

**UNIVERSIDADE FEDERAL DE ALAGOAS**  
**INSTITUTO DE COMPUTAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM**  
**COMPUTACIONAL DE CONHECIMENTO**

**ALYSON LEANDRO COSTA OLIVEIRA**

**O PROBLEMA DE *ENERGY HOLE* EM REDES DE SENSORES  
SEM FIOS: A INFLUÊNCIA DA LOCALIZAÇÃO DOS NÓS  
SENSORES NO CONSUMO DE ENERGIA**

**MACEIÓ**

**2013**

**ALYSON LEANDRO COSTA OLIVEIRA**

**O PROBLEMA DE *ENERGY HOLE* EM REDES DE SENSORES  
SEM FIOS: A INFLUÊNCIA DA LOCALIZAÇÃO DOS NÓS  
SENSORES NO CONSUMO DE ENERGIA**

Dissertação apresentada ao Instituto de Computação da Universidade Federal de Alagoas, como parte das exigências do Programa de Pós-Graduação em Modelagem Computacional de Conhecimento, para obtenção do título de Mestre.

Orientador: Prof. Dr. Alejandro C. Frery

Coorientador: Prof. Dr. Heitor S. Ramos

**MACEIÓ**

**2013**

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**  
**Bibliotecário Responsável: Valter dos Santos Andrade**

O48p Oliveira, Alyson Leandro Costa.  
O problema de energy hole em redes de sensores sem fios: a influência da  
Localização dos nós sensores no consumo de energia / Alyson Leandro Costa  
Oliveira. – 2013.  
110 f. : il.

Orientador: Alejandro C. Frery.  
Coorientador: Heitor S. Ramos.  
Dissertação (mestrado em Modelagem Computacional de Conhecimento) –  
Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2013.

Bibliografia: f. 00-00.  
Apêndices: f. 00-00.

1. Energia – Consumo. 2. Redes de computação. 3. Sensores sem fio.  
4. Sistemas de comunicação sem fio. 5. Modelagem computacional.  
6. Controle de topologia. 7. Energy hole. 8. Processos pontuais espaciais.  
8. Sinalgo. 9. Método de Monte Carlo. I. Título.

CDU: 004.738.5





Membros da Comissão Julgadora da Dissertação de Mestrado de Alyson Leandro Costa Oliveira, intitulada: “O Problema de *Energy Hole* em Redes de Sensores sem Fio: a Influência da Localização no Consumo de Energia”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas em 20 de dezembro de 2013, às 11h00min, na sala de aula do Mestrado em Modelagem Computacional de Conhecimento da UFAL.

### COMISSÃO JULGADORA

**Prof. Dr. Alejandro Cesar Frery**

UFAL – Instituto de Computação

Orientador

**Prof. Dr. Heitor Soares Ramos Filho**

UFAL – Instituto de Computação

Orientador

**Prof. Dr. André Luiz Lins de Aquino**

UFAL – Instituto de Computação

Examinador

**Prof. Dr. Leandro Aparecido Villas**

UNICAMP – Instituto de Computação

Examinador



## AGRADECIMENTOS

A Deus tudo que me fora concedido e a fonte de sustento espiritual.

Ao meu pai, Evilazio, o incentivo e o apoio desde o início da minha caminhada profissional e acadêmica.

À minha mãe, Claudinete, o ensino da paciência e humildade.

Aos meus irmãos e minhas irmãs, Victor, Alan, Kyccia e Charmane.

À minha esposa, Dani, o amor, o carinho, a paciência (principalmente com os estudos em horários alternativos...), a espera (principalmente quando tive que me ausentar), a dedicação, o apoio, o incentivo e a disponibilidade para escutar minhas angústias, que foram fundamentais nessa fase da minha vida.

Aos colegas do LaCCAN, especialmente Rian, Paulo e Tony, as conversas e o aprendizado diário.

Aos colegas da UFMG, a acolhida na universidade e na cidade de Belo Horizonte.

Aos amigos de apartamento, Bruno, Lorena e, especialmente, Marcelo. Jamais esquecerei os momentos de troca de experiências vivido naqueles pouco mais de cinco meses.

À Karina, o carinho e a atenção com que me recebeu em sua casa.

Aos colegas do LCCV, em especial Rose, professor William e professor Leonardo, a compreensão e o incentivo para a finalização dessa etapa.

Ao professor Loureiro, a receptividade e os conselhos.

Ao professor e meu coorientador, Heitor, a mentoria profissional, acadêmica e de vida que já duram nove anos.

Ao professor e meu orientador, Alejandro, a paciência, os conselhos, os debates dos assuntos extracurriculares, a cobrança, o auxílio na formação de opinião, o ajudar a pensar, enfim, a orientação que permitiu a construção deste trabalho.

E, por fim, a todos os professores do Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da UFAL a contribuição que deram a minha formação profissional e acadêmica, aos colegas de disciplinas e às demais pessoas que participaram direta ou indiretamente na consolidação desse trabalho.

## RESUMO

Apresentamos as Redes de Sensores sem Fios (RSSFs) como uma ferramenta de comunicação e processamento de informações. Com o propósito de buscar uma alternativa para aproveitar ao máximo esse potencial, exploramos um dos principais desafios impostos pelas configurações desse tipo de rede: o consumo de energia. Mais especificamente, examinamos um problema conhecido na literatura como Problema do *Energy Hole*, com o objetivo de avaliar o comportamento do consumo energético entre os diversos nós sensores espalhados em um determinado ambiente de interesse. Baseado em: 1) RSSFs homogêneas e com comunicação por meio de múltiplos saltos; 2) uma modelagem de RSSFs, abordando algumas abstrações; e 3) definições de processos pontuais; propomos um processo pontual não-homogêneo para avaliar o que entendemos ser um dos pontos de principal ação no consumo de energia de uma RSSFs: a localização dos nós sensores e sua influência nas atividades de comunicação. Devido a complexidade de uma avaliação analítica, propomos e implementamos um experimento Monte Carlo que permite avaliar quantitativamente fatores como a relação entre a energia consumida ao longo de uma atividade de monitoramento e a densidade de nós sensores em uma região, assim como também a relação entre a taxa de cobertura e os cenários resultantes da aplicação do processo pontual proposto.

**Palavras-chaves:** redes de sensores sem fios. modelagem computacional. consumo de energia. controle de topologia. energy hole. processos pontuais espaciais. sinalgo. método monte carlo.

## ABSTRACT

Wireless Sensor Networks (WSNs) are presented as a communication and information processing tool. In order to find a way to make the most of their potential, we explore one of the central challenges imposed by the technical features of such networks: energy consumption. In particular, we explore the Energy Hole problem, in order to evaluate the behavior of energy consumption among sensor nodes scattered in an environment. Based on: 1) homogeneous WSNs and multihop communication; 2) a model for with some degree of abstraction; and 3) spatial point processes; we propose an inhomogeneous spatial point process to evaluate one of the most important factors in energy consumption: the distribution of sensors nodes and its influence on communication. Due to the complexity of dealing with this problem from an analytical viewpoint, a Monte Carlo experiment is devised to evaluate this influence. In particular, we study the relationship between energy expended and the sensor nodes density, and the relationship between the coverage and the resulting instances from the application of the proposed spatial point process.

**Keywords:** wireless sensor networks. computing modeling. energy consumption. topology control . energy hole. spatial point processes. sinalgo. monte carlo experiment.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de RSSFs. . . . .	24
Figura 2 – Exemplo de região de monitoramento com nós sensores distribuídos e alguns dos possíveis caminhos de comunicação. . . . .	30
Figura 3 – Exemplo de região de monitoramento com nós sensores distribuídos, alguns dos possíveis caminhos de comunicação e o problema do <i>energy hole</i> . . . . .	31
Figura 4 – Exemplos de posicionamento de maneira aleatória uniforme . . . . .	33
Figura 5 – Exemplos de posicionamento em um processo atrativo . . . . .	33
Figura 6 – Exemplos de posicionamento em um processo repulsivo . . . . .	34
Figura 7 – Exemplos de posicionamento de maneira aleatória uniforme e com nós heterogêneos . . . . .	35
Figura 8 – Exemplos de posicionamento em um processo atrativo . . . . .	35
Figura 9 – Tempos de chegada $T_i$ e entre chegadas $S_i$ . . . . .	37
Figura 10 – Região de contagem $N(B) = 2$ . . . . .	38
Figura 11 – Área alvo $A = C_r^p$ , para aplicação de intensidade $a\eta$ , e o restante da região total de interesse $B = A^c \cap C_r^p$ , para aplicação de intensidade $\eta$ . . . . .	42
Figura 12 – Exemplos de processos pontuais espaciais $C(n, a)$ . . . . .	43
Figura 13 – Exemplos de processos pontuais espaciais $M^2P^2$ . . . . .	46
Figura 14 – Caracterização do problema do <i>energy hole</i> . . . . .	49
Figura 15 – Cenários para distribuição aleatória uniforme (a, d) e o aumento da densidade de nós ao redor da estação base a partir do processo $M^2P^2$ (b, c) e nossa proposta (e, f). . . . .	50
Figura 16 – Exemplos de funções de intensidade. . . . .	51
Figura 17 – Comportamento da função intensidade $\kappa$ para alguns valores do parâmetro $a$ . . . . .	53
Figura 18 – Comportamento do processo $I$ para o parâmetro de intensidade $a = \{0, 1, 2\}$ e quantidade de pontos $n = \{100, 300, 500\}$ . . . . .	55
Figura 19 – Diagrama esquemático de RSSFs. . . . .	58
Figura 20 – Diagrama esquemático de RSSFs (escopo de atuação). . . . .	59
Figura 21 – Distribuição de dez nós sensores e suas respectivas áreas de comunicação. . . . .	63
Figura 22 – Impacto da densidade de nós no consumo de energia ao longo de uma RSSFs. . . . .	74
Figura 23 – Impacto da densidade de nós em regiões próximas à estação base no consumo de energia ao longo de uma RSSFs. . . . .	75
Figura 24 – Impacto da densidade de nós em regiões próximas à estação base no consumo de energia ao longo de uma RSSFs. . . . .	76



Figura 25 – Impacto do processo $I$ na cobertura da região de monitoramento. . . .	78
Figura 26 – Comportamento da dispersão de nós sensores quando aplicado o processo $I$ , para o parâmetro de intensidade $a = \{0, 1, 2\}$ e quantidade de nós $n = 500$ . . . . .	79
Figura 27 – Avaliação do resultado da aplicação do processo $I$ sob a ótica do consumo médio de energia e do <i>sinkbetweenness</i> . . . . .	80
Figura 28 – Avaliação do resultado da aplicação do processo $I$ sob a ótica do consumo médio de energia e do <i>sinkbetweenness</i> , para uma região particionada em 3 discos de saltos de comunicação. . . . .	81
Figura 29 – Avaliação do resultado da aplicação do processo $I$ sob a ótica do consumo médio de energia e do <i>sinkbetweenness</i> , para uma região particionada em 5 discos de saltos de comunicação. . . . .	81
Figura 30 – Análise comparativa do <i>sinkbetweenness</i> para o processo $I$ e o processo $M^2P^2$ . . . . .	82
Figura 31 – Análise comparativa do <i>sinkbetweenness</i> para o processo $I$ e o processo $M^2P^2$ , para uma região particionada em 3 discos de saltos de comunicação. . . . .	84
Figura 32 – Análise comparativa do <i>sinkbetweenness</i> para o processo $I$ e o processo $M^2P^2$ , para uma região particionada em 5 discos de saltos de comunicação. . . . .	84
Figura 33 – Análise comparativa para o impacto do processo $I$ e do processo $M^2P^2$ na cobertura da região de monitoramento. . . . .	85
Figura 34 – Análise comparativa para o impacto do processo $I$ e do processo $M^2P^2$ na cobertura da região de monitoramento, para um raio de sensoriamento $r_s = 2,5m$ . . . . .	86
Figura 35 – Análise comparativa para o impacto do processo $I$ e do processo $M^2P^2$ na cobertura da região de monitoramento, para um raio de sensoriamento $r_s \cong 4,17m$ . . . . .	87

## LISTA DE TABELAS

Tabela 1	– <i>MD</i> para a média amostral. . . . .	98
Tabela 2	– <i>MD</i> para o desvio padrão amostral. . . . .	99
Tabela 3	– <i>MD</i> para o coeficiente de autocorrelação amostral. . . . .	99
Tabela 4	– <i>MD</i> para o estatístico $F$ de ANOVA. . . . .	100
Tabela 5	– <i>MD</i> para regressão não linear, iniciando em <i>Start 1</i> e configuração padrão. . . . .	102
Tabela 6	– <i>MD</i> para regressão não linear, iniciando em <i>Start 2</i> e configuração padrão. . . . .	104
Tabela 7	– <i>MD</i> para regressão não linear, iniciando em <i>Start 1</i> e configuração personalizada. . . . .	106
Tabela 8	– <i>MD</i> para regressão não linear, iniciando em <i>Start 2</i> e configuração personalizada. . . . .	107
Tabela 9	– <i>MD</i> menores para o coeficiente $\hat{\beta}$ e o desvio padrão residual RSD em R aplicando regressão não linear. . . . .	110

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Motivação</b>	<b>17</b>
<b>1.2</b>	<b>Definição do problema</b>	<b>19</b>
<b>1.3</b>	<b>Proposta</b>	<b>20</b>
<b>1.4</b>	<b>Metodologia</b>	<b>21</b>
<b>2</b>	<b>FUNDAMENTOS E TRABALHOS RELACIONADOS</b>	<b>23</b>
<b>2.1</b>	<b>Redes de sensores sem fios</b>	<b>23</b>
2.1.1	Representação de uma rede de sensores sem fios	24
2.1.2	Organização da rede e comunicação de nós sensores	25
2.1.3	Problemática de energia das RSSFs	27
2.1.3.1	O problema do <i>energy hole</i>	29
<b>2.2</b>	<b>Processos pontuais</b>	<b>32</b>
2.2.1	Definindo processos pontuais	36
<b>2.3</b>	<b>Modelos de interesse: processos pontuais não-homogêneos</b>	<b>41</b>
2.3.1	O Processo pontual $C$	41
2.3.2	O Processo pontual $M^2P^2$	44
<b>3</b>	<b>PROPOSTA: UM MODELO DE DISPERSÃO DE NÓS PARA BANCAMENTO DE CARGA</b>	<b>47</b>
<b>3.1</b>	<b>Princípio</b>	<b>48</b>
<b>3.2</b>	<b>Modelagem</b>	<b>51</b>
3.2.1	Estudo da função de intensidade	51
3.2.2	Processo pontual não-homogêneo proposto: o processo $I$	52
<b>3.3</b>	<b>Alguns exemplos</b>	<b>54</b>
<b>4</b>	<b>METODOLOGIA</b>	<b>57</b>
<b>4.1</b>	<b>Ingredientes da simulação</b>	<b>58</b>
4.1.1	Modelos	60
4.1.1.1	Área de monitoramento	60
4.1.1.2	Distribuição de nós sensores	61
4.1.1.3	Características operacionais de nós sensores	61
4.1.1.3.1	Raio de comunicação	62
4.1.1.3.2	Raio de Sensoriamento	62
4.1.2	Análise do ambiente de simulação	64
4.1.2.1	Plataforma de simulação	65

4.1.2.2	Plataforma de análise de dados . . . . .	66
<b>4.2</b>	<b>Avaliação . . . . .</b>	<b>67</b>
4.2.1	Estudo Monte Carlo . . . . .	67
4.2.2	Cenários de estudo . . . . .	68
4.2.2.1	Relação entre energia consumida e densidade de nós . . . . .	68
4.2.2.2	Influência da cobertura . . . . .	69
4.2.2.3	Relação entre energia consumida e <i>sinkbetweenness</i> . . . . .	70
4.2.2.4	Análise comparativa entre processo <i>I</i> e processo $M^2P^2$ . . . . .	71
4.2.3	Ferramentas computacionais . . . . .	71
4.2.4	Detalhes de implementação . . . . .	72
<b>5</b>	<b>RESULTADOS . . . . .</b>	<b>73</b>
<b>5.1</b>	<b>Energia Consumida Vs. Densidade de Nós . . . . .</b>	<b>73</b>
<b>5.2</b>	<b>Cobertura . . . . .</b>	<b>76</b>
<b>5.3</b>	<b><i>Sinkbetweenness</i> . . . . .</b>	<b>79</b>
<b>5.4</b>	<b>Análise comparativa entre Processo <i>I</i> e Processo <math>M^2P^2</math> . . . . .</b>	<b>80</b>
5.4.1	<i>Sinkbetweenness</i> . . . . .	80
5.4.2	Cobertura . . . . .	83
<b>6</b>	<b>CONCLUSÃO . . . . .</b>	<b>89</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>91</b>
	<b>APÊNDICE A – AMBIENTE DE SIMULAÇÃO: UMA AVALIAÇÃO DA PRECISÃO NUMÉRICA . . . . .</b>	<b>97</b>
<b>A.1</b>	<b>Metodologia de avaliação da precisão numérica . . . . .</b>	<b>97</b>
<b>A.2</b>	<b>Resultados da avaliação . . . . .</b>	<b>98</b>
<b>A.3</b>	<b>Conclusões da avaliação . . . . .</b>	<b>108</b>

# 1 INTRODUÇÃO

Nos últimos anos, as Redes de Sensores sem Fios (RSSFs) vêm se destacando no meio científico não só pelo potencial de comunicação e processamento de informações que apresentam, mas também pelos desafios impostos a partir das características do elemento principal da rede: o nó sensor.

Esses nós são elementos compostos basicamente por uma unidade de processamento, uma unidade de sensoriamento, bateria e por uma unidade de comunicação, todos reunidos, geralmente, em um espaço reduzido, o que impõe diversas limitações aos recursos disponíveis (AKYILDIZ et al., 2002; YICK; MUKHERJEE; GHOSAL, 2008). A energia é um recurso chave para que um nó consiga efetuar as tarefas que lhe são atribuídas (RAGHUNATHAN et al., 2002; SADLER, 2005) e será estudada nesse trabalho.

## 1.1 Motivação

De fato, observamos na literatura um crescimento contínuo de estudos relacionados as RSSFs. É interessante perceber que essas estruturas são utilizadas para a aquisição de dados e informações, para a posterior geração de conhecimento, em estudos relacionados a diversos fenômenos da natureza como, por exemplo, comportamento de animais em seus *habitats*, medições de temperatura, pressão, humidade e iluminação e comportamento de rios e oceanos; e a diversas áreas de conhecimento como a biologia, a química, a física e as engenharias, dentre outras.

Também é fato que a estrutura proporcionada pelas RSSFs depende de um elemento com características físicas e computacionais especiais. Esse elemento, o nó sensor, apresenta, dentre outros fatores, uma limitação de energia. A percepção e o entendimento dessa limitação são fundamentais para a proposição de qualquer solução com o objetivo de proporcionar melhorias e avanços na área de Redes de Sensores sem Fios. O que estamos querendo demonstrar é que a principal motivação do nosso trabalho foi a proposição de um estudo que esteja diretamente relacionado a minimização de problemas decorrentes da limitação de energia dos nós sensores. Entretanto, como veremos alguns exemplos da Seção 2, a área de RSSFs é muita rica e multidisciplinar e, especificamente falando em energia, é preciso delimitar o escopo de atuação.

O principal poder das RSSFs está na característica de aquisição de informações de forma distribuída, através da interação de vários nós sensores espalhados sobre uma determinada área de interesse. Diante dessa característica e da limitação energética que está sendo discutida, é fácil pensar em duas abordagens gerais para tratar o problema de energia: aumentar (inicialmente ou ao longo do tempo de funcionamento da rede) a

quantidade de energia disponível em cada nó sensor e diminuir a quantidade de energia gasta ao longo do tempo de atividade da rede. Podemos perceber que o desafio se torna ainda maior quando a abordagem se resume a apenas uma, pois, conforme é explanado, por exemplo, em Culler, Estin e Srivastava (2004) e Liu et al. (2006) e Dardari et al. (2007), limitações de tamanho dos nós sensores, limitações geográficas de certas áreas de estudo e os objetivos de uma determinada aplicação tornam muito difíceis a realização de recargas de energia nas baterias dos nós sensores. Assim, para que uma RSSFs consiga se apresentar ativa para realizar seus objetivos pela maior quantidade de tempo possível, a segunda abordagem geral – a diminuição da quantidade de energia gasta na rede como um todo – se torna mais viável. Dessa forma, é necessário que identifiquemos quais os principais fatores responsáveis pelo consumo de energia em um nó sensor.

Em Tilak, Abu-Ghazaleh e Heinzelman (2002) e Culler, Estin e Srivastava (2004), por exemplo, observamos que as tarefas de comunicação são responsáveis pela maior quantidade de energia despendida por um nó sensor. Já em trabalhos mais recentes como o de Anastasi et al. (2009), podemos ver que, aliado as tarefas básicas de comunicação (transmissão e recepção), a maneira como os nós sensores se comunicam entre si e a disposição dos nós na área de estudo são os principais causadores do gasto de energia em RSSFs.

