, and Future Directions', John Wiley & Sons, Inc., pp. 221–256. URL <http://dx.doi.org/10.1002/0471755591.ch9>.
- Gilbert, E. N. (1961), 'Random plane networks', *Journal of the Society for Industrial and Applied Mathematics* **9**(4), 533–543.
- Glauche, I., Krause, W., Sollacher, R. & Greiner, M. (2003), 'Continuum percolation of wireless ad hoc communication networks', *Physica A-Statistical Mechanics and Its Applications* **325**, 577–600.
- Gross, J. L. & Yellen, J., eds (2003), *Handbook of Graph Theory*, Discrete Mathematics and its Applications, CRC Press.
- Gupta, P. & Kumar, P. (2000), 'The capacity of wireless networks', *IEEE Transactions on Information Theory* **46**(2), 388–404.
- Harary, F. & Gupta, G. (1997), 'Dynamic graph models', *Mathematical and Computer Modelling* **25**(7), 79–87.
- Heinzelman, W., Chandrakasan, A. & Balakrishnan, H. (2002), 'An application-specific protocol architecture for wireless microsensor networks', *IEEE Transaction on Wireless Communication* **1**, 660–670.
- Heinzelman, W. R., Chandrakasan, A. & Balakrishnan, H. (2000), Energy-efficient communication protocol for wireless microsensor networks, *in* 'Proceedings of the 33rd Hawaii International Conference on System Sciences', IEEE Computer Press, pp. 1–10.
- Huber, P. J. (1981), *Robust Statistics*, Wiley, New York.
- Johnstone, I., Nicholson, J., Shehazad, B. & Slipp, J. (2007), Experiences from a wireless sensor network deployment in a petroleum environment, *in* 'Proceedings of the 2007 International Conference on Wireless Communications and Mobile Computing, Honolulu, Hawaii', ACM, New York, USA, pp. 382–387.
- Kampstra, P. (2008), 'Beanplot: A boxplot alternative for visual comparison of distributions', *Journal of Statistical Software, Code Snippets* **28**(1), 1–9. URL <http://www.jstatsoft.org/v28/c01/>.
- Keeling, K. B. & Pavur, R. J. (2007), 'A comparative study of the reliability of nine statistical software packages', *Computational Statistics & Data Analysis* **51**(8), 3811–3831.
- Kim, H. & Youn, H. Y. (2006), Finding the number of clusters minimizing energy consumption of wireless sensor networks, *in* 'Computational Science and Its Applications - ICCSA 2006, Pt 3', Vol. 3982, pp. 852–861.
- Knüsel, L. (1989), 'Computation of statistical distributions'. URL <http://www.stat.uni-muenchen.de/~knuesel/>, última consulta em fevereiro de 2009.
- Knüsel, L. (1998), 'On the accuracy of statistical distributions in Microsoft Excel 97', *Computational Statistics & Data Analysis* **26**(3), 375–377.

- Krishnamurthy, L., Adler, R., Buonadonna, P., Chhabra, J., Flanigan, M., Kushalmager, N., Nachman, L. & Yarvis, M. (2005), Design and deployment of industrial sensor networks: experiences from a semiconductor plant and the North Sea, *in* 'Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (Sensys), San Diego, CA', ACM, New York, USA, pp. 64–75.
- Marsaglia, G. (1998), 'The diehard battery of tests of randomness'. URL <http://www.stat.fsu.edu/pub/diehard>, última consulta em fevereiro de 2009.
- Matsumoto, M. & Nishimura, T. (1998), 'Mersenne-Twister: a 623-dimensionally equidistributed uniform pseudorandom number generator', *ACM Transactions on Modeling and Computer Simulation* **8**(1), 3–30.
- McCullough, B. D. (1998), 'Assessing the reliability of statistical software: Part I', *American Statistician* **52**(4), 358–366.
- McCullough, B. D. (1999), 'Assessing the reliability of statistical software: Part II', *American Statistician* **53**(2), 149–159.
- McCullough, B. D. & Heiser, D. A. (2008), 'On the accuracy of statistical procedures in Microsoft Excel 2007', *Computational Statistics & Data Analysis* **52**(10), 4570–4578.
- McCullough, B. D. & Wilson, B. (1999), 'On the accuracy of statistical procedures in Microsoft Excel 97', *Computational Statistics & Data Analysis* **31**(1), 27–37.
- Meester, R. & Roy, R. (1996), *Continuum Percolation*, Cambridge University Press.
- Nakamura, E. F., Loureiro, A. A. F. & Frery, A. C. (2007), 'Information fusion for wireless sensor networks: Methods, models, and classifications', *ACM Computing Surveys* **39**(3), article 9, 55 pages.
- Narayanaswamy, S., Kawadia, V., Sreenivas, R. S. & Kumar, P. R. (2002), Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the COMPOW protocol, *in* 'Proceedings of European Wireless', pp. 156–162.
- National Institute of Standards and Technology (n.d.), 'Statistical reference datasets'. URL <http://www.itl.nist.gov/div898/strd/general/dataarchive.html>, última consulta em fevereiro de 2009.
- Noh, J. H., Lee, B. J., Ha, N. K. & Han, K. J. (2005), Routing protocols based on super cluster header in wireless sensor network, *in* 'Proceedings of the 4th International Conference on Networking', Vol. 3420, Springer-Verlag, Berlin, pp. 731–739.
- Panchapakesan, P. & Munjunath, D. (2001), On the transmission range in dense ad hoc radio networks, *in* 'Proceeding of IEEE Signal Processing Communication', IEEE Computer Press, pp. 1–5.
- Penrose, M. D. (1997), 'The longest edge of the random minimal spanning tree', *Annals of Applied Probability* **7**(2), 340–361.
- Penrose, M. D. (1999), 'On k-connectivity for a geometric random graph', *Random Structures & Algorithms* **15**(2), 145–164.