Detalhando um pouco mais o que entendemos por comunicação em RSSFs, estamos afirmando que características como o raio de comunicação, responsável por delimitar o alcance de comunicação de um nó com outros nós em uma rede, e principalmente, a forma como os dados sensorizados trafegam de um nó para outro nó são aspectos fundamentais na proposta de soluções de minimização de consumo de energia em uma RSSFs.

Detalhando ainda mais a maneira como os nós se comunicam, Li e Mohapatra (2007) destacam que as tarefas de sensoriamento e posterior aquisição de dados em RSSFs são feitas, geralmente, de duas maneiras: por único salto e por múltiplos saltos. Na comunicação através de único salto (*single-hop*) qualquer nó de uma RSSFs, após medir um certo fenômeno em estudo, envia suas informações diretamente a um nó central na rede, mais conhecido como nó sorvedouro ou estação base. É a partir desse nó que um usuário pode ter acesso às informações coletadas pelos nós da rede.

Já na comunicação por múltiplos saltos (*multihop*) é atribuído um papel extra aos nós sensores, o papel de retransmissor (*relay*) de dados, fazendo com que os nós sensores não precisem enviar seus dados diretamente à estação base e podendo, para isso, ajustar as configurações de seu rádio para que possam transmitir e receber dados em uma potência mais baixa, o que é apresentando por Santi (2005) como um fator bastante significativo para a economia de energia.

Até o momento apresentamos o nosso interesse na busca de uma solução que atue na minimização do problema de limitação de energia em Redes de Sensores sem Fios,



abordando como um dos principais fatores a forma como os nós sensores se comunicam. Explicaremos na próxima sessão que esse cenário, além de ter forte relação com a característica de deposição dos nós sensores ao longo de uma região de sensoriamento, apresenta um problema específico.

## 1.2 Definição do problema

Um efeito adverso que a comunicação através de múltiplos saltos ocasiona está ligado diretamente à maneira como os nós estão distribuídos no ambiente de monitoramento, tanto em termos de localização quanto em termos de quantidade. É fácil perceber que o papel de retransmissor que um nó assume na comunicação por múltiplos saltos pode ter um efeito ruim para o tempo de vida da rede como um todo. Como toda a coleta de dados é direcionada a um ponto central da rede (o nó sorvedouro), os algoritmos de roteamento aplicados à rede fazem com que cada nó busque um outro nó que está cada vez mais próximo da estação base para que a informação sensoriada seja percebida. Dessa forma, é intuitiva a percepção de que os nós que estão localizados a um salto da estação base gastam, em média, mais energia que os nós localizados a dois saltos da estação base, que por sua vez gastam mais energia que os nós localizados a três saltos, e assim por diante.

Acompanhando as descrições realizadas por Li e Mohapatra (2007), podemos observar que esse problema é sumarizado como um problema de consumo não-balanceado de energia entre os nós de uma RSSFs. Ainda de acordo com esse trabalho, os autores afirmam que esse perfil de consumo energético pode levar a uma diminuição significativa do tempo de vida de uma RSSFs, pois o gasto de energia médio dos nós localizados nas regiões mais próximas da estação base leva a “morte” prematura de tais nós, fazendo com que os outros nós que ainda possuem potencial energético para efetuar as tarefas de comunicação percam conectividade com a estação base. Esse “buraco” na rede justifica o nome pelo qual esse problema é mais conhecido na literatura, o problema do *energy hole*.

Além de todos os aspectos já relatados até aqui, os estudos de Medidi e Zhou (2007) e Wu, Chen e Das (2008) expõem outro fator de impacto no consumo de energia de uma RSSFs, e principalmente, no problema do *energy hole*. Em vários estudos de RSSFs (que devido a grande dificuldade de se obter resultados práticos muitas vezes a comunidade científica recorre às simulações) o modelo de dispersão dos nós em uma região de estudo é feito seguindo uma distribuição espacial aleatória uniforme. Do ponto de vista de resultados de simulação, Andel e Yasinsac (2006) já haviam apontado essa prática como uma das suposições imprecisas feitas em simulações, o que pode levar a resultados errados. Mas além disso, do ponto de vista de solução para o problema do *energy hole*, Medidi e Zhou (2007) e Wu, Chen e Das (2008), além de Liu (2006) e Zhang e Wu (2011), apontam como um importante fator a utilização de uma distribuição de nós não-uniforme para

que o problema possa ser minimizado, visto que todos eles apontam como ideia básica o aumento da densidade de nós nas regiões próximas à estação base, propiciando que mais nós possam realizar o papel de retransmissor.

### 1.3 Proposta

Diante do cenário exposto e partindo sempre das premissas relacionadas à influência das formas de comunicação e distribuição dos nós sensores em uma RSSF, no nosso trabalho propomos um modelo de distribuição dos nós sensores, que nos permite avaliar o consumo de energia de uma rede de sensores.

Em particular, o modelo de dispersão tem como objetivo principal fazer com que os nós sensores ao longo da rede tenham um consumo de energia balanceado, através da deposição de uma quantidade maior de nós sensores nas regiões que são mais próximas à estação base (sem alterar a quantidade total de nós sensores ao longo da rede), a fim de que o problema da “morte prematura” dos nós que se localizam nessas regiões seja minimizado. Essa finalidade está fundamentada também no fato de alguns trabalhos terem demonstrado que o modelo que rege a dispersão dos nós ser muito mais importante do que puramente representar os nós em um determinado espaço para se obter um cenário de simulação. Além de protocolos e algoritmos propostos precisarem ser testados em diferentes situações, cenários e topologias, diferentes tipos de distribuição dos nós tornam os cenários de simulação mais realistas. Adicionalmente, Younis e Akkaya (2008) apresentam algumas estratégias de distribuição de nós para redes de sensores sem fios, sugerindo que o modelo de distribuição pode ser otimizado em funções de algumas propriedades da rede que desejem ser alcançadas como tempo de vida, conectividade e cobertura.

Essas duas últimas propriedades podem sofrer impactos significantes devido ao fato de aumentar a densidade de nós em uma determinada região da rede. Para deixar claro o que estamos dizendo, no nosso trabalho a quantidade de nós sensores distribuídos em uma RSSFs é controlado. Assim, quando aumentamos a concentração de nós em uma determinada região nas proximidades da estação base a probabilidade dos nós terem pelo menos um vizinho vai se tornando cada vez mais alta. Porém o oposto acontece à medida que essas regiões de deposição de nós vão ficando mais distantes da estação base. Já com relação à cobertura, como apontado em Sadler (2005), Santi (2005), Liu (2006), Kumar, Lai e Balogh (2008) e Li et al. (2008), aumentar a densidade de nós em uma região só fará com que haja, dentre outros fatores, redundância de informações sensoriadas pelos nós, e diminuir a quantidade de nós pode levar a regiões onde as informações medidas são insuficientes para determinar um determinado evento que esteja acontecendo. Portanto, é importante frisar que, em nosso trabalho, o modelo proposto apenas leva em consideração cenários nos quais a rede apresenta uma taxa de conectividade e cobertura acima de um

determinado limiar, para que seja possível mostrar que o fato de distribuir os nós sensores através de um modelo diferente do usual proporcionará um aumento do tempo de vida útil da rede sem afetar conectividade e cobertura.

Correndo o risco de parecer que estamos adiantando a metodologia do nosso trabalho, um outro aspecto importante da nossa proposta é a representação esquemática para uma Rede de Sensores sem Fios, dando continuidade a modelagem apresentada no trabalho de Frery et al. (2010). Damos destaque a esse aspecto nesta seção por se tratar de mais uma contribuição para o fortalecimento dessa modelagem, a qual também é a base para a definição das abstrações que utilizaremos na nossa abordagem.

Para tanto, seguiremos uma modelagem composta pelos seguintes aspectos:

- a) região de monitoramento;
- b) distribuição dos nós sensores;
- c) características de operação dos nós sensores.

Diferentemente do trabalho de Frery et al. (2010), o qual trata de aspectos relacionados a agrupamentos e fusões de dados, tendo como métrica principal de avaliação a inferência que o usuário pode fazer sobre os dados originais do fenômeno em observação quando estão disponíveis os dados amostrados pelos nós sensores e coletados através da estação base, no nosso trabalho nós não seguiremos as modelagens do sinal amostrado e do sinal reconstruído. Com isso, apesar de nos preocuparmos com características importantes de uma RSSFs como conectividade e cobertura, não apresentaremos um estudo do impacto da nossa solução em confronto com a qualidade dos dados percebidos e transmitidos à estação base.

## 1.4 Metodologia

Como exposto anteriormente, dada a dificuldade em obter resultados puramente analíticos a partir da formalização das Redes de Sensores sem Fios, será adotada uma abordagem de utilização de técnicas de simulação. Como estamos falando em modelar uma porção do mundo real para obter resultados de avaliação da nossa proposta, é importante frisar a nossa preocupação com os erros e imprecisões os quais as simulações, como qualquer outro método ou técnica, estão sujeitas. Dessa forma, no nosso trabalho seguimos uma organização similar a que pode ser encontrada em Alencar-Neto (2007). Em nossa discussão, voltamos a atenção para três ingredientes básicos de simulação: os modelos (abstrações), método de Monte Carlo (ROBERT; CASELLA, 2000), ferramenta que tem se tornado fundamental no estudos na área e RSSFs, e às plataformas de simulação e análise de dados.

Especificamente sobre as plataformas de simulação e análise de dados, realizamos uma avaliação crítica detalhada com o objetivo de escolher, de fato, recursos computacionais

que nos dessem segurança quanto a precisão numérica das simulações e avaliações. Entendemos que essa é uma contribuição importante do nosso trabalho, pois, como acabamos de afirmar, simulações estão sujeitas a erros e imprecisões, e muitas dessas falhas podem estar relacionadas a uma má escolha das plataformas de simulação e análise (ANDEL; YASINSAC, 2006).

No que diz respeito especificamente à metodologia de avaliação, dentre outros aspectos, merece um destaque a análise comparativa em face a alguns resultados obtidos por Li e Mohapatra (2007), e com relação a um modelo de dispersão de nós chamado M<sup>2</sup>P<sup>2</sup> (detalhes na Seção 2.3.2), o qual também é uma contribuição encontrada em Frery et al. (2010) e tendo sua análise comparativa justificada pelo fato de se tratar de um processo também aplicável ao problema do *energy hole* (para alguns cenários específicos).

O restante desse trabalho está estruturado como segue:

- a) **Seção 2** – trata de definições importantes para o estudo de Redes de Sensores sem Fios, focando principalmente na definição da problemática de energia e do problema do *energy hole*, além de apresentar definições e notações relacionadas a processos pontuais, com destaque para os processos pontuais não-homogêneos de interesse para a nossa proposta;
- b) **Seção 3** – apresenta detalhes sobre a nossa proposta, ou seja, discute aspectos relacionados ao modelo de distribuição de nós não-uniforme proposto;
- c) **Seção 4** – descreve a metodologia utilizada em detalhes, apresentado a construção da simulação e os métodos de avaliação;
- d) **Seção 5** – exhibe os resultados obtidos por meio dos experimentos Monte Carlo;
- e) **Seção 6** – conclui a dissertação, discutindo aspectos relevantes dos resultados e apontando algumas questões para trabalhos futuros;
- f) **Apêndice A** – apresenta um detalhado estudo sobre a avaliação da precisão numérica de plataformas de simulação e análise de dados (proposta, metodologia e resultados abordados em Almiron et al. (2010)).

## 2 FUNDAMENTOS E TRABALHOS RELACIONADOS

### 2.1 Redes de sensores sem fios

Uma Rede de Sensores sem Fios (RSSFs) é uma ferramenta de sensoriamento distribuído composta por dezenas, centenas ou mais nós sensores que têm capacidade de monitorar, coletar, processar e transmitir informações à respeito de um determinado fenômeno/região a ser analisado. Esse tipo de rede pode ser aplicada no monitoramento, rastreamento, coordenação e processamento em diferentes contextos como, por exemplo, no controle de condições ambientais em uma floresta ou oceano, no monitoramento de tráfego de uma rodovia, na área de engenharia em monitoramentos e modelagens de estruturas e em monitoramentos de extrações de petróleo, e, até mesmo, no monitoramento de órgãos do corpo humano.

Apesar dos grandes avanços tecnológicos que têm ocorrido nas áreas de microprocessadores e comunicação sem fios (o que permitiu a idealização e a construção de RSSFs), esse poder de observação e determinado grau de controle do mundo real ainda apresentam diversas exigências e desafios, o que tem despertado o interesse de vários segmentos da comunidade científica.

Nesse trabalho, dentre outras questões importantes ligadas ao estudo de RSSFs, destacamos:

- a) as características de operação dos nós sensores, mais especificamente o modo como os nós criam *links* de comunicação entre si;
- b) a dispersão e organização dos nós no espaço geográfico onde são medidos os fenômenos;
- c) o consumo de energia ocasionado pelas tarefas realizadas pelos nós sensores, principalmente as tarefas de comunicação;
- d) a representação e o estudo de uma RSSFs através da estratégia de simulação.

Para mais detalhes sobre as principais características, desafios e soluções abordadas em RSSFs, recomendamos a leitura dos trabalhos de Akyildiz et al. (2002), Chong e Kumar (2003), Culler, Estin e Srivastava (2004) e Yick, Mukherjee e Ghosal (2008).

Dando um destaque especial para a última questão listada, percebemos que uma das dificuldades que pesquisadores têm encontrado é o fato da inviabilidade de desenvolvimento e implementação de técnicas e estudos práticos devido ao alto custo de implantação de uma RSSF. Como podemos observar na literatura, uma alternativa para isso tem sido a idealização e construção de técnicas capazes de representar e simular o comportamento das

RSSFs. Nesse trabalho, na Seção 2.1.1, será apresentada uma forma de representação de uma RSSF baseada no trabalho de Frery et al. (2010), em que é apresentado o funcionamento de uma RSSF no contexto de processamento de sinais. Na Seção 4, mostraremos a metodologia aplicada para tornar possível o estudo proposto através dessa modelagem de uma RSSF.

Adicionalmente, mostraremos nas próximas seções, em especial nas Seções 2.1.2 e 2.1.3, que utilizaremos esse modelo de representação para analisar a problemática do *energy hole*, introduzido na Seção 1. Serão debatidas as características desse problema e a motivação da nossa análise baseada nas questões listadas anteriormente.

### 2.1.1 Representação de uma rede de sensores sem fios

Para um melhor entendimento do contexto cujo nosso estudo está inserido, aproveitamos o diagrama (Figura 1) apresentado por Frery et al. (2010). Esse diagrama tem o objetivo de descrever uma Rede de Sensores sem Fios como um processo de amostragem/reconstrução no domínio de processamento de sinais. O entendimento dessa modelagem é importante não só para a delimitação do escopo de estudo, mas também é fundamental para a delimitação de como colocar em prática aspectos relacionados à nossa proposta e à metodologia abordada.

O diagrama representa uma RSSF coletando informações,

Figura 1 – Diagrama de RSSFs.

$$\begin{array}{ccccccc}
 \mathcal{N} & \xrightarrow{F} & \mathbf{V}^* & \xrightarrow{S} & \mathbf{V} & \xrightarrow{\Psi} & \mathbf{V}' & \xrightarrow{\hat{F}} & \mathbf{V}^* \\
 & & \begin{array}{c} \downarrow R^* \\ D^* \end{array} & & & & & & \begin{array}{c} \downarrow R' \\ D' \end{array}
 \end{array}$$

Fonte: Frery et al. (2010).

em que  $\mathcal{N}$  representa a natureza, isto é, a região onde serão feitos os estudos e medições,  $F$  representa o fenômeno de interesse e  $\mathbf{V}^*$  é o domínio espacial e temporal que representa o conjunto de medidas do fenômeno de interesse. Se fosse possível a completa observação de  $F$ , teríamos  $\mathbf{V}^*$  contendo dados ideais. Assim, conforme mostra o diagrama, poderíamos construir um conjunto de regras ideais  $R^*$  que nos levariam a tomar decisões, também ideais,  $D^*$ .

Em uma abordagem que se aproxime mais da realidade, podemos extrair informações através de uma rede composta por  $n$  nós sensores representados por  $S = \{S_1, \dots, S_n\}$ , onde o conjunto de nós é distribuído em uma determinada região a ser monitorada e cada nó tem a capacidade de observar e transmitir seu valor de amostragem. A partir das observações de cada nó sensor, é obtido o conjunto espacial e temporal de amostras  $\mathbf{V}$ .



Como em muitos casos não é interessante reportar todo o conjunto amostrado  $\mathbf{V}$  aos usuários da rede responsáveis pelas tomadas de decisões, pois isso pode ser muito custoso em termos energéticos, algumas técnicas de otimização  $\Psi$  podem ser utilizadas. Tais técnicas podem ser operações de agrupamento e/ou agregação que levem o conjunto  $\mathbf{V}$  a um conjunto  $\mathbf{V}'$  com menor quantidade de dados mas sem perda significativa da qualidade.

Li e Mohapatra (2007) foram pioneiros na identificação e modelagem do fenômeno denominado *energy hole*. Essa modelagem foi feita sem o uso de técnicas de agrupamento, possivelmente para não introduzir mais uma variável e, com isso, observar da forma mais direta possível o problema em questão.

Dessa forma, abstraindo as representações de  $\Psi$  e  $\mathbf{V}'$ , uma estação central, que recebe as informações  $\mathbf{V}$  percebidas pelo conjunto de nós sensores  $S$  e diante de alguma técnica de reconstrução, poderia fazer alguma inferência  $\hat{F}$  sobre o fenômeno em observação original  $F$ . A partir de  $\hat{F}$ , os usuários da rede poderiam construir um conjunto de regras  $R'$  que levem a um conjunto de decisões  $D'$ .

É importante observar que para a obtenção de vantagens na etapa de reconstrução, um fator fundamental é a técnica de reconstrução a ser aplicada, a qual será responsável pela identificação das informações contidas no conjunto de amostras fornecido pelos nós sensores. Uma etapa de reconstrução do fenômeno amostrado realizada de forma satisfatória, o que depende também da escolha das estratégias utilizadas nas etapas anteriores mostradas nessa seção, proporcionaria aos usuários a tomada de decisões  $D'$  bem próximo das decisões ideais  $D^*$ .

### 2.1.2 Organização da rede e comunicação de nós sensores

Agora que já conhecemos como uma RSSFs pode ser esquematizada, precisamos entender melhor alguns detalhes que proporcionam aos nós sensores a habilidade de executar a tarefa de sensoriamento colaborativa para a qual são destinados.

Existem dois fatores que influenciam diretamente na maneira como os nós irão realizar as suas tarefas de comunicação para que o objetivo de aquisição de informações da rede como um todo seja alcançado: o modo como os nós criam *links* de comunicação entre si e a dispersão e organização dos nós no espaço geográfico onde são medidos os fenômenos.

Primeiramente, quando falamos em criação de *links* de comunicação, precisamos pensar que uma característica fundamental para que caminhos de comunicação entre nós sensores sejam contruídos está relacionada ao potencial de seus rádios de comunicação. Como já foi falado anteriormente, cada nó sensor possui uma unidade de comunicação que o habilita a transmitir e receber mensagens de outros nós a partir de um meio compartilhado

de transmissão (no caso das RSSFs, o ar). Através desse aspecto, além de outros aspectos não menos importantes como a aplicação executada na rede e a aquisição de dados de maneira periódica ou disparada por eventos, por exemplo, é que surgem os protocolos de roteamento, responsáveis por determinar os melhores caminhos entre um nó sensor e uma estação base. No nosso trabalho, é suficiente o entendimento de como o rádio de comunicação dos nós sensores podem influenciar no objetivo final de uma RSSF. Portanto, para mais detalhes sobre estudos de protocolos de roteamento em RSSFs, recomendamos a leitura dos trabalhos de Intanagonwiwat et al. (2003) e Al-Karaki e Kamal (2004).

Diante do que foi exposto até o momento e baseado nos diversos estudos relacionados à comunicação em RSSFs, podemos dizer que existem duas formas de comunicação entre os nós sensores: o roteamento de um salto e o roteamento multissalto. No primeiro caso, os nós de uma RSSFs transmitem os dados sensorizados e processados diretamente para a estação base. Isso é possível por meio de nós sensores que possuam rádios de comunicação com características de potência de transmissão suficiente para criar um *link* de comunicação com a estação base independente de sua localização na área de monitoramento. A grande desvantagem desse tipo de comunicação é que vários nós precisam gastar uma quantidade relativamente alta de energia, pois sabemos que a capacidade de transmissão de uma mensagem por um nó sensor está diretamente ligada a potência de sinal que este utiliza através do seu rádio de comunicação.

Já o segundo caso pode ser encarado como uma solução para o problema de gasto energético excessivo proveniente do modo de comunicação por um salto. Nesse modo de comunicação, um nó sensor, além de realizar as tarefas de sensoriamento e processamento de dados, possui o papel de retransmissor de dados, ou seja, cada nó que compõe a rede utiliza o seu potencial de comunicação não só para transmitir os dados que ele mesmo mede do ambiente, mas também para retransmitir os dados provenientes de medições realizadas por outros nós. Acontece que os nós sensores podem ajustar seus rádios de comunicação a uma potência relativamente baixa com o objetivo de transmitir suas informações a nós próximos (vizinhos), economizando, assim, uma quantidade considerável de energia se comparada a energia gasta no modo de comunicação através de único salto. Entretanto, apesar dos ganhos consideráveis, é a partir desse tipo de comunicação que surge o problema do *energy hole*. Esse problema e suas particularidades para o consumo de energia em RSSFs serão abordados na Seção 2.1.3.

Já com relação aos aspectos de distribuição e organização de nós em uma RSSFs, é importante observar que a definição do modo da comunicação de dados também está diretamente relacionado a esse fator, pois a quantidade e o posicionamento dos nós são importantes na definição das potências de comunicação utilizadas para efetuar transmissões e recepções de dados. Portanto, em estudos de RSSFs, é importante que também a estratégia de dispersão de nós seja levada em consideração. Entretanto é comum encontrarmos estudos

que representam a dispersão de nós sensores seguindo apenas uma distribuição espacial aleatória uniforme.

Andel e Yasinsac (2006), além de outros autores, também revelam em seus estudos que esse tipo de suposição é impreciso, o que pode levar simulações a produzirem resultados errôneos.

A reação a essas afirmações tem sido confirmada nos últimos anos, pois alguns estudos vem procurando representar as redes de sensores sem fios de maneira a aproximar cada vez mais os cenários de simulação a retratar com fidelidade cenários reais. Adicionalmente, trabalhos como o de Younis e Akkaya (2008) têm demonstrado, por exemplo, que as características de dispersão de nós sensores são tão importantes em RSSFs quanto a preocupação com características de comunicação, que o modelo que rege a dispersão de nós é muito mais importante do que puramente representar os nós em um determinado espaço para se obter um cenário de simulação e que o modelo de distribuição pode ser otimizado em funções de algumas certas propriedades da rede que desejem ser alcançadas.

A explanação sobre esses dois aspectos, a comunicação e a distribuição de nós em RSSFs, se deve ao fato de procurarmos entender com mais detalhes alguns aspectos importantes para a execução do principal objetivo de uma RSSFs. Porém, o nosso principal objetivo foi dar uma introdução à motivação do nosso trabalho. Como é possível observar diante de tudo que foi exposto, a preocupação com os diversos aspectos relacionados as RSSFs se deve, principalmente, a um problema inerente às limitações de bateria impostas aos nós sensores e, conseqüentemente, a uma RSSF como um todo. Realizar estudos para o entendimento e a proposição de soluções em modelos de comunicação e dispersão de nós em RSSFs está quase sempre ligado à problemática do consumo de energia.

Como veremos na próxima seção (Seção 2.1.3), várias são as estratégias e soluções propostas para atacar esse problema. No nosso trabalho fazemos uma descrição da interligação entre as características de um modelo de distribuição de nós e o modo de comunicação multissalto. Mostraremos que uma RSSFs com tais parâmetros pode estar sujeita a um consumo de energia não-balanceado entre seus nós e, como consequência mais drástica, pode ter o seu tempo de vida reduzido consideravelmente.

### 2.1.3 Problemática de energia das RSSFs

Como já vimos, uma das características de um nó sensor é ser composto por um unidade de bateria, responsável pelo fornecimento de energia para execução das tarefas de sensoriamento, processamento e comunicação. Como também já é de nosso conhecimento, essa fonte de energia tem sérias limitações, pois a quantidade de energia disponível é relativamente pequena e na grande maioria dos casos é muito difícil ou até mesmo impossível realizar recargas periódicas, visto que os nós sensores de uma RSSFs

podem estar localizados em áreas de difícil acesso. Portanto, não é por acaso que Sadler (2005) e Raghunathan, Ganeriwal e Srivastava (2006) afirmam que energia em RSSFs é um recurso escasso e precioso e que, devido a isso, todos os aspectos de RSSFs devem ser pensados e projetados com o objetivo de preservar energia.

Entretanto, sabemos que uma RSSF tem por objetivo monitorar uma região e/ou fenômeno e isso deve ser realizado, muitas vezes, por longos períodos de tempo, dependendo principalmente de requisitos impostos pela aplicação que esteja sendo executada. Assim, como bem é questionado por Anastasi et al. (2009), *“como balancear o consumo do recurso energético disponível?” Como manter uma rede ativa o máximo de tempo possível quando os recursos de energia são limitados a uma pequena bateria e os nós sensores precisam estar sempre se comunicando entre si, através de rádio, para processar as mudanças ocorridas em um ambiente?*

Os trabalhos de Sadler (2005) e Anastasi et al. (2009) são pontos de partida para possíveis respostas para essas perguntas. Um dos fatores mais importantes desses trabalhos é a afirmação de que a proposição de soluções que produzam resultados interessantes para o prolongamento do tempo de vida de uma RSSF começa pela busca, primeiramente, do entendimento de como a energia é gasta por um nó ao realizar suas tarefas. Há um consenso na literatura apontando para os dispositivos de rádio e, conseqüentemente, para as tarefas de comunicação como responsáveis pelo maior consumo de energia em uma RSSF. Essa afirmação pode ser confirmada em Pottie e Kaiser (2000), Raghunathan et al. (2002) e Culler, Estin e Srivastava (2004).

Para reforçar mais ainda essa afirmação, vários são os trabalhos que apontam para estratégias de economia de energia que atuam diretamente nas tarefas de comunicação dos nós sensores. Essas estratégias incluem, por exemplo, propostas que levem à redução do tráfego de dados na rede (seja pela utilização de técnicas de agrupamento e/ou fusão de dados ou técnicas de otimização de rotas do nós sensores até o nó sorvedouro), que considerem o desligamento periódico de alguns nós da rede (quando esses não estiverem transmitindo ou recebendo dados) e que controlem de forma dinâmica características particulares dos rádios de nós sensores. Para mais detalhes à respeito dos termos citados aqui, recomendamos os trabalhos de Nakamura, Loureiro e Frery (2007) e Alencar-Neto (2007), além dos trabalhos por eles referenciados, em especial Al-Karaki e Kamal (2004), Demirkol, Ersoy e Alagoz (2006), Reis et al. (2007), Yoon e Shahabi (2007) e Akkaya, Demirbas e Aygun (2008).

Entendemos também que uma outra estratégia pode trazer resultados positivos à proposição de soluções voltadas à problemática em questão. Como demonstrado por Santi (2005), uma maneira de agir é mudando um pouco de foco, isto é, podemos trabalhar não somente com as características físicas dos nós sensores ou, mais especificamente, com as características físicas das estruturas de rádio, podemos trabalhar também na maneira como

os nós se organizam geograficamente na área de monitoramento. De certa forma, isso é relativamente evidente, pois um dos fatores que contribuem para uma comunicação ter ou não sucesso e ser ou não efetiva está relacionado à distância entre os agentes comunicantes.

Veremos em mais detalhes que a nossa abordagem tem essa visão. Essa teve sua motivação iniciada por trabalhos como o de Almiron (2009), o qual apresenta, dentre outros estudos, a influência do posicionamento de nós sensores em RSSFs em características importantes para uma rede como conectividade e cobertura. Nesse trabalho, é colocado como um ponto a ser melhor estudado o impacto de metodologias de deposição de nós na longevidade de uma RSSFs. Ramos et al. (2011) apresenta um dos trabalhos que podem ser classificados nessa linha de pesquisa, como veremos em um pouco mais de detalhes na Seção 2.3.2.

Para um conhecimento mais detalhado da problemática de energia em RSSFs e de algumas das principais soluções propostas na literatura, recomendamos ao leitor os trabalhos de Sohrabi et al. (2000), Chang e Tassiulas (2004), Hoesel et al. (2004), Ye, Heidemann e Estrin (2004), Raghunathan, Ganeriwal e Srivastava (2006) e Nakamura, Loureiro e Frery (2007).

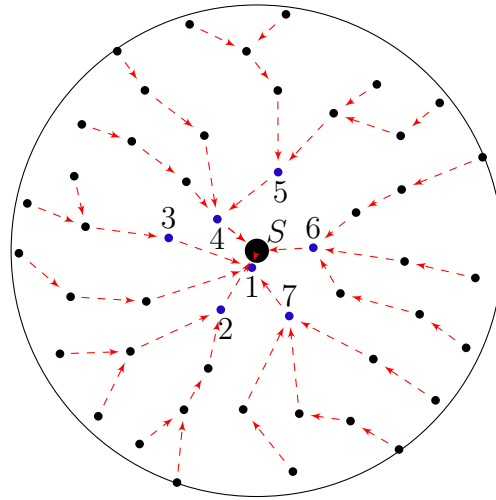
#### 2.1.3.1 O problema do *energy hole*

No contexto de redes de sensores sem fios que abordamos nesse trabalho (introduzido nas Seções 2.1.1 e 2.1.2), mais especificamente nos casos em que existe comunicação multissalto e uma distribuição aleatória uniforme dos nós sensores, existe um problema específico que também está diretamente ligado à problemática de energia e tempo de vida em RSSFs. Esse problema é mais conhecido na literatura como problema do *energy hole*.

Para um melhor entendimento de como esse problema se apresenta, suponhamos que a Figura 2 representa uma RSSF, com seus nós espalhados na região de monitoramento segundo um modelo de distribuição aleatório uniforme (Na Figura 2, cada ponto representa um nó sensor, sendo a estação base o maior deles e apontado por  $S$ ). Os caminhos – fluxo de dados da rede –, representados pelas setas vermelhas, indicam como cada nó sensor irá se comunicar com um de seus vizinhos – baseado em um roteamento multissalto –, fazendo com que o dados sensorizados cheguem até a estação base. Os nós em destaque – azuis e numerados de 1 a 7 – são alguns dos nós mais próximos da estação base e que fazem mais o papel de *relay*, em média).

Diante das características provenientes do modo de comunicação multissalto, em que todos os dados provenientes das tarefas de sensoriamento dos nós sensores chegam até a estação base de forma indireta, podemos, de maneira intuitiva, observar que os nós sensores localizados nas regiões mais próximas à estação base efetuam, em média, a tarefa de redirecionamento uma quantidade maior de vezes que os nós sensores localizados em regiões mais distantes. Após sucessivas fases de sensoriamento e posterior envio de dados

Figura 2 – Exemplo de região de monitoramento com nós sensores distribuídos e alguns dos possíveis caminhos de comunicação.



Fonte: Autor.

à estação base e como estamos falando de comunicação em múltiplos saltos, também é intuitivo perceber que esse consumo não-balanceado de energia ao longo da rede fará com que os nós dispostos nas regiões a um salto da estação base gastem completamente a energia disponível antes dos nós dispostos a dois saltos da estação base. Da mesma maneira, os nós localizados a dois saltos da estação base irão gastar sua energia mais rapidamente que os nós localizados a três saltos da estação base e assim sucessivamente.

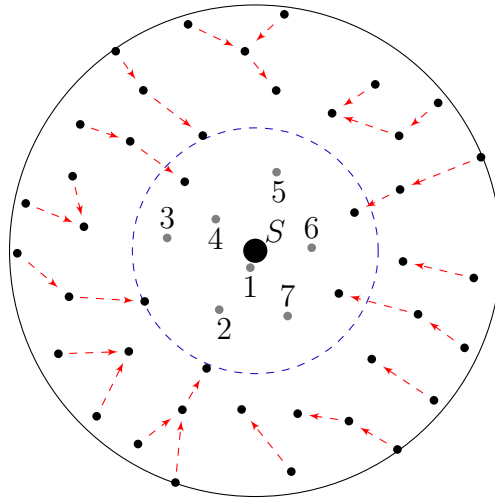
Como mostra a representação do círculo pontilhado em azul na Figura 3 (trata-se de uma representação similar a Figura 2), esse comportamento acarretará na formação de um “buraco” entre a estação base e os nós com energia remanescente no momento em que não existirem mais nós a um salto da estação base com capacidade energética para efetuar a tarefa de retransmissão (na Figura 3, os nós em destaque, cinzas e numerados de 1 a 7), ou seja, os nós localizados nas regiões mais distantes da estação base, que ainda possuem potencial energético, simplesmente não poderão criar rotas de comunicação que possam levar os dados sensorizados até a estação base, fazendo com que a rede ainda possua potencial energético para operar mas tenha seu tempo de vida reduzido drasticamente.

Na literatura podemos encontrar diversos trabalhos que caracterizam o problema do *energy hole* e propõe soluções para sua minimização. Dentre estes trabalhos, destacamos Li e Mohapatra (2007), Medidi e Zhou (2007) e Wu, Chen e Das (2008).

Em Li e Mohapatra (2007) é apresentado uma descrição do problema de *energy hole* por intermédio de um modelo matemático. A partir do modelo proposto são realizadas análises de alguns métodos apresentados na literatura com o objetivo de avaliar se esses podem ser aplicados como forma de minimizar o problema. A conclusão é que técnicas



Figura 3 – Exemplo de região de monitoramento com nós sensores distribuídos, alguns dos possíveis caminhos de comunicação e o problema do *energy hole*.



Fonte: Autor.

de agregação de dados e adição de nós sensores ao longo da rede podem ser utilizadas de maneira satisfatória. Contudo, o crescimento quantitativo de nós sensores só terá um efeito razoável se a estratégia for a inclusão de alguns nós com capacidade diferenciada de computação e armazenamento de energia, criando uma rede heterogênea, e/ou adicionar nós de mesma capacidade que os nós já existentes em regiões mais críticas (maior proximidade da estação base) para que os nós dividam a carga de mensagens retransmitidas.

Já em Wu, Chen e Das (2008), o objetivo da solução proposta é justamente a de aumentar a densidade de nós, como avaliado positivamente em Li e Mohapatra (2007), tornando a distribuição não-uniforme. Na estratégia proposta, o aumento da densidade de nós é feito de maneira proporcional a proximidade com a estação base, isto é, alguns nós são adicionados de modo a se posicionarem a um salto da estação base, outros nós, em quantidade menor que a anterior (em média), são adicionados em regiões a dois saltos da estação base e assim por diante. A tentativa é tentar se aproximar de uma dispersão de nós que seja contínua ao longo da extensão da área de interesse.

Também com o mesmo objetivo de evitar a morte prematura de uma RSSF por meio da minimização do problema do *energy hole*, o trabalho de Medidi e Zhou (2007), se comparado ao trabalho de Wu, Chen e Das (2008), apresenta uma técnica adicional: a atribuição da técnica de *duty cycle*. Em poucas palavras, esse tipo de estratégia segue a ideia de que os nós sensores de uma RSSF respondem a regras de ativação/desativação de seus rádios de comunicação com o objetivo de poupar energia nos momentos em que não forem efetuadas tarefas de recepção/transmissão. Com isso, além de se preocupar com a utilização de técnicas de distribuição de nós de maneira não-uniforme, a estratégia

proposta tem o foco de minimizar problemas como colisão, pois uma maior quantidade de nós passa a dividir o meio de comunicação sem fio, e escuta ociosa, que devido às características do meio de transmissão compartilhado faz com que os nós de uma RSSFs desperdicem energia na recepção de mensagens de outros destinatários.

No nosso trabalho, seguimos a ideia abordada nos três trabalhos apresentados anteriormente, a de aumentar a densidade de nós sensores nas regiões próximas à estação base, mas, assim como em Li e Mohapatra (2007) e Wu, Chen e Das (2008), não tratamos de problemas como redundância da informação sensoriada, colisões e escuta ociosa.

Assim, na nossa metodologia, propomos, a partir do modelo de estudo de RSSFs proposto em Frery et al. (2010), um modelo de dispersão de nós que nos permite avaliar qual o impacto dessa estratégia na minimização do problema do *energy hole*. Esse modelo, que será melhor detalhado na Seção 4.1.1.2, proporciona que os nós sensores sejam distribuídos ao longo de uma RSSF de maneira não-uniforme e contínua, sem deixar de lado alguns parâmetros que, como já dissemos, são extremamente importantes em RSSFs, como cobertura e conectividade.

Na Seção 4, apresentaremos como foi realizada a junção da estratégia de Frery et al. (2010) e o modelo de dispersão proposto para que o nosso objetivo de avaliação fosse alcançado. Já no restante desta seção serão apresentadas mais alguns conceitos e definições de fundamental importância para o entendimento das ferramentas e estratégias utilizadas em nossa metodologia.

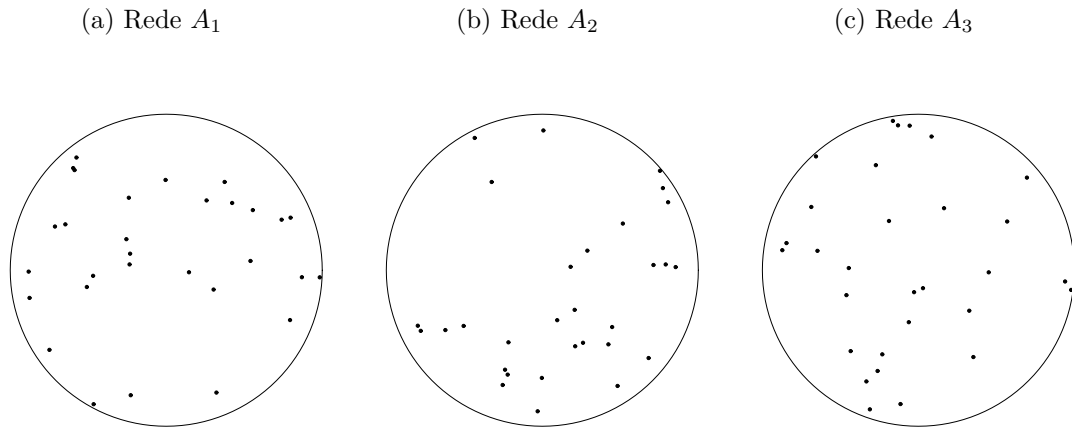
## 2.2 Processos pontuais

Um dos principais fatores relacionados ao estudo do consumo de energia em RSSFs é a maneira como os nós estão dispostos em uma região de estudo.

Realizando um estudo geral na literatura para avaliar como se deu a evolução dos modelos de dispersão de nós, utilizados para o estudo de RSSFs, podemos observar que vários estudos, principalmente os primeiros, basearam-se no posicionamento regular dos nós sensores no espaço. Consultar Reis et al. (2007) para alguns exemplos. Porém já é consenso que, apesar de terem sido e ainda serem bastante úteis para o estudo de determinados aspectos em RSSFs, tais modelos têm diversas limitações.

Observamos ainda que outra técnica bastante utilizada para modelagem de dispersão de nós é a distribuição espacial aleatória uniforme. Consultar Heinzelman, Chandrakasan e Balakrishnan (2002) e Reis et al. (2007) para alguns exemplos. É uma técnica muito frequentemente encontrada na literatura, na qual as coordenadas do posicionamento de cada nó sensor são modeladas como um par de variáveis aleatórias uniformes independentes. A Figura 4 ilustra esta situação com três redes,  $A_1$ ,  $A_2$  e  $A_3$ , com 30 nós sensores cada.

Figura 4 – Exemplos de posicionamento de nós sensores de maneira aleatória uniforme.

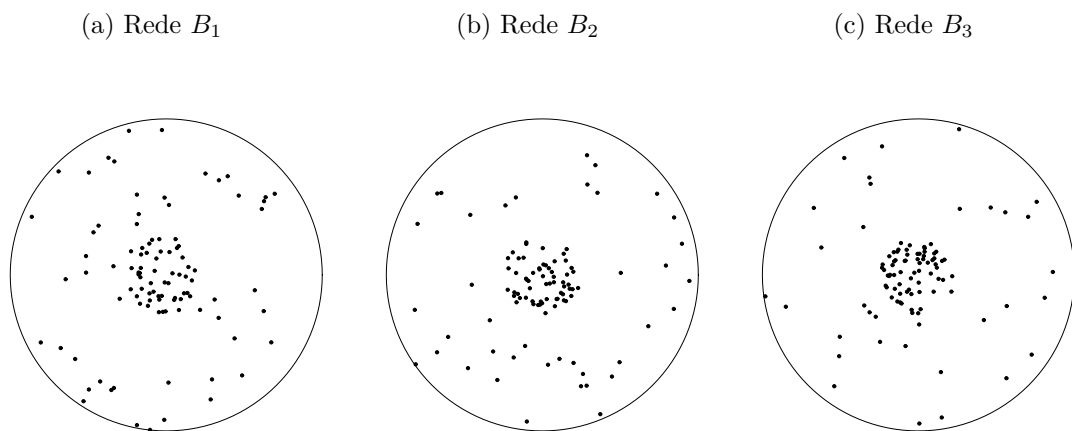


Fonte: Autor.