- Qian, Y., Zhou, J. F., Qian, L. P. & Chen, K. S. (2006), Prolonging the lifetime of wireless sensor network via multihop clustering, *in* 'Next Generation Teletraffic and Wired/Wireless Advanced Networking', Vol. 4003, pp. 118–129.
- Rappaport, T. S. (2002), *Wireless Communications: Principles and Practice*, second ed., Prentice Hall.
- Reis, I. A., Câmara, G., Assunção, R. & Vieira Monteiro, A. M. (2007), Data-aware clustering for geosensor networks data collection, *in* 'XIII Simpósio Brasileiro de Sensoriamento Remoto - Florianópolis', INPE, pp. 6059–6066.
- Römer, K. & Mattern, F. (2004), 'The design space of wireless sensor networks', *IEEE Wireless Communications* **11**, 54–61.
- Santi, P. (2005), 'Topology control in wireless ad hoc and sensor networks', *ACM Computing Surveys* **37**(2), 164–194.
- Santi, P. & Blough, D. (2002), An evaluation of connectivity in mobile wireless ad hoc networks, *in* 'Proceedings of IEEE Dependable Systems and Networks', pp. 89–98.
- Santi, P. & Blough, D. (2003), 'The critical transmitting range for connectivity in sparse wireless ad hoc networks', *IEEE Transactions on Mobile Computing* **2**(1), 25–39.
- Santi, P., Blough, D. & Vainstein, F. (2001), A probabilistic analysis for the range assignment problem in ad hoc networks, *in* 'Proceedings of ACM Mobihoc', pp. 212–220.
- Santoro, N. (2007), *Design and Analysis of Distributed Algorithms*, Parallel and Distributed Computing, John Wiley & Sons, Inc.
- Simon, G., Maroti, M., Ledeczi, A., Balogh, G., Kusy, B., Nadas, A., Pap, G., Sallai, J. & Frampton, K. (2004), Sensor network-based countersniper system, *in* 'Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (Sensys)', ACM, Baltimore, MD, USA, pp. 1–12.
- Simon, S. D. & Lesage, J. P. (1989), 'Assessing the accuracy of ANOVA calculations in statistical software', *Computational Statistics & Data Analysis* **8**(3), 325–332.
- Sohraby, K., Minoli, D. & Znati, T. (2007), *Wireless Sensor Networks: Technology, Protocols, and Applications*, John Wiley & Sons, Inc.
- Tan, H. O. & Korpeoglu, I. (2003), 'Power efficient data gathering and aggregation in wireless sensor networks', *Sigmod Record* **32**(4), 66–71.
- Tolle, G., Polastre, J., Szewczyk, R., Culler, D., Turner, N., Tu, K., Burgess, S., Dawson, T., Buonadonna, P., Gay, D. & Hong, W. (2005), A macroscope in the redwoods, *in* 'Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (Sensys)', San Diego, CA, USA, pp. 51 – 63. ACM.
- Vasilescu, I., Kotay, K., Rus, D., Dunbabin, M. & Corke, P. (2005), Data collection, storage, retrieval with an underwater sensor network, *in* 'Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (Sensys)', ACM, San Diego, CA, USA, pp. 154–165. URL <http://doi.acm.org/10.1145/1098918.1098936>.

- Velho, L., Frery, A. C. & Gomes, J. (2008), *Image Processing for Graphics and Vision*, Texts in Computer Science, 2 ed., Springer.
- Wang, A., Heinzelman, W. B., Sinha, A. & Chandrakasan, A. P. (2001), 'Energy-scalable protocols for battery-operated microsensor networks', *Journal of VLSI Signal Processing Systems for Signal Image and Video Technology* **29**(3), 223–237.
- Watts, D. & Strogatz, S. (1998), 'Collective dynamics of small-world networks', *Nature* **393**(6684), 440–442.
- Werner-Allen, G., Lorincz, K., Welsh, M., Marcillo, O., Johnson, J., Ruiz, M. & Lees, J. (2006), 'Deploying a wireless sensor network on an active volcano', *IEEE Internet Computing* **10**, 18–25.
- Yick, J., Mukherjee, B. & Ghosal, D. (2008), 'Wireless sensor network survey', *Computer Networks* **52**(12), 2292–2330.
- Zhang, P., Sadler, C., Lyon, S. & Martonosi, M. (2004), Hardware design experiences in ZebraNet, in 'Proceedings of the Second International Conference on Embedded Networked Sensor Systems (Sensys), Baltimore, MD'.
- Zhao, F., Liu, J., Liu, J., Guibas, L. & Reich, J. (2003), 'Collaborative signal and information processing: An information-directed approach', *Proceedings of The IEEE* **91**(8), 1199–1209.

Esta dissertação foi preparada em \LaTeX (pdfTeX, Version 3.141592-1.40.3) utilizando Kile como editor. As referências bibliográficas foram administradas com \BIBTeX com auxílio de JabRef (versão 2.3); múltiplas citações foram ordenadas cronologicamente. As figuras foram criadas com TikZ, PGF e R. Utilizamos fonte Utopia para o texto e a família Euler para as equações. Em versões prévias, a numeração de capítulos, seções, equações, tabelas, figuras e páginas começavam em zero seguindo a filosofia de Dijkstra (1982), mas isso não foi mantido por não seguir o padrão da ABNT. O ambiente de trabalho foi Linux Ubuntu Intrepid.