Apesar dessa técnica propiciar maior realismo, algumas situações da realidade ainda não podem ser representadas. Conforme é explicado em Almiron (2009), o trabalho de Frery et al. (2008) identifica alguns cenários de interesse no estudo das RSSFs que podem ser ilustrados pelos seguintes exemplos:

- a) uma região de monitoramento cuja superfície em que os nós sensores são posicionados apresenta um declive fortemente marcado, mantendo os nós sensores principalmente em um centro de baixo nível, o que é denominado como processo atrativo. A Figura 5 ilustra esta situação com três redes,  $B_1$ ,  $B_2$  e  $B_3$ , com 100 nós sensores cada;

Figura 5 – Exemplos de posicionamento de nós sensores em um processo atrativo

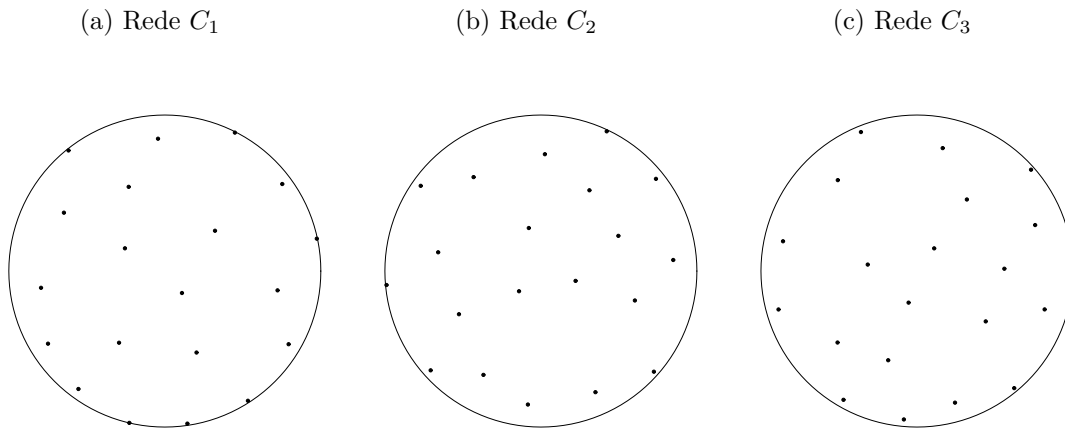


Fonte: Autor.

- b) uma região de monitoramento para a qual fosse possível ter controle sobre o

posicionamento dos nós, ou seja, na possibilidade de intervenção humana no momento de distribuição de nós sensores. Nesse caso, a intenção seria a de colocar os nós o mais longe possível uns dos outros, evitando ter dois ou mais nós em uma área reduzida da região total a ser monitorada. Essa situação é denominada processo repulsivo e é ilustrada pela Figura 6, em que são apresentadas três redes,  $C_1$ ,  $C_2$  e  $C_3$ , com 20 nós sensores cada.

Figura 6 – Exemplos de posicionamento de nós sensores em um processo repulsivo



Fonte: Autor.

Para tratar de ambos os casos, além do modo de posicionamento de nós sensores seguindo uma distribuição uniforme aleatória, Frery et al. (2008) propõem o processo  $C$ . Esse processo foi utilizado nos trabalhos de Alencar-Neto (2007) e Almiron (2009) e Frery et al. (2010).

Adicionalmente, uma evolução desse processo foi proposta por Ramos et al. (2011). Nessa abordagem, o processo de distribuição de nós sensores é denominado  $M^2P^2$  e foi proposto com o objetivo, dentre outros, de adequar a modelagem de distribuição de nós a mais cenários, principalmente no que diz respeito aos cenários para estudos de RSSFs heterogêneas e hierárquicas (AKYILDIZ et al., 2002; CULLER; ESTIN; SRIVASTAVA, 2004).

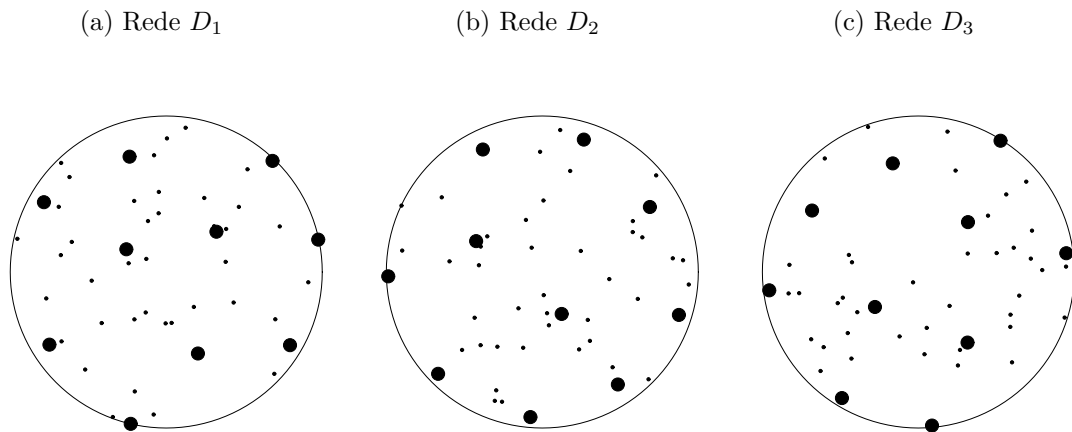
Em RSSFs desse tipo existem nós sensores com características diferenciadas de processamento, armazenamento e comunicação. No processo  $M^2P^2$  esse aspecto é caracterizado com a diferenciação de dois tipos de nós, os sensores-L e os sensores-H, definidos em Ramos et al. (2011) como *L-sensors* e *H-sensors*, respectivamente.

Quanto a distribuição de nós sensores em uma região de monitoramento, além dos cenários retratados pelo processo  $C$ , o processo  $M^2P^2$  pode modelar as seguintes situações:

- a) deseja-se que os nós sensores sejam distribuídos de maneira aleatória e inde-

pendentes uns dos outros, entretanto há a necessidade (e a possibilidade) de distribuir alguns poucos nós em áreas estratégicas. A Figura 7 ilustra esta situação com três redes,  $D_1$ ,  $D_2$  e  $D_3$ , com 50 nós sensores cada, sendo 10 desses do tipo sensores-H (em destaque);

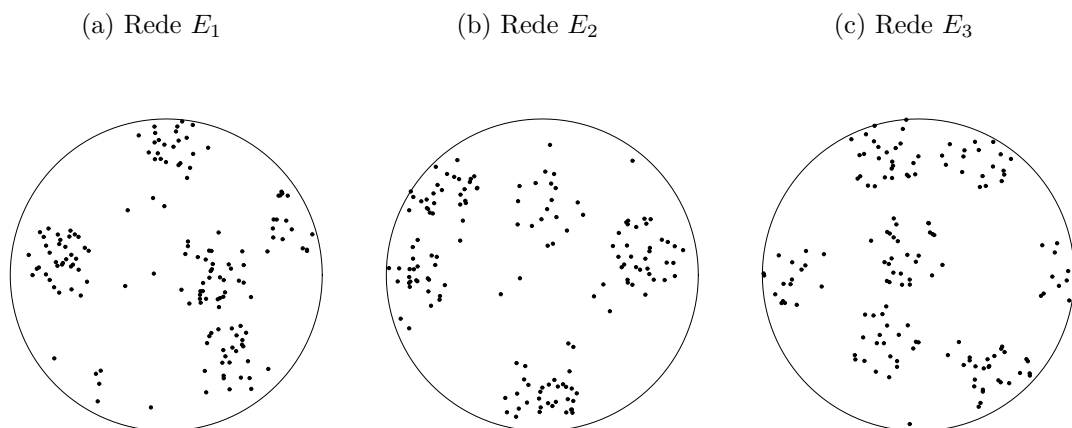
Figura 7 – Exemplos de posicionamento de nós sensores de maneira aleatória uniforme e com nós heterogêneos



Fonte: Autor.

b) uma região de monitoramento cuja superfície em que os nós sensores são posicionados apresenta vários declives fortemente marcados, ou seja, assim como no processo  $C$ , há um processo de atratividade, porém em mais de uma área. A Figura 8 retrata essa situação, em que são apresentadas três redes,  $E_1$ ,  $E_2$  e  $E_3$ , com 150 nós sensores cada.

Figura 8 – Exemplos de posicionamento de nós sensores em um processo atrativo



Fonte: Autor.

Mais detalhes sobre como as abstrações (modelos) dos processos  $C$  e  $M^2P^2$  são construídas serão abordados na Seção 2.3.

Até o momento, reservamos um espaço com o objetivo principal de mostrar para o leitor que o estudo de modelos de distribuição dos nós, na grande maioria dos trabalhos em RSSFs, tem como finalidade maior a modelagem de situações do mundo real que sejam de interesse para o estudo em RSSFs. Em outras palavras, como já foi dito anteriormente, o objetivo principal foi apresentar um breve histórico do que vem sendo feito na literatura. Apesar de, no nosso trabalho, esse também ser um dos objetivos, nós utilizaremos a modelagem de distribuição de nós sensores em um outro contexto.

Existem aplicações e situações em RSSFs que exigem um determinado grau de controle no momento de distribuição dos nós da região de interesse, seja, por exemplo, pelo fato de existirem subregiões que necessitem de uma maior vigilância e/ou urgência no monitoramento ou pelo simples fato de um algoritmo de roteamento exigir que uma quantidade maior de nós esteja presente em determinados locais. No contexto do nosso trabalho, realizamos uma breve introdução na Seção 1 ao problema do *energy hole* e a principal ideia para se minimizar esse problema, que é aumentar a concentração de nós sensores dispostos nas localidades mais próximas ao nó sorvedouro.

Neste trabalho apresentamos um modelo de dispersão dos nós que descreve a situação em que um projetista de uma RSSF pode, de alguma forma e com base em determinadas características da rede, controlar a concentração de nós em determinadas regiões. Esse controle tem por objetivo projetar uma rede em que o problema do *energy hole* seja reduzido, conseguindo um tempo maior de vida se comparado ao de uma RSSFs projetada para que os nós sejam distribuídos de forma puramente aleatória ou, até mesmo, de maneira regular.

No que resta desta seção, iremos apresentar algumas formulações matemáticas de processos pontuais (seção 2.2.1), as quais serão a base para a apresentação do modelo de dispersão de nós proposto nesse trabalho (seção 4.1.1.2), assim como também para a apresentação de alguns detalhes dos modelos propostos em Frery et al. (2008), para o processo  $C$ , e em Ramos et al. (2011), para o processo  $M^2P^2$  (seção 2.3).

### 2.2.1 Definindo processos pontuais

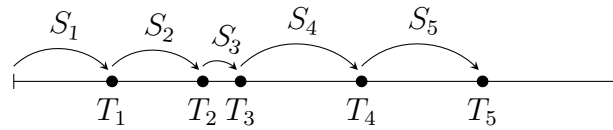
Esta seção é baseada no trabalho de Almiron (2009).

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 9).

Figura 9 – Tempos de chegada  $T_i$  e entre chegadas  $S_i$ .



Fonte: Almiron (2009).

Uma outra alternativa é estudar o processo de contagem acumulado, definido a seguir.

**Definição 1 (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*

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

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

$$\mathbb{1}[B] = \begin{cases} 1 & \text{se } B \equiv \text{TRUE}, \\ 0 & \text{caso contrário.} \end{cases} \quad (2)$$

Analogamente, é útil utilizar o processo de contagem por intervalos (ver Definição 2). O intervalo de contagem  $N[a, b)$  registra o número de pontos no intervalo  $[a, b)$ .

**Definição 2 (Intervalos de contagem)** *Seja  $N_t$  o processo de contagem acumulado, definimos o processo de contagem no intervalo  $[a, b)$  como*

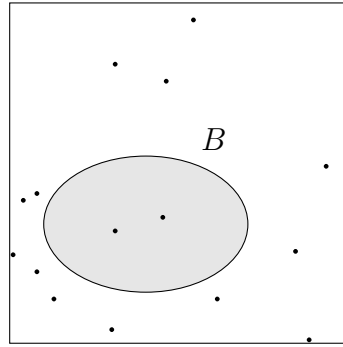
$$N[a, b) = N_b - N_a, \quad (3)$$

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 10). As variáveis  $N(B)$  têm propriedades

Figura 10 – Região de contagem  $N(B) = 2$ .



Fonte: Almiron (2009).

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 3 (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.  $A_1 \cap A_2 = \emptyset \Rightarrow N(A_1 \cup A_2) = N(A_1) \cup N(A_2)$ ,
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  $R(P)$ .

Geralmente consideramos que o processo pontual é localmente finito e simples. Estes conceitos são apresentados nas Definições 4 e 5, respectivamente.

**Definição 4 (Processo pontual localmente finito)** *Um processo pontual  $P$  definido por  $N(\cdot)$  é dito localmente finito, se*

$$\Pr(N(B) < \infty) = 1, \quad (4)$$

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



Em outras palavras, a Equação 4 diz que com probabilidade 1 qualquer região limitada somente contém um número finito de pontos.

**Definição 5 (Processo pontual simples)** *Um processo pontual definido por  $N(\cdot)$  é dito simples, se*

$$\Pr(N(\{x\}) \leq 1) = 1, \quad (5)$$

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

De acordo com a Equação 5, em um processo pontual simples, com probabilidade 1, não existem dois pontos coincidentes.

Diante das definições expostas, já podemos definir o processo pontual de Poisson em duas dimensões.

**Definição 6 (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  $\eta$  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:

**Teorema 1 (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 que  $N(W) = n$ , a distribuição condicional de  $N(B)$  para  $B \subseteq W$  segue uma distribuição binomial, isto é,*

$$\Pr(N(B) = k \mid N(W) = n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad (6)$$

em que a probabilidade de ocorrência é dada por  $p = \xi(B)/\xi(W)$ . Além disso, 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 círculo  $C_r^p$ , onde  $r \in \mathbb{R}_+$  indica o raio e o centro é dado pelo ponto  $p \in \mathbb{R}^2$ . 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 coletivamente independentes e uniformemente distribuídas em  $W$ .

Para estudar outros tipos de processos, nos quais a intensidade não é a mesma em qualquer região, introduzimos na definição 7 o conceito de função de intensidade.

**Definição 7 (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 proceso geral podemos obter outros processos de interesse.

**Definição 8 (Processo pontual geral de Poisson)** *Seja  $S$  um espaço, e  $\Lambda$  uma medida sobre  $S$ . O processo pontual geral de Poisson sobre  $S$  com intensidade  $\Lambda$  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 8, quando  $\eta$  é constante sobre  $W$ .

**Definição 9 (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 no compacto  $B \subset \mathbb{R}^2$  e intensidade  $\Lambda(B) = \int_B \eta(u) du$  na definição 8.*

Em outras palavras, nossa proposta consiste em descrever a localização dos sensores através de um processo de Poisson não-homogêneo. Com isso, temos plena liberdade para escolher a intensidade do processo, e o faremos de tal sorte que haverá, em média, uma concentração maior de nós em volta de um ponto específico (no contexto de RSSFs, a estação base).

Detalharemos o processo de poisson não-homogêneo de interesse, isto é, a proposição do modelo de dispersão de nós sensores na Seção 3. Antes, veremos na Seção 2.3 aspectos relacionados aos modelos que foram fundamentais para a construção da nossa proposta.

## 2.3 Modelos de interesse: processos pontuais não-homogêneos

Conforme iniciamos a discussão na Seção 2.2, podemos encontrar algumas técnicas de modelagem que representem a distribuição de nós sensores em um espaço. Vimos que a evolução desses modelos no estudo de RSSFs resultou em abstrações que refletem cada vez mais situações próximas da realidade. Dois exemplos dessa evolução são os processos  $C$  e  $M^2P^2$ .

O processo  $C$  não foi concebido com o objetivo de estudar problemas como o de *energy hole*. A ideia principal, conforme podemos concluir dos diversos trabalhos que já o utilizaram (ALENCAR-NETO, 2007; FRERY et al., 2008; ALMIRON, 2009; FRERY et al., 2010), era a de tornar o estudo de RSSFs mais abrangente e com resultados de simulações mais precisos, visto que esse processo fazia parte da implementação de um conjunto de modelos responsáveis por criar uma representação para RSSFs do ponto de vista de processamento de sinais.

Já o processo  $M^2P^2$  foi construído com outros objetivos, apesar de podermos atribuir-lhe a ideia de evolução do processo  $C$ . Dentre outras ideias, a caracterização do problema de *energy hole* é um dos focos dos criadores do processo, em Ramos et al. (2011).

Com essas ideias, aliadas aos estudos de Li e Mohapatra (2007), propomos um novo método de dispersão de nós com o objetivo de avaliar (e possivelmente minimizar) o efeito do problema de *energy hole* em RSSFs.

Devido à importância dos processos  $C$  e  $M^2P^2$  para o nosso trabalho e como já conhecemos algumas definições de processos pontuais, reservamos o restante desta seção para apresentar as modelagens que dão suporte a esses processos.

### 2.3.1 O Processo pontual $C$

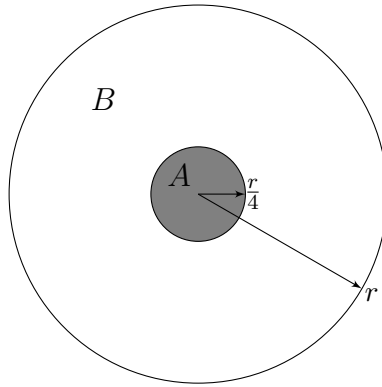
Esta seção é baseada nas definições feitas por Almiron (2009), as quais tiveram como suporte o trabalho de Frery et al. (2008).

O processo  $C$  é definido como  $C(n, a)$ , em que  $n$  é a quantidade de nós que devem ser posicionados em uma determinada região, e  $a$  é o parâmetro que determina a intensidade do processo. Como região de interesse, o trabalho original de Frery et al. (2008) define como sendo  $W = [0, 100]^2$ . Sem perda de generalidade, Almiron (2009) define a área como sendo  $W = [0, 1]^2$  e para efeitos de coerência com as definições apresentadas na Seção 2.2.1, no nosso trabalho definimos a região de interesse  $W$  como sendo um círculo  $C_r^p$ , em que  $r \in \mathbb{R}_+$  indica o raio, e o centro é dado pelo ponto  $p \in \mathbb{R}^2$ .

Diante dessas definições e das características apresentadas na Seção 2.2, podemos compor o processo  $C(n, a)$  com três processos pontuais:

1. o processo pontual espacial Binomial  $B(n)$ , introduzido no teorema 1,
2. o processo pontual espacial SSI de Matérn  $M(n, r(p))$ , não definido nesse trabalho mas com detalhes em Frery et al. (2008) e Baddeley (2007),
3. o processo pontual espacial  $S(n, a, \eta)$  composto por dois processos Poisson, um aplicado com intensidade  $a\eta$  sobre uma área alvo (subconjunto da região de interesse e representada por  $C_{\frac{r}{4}}^p$ , por exemplo), e outro com intensidade  $\eta$  na região  $(C_{\frac{r}{4}}^p)^c \cap C_r^p$  (ver figura 11).

Figura 11 – Área alvo  $A = C_{\frac{r}{4}}^p$ , para aplicação de intensidade  $a\eta$ , e o restante da região total de interesse  $B = A^c \cap C_r^p$ , para aplicação de intensidade  $\eta$ .



Fonte: Autor.

A partir dessa composição, o processo  $C(n, a)$  é definido como:

**Definição 10 (Processo pontual espacial  $C(n, a)$ )** *Sejam  $n \in \mathbb{N}$  e  $a \in \mathbb{R}$ ; o processo pontual espacial  $C(n, a)$  é*

$$C(n, a) = \begin{cases} M(n, r(e^a)) & \text{se } a < 0, \\ B(n) & \text{se } 0 \leq a \leq 1, \\ S(n, a) & \text{se } a > 1. \end{cases} \quad (7)$$

Para ilustrar a execução desse processo, reproduzimos (com as adequações necessárias para que o processo fosse aplicável, sem perda de generalidade, a uma região de interesse representada por um círculo, diferentemente da proposta original que se utilizava de uma região quadrada) os exemplos encontrados em Almiron (2009).

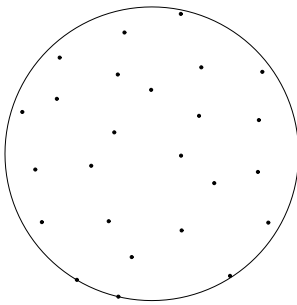
As figuras 12a, 12b e 12c apresentam alguns exemplos de instanciações de processos denominados repulsivos, com  $a = -30$ . Exemplos para processos independentes podem ser vistos nas figuras 12d, 12e e 12f. Finalmente, nas figuras 12g, 12h e 12i são apresentadas

implantações de processos chamados atrativos, com  $a = 30$ . Para os três casos, a representação da estação base é sempre feita pelo nó sensor que se encontra mais ao centro da região de deposição dos nós.

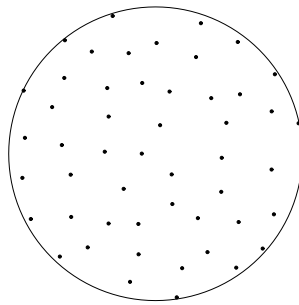
Damos um destaque especial para as Figuras 12g, 12h e 12i, pois esses cenários têm forte importância para a idealização do modelo que propomos na Seção 4.1.1.2.

Figura 12 – Exemplos de processos pontuais espaciais  $C(n, a)$ .

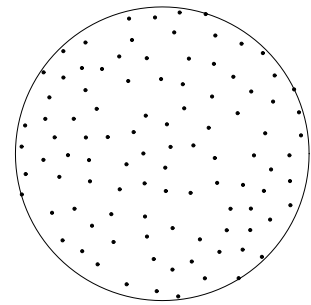
(a)  $n = 25$  e  $a = -30$



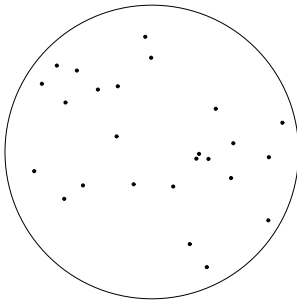
(b)  $n = 50$  e  $a = -30$



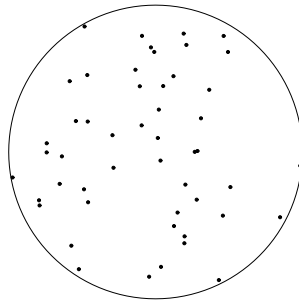
(c)  $n = 100$  e  $a = -30$



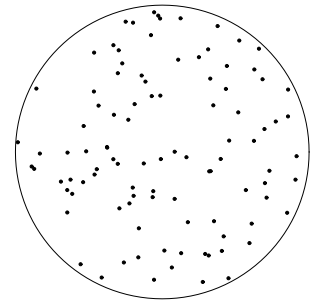
(d)  $n = 25$  e  $a = 0$



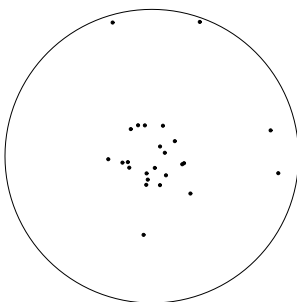
(e)  $n = 50$  e  $a = 0$



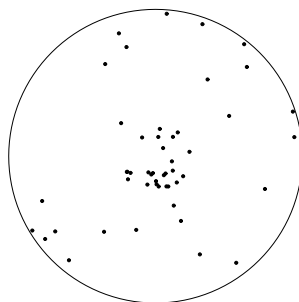
(f)  $n = 100$  e  $a = 0$



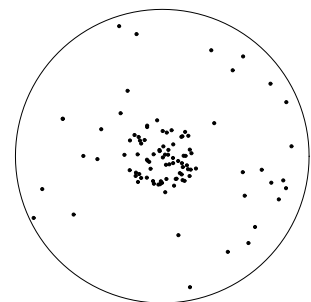
(g)  $n = 25$  e  $a = 30$



(h)  $n = 50$  e  $a = 30$



(i)  $n = 100$  e  $a = 30$



Fonte: Autor.

### 2.3.2 O Processo pontual $M^2P^2$

Esta seção é baseada nas definições feitas por Ramos et al. (2011) para o *Multilevel Marked Point Process*, resumidamente nomeado por  $M^2P^2$ .

Partindo principalmente de definições como as que podem ser encontradas no nosso trabalho na Seção 2.2.1, especialmente a definição 9, Ramos et al. (2011) discutem à respeito da proposição de um processo de distribuição de nós que possa ser utilizado, no contexto de RSSFs, para especificar algumas regiões, dentro da região total de interesse, que estariam aptas a receber uma maior concentração de nós sensores, usando um parâmetro de intensidade para esse fim.

Até aqui, percebemos uma similaridade com a ideia do processo pontual  $C$ . Uma das diferenças, dentre as mais importantes, é que o processo  $M^2P^2$  adiciona ao seu modelo o conceito de redes heterogêneas, classificando os nós sensores a serem distribuídos na rede em dois tipos: nós sensores- $L$ , nós considerados comuns, em poder de processamento, armazenamento e energia, e os nós sensores- $H$ , nós diferenciados, com características mais avançadas em relação aos nós comuns (os nós foram originalmente nomeados de *L-sensors* e *H-sensors*, respectivamente).

Podemos perceber que essa proposição explica alguns objetivos e alguns cenários que adiantamos na Seção 2.2. Dentre esses objetivos, Ramos et al. (2011) enfatizam que a ideia é escolher uma função de intensidade tal que proporcione uma maior concentração de nós sensores próximos à estação base e aos nós sensores- $H$  para que, com isso, o problema de *energy hole* seja minimizado.

Para isso, a proposição consiste em, primeiramente, localizar uma quantidade  $m$  de nós sensores- $H$  na região de interesse (sem perda de generalidade, continuaremos representando a região de interesse como um círculo  $C_r^p$ ) e, depois, distribuir uma quantidade  $n - m$ , sendo  $n$  a quantidade total de nós sensores a serem distribuídos, dos nós sensores restantes (nós sensores- $L$ ) “próximos” aos  $m$  nós sensores distribuídos anteriormente.

Denotando as coordenadas dos  $m$  nós sensores- $H$  por

$$\mathbf{h} = \{(hx_1, hy_1), (hx_2, hy_2), \dots, (hx_m, hy_m)\}, \quad (8)$$

podemos considerar a função de intensidade como:

$$\lambda(x, y) = \begin{cases} a & \text{se } d\{(x, y), (hx_i, hy_i)\} \leq rcom, 1 \leq i \leq m, \\ 1 & \text{caso contrário.} \end{cases} \quad (9)$$

em que  $a \geq 1$  é o parâmetro de atratividade,  $d$  é qualquer medida de distância e  $rcom$  é o raio de comunicação atribuído aos nós sensores- $L$ .

Com isso, Ramos et al. (2011) denotam a porção do processo pontual  $M^2P^2$  referente à distribuição dos nós sensores- $L$  como sendo  $\Lambda(n - m, a, h)$ . Já para a distribuição de nós

sensores- $H$ , primeiramente é interessante que analisemos a justificativa utilizada: em Redes de Sensores sem Fios heterogêneas, nós sensores- $H$  são úteis porque, como têm maior capacidade computacional que os nós sensores- $L$ , fornecem atalhos de comunicação e, com isso, diminuem a quantidade de saltos quando informações são enviadas até a estação base. Como essas características fazem com que nós desse tipo sejam mais custosos que nós do tipo sensores- $L$ , é desejável que sua distribuição seja feita de maneira controlada, tal que a quantidade de nós próximos uns aos outros seja a menor possível e procurando minimizar, portanto, a quantidade de nós sensores- $H$  necessários para criar os devidos atalhos.

Para conseguir modelar essa situação, Ramos et al. (2011), assim como Frery et al. (2008), utiliza o processo pontual estocástico SSI de Matérn (BADDELEY, 2007), para que os nós sensores- $H$  sejam distribuídos seguindo um modelo repulsivo. Com isso, a porção do processo pontual  $M^2P^2$  que se refere à distribuição dos nós sensores- $H$  pode ser denotada como sendo  $H(m, 2rin)$ , em que  $rin$  é o raio de inibição responsável pela caracterização da repulsividade entre os nós.

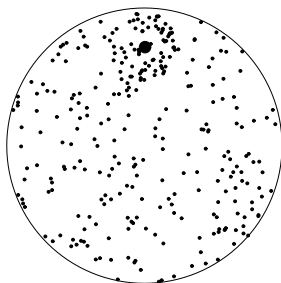
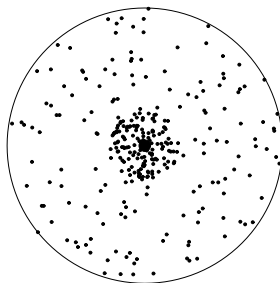
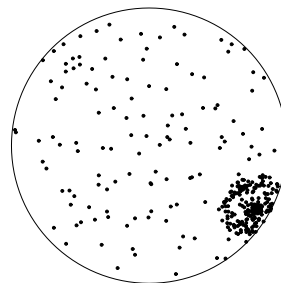
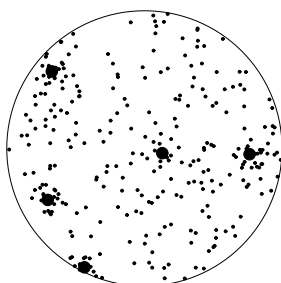
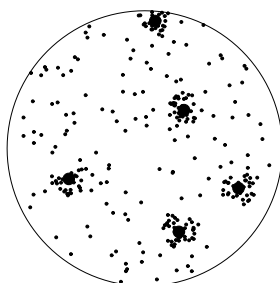
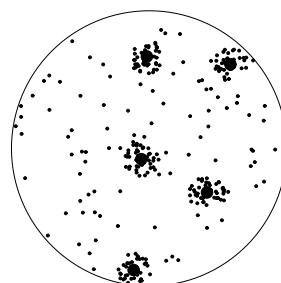
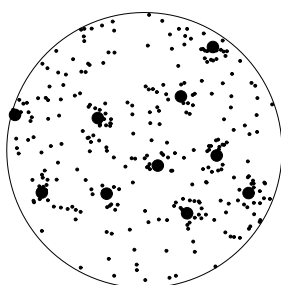
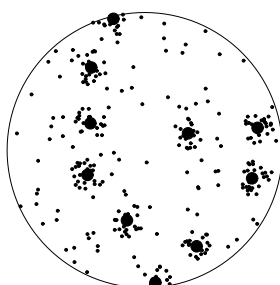
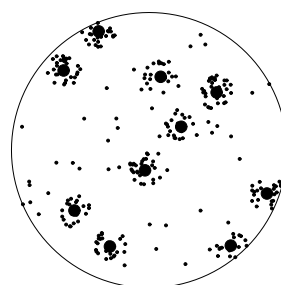
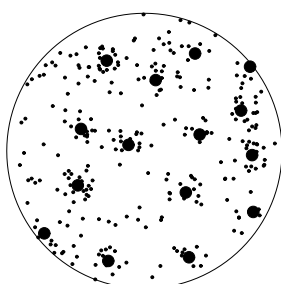
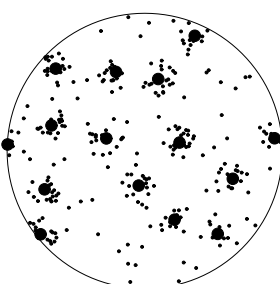
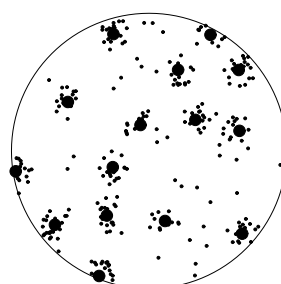
Diante das formulações apresentadas para os processos  $\Lambda$  e  $H$ , o processo pontual  $M^2P^2$  pode ser definido como:

**Definição 11 (Processo pontual espacial  $M^2P^2$ )** *Considere um número  $m \geq 1$  de nós sensores- $H$  de um total de  $n > m$  nós sensores, a intensidade  $a \geq 1$  de nós sensores- $L$  em um círculo de raio  $rcom > 0$  centrado em cada nó sensor- $H$  e raio de inibição  $rin > 0$  entre os nós sensores- $H$ . Assim,  $M^2P^2$  é um processo composto de  $m$  amostras de  $H(m, 2rin)$  e  $n - m$  amostras de  $\Lambda(n - m, a, h)$ .*

Para ilustrar essa definição, recorreremos a alguns exemplos semelhantes aos apresentados em Ramos et al. (2011). Além dos cenários seguindo uma distribuição uniforme e independente, o processo é capaz de simular os cenários, dentre outros, apresentados na Figura 13. As Figuras 13a, 13b e 13c apresentam exemplos de instanciações para o caso em que  $m = 1$ , em que o único nó sensor- $H$  distribuído faz o papel de estação base. Para os casos em que  $m > 1$  (exemplos seguintes), a estação base é representada pelo nó sensor- $H$  que estiver geograficamente mais próximo do ponto central da região de dispersão dos nós.

Mais alguns exemplos podem ser vistos nas figuras 13d, 13e e 13f, para processos com uma quantidade de nós sensores- $H$   $m = 5$ . Já nas Figuras 13g, 13h e 13i são apresentadas implantações de processos para o caso em que  $m = 10$ . Finalmente, são apresentados cenários para a quantidade de nós sensores- $H$   $m = 15$  nas Figuras 13j, 13k e 13l. Em todos os exemplos da Figura 13 temos  $n = 300$  nós sensores distribuídos. Para os valores de atratividade  $a$  utilizados, Ramos et al. (2011) os denomina em sua avaliação como levemente atrativo ( $a = 5$ ), bastante atrativo ( $a = 15$ ) e fortemente atrativo ( $a = 30$ ).

Damos um destaque especial para a Figura 13b, pois esse é um cenário de extrema importância para o modelo que propomos na Seção 4.1.1.2.

Figura 13 – Exemplos de processos pontuais espaciais  $M^2P^2$ .(a)  $m = 1$  e  $a = 5$ (b)  $m = 1$  e  $a = 15$ (c)  $m = 1$  e  $a = 30$ (d)  $m = 5$  e  $a = 5$ (e)  $m = 5$  e  $a = 15$ (f)  $m = 5$  e  $a = 30$ (g)  $m = 10$  e  $a = 5$ (h)  $m = 10$  e  $a = 15$ (i)  $m = 10$  e  $a = 30$ (j)  $m = 15$  e  $a = 5$ (k)  $m = 15$  e  $a = 15$ (l)  $m = 15$  e  $a = 30$ 

Fonte: Autor.



### 3 PROPOSTA: UM MODELO DE DISPERSÃO DE NÓS PARA BALANCEAMENTO DE CARGA

Do que foi explanado até momento, podemos resumir que:

- a) no contexto de Redes de Sensores sem Fios, entendemos que a principal limitação de um nó sensor que deve ser explorada tem relação com a maneira como ele utiliza sua energia;
- b) a exploração dessa limitação pode ser associada a diversos fatores como, por exemplo, forma de comunicação entre os nós sensores (único salto ou multissalto), algoritmos de roteamento, distribuição de nós em um espaço de interesse e redes compostas por nós sensores de diferentes características;
- c) a associação observada é que tais fatores contribuem diretamente para um melhor ou pior aproveitamento do recurso energético de um nó sensor e, consequentemente, são determinantes para o tempo de vida de uma RSSFs;
- d) dessa forma, existem algumas abordagens possíveis para a mitigação desse problema. Uma delas é a observação e atuação em um problema específico, denominado na literatura por *energy hole*.

Diante dessas informações, decidimos que nossa proposta atuaria em um escopo voltado a RSSFs com características de comunicação multissalto e com a presença de nós sensores homogêneos, ou seja, com nós sensores configurados de maneira similar, muito próximos da igualdade.

Estabelecido esse cenário, entendemos que, dentre os fatores apresentados, a distribuição espacial dos nós têm um papel de destaque para o acelerado gasto de energia de um nó sensor. Portanto, decidimos explorar a limitação energética em RSSFs abordando o problema de *energy hole*, a partir de um dos principais encaminhamentos fornecidos pelo trabalho de Li e Mohapatra (2007), a sugestão de um modelo de dispersão de nós que levasse ao um balanceamento do consumo energético entre os nós participantes das atividades de sensoriamento de uma região de interesse.

No que diz respeito a questão do balanceamento de carga, uma das suposições feitas por Li e Mohapatra (2007) era que o aumento da densidade de nós sensores nas regiões próximas a estação base seria um ponto de partida para chegar a esse objetivo. Se observamos o que foi apresentado na Seção 2.3.2, a proposta do processo pontual  $M^2P^2$  é aderente a essa concepção, para o caso dos cenários atrativos e considerando apenas a presença de um nó sensor  $H$ . Entretanto, apesar dos resultados de avaliação apresentarem uma melhoria significativa do tempo de vida da rede, o aumento de densidade é realizado apenas para a região que contém nós a um salto de roteamento da

estação base. Tal fato nos levou aos seguintes questionamentos: *apesar de minimizar o problema na região a um salto da estação base, a proposta não poderia estar "empurrando" o problema para a região seguinte, ou seja, a dois saltos? O aumento de densidade com o devido cuidado das demais regiões (a qualquer salto de comunicação da estação base) poderia levar a um maior balanceamento de carga e, conseqüentemente, a uma elevação no tempo de vida de uma RSSFs?*

É partindo desses questionamentos que propomos a modelagem de um processo pontual não-homogêneo (conceitualmente similar aos processos  $C$  e  $M^2P^2$ ) que representa a dispersão de nós sensores em uma área de monitoramento através do aumento da concentração de nós ao redor da estação base, mas considerando também que a distribuição precisa ser controlada ao longo dessa região para que o balanceamento de carga seja intensificado.

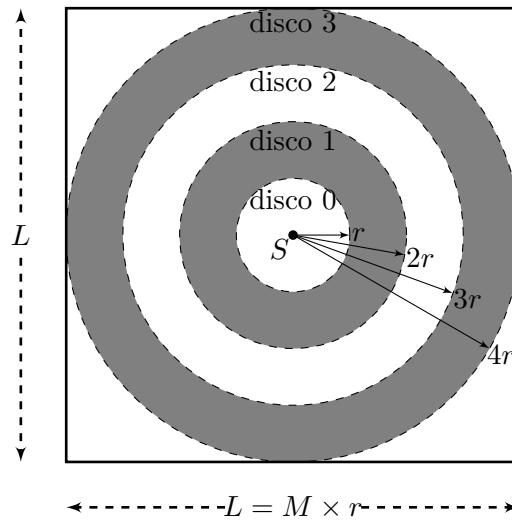
Contudo, após determinada uma proposta, seria necessário reproduzi-la de alguma maneira. Como apresentamos, os ensinamentos de Alencar-Neto (2007) e Frery et al. (2010) são fundamentais para determinar a proposta de modelagem e avaliação. Veremos em mais detalhes na Seção 4 como tais ensinamentos são explorados e, a partir daí, podemos afirmar que uma das contribuições do nosso trabalho é também mostrar a aplicação de mais um modelo e sua respectiva avaliação à abordagem de estudo de RSSFs proposta nesses trabalhos.

Na seções que seguem, discutimos sobre a concepção, a modelagem e alguns exemplos da nossa proposta.

### 3.1 Princípio

Antes de continuarmos a discussão a respeito de como os processos  $C$  e  $M^2P^2$  influenciaram na nossa proposta, abordaremos a caracterização do problema do *energy hole* apresentada por Li e Mohapatra (2007). Além de também ter tido bastante influência na nossa proposição, essa caracterização será essencial para as discussões que serão apresentadas nos Capítulos 4 e 5, isto é, tanto para a apresentação da nossa metodologia quanto para a avaliação dos resultados obtidos na simulação.

Conforme adiantamos na seção 2.1.3, em Li e Mohapatra (2007) é apresentado uma caracterização para o problema de *energy hole* através de um modelo matemático, do qual são feitas algumas avaliações analíticas das quantidades de mensagens e energia consumida pelos nós ao longo de uma RSSF homogênea e inserida em um contexto de comunicação através de múltiplos saltos. Abstraindo alguns detalhes desse modelo, observemos a figura 14.

Figura 14 – Caracterização do problema do *energy hole*.

Fonte: Adaptado de Li e Mohapatra (2007).

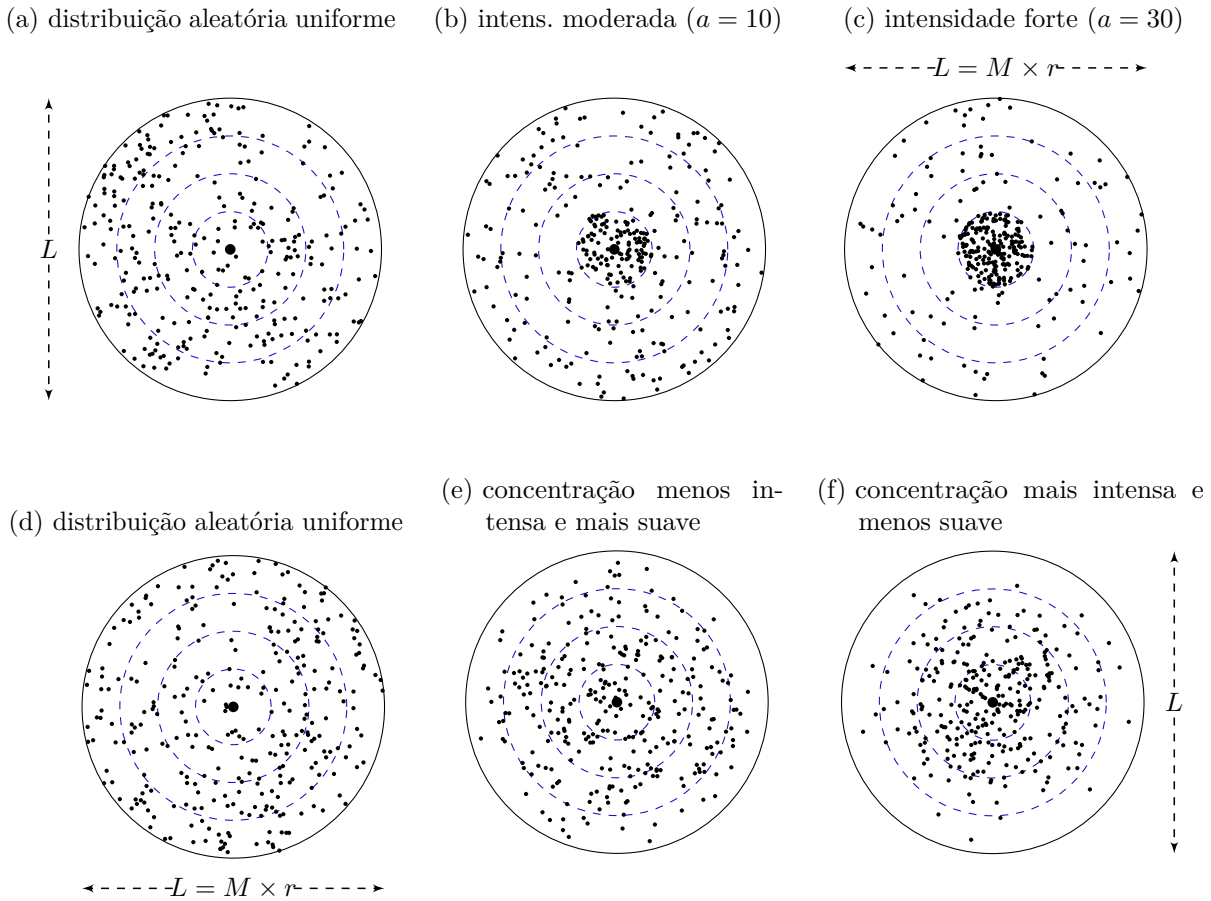
Na caracterização proposta, foi considerada uma RSSF com nós sensores homogêneos e distribuídos de maneira aleatória e uniforme em uma região quadrada de área  $A = [0, L] \times [0, L] = [0, L]^2$ , em que  $L = M \times r$ . Dessa maneira, e assumindo que a estação base ( $S$ ) se encontra posicionado no centro da região proposta, a área pode ser inteiramente dividida em  $\frac{M}{2}$  discos concêntricos com uma distância de  $r$  metros.

A partir daqui, vejamos as Figuras 15a, 15b e 15c. Essas representam nada mais do que alguns cenários do modo atrativo do processo  $M^2P^2$  sob a ótica da caracterização apresentada na Figura 14 (preferimos não representar o modo atrativo do processo  $C$  por ser visualmente similar a esse conjunto de figuras, dessa forma as mesmas conclusões se aplicam também a esse processo). Para os três casos utilizamos uma quantidade  $n = 300$  de nós sensores, uma região circular de diâmetro  $L$  e discos concêntricos de raio  $r$ ,  $2r$ ,  $3r$  e  $4r$  (representados pelas linhas azuis tracejadas). A diferença fica por conta da modificação do parâmetro  $a$  (mais detalhes na definição 11), em que  $a = 0$  – distribuição aleatória uniforme – para a Figura 15a,  $a = 10$  para a Figura 15b e  $a = 30$  para a Figura 15c.

Como já havíamos adiantado, e agora esquematicamente, percebemos que a intervenção na concentração dos nós está restrita a região do círculo de raio  $r$ , independente do fator de atratividade atribuído ao parâmetro  $a$ . Essa é a região em que os nós sensores estão a um salto de comunicação da estação base, supondo que seus respectivos raios de comunicação equivalem a uma distância máxima  $r$ .

Devido a essa restrição é que questionamos se o problema não estaria sendo prolongado para as demais regiões, como podemos observar, por exemplo, no disco de raio  $2r$ , em que a densidade de nós parece diminuir, contribuindo para o surgimento de um “problema de *energy hole*” associado aos nós sensores localizados a dois saltos da

Figura 15 – Cenários para distribuição aleatória uniforme (a, d) e o aumento da densidade de nós ao redor da estação base a partir do processo  $M^2P^2$  (b, c) e nossa proposta (e, f).



Fonte: Autor.

estação base, além de possivelmente ser prejudicial ao fator de cobertura da região de monitoramento.

Dessa forma, a nossa proposta tenta investigar esse problema fazendo com que a dispersão dos nós sensores seja na forma apresentada nas Figuras 15e e 15f. Considerando sensores com as mesmas características e uma RSSF com os mesmos parâmetros utilizados para a representação do modo atrativo do processo  $M^2P^2$ , podemos observar que a intensificação de nós sensores ao redor da estação base é feita de maneira mais suave, se compararmos as Figuras 15e e 15f com as Figuras 15b e 15c (a Figura 15d representa apenas mais um caso de distribuição aleatória uniforme e está sendo apresentada para que as comparações entre cenários fique mais clara). Para esses casos também são apresentados possíveis contribuições negativas a respeito do fator de cobertura da rede, porém associadas as regiões de fronteira da área de interesse (Figura 15f).

Partindo desse começo de análise, buscamos uma maneira de modelar esses cenários. Na seção seguinte apresentaremos essa modelagem.

## 3.2 Modelagem

De posse das suposições iniciais que mostramos na seção anterior, precisamos retornar, primeiramente, ao que foi exposto no final da Seção 2.2.1.

Para chegarmos a uma modelagem dos cenários apresentados nas Figuras 15d, 15e e 15f, podemos dizer que nossa proposta consiste em descrever a localização dos sensores através de um processo de Poisson não-homogêneo. Em outras palavras, buscávamos a definição de um processo pontual no qual a intensidade não fosse a mesma em subconjuntos de uma região de dispersão. Nesse caso, nossa busca se concentrou, conforme nos indicou as definições 7, 8 e 9, em uma função que denotasse essa intensidade e que também representasse a ideia de haver, em média, uma concentração maior de nós sensores em volta de um ponto específico.

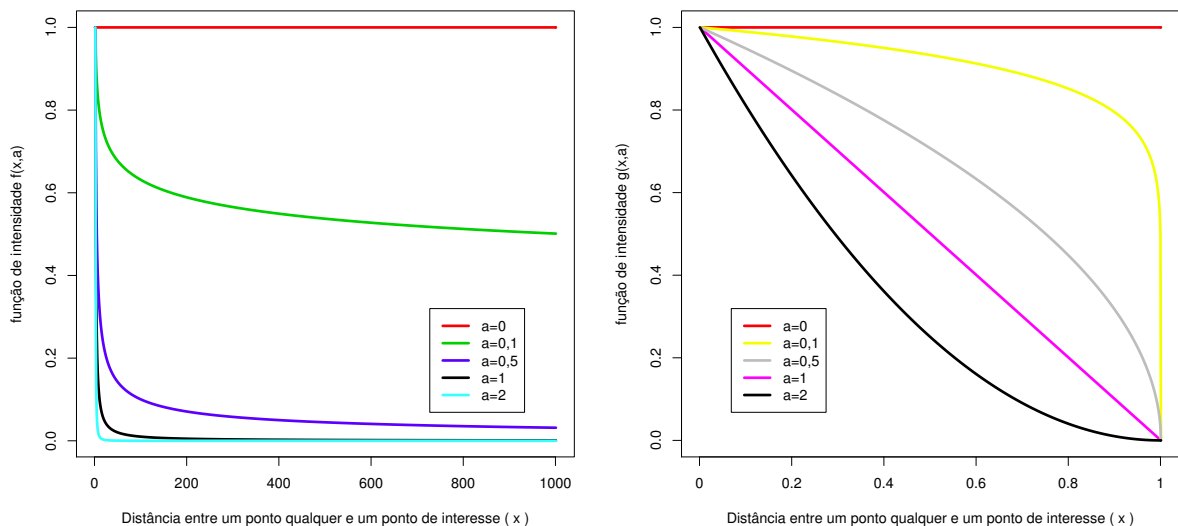
### 3.2.1 Estudo da função de intensidade

Detalhando um pouco mais a busca por essa “função de intensidade” (resumidamente com essa denominação a partir daqui), precisávamos basicamente de dois atributos: 1) que atuasse em função da distância entre um ponto qualquer até um ponto de interesse e 2) que sua intensidade fosse variável. É fácil perceber que o primeiro atributo é essencial pois é o desafio a ser alcançado. Já a segunda característica é necessária para que pudéssemos avaliar diversos cenários, já que nos propomos a fazer uma avaliação através de simulações. Vejamos na Figura 16 o exemplo de duas funções que representam essa ideia.

Figura 16 – Exemplos de funções de intensidade.

(a)  $f(x, a) = 1/x^a$

(b)  $g(x, a) = (1 - x)^a$



Fonte: Autor.

Os gráficos apresentados na Figura 16a representam uma função  $f(x, a): \mathbb{R}_+ \rightarrow \mathbb{R}_+$ . Já a Figura 16b apresenta os gráficos de uma função  $g(x, a): [0, 1] \rightarrow [0, 1]$ . Ambas as funções possuem parâmetros com os mesmos significados:  $x$  é a distância de um ponto qualquer a um ponto de interesse e  $a \geq 0$  é um fator de intensidade. Para a Figura 16a, apresentamos apenas uma porção dos gráficos para uma melhor visualização e comparação entre o comportamento das funções  $f$  e  $g$ . Também por motivos de comparação, estabelecemos para ambas as funções o mesmo conjunto de valores para o parâmetro  $a$ .

Podemos interpretar os gráficos afirmando que existe uma probabilidade (valor das funções  $f$  e  $g$ ) associada a localização de um ponto qualquer no espaço. Para ambas as figuras, quando o parâmetro  $a = 0$  (curva em vermelho) estamos diante de um processo pontual binomial, pois o valor de probabilidade associado é o mesmo independente da distância, ou seja, a localização dos pontos acontece de maneira independente. Já para os gráficos representados pela variação crescente do parâmetro  $a$ , vemos que os valores das funções são diferentes. Na Figura 16a, podemos observar que as curvas têm um comportamento similar, o que varia é a região onde acontecerá a maior probabilidade de localização de pontos. Quanto maior o valor do parâmetro  $a$ , menor é a região, ou seja, quando o valor do fator de intensidade é crescente, a localização dos pontos num possível processo pontual não-homogêneo volta sua atenção para uma região cada vez mais próxima de um ponto de referência. Esse comportamento pode ser representado pela área delimitada pelo ponto em que a curva muda de direção.

No que diz respeito aos gráficos da função  $g$ , na Figura 16b, o mesmo comentário da função  $f$  é válido, porém as regiões de aumento de concentração de pontos são maiores, fazendo com que o aumento de densidade de pontos seja mais suave e espalhada.

Um fator também importante para determinação da nossa função de intensidade é que o espaço de distância precisaria ser limitado, em outras palavras, precisávamos que a convergência entre os valores de probabilidade 0 e 1 fosse, de certa forma, rápido. Em ambas as funções percebemos esse desempenho para os valores do parâmetro  $a \geq 1$ , mas para a função  $f$  isso não acontece para os valores de  $a \leq 1$ .

Diante do que foi exposto para esses exemplos de funções, escolhemos a função  $g$  como base para a nossa modelagem.

### 3.2.2 Processo pontual não-homogêneo proposto: o processo $I$

Diante de algumas caracterizações que serão abordadas a partir da Seção 4 e de posse de uma função de intensidade  $g$ , podemos definir, de maneira mais geral, que uma função de intensidade  $\kappa$  é tal que

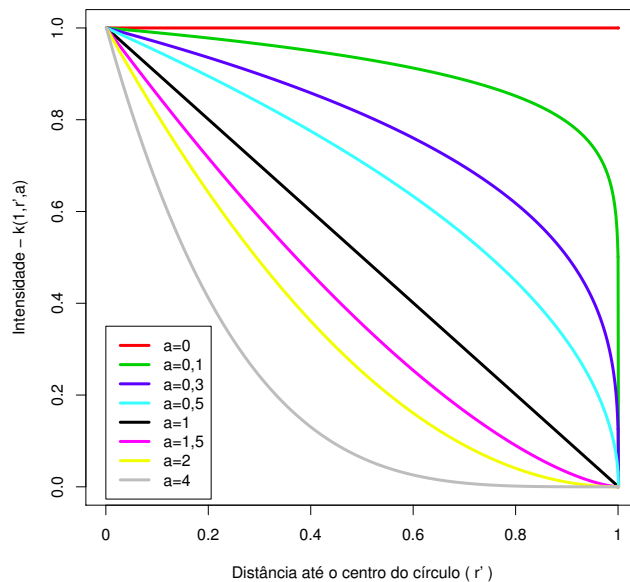
$$\kappa(r, r', a) = \frac{(r - r')^a}{r^a} \quad (10)$$

em que  $r$  representa o raio da área de monitoramento definida (ver Seção 4.1.1 para mais detalhes),  $r' = \sqrt{x_i^2 + y_i^2}$ , com  $x_i$  e  $y_i$  sendo os valores das coordenadas de um ponto definidos por variáveis aleatórias independentes; e  $a \geq 0$  é o fator principal para determinar a maior concentração de pontos em determinadas regiões da área de monitoramento.

Assim, propomos um processo pontual  $I(n, \kappa)$  que descreve a localização espacial de  $n$  pontos, na área de monitoramento definida para o nosso estudo, através da função de intensidade  $\kappa$ , cujo domínio é a dimensão dessa região de interesse.

Para entender como essa abordagem funciona, vejamos o comportamento da função  $\kappa$ , na Figura 17, para o caso particular em que  $r = 1$ , isto é, para o caso em que a área de monitoramento é um círculo de raio unitário.

Figura 17 – Comportamento da função intensidade  $\kappa$  para alguns valores do parâmetro  $a$ .



Fonte: Autor.

Na figura, temos um gráfico em que o eixo das abscissas representa a distância entre um ponto qualquer posicionado dentro do círculo até o centro do mesmo. Já o eixo das ordenadas representa os valores para função intensidade  $\kappa$ . O que queremos dizer é que essa função  $\kappa$ , quando aplicada ao processo pontual proposto,  $I(n, \kappa)$ , na prática, determina com que probabilidade um ponto estará localizado em uma determinada região da área de monitoramento, ou ainda, podemos dizer que as regiões com maior valor de intensidade representam áreas que, em média, têm uma maior concentração de pontos.

Com base nisso, observamos que podemos modelar a abordagem clássica de distribuição de pontos em um determinado espaço, isto é, a distribuição espacial totalmente aleatória. Esse caso está representado pela curva  $a = 0$ , a qual mostra que qualquer ponto na área de monitoramento tem a mesma intensidade.

Já quando variamos o valor do parâmetro  $a$ , podemos ver que o valor da função de intensidade decresce a medida que a distância de um ponto ao centro do círculo aumenta. É importante destacar que há uma diferença importante para a variação entre os intervalos  $0 < a < 1$  e  $1 < a < \infty$ . As variações de intensidade no primeiro intervalo são mais suaves que as variações de intensidade no segundo intervalo. Isso é um fator importante para o nosso estudo, pois com a análise exploratória prévia do comportamento dessa função (como podemos observar na Seção 3.2.1) aplicado ao processo  $I$ , já foi possível observar que para valores muito elevados de  $a$  não é possível adquirir cenários de simulação com níveis aceitáveis de cobertura da área de monitoramento. Conforme declaramos na seção introdutória desse trabalho, só utilizaremos cenários que apresentem uma taxa de cobertura acima de um determinado limiar.

Na seção que segue, abordaremos exemplos gráficos do comportamento do processo pontual não-homogêneo  $I$ .

### 3.3 Alguns exemplos

A Figura 18 mostra os resultados da aplicação da função  $g$  na construção de cenários de simulação obtidos através do processo  $I$ .

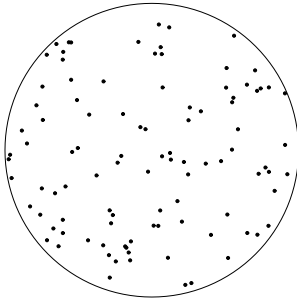
Podemos perceber, conforme descrevemos anteriormente quando explanávamos sobre o estudo prévio da função de intensidade (Seção 3.2.1), que ao analisarmos as representações por linha (por exemplo as Figuras 18d, 18e e 18f) o incremento no valor do parâmetro  $a$  resulta em uma maior densidade de pontos nas proximidades do centro do círculo. Se analisarmos através das representações por coluna (por exemplo as Figuras 18b, 18e e 18h), quando o parâmetro  $a$  tem o mesmo valor a intensidade do processo é a mesma, independente da variação da quantidade de pontos a serem alocados. Com isso, finalmente conseguimos uma representação semelhante às ideias propostas por Li e Mohapatra (2007).

De posse da nossa proposta apresentaremos na próxima seção a metodologia que utilizamos para avaliar o comportamento desse processo pontual no contexto de RSSFs, isto é, mostraremos em mais detalhes como utilizamos os ensinamentos de Alencar-Neto (2007) Frery et al. (2010) no que dizem respeito à representação esquemática de uma RSSFs e como essa abstração pode, adicionalmente a outras definições, ser traduzida em modelos possíveis de serem utilizados em nossas simulações e estudos comparativos.

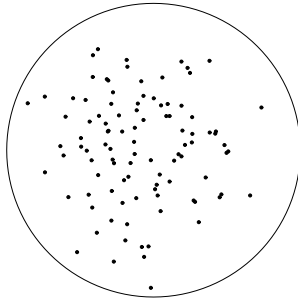


Figura 18 – Comportamento do processo  $I$  para o parâmetro de intensidade  $a = \{0, 1, 2\}$  e quantidade de pontos  $n = \{100, 300, 500\}$ .

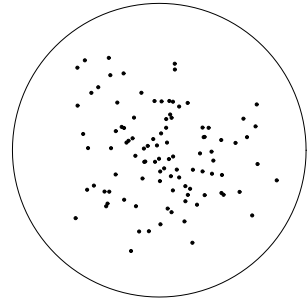
(a)  $a = 0$  e  $n = 100$



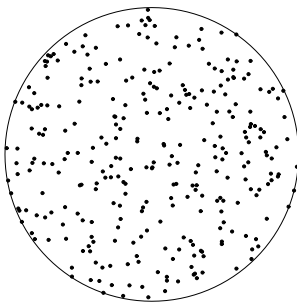
(b)  $a = 1$  e  $n = 100$



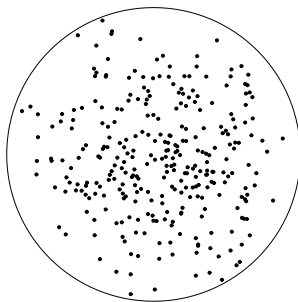
(c)  $a = 2$  e  $n = 100$



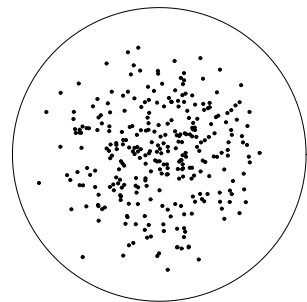
(d)  $a = 0$  e  $n = 300$



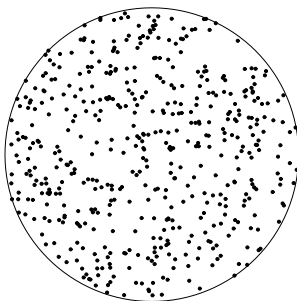
(e)  $a = 1$  e  $n = 300$



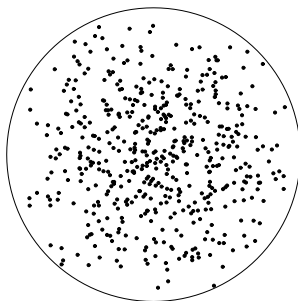
(f)  $a = 2$  e  $n = 300$



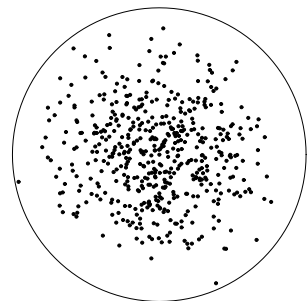
(g)  $a = 0$  e  $n = 500$



(h)  $a = 1$  e  $n = 500$



(i)  $a = 2$  e  $n = 500$



Fonte: Autor.



## 4 METODOLOGIA

Como já relatamos, nas últimas décadas o interesse por estudos na área de RSSFs cresceu significativamente. Tanto interesse se deve, dentre outras questões, ao fato de redes desse tipo poderem ser aplicadas como ferramenta para soluções dos mais variados problemas de monitoramento e aquisição de dados.

De fato, a aplicação de RSSFs pode trazer diversos benefícios. Entretanto, ainda se fazem necessários diversos aprimoramentos que levem a uma maior difusão do seu uso. Tais melhoramentos são os principais alvos de estudo da comunidade científica.

Contudo, apesar das RSSFs permitirem uma rápida distribuição devido a não necessitarem de uma infraestrutura fixa pré-estabelecida, na maioria dos casos é muito difícil que os estudos nessa área sejam dirigidos por métodos analíticos ou experimentações de bancada. Por esse motivo é que a utilização de técnicas de simulação têm se apresentado uma ferramenta altamente valiosa (ANDEL; YASINSAC, 2006). Pesquisadores têm utilizado simulações para diversos objetivos, como por exemplo na comparação de múltiplas alternativas sob uma grande quantidade de condições, e na avaliação do comportamento de modelos, algoritmos e/ou protocolos antes da sua efetiva implantação em um sistema ou aplicação.

Porém, como qualquer outro método ou técnica, simulações também estão sujeitas a erros. Em muitos casos, tais erros podem ser provenientes de falhas nas construções de modelos ou até mesmo em uma má escolha das plataformas para simulação e para análise, o que podem ocasionar análises de dados inapropriadas e produção de resultados incorretos. Para efetiva implementação das ideias propostas nesse trabalho, nós nos preocupamos com esses aspectos.

No que diz respeito às falhas nas construções de modelos, podemos interpretar tais falhas como, por exemplo, suposições imprecisas, que levam os modelos a retratarem características errôneas do ambiente a ser simulado. Em nosso trabalho, aproveitamos o modelo proposto para, além de apresentar uma solução para o problema do *energy hole* e o consumo não balanceado de energia em RSSFs com nós sensores se comunicando por múltiplos saltos, difundir as ideias propostas por Frery et al. (2010) e por Ramos et al. (2011), dentre outros. Essas ideias buscam corrigir um dos pontos observados por Andel e Yasinsac (2006), sobre os resultados imprecisos obtidos em simulações na área de RSSFs devido, dentre outros fatores, à utilização de modelos de dispersão de nós regidos somente por uma distribuição espacial aleatória uniforme. Veremos mais detalhes sobre a metodologia empregada para a implementação do modelo de dispersão proposto na Seção 4.1.1.

Já com relação às plataformas de simulação e análise empregadas, tentamos evitar (ou minimizar) problemas como os apontados em Bustos e Frery (2005), Andel e Yasinsac (2006), Almiron, Almeida e Miranda (2009) e Almiron et al. (2010), fazendo com que a avaliação crítica desses recursos computacionais empregados sejam parte integrante do nosso estudo. Mais detalhes serão abordados na Seção 4.1.2 e no Apêndice A.

## 4.1 Ingredientes da simulação

Já que declaramos que um dos objetivos do nosso trabalho é a preocupação com a maneira como modelamos a porção do mundo real na qual o nosso trabalho está inserido, é grande a importância de deixarmos claro como essa modelagem e organização será realizada. Relembrando um dos nossos destaques no Capítulo 2, em especial a Seção 2.1.1, apresentamos uma forma de representação de uma RSSFs (Figura 19) baseada no trabalho de Frery et al. (2010). Mas como traduzir essa modelagem em um ambiente de simulação? Para responder a essa pergunta, além de continuarmos a análise do trabalho de Frery et al. (2010), recorreremos a abordagem feita por Alencar-Neto (2007).

Figura 19 – Diagrama esquemático de RSSFs.

$$\begin{array}{ccccccc}
 \mathcal{N} & \xrightarrow{F} & \mathbf{V}^* & \xrightarrow{S} & \mathbf{V} & \xrightarrow{\Psi} & \mathbf{V}' & \xrightarrow{\hat{F}} & \mathbf{V}^* \\
 & & R^* \downarrow & & & & & & R' \downarrow \\
 & & D^* & & & & & & D'
 \end{array}$$

Fonte: Frery et al. (2010).

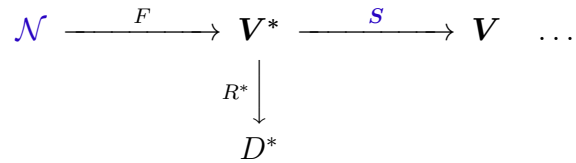
No nosso trabalho seguimos uma organização bem semelhante a que foi proposta em Alencar-Neto (2007). Para traduzir a nossa preocupação em obtermos um ambiente de simulação que tente ao máximo evitar imprecisões e, conseqüentemente, falhas nas análises as quais estamos nos propondo a fazer, introduzimos os primeiros ingredientes da nossa metodologia: os modelos.

Segundo Alencar-Neto (2007), o primeiro passo para a proposição de uma abordagem de observação de um determinado fenômeno consiste em propor abstrações, ou seja, hipóteses que, por serem simplificações, redundam em modelos tratáveis, para descrever cada elemento que se julgue importante desse fenômeno de interesse. No nosso trabalho os elementos identificados como importantes são:

- a) a área sob monitoramento;
- b) a distribuição no espaço dos nós sensores;
- c) as características de operação dos nós sensores.

Não esquecendo da porção do diagrama esquemático de Frery et al. (2010) que estamos atuando, tais elementos são os destacados em azul na Figura 20, em que  $\mathcal{N}$  representa o elemento “área sob monitoramento”, e os elementos “distribuição de nós sensores no espaço” e suas “características operacionais” são representados pelo conjunto de nós sensores  $S = \{S_1, \dots, S_n\}$ .

Figura 20 – Diagrama esquemático de RSSFs (escopo de atuação).



Fonte: Adaptado de Frery et al. (2010).

Diferentemente dos trabalhos que utilizamos como base para essa identificação, não tratamos de aspectos relacionados a agrupamento e fusão de dados. Apesar de toda a importância demonstrada pelos autores, preferimos não adotar a mesma abordagem pois em algumas análises exploratórias prévias observamos que estratégias desses tipos, de fato, “mascaram” o problema do *energy hole*, ponto chave de análise do nosso trabalho.

A sequência que damos aos trabalhos de Alencar-Neto (2007) e Frery et al. (2010) é essencial para os objetivos do nosso trabalho. Com essa abordagem, conseguimos aproveitar alguns aspectos como:

- a) a independência dos elementos que formam o modelo, o que o torna bastante flexível;
- b) o funcionamento de uma RSSF no contexto de processamento de sinais, o que facilita a construção das abstrações citadas por Alencar-Neto (2007);
- c) a estratégia para avaliar o impacto de diversos fatores em uma RSSF, adequando a nossa realidade que é a avaliação de um modelo de dispersão de nós.

Mais detalhes sobre cada um dos elementos destacados aqui, e como a partir de suas abstrações poderemos chamá-los de modelos, poderão ser encontrados na Seção 4.1.1.

Contudo, a simplificação do mundo real não está relacionada à simplicidade da implementação de um modelo. Adicionalmente, o fato de não ser desejável ou até mesmo possível realizar experimentações de bancada e/ou utilizar ferramentas analíticas muitas vezes não implica na realização de simulações em um tempo considerado aceitável. Graças à natureza essencialmente estocástica dos modelos formulados, é possível extrair informações utilizando experimentos Monte Carlo (ALENCAR-NETO, 2007). Esse detalhe é a principal motivação pela qual decidimos utilizar a abordagem de Monte Carlo em nosso estudo

e, como poderemos ver em mais detalhes na Seção 4.2.1, essa ferramenta tem sido de fundamental importância para os estudos na área de RSSFs.

Dado que já possuímos os modelos descritos e a decisão de utilizar uma abordagem Monte Carlo, precisamos saber qual o ambiente computacional que será utilizado para implementá-los. Dessa forma, um outro ingrediente importante para nossas simulações é a escolha das plataformas computacionais. Podemos identificar duas principais plataformas de simulação Monte Carlo que, frequentemente, se confundem em uma: a plataforma para realizar, de fato, as simulações, e a plataforma onde os resultados são analisados. Como plataforma de análise de resultados utilizamos o ambiente de desenvolvimento e *software* estatístico R (R CORE TEAM, 2013). Já como plataforma de simulação, utilizamos o simulador *Sinalgo* (DISTRIBUTED COMPUTING GROUP, 2007), disponibilizado na plataforma de desenvolvimento Java (JAVA DEVELOPMENT TEAM, 2013), em conjunto com o ambiente R.

Como já antecipamos, a escolha dessas plataformas são parte integrante desse trabalho. Portanto, sugerimos ao leitor a Seção 4.1.2 como ponto de partida para mais detalhes de como chegamos a escolha dessas plataformas.

### 4.1.1 Modelos

Na seção anterior falamos da importância de dar continuidade à metodologia proposta em Alencar-Neto (2007) e Frery et al. (2010). Dissemos que um dos fatores relevantes e que torna possível essa continuidade é a flexibilidade da modelagem proposta.

Valendo-se disso, identificamos os elementos importantes para a construção do nosso modelo de simulação e descartamos aqueles que avaliamos desnecessários. Além disso, essa flexibilidade permitiu que modificássemos a construção das abstrações (modelagem dos elementos) de acordo com as necessidades previstas e com as particularidades do nosso estudo, sem perder a generalidade da metodologia.

Nas seções seguintes, apresentamos as abstrações propostas para os elementos do nosso modelo de simulação.

#### 4.1.1.1 Área de monitoramento

Diferentemente de grande parte dos trabalhos em RSSFs, que apontam como área de monitoramento uma região quadrada, aqui consideramos nos nossos experimentos que os sensores estão distribuídos espacialmente em uma região circular de raio  $R = 50$  m. Essa ideia surgiu a partir da leitura do trabalho de Li e Mohapatra (2007), pioneiro na caracterização e modelagem do problema de *energy hole*. O objetivo dessa diferenciação foi demonstrar com maior qualidade os resultados obtidos através do comportamento do modelo proposto de dispersão de nós sensores.

Nesse ponto, também diferenciamos do trabalho de Alencar-Neto (2007), porém utilizamos as mesmas simplificações propostas por esse, pois não consideramos aspectos como relevo ou quaisquer objetos que representem obstáculos limitantes da área de percepção dos nós sensores e o raio de comunicação entre esses. Dessa maneira, não agregamos complexidade e fatores extras à modelagem da região de monitoramento, além de focarmos em características técnicas dos nós sensores como elementos determinantes para caracterização das áreas de percepção e raios de comunicação.

#### 4.1.1.2 Distribuição de nós sensores

Essa abstração está fortemente relacionada também com o foco da proposta do nosso trabalho.

Já vimos no Capítulo 3 como abordamos essa abstração e a traduzimos em um modelo. Recomendamos a leitura do referido capítulo para esclarecimentos de como propomos a modelagem através de um processo pontual não-homogêneo  $I(n, \kappa)$ , o qual descreve a localização espacial de  $n$  pontos através da função de intensidade  $k$ , cujo domínio é a dimensão da área de monitoramento definida na seção anterior. A função de intensidade  $\kappa$  (Equação 3.2.2) é tal que

$$\kappa(r, r', a) = \frac{(r - r')^a}{r^a}$$

em que  $r$  representa o raio da área de monitoramento circular,  $r' = \sqrt{x_i^2 + y_i^2}$ , com  $x_i$  e  $y_i$  sendo os valores das coordenadas de um ponto definidos por variáveis aleatórias independentes; e  $a \geq 0$  é o fator principal para determinar a maior concentração de pontos em determinadas regiões da área de monitoramento.

Com relação à implementação dos conceitos de processos pontuais aplicados a esse trabalho e, conseqüentemente, ao modelo de dispersão de nós sensores proposto, utilizamos a função `rpoint`, para a distribuição dos pontos, e objetos do tipo `owin`, para delimitar a área de distribuição desses pontos. Todos os recursos utilizados para a implementação desse modelo estão implementados no pacote `spatstat`, que está disponível para a plataforma R.

#### 4.1.1.3 Características operacionais de nós sensores

Nesta seção, apresentamos uma abstração para modelar o tipo de nó sensor que utilizamos em nossas simulações. São discutidas duas características desses nós: raio de comunicação e raio de sensoriamento (também comumente denominado de raio de percepção).

#### 4.1.1.3.1 Raio de comunicação

Para a discussão sobre as características operacionais dos nós sensores e para efeitos de discussões dos resultados obtidos em simulação, principalmente os resultados que dizem respeito à influência do modelo de dispersão de nós em relação ao balanceamento do consumo energético ao longo da rede, precisaremos de características comparáveis às apresentadas por Li e Mohapatra (2007).

Na seção 3.1 apresentamos a caracterização para o problema de *energy hole* através de um modelo matemático proposto em Li e Mohapatra (2007). Diante disso, é importante perceber que, considerando a região de monitoramento que propusemos para o nosso estudo, podemos definir no nosso modelo que o raio de comunicação é dado por  $r_c = \frac{L}{M}$  ( $L$  e  $M$  conforme definições apresentadas no Capítulo 3.1 e representação na Figura 14).

A partir da definição desse raio de comunicação  $r_c$ , cada nó sensor  $S_i$  terá um alcance de comunicação equivalente a um círculo de raio  $r_c$  e centrado no próprio nó sensor. Como simplificação do modelo, a comunicação pode ser realizada com a mesma intensidade e qualidade em qualquer ponto dentro desse círculo. Outro fator importante é que o número de vizinhos do sensor  $S_i$  é dado por

$$N(H_i) = \sum_{j=1}^n \mathbb{1}[S_j], \quad (11)$$

para todo  $j \neq i \in [1, n]$ , com  $n > 0$  representando a quantidade de nós sensores distribuídos e  $\mathbb{1}[S_j]$  sendo a função indicadora definida por

$$\mathbb{1}[S_j] = \begin{cases} 1 & \text{se } S_j \text{ está posicionado na área de comunicação de } S_i, \\ 0 & \text{caso contrário.} \end{cases} \quad (12)$$

Em outras palavras, podemos dizer que cada nó sensor terá como vizinho aquele nó sensor que estiver dentro de sua área de comunicação. Na figura 21, podemos ver um exemplo dessa abstração.

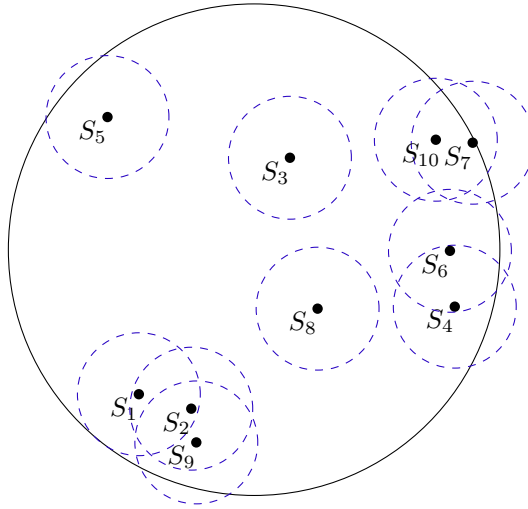
#### 4.1.1.3.2 Raio de Sensoriamento

Da mesma forma que delimitamos os aspectos importantes para a definição de raio de comunicação para o nosso estudo, é de igual importância a apresentação de características relevantes para a definição de raio de sensoriamento.

No nosso trabalho, apesar de não termos o compromisso de apresentar resultados que confrontem a influência do modelo de dispersão de nós com a qualidade do dado sensoriado e transmitido até a estação base, precisamos delimitar tais características pois o raio de sensoriamento terá grande importância para o entendimento do estudo simplificado de cobertura de uma RSSF, diante do nosso modelo de dispersão, que propusemos em nossas



Figura 21 – Distribuição de dez nós sensores e suas respectivas áreas de comunicação.



Fonte: Adaptado de Alencar-Neto (2007).

simulações (cujo os resultados são apresentados no Capítulo 5). Nesse contexto, decidimos pela manutenção de características comparáveis às apresentadas por Alencar-Neto (2007) e Frery et al. (2010).

Nos trabalhos de Alencar-Neto (2007) e Frery et al. (2010) são definidos duas abstrações para definir os nós sensores a serem utilizados em suas simulações. Na primeira delas, uma função característica para o nó sensor  $S_i$  é definida por uma função indicadora  $k_i: \mathbb{R}^2 \rightarrow \mathbb{R}_+$  tal que

$$k_i(x, y) = \mathbb{1}_{V_i}(x, y) \quad (13)$$

em que  $V_i$  é a célula de Voronoi (AURENHAMMER, 1991) associada ao nó sensor  $S_i$ , e  $\mathbb{1}_A(x)$  é a função indicadora do conjunto  $A$ , isto é,  $\mathbb{1}_A(x) = 1$  se  $x \in A$ , e  $\mathbb{1}_A(x) = 0$  caso contrário.

Na segunda abstração, que possui maior apelo físico do que a primeira (ALENCAR-NETO, 2007), a área de sensoriamento de cada nó sensor é caracterizada pela função de resposta radial do seu fotossensor, definida como a função indicadora de uma área circular em volta do sensor. Assim, assumindo um único modelo de fotossensor e como consequência uma única função de resposta radial, foram modeladas as áreas de sensoriamento  $A_i$  dos nós sensores  $S_i$  como sendo discos de áreas iguais definidos pelo mesmo raio de sensoriamento  $r_s$ . Nessa abstração, a função característica do nó sensor  $S_i$  é definida por uma função  $k_i: \mathbb{R}^2 \rightarrow \mathbb{R}_+$  tal que

$$k_i(x, y) = \begin{cases} 1 & \text{se } \sqrt{(x_i - x)^2 + (y_i - y)^2} \leq r_s, \\ 0 & \text{caso contrário.} \end{cases} \quad (14)$$

De forma a tornar as abstrações comparáveis, Alencar-Neto (2007) e Frery et

al. (2010) chamam a atenção para a modelagem de um nó sensor que possui a mesma sensibilidade em toda a sua área de sensoriamento em ambas as abstrações. No primeiro caso, a área de sensoriamento  $A_i$  do nó sensor  $S_i$  corresponde à célula de Voronoi definida por sua localização. Já no segundo caso, a área de sensoriamento  $A_i$  é um círculo de raio  $r_s$  centrado no nó sensor  $S_i$ . Exemplificando de maneira mais prática, como a dispersão de 100 nós sensores em uma área de monitoramento idêntica a descrita na Seção 4.1.1.1, e o particionamento dessa área em tantas células de Voronoi quanto a quantidade de nós sensores, é possível assumir que, em média, a área de sensoriamento de um nó sensor corresponde a aproximadamente  $78,54 \text{ m}^2$ . Assim, equalizando com a primeira abstração, o raio de sensoriamento para a segunda abstração pode ser definido como  $r_s = \sqrt{78,54/\pi} \approx 5 \text{ m}$ . Para a abstração e construção do nosso modelo, utilizaremos apenas o segundo caso.

Quanto à quantificação do raio de sensoriamento na nossa abordagem, como não precisaremos tornar nossa abstração comparável com a modelagem do primeiro caso, e como o principal objetivo da utilização do raio de sensoriamento é apenas medir o percentual de cobertura percebido pelo conjunto de nós sensores espalhados na área de monitoramento, utilizaremos valores para o raio de sensoriamento  $r_s$  de maneira semelhante à abordagem de Liu (2006), o qual torna o raio de sensoriamento proporcional ao raio de comunicação.

#### 4.1.2 Análise do ambiente de simulação

Na seção anterior, apresentamos os primeiros ingredientes para responder à pergunta feita na Seção 4.1, como traduzir a modelagem de RSSF proposta em um ambiente de simulação? A partir do que discutimos até o momento neste capítulo e seguindo as afirmações de Almiron (2009), vimos que, primeiramente seria necessário selecionar os elementos e as propriedades físicas de interesse do nosso estudo, e posteriormente seria necessário que esses elementos e propriedades fossem modelados matematicamente com um certo nível de abstração. Para atingir esse objetivo, utilizamos ingredientes estocásticos para que tais abstrações (modelos) fossem dotados do maior realismo possível.

Diante desses modelos, seria possível trabalhar de forma analítica, mas nem sempre a simplificação do mundo real está relacionada à simplicidade da implementação e a opção de utilização de ferramentas analíticas nem sempre é a mais aceitável. Como também é exposto por Almiron (2009), nesses casos, o uso de simulação estocástica é a ferramenta mais empregada.

Com isso e para que finalmente possamos responder ao questionamento sobre como construir um ambiente de simulação para o estudo proposto nesse trabalho, precisamos agora implementar os modelos apresentados. Nesse contexto, um significado para o termo “implementação” pode ser o interesse na representação de um modelo em estruturas que permitam sua análise e ilustrem seu funcionamento, através da geração de diversos

cenários, sejam esses típicos e/ou extremos. Dessa forma, e para deixar mais claro o que estamos querendo apresentar, um dos últimos ingredientes necessários para a construção do ambiente de simulação é o ambiente computacional. Sem ainda levar em consideração a abordagem Monte Carlo, último, mas não menos importante, ingrediente de nossa simulação, o ambiente computacional pode ser dividido em duas principais plataformas: a plataforma para realizar, de fato, as simulações, e a plataforma em que os resultados são analisados.

Pensando nos problemas abordados por Bustos e Frery (2005), Andel e Yasinsac (2006), Almiron, Almeida e Miranda (2009) e Almiron et al. (2010), especialmente os que dizem respeito aos erros originados pelo usuário (falta de conhecimento sobre as funções e configurações da plataforma de simulação) e aos erros oriundos da plataforma de análise (erro de arredondamento, de acumulação, de truncamento e de cancelamento, por exemplo), decidimos que a escolha dessas plataformas seria parte integrante, e muito importante, do nosso estudo.

Portanto, reservamos o restante desse capítulo para realizar comentários sobre essa escolha e aproveitamos para apresentar os principais resultados obtidos nos trabalhos de Almiron, Almeida e Miranda (2009) e Almiron et al. (2010).

#### 4.1.2.1 Plataforma de simulação

Partindo do contexto principal do nosso trabalho, a área de RSSFs, e conforme é apontado por Almiron (2009), atualmente existem diversos simuladores que podem ser utilizados para implementação de novos modelos. Na realidade, esses simuladores já têm diversos modelos implementados e que podem ser utilizados para auxiliar o estudo de novos modelos, seja com o objetivo de complementar a abstração de fases do estudo ou até mesmo com o objetivo de confrontar modelos distintos mas que são solução para um mesmo conjunto de desafios.

Contudo, em Andel e Yasinsac (2006) podemos encontrar uma observação importante: é impossível simular todos os aspectos e situações de uma rede móvel *ad hoc*. Sem perda de generalidade, isso também é válido para as RSSFs. Essa observação nos leva a constatar que a realidade é complexa e escolher um nível de abstração desejável, ou até mesmo “correto”, pode fazer com que os resultados obtidos em simulação não sejam válidos. Um resultado importante que corrobora essa ideia é indicado em Almiron (2009) e pode ser encontrado em mais detalhes em Cavin, Sasson e Schiper (2002). Nesse trabalho, são apontadas inconsistências nos modelos implementados em vários simuladores amplamente utilizados.

Como as abstrações que originam os modelos partem das ideias e estudos diretamente relacionados às decisões de usuários, uma maneira de minimizar esse fator de erro é a correta delimitação do escopo de estudo. No nosso trabalho, fazemos isso enfatizando que

o modelo proposto se aplica às RSSFs homogêneas, que a comunicação entre os nós sensores ocorre por múltiplos saltos e que o objetivo é a minimização do problema de *energy hole*. Além disso buscamos incluir, o quanto for possível, a preocupação com problemas que possam surgir a partir do modelo proposto.

Mesmo que seja possível tratar de todos esses aspectos, Almiron (2009) afirma ainda que mesmo que os resultados sejam válidos na simulações, inclusive tão válidos que não conseguiríamos diferenciá-los de resultados obtidos em experiências reais, podemos não estar livres de fontes de erro em nossas conclusões. Isso pode ocorrer porque por mais que um modelo esteja próximo da realidade, esse sempre será resultado de uma abstração e erros estarão associados, por menor que sejam. Entretanto um fator que pode ser crítico é a fase posterior à simulação: a análise dos resultados (análise dos dados). Como se tratam de análises estatísticas, torna-se necessária a utilização de uma plataforma com alto grau de precisão numérica. Veremos mais detalhes sobre esse outro fator na próxima seção.

#### 4.1.2.2 Plataforma de análise de dados

De posse dos resultados obtidos através de simulação, o próximo passo é aplicar técnicas estatísticas que forneçam respostas aos questionamentos produzidos durante a elaboração dos modelos implementados.

Como bem afirma Almiron (2009), a aplicação dessas técnicas implica em calcular reiteradamente funções estatísticas simples além de operações numericamente mais complexas. Erros numéricos de pequena magnitude nas funções estatísticas mais simples podem gerar conclusões totalmente erradas. Portanto, a precisão numérica deve ser um fator decisivo no momento de escolher uma plataforma para análise de dados.

Entretanto, a escolha dessas plataformas tem sido guiada por considerações mais operacionais, tais como a facilidade de leitura e gerenciamento de dados, e a capacidade de geração de relatórios. Por esta razão, frequentemente são empregadas planilhas eletrônicas para a análise dos resultados de simulação. Não que esses não sejam fatores importantes, mas entendemos que um trabalho que se propõe a fazer análise estatística responsável deve sempre levar em consideração, acima de outros fatores, a precisão numérica.

Diante dessa ideia, apresentamos um resumo das análises feitas em Almiron, Almeida e Miranda (2009) e Almiron et al. (2010).

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 e Wilson (1999), McCullough e Heiser (2008), Altman (2002), Altman, Gill e McDonald (2007), Bustos e Frery (2005) e Keeling e Pavur (2007).

Os trabalhos de Almiron, Almeida e Miranda (2009) e Almiron et al. (2010) mostram os resultados a aplicação desse protocolo a diversas plataformas de análise de dados de uso frequente, dentre essas plataformas estão as planilhas eletrônicas como `Open Office Calc`, `Microsoft Excel`, `Gnumeric`, `NeoOffice` e `GNU OLeo`, e algumas plataformas estatísticas mais completas como `Octave`, `Ox`, `Phyton` e `R`.

## 4.2 Avaliação

Nessa seção, detalhamos como conduzimos as nossas simulações a partir dos modelos apresentados e referenciados na Seção 4.1.1. Também é reforçada a escolha das plataforma de simulação de análise de dados, escolha essa detalhada na Seção 4.1.2, assim como são apresentados detalhes gerais sobre o ambiente de simulação como sistema operacional e ferramentas adicionais.

Além desses fatores, apresentamos os cenários de avaliação do modelo proposto no Capítulo 3, os quais servirão como base para a discussão dos resultados demonstrados no Capítulo 5.

### 4.2.1 Estudo Monte Carlo

Assim como para a formatação dos modelos utilizados neste trabalho a partir de algumas abstrações e tomando como base os trabalhos de Alencar-Neto (2007), Almiron (2009) e Frery et al. (2010), recorreremos a mais uma prática encontrada nesses trabalhos: o estudo Monte Carlo.

Assim como para esses trabalhos, e em outros encontrados na literatura, vemos que o método Monte Carlo tem se tornado uma ferramenta indispensável para o estudo de RSSFs, pelas razões já citadas ao longo do nosso trabalho, mas especialmente pelo fato de não ser trivial a aquisição de resultados puramente analíticos a partir das diversas formulações propostas para os estudos desse tipo de rede de comunicação.

No caso da formulação apresentada aqui, abrangendo abstrações e modelos voltados a região de monitoramento, a dispersão de nós sensores nessa área e as características operacionais desses elementos, estruturamos um estudo Monte Carlo utilizando um ou mais fatores para cada uma dessas variáveis.

Definida uma região de monitoramento com base nos conceitos apresentados nas Seções 2.2.1 e 4.1.1.1, ou seja, dada uma área circular de raio  $R = 50$  m, para cada um dos cenários de estudo (que detalharemos na Seção 4.2.2), e também com base nos

trabalhos de Alencar-Neto (2007) e, principalmente, Ramos et al. (2011), temos as seguintes características para os fatores do nosso estudo:

- a) quantidade de nós sensores é dada por  $n \in \{500, 700, 1000, 1500, 2000\}$ ;
- b) distribuição espacial dos nós sensores para o processo pontual  $I(n, \kappa)$  é dada tal que  $\kappa(R, r_c, a)$ , com  $a \in \{0, 0.3, 0.6, 1, 1.5, 2, 4\}$ ;
- c) raio de comunicação  $r_c \in \left\{ \frac{R}{5}, \frac{R}{4}, \frac{R}{3} \right\} \cong \{10, 12.5, 16.7\}$  m;
- d) raio de sensoriamento  $r_s \in \left\{ \frac{r_c}{5}, \frac{r_c}{4}, \frac{r_c}{3} \right\} \cong \{2.5, 3.12, 4.17\}$  m, para o caso em que  $r_c = 12.5$  m.

Para todas as comparações entre os fatores, foi utilizada uma quantidade de 100 replicações. Entendemos, em estudos preliminares, como sendo uma quantidade suficiente para ter boas estimativas.

Dessa forma, conforme mostraremos no Capítulo 5, esse estudo Monte Carlo tem como objetivo a aquisição de informação para que possamos avaliar o comportamento do modelo de distribuição de nós proposto sob a ótica principal do balanceamento de carga ao longo da área de monitoramento/dispersão.

## 4.2.2 Cenários de estudo

Para complementar o estudo Monte Carlo demonstrado na seção anterior (4.2.1) e iniciar a análise dos resultados de simulação, é importante primeiramente o entendimento de como será a abordagem de avaliação. Nessa Seção, ampliamos a visão sobre o estudo Monte Carlo, explicando mais alguns fatores importantes utilizados em nossas simulações e apresentando os cenários de avaliação, os quais são fundamentais para o entendimento dos resultados abordados no Capítulo 5.

### 4.2.2.1 Relação entre energia consumida e densidade de nós

Conforme relatamos no Capítulo 3, precisamos avaliar o modelo proposto de dispersão de nós sensores sob o ponto de vista da quantidade de energia despendida ao longo de um determinado período de tempo, já que temos com uma das hipóteses o balanceamento de carga. Portanto, era fundamental a utilização de um modelo de energia nas simulações.

Na Seção 5.1, observamos o comportamento do processo pontual  $I(n, \kappa)$  no que diz respeito à relação entre o gasto energético e a densidade de nós espalhados na região de monitoramento. Para esse cenário, fazemos um comparativo com os resultados obtidos no estudo de Li e Mohapatra (2007). Para isso, utilizamos o mesmo modelo de energia apresentado nesse trabalho. De forma simplificada, para cada nó sensor são considerados apenas os gastos energéticos provenientes das tarefas diretas de comunicação, ou seja, a

energia despendida com tarefas de transmissão e de recepção de dados, e da tarefa de sensoriamento. Gastos com “escuta ociosa” do meio de comunicação, por exemplo, são ignorados. Portanto, o modelo pode ser definido como:

$$E_{S_x} = \alpha \times \Theta \quad (15)$$

$$E_{T_x} = (\beta_1 + \beta_2 \times r_c^\Phi) \times \Theta \quad (16)$$

$$E_{R_x} = \gamma \times \Theta \quad (17)$$

em que  $E_{S_x}$  é a energia despendida com a tarefa de sensoriamento,  $E_{T_x}$  é a energia gasta com a tarefa de transmissão e  $E_{R_x}$  é o gasto energético relacionado a tarefa de recepção. Para todos os casos  $\Theta$  é a taxa de dados (medida em bits/segundo) transmitidos/recebidos por cada nós sensor. Assim como exposto por Li e Mohapatra (2007),  $r_c^\Phi$  é um fator utilizado para representar a perda decorrente do meio de comunicação (na literatura é mais conhecido como *path loss*) e está diretamente ligado ao poder de alcance de transmissão/recepção de um nó sensor. Tipicamente  $\Phi$  recebe valores 2 ou 4. Já  $\beta_1$ ,  $\beta_2$  e  $\gamma$  são constantes que podem assumir os seguintes valores, para os padrões atuais de tecnologias de nós sensores:

- a)  $\alpha = 60 \times 10^{-9}$  J/bit;
- b)  $\beta_1 = 45 \times 10^{-9}$  J/bit;
- c)  $\beta_2 = 10 \times 10^{-12}$  J/bit/m<sup>2</sup>, quando  $\Phi = 2$ ;
- d)  $\beta_2 = 0.001 \times 10^{-12}$  J/bit/m<sup>2</sup>, quando  $\Phi = 4$ ;
- e)  $\gamma = 135 \times 10^{-9}$  J/bit.

A partir das Equações 15, 16 e 17, definimos a energia total gasta por cada nó sensor durante um determinado período de tempo como sendo:

$$E_{total} = \sum_{i=1}^T E_{S_{x_i}} + E_{T_{x_i}} + E_{R_{x_i}} \quad (18)$$

em que  $i = \{1, 2, 3, \dots, T\}$  representa os períodos de tempo de uma simulação.

#### 4.2.2.2 Influência da cobertura

De acordo com estudos prévios e uma característica que pode ser observada com certa obviedade, sabemos que nosso modelo proposto possui forte influência na porção de área percebida pelo conjunto de nós com relação a região total de monitoramento. Assim, também se tornou importante a avaliação desse comportamento.

De maneira simples, confrontamos a taxa de cobertura provenientes da alternância de valores para os fatores quantidade de nós sensores  $n$ , intensidade  $a$  para a função  $\kappa$

do processo pontual  $I(n, \kappa)$  e raio de sensoriamento  $r_s$  (para esse último, simplificamos a definição de seu valor, tomando como referência o valor para o raio de comunicação  $r_c = 12.5$  m e seguindo uma diretriz exposta em Li e Mohapatra (2007)).

#### 4.2.2.3 Relação entre energia consumida e *sinkbetweenness*

Conforme já foi afirmado neste trabalho, algumas das bases para a construção da nossa proposição foram os conceitos relacionados a processos pontuais, principalmente o conjunto de conceitos que nos levaram a definição de processos pontuais não-homogêneos. Nesse contexto, apresentamos os processos  $C$  e  $M^2P^2$ .

Assim, um dos nossos objetivos é apresentar resultados para uma análise comparativa entre o modelo proposto  $I$  e esses modelos (podemos dizer que o processo  $C$  se tornou um caso particular do processo  $M^2P^2$ , portanto para a análise comparativa será abordado apenas o segundo processo).

Para que isso se tornasse possível, resgatamos um relação entre medidas apresentada em Ramos et al. (2011), principalmente no que se refere a métrica topológica denominada *sinkbetweenness*.

Ramos et al. (2011) fazem um paralelo entre medidas de centralidade e a energia gasta por um nó sensor devido as tarefas de comunicação, especialmente a tarefa relacionada ao papel de retransmissor que um nó sensor tem que realizar em uma rede de sensores sem fios organizada em múltiplos saltos de comunicação. Aqui o raciocínio é: em geral, quanto mais central um nó sensor está localizado mais dados esse terá que transmitir/retransmitir. Assim, uma parte do estudo desse trabalho é investigar se alguma medida de centralidade (*betweenness, eigenvector, closeness, degree, Google page rank, constraints, hubscore e authority*) está relacionada com o total de pacotes transmitidos pelos nós localizados a um salto de comunicação da estação base.

Além dos medidas listadas, em Ramos et al. (2011) há a proposição de uma medida denominada ***Sinkbetweenness***. Essa medida é representada por um subconjunto do conjunto total de caminhos de custo mínimo de comunicação possíveis entre todos os nós sensores em uma RSSF. Desse subconjunto fazem parte apenas os caminhos de comunicação que passam obrigatoriamente pela estação base (o nó *Sink*), já que todos os nós sensores precisam direcionar as informações sensoriadas para esse nó central.

Os resultados desse estudo mostram que, independente da localização geográfica da estação base na região de monitoramento, a métrica *sinkbetweenness* se apresenta como a mais representativa dentre as estudadas, além de apresentar uma alta correlação com a quantidade de pacotes transmitidos, e conseqüentemente, com o gasto de energia, confirmado a hipótese levantada pelos autores. Aproveitando esses resultados, utilizamos essa medida para realizar um paralelo entre o nosso modelo de dispersão proposto e um



modelo aplicado ao problema do *energy hole* (conforme já apresentamos, o processo pontual  $M^2P^2$ ).

Antes, porém, apresentamos resultados na Seção 5.3 sobre um cenário que procura discutir a relação existente, também, entre o *sinkbetweenness* e a quantidade de energia gasta pelos nós sensores de acordo com o modelo de energia utilizado para este trabalho (ver Seção 4.2.2.1).

#### 4.2.2.4 Análise comparativa entre processo $I$ e processo $M^2P^2$

Finalizando os cenários de estudo para avaliação do nosso modelo, chegamos a um dos cenários mais importantes: a análise comparativa com o modelo de dispersão de nós sensores  $M^2P^2$ .

Assim como para o modelo de dispersão  $I$ , nas nossas simulações envolvendo o processo  $M^2P^2$ , utilizamos os mesmo valores para os fatores do estudo Monte Carlo apresentado na Seção 4.2.1. A exceção fica por conta do parâmetro de intensidade do processo, em que os valores são  $a \in 1, 5, 10, 15, 30$ , respeitando os valores atribuídos a esse parâmetro nos estudos de Ramos et al. (2011).

Nesse cenário, além da comparação do consumo de energia feito através do paralelo com a medida de *sinkbetweenness*, apresentamos uma análise comparativa para taxa de cobertura.

### 4.2.3 Ferramentas computacionais

Finalmente, formulados os modelos, definida a estrutura da simulação por meio de um estudo Monte Carlo e apresentados os cenários para análise dos resultados, resumimos nesta seção as plataformas computacionais utilizadas.

Conforme já explicamos nesse trabalho, inclusive com detalhes na Seção 4.1.2 (ver também Apêndice A), nossa implementação tem como centro a plataforma computacional R (R CORE TEAM, 2013), versão 3.0.1.

Também foram utilizadas de maneira integrada o seguinte conjunto de ferramentas e plataformas:

- a) Ubuntu (CANONICAL LTD., 2013), versão 12.10 (Quantal Quetzal), sistema operacional;
- b) Sinalgo (DISTRIBUTED COMPUTING GROUP, 2007), versão 0.75.3, plataforma de simulação de algoritmos distribuídos ;
- c) Java (JAVA DEVELOPMENT TEAM, 2013), versão 7 update 21, plataforma de desenvolvimento;

- d) Eclipse (THE ECLIPSE FOUNDATION, 2013), versão 3.8.0, ambiente integrado de desenvolvimento de software.

#### 4.2.4 Detalhes de implementação

Para facilitar a compreensão das discussões feitas no capítulo a seguir (Capítulo 5), nós resumimos algumas suposições aceitáveis para classe de redes de sensores sem fios regidas por comunicação por meio de múltiplos saltos:

- a) cada nó sensor, nas simulações implementadas nesse trabalho, gera dados continuamente a uma taxa constante e envia a uma estação base comum, através de múltiplos saltos. Essa geração de dados representa o dado sensoriado;
- b) todos os nós sensores apresentam o mesmo raio de comunicação;
- c) a camada de acesso ao meio de comunicação (na literatura referenciada como *MAC layer*) é considerada ideal, isto é, a gerência do agendamento de transmissão é perfeita, não havendo perda de mensagens por eventos de colisão;
- d) para todas as replicações do estudo Monte Carlo, a RSSFs apresentada é considerada conexa, ou seja, todos os nós tem um caminho possível para transmitir seus dados sensorizados até a estação base;
- e) quanto ao roteamento das mensagens, cada nó sensor transmite sua medição ao nó vizinho mais próximo. Caso exista mais de um nó na vizinhança, um candidato é escolhido de forma aleatória.

## 5 RESULTADOS

Conforme já relatamos ao longo desse trabalho, a abordagem que decidimos utilizar, para as avaliações do problema de *energy hole* e da eficácia do modelo de dispersão de nós sensores proposto, foi a simulação. Essas simulações foram construídas seguindo as diretrizes de modelagem de uma RSSF apontadas em Alencar-Neto (2007) e posteriormente fundamentadas em Frery et al. (2010). A partir dessa modelagem, implementamos nossas avaliações através de uma abordagem Monte Carlo.

### 5.1 Energia Consumida Vs. Densidade de Nós

Nossa primeira avaliação tem o objetivo de confirmar alguns resultados apresentados em Li e Mohapatra (2007), o qual avalia, dentre outros aspectos, o impacto da densidade de nós no consumo de energia ao longo de uma RSSF. Para isso, ainda no trabalho citado, foram realizadas simulações com uma região de monitoramento medindo  $2000 \times 2000$  m, com o número de nós sensores distribuídos variando entre os valores  $n = \{500, 600, 700, 1000, 1500, 2000\}$  e com um tempo de duração de 2000 s. Para efeitos de análise comparativa entre a estratégia de dispersão de nós adotadas no nosso trabalho e a estratégia utilizada no trabalho de Li e Mohapatra (2007), foram realizadas algumas compatibilizações nas implementações das nossas simulações, obedecendo as características e medidas utilizadas no modelo apresentado na Seção 4.1.1.2 (Figura 14), além de utilizarmos os valores apresentados na Seção 4.2.1.

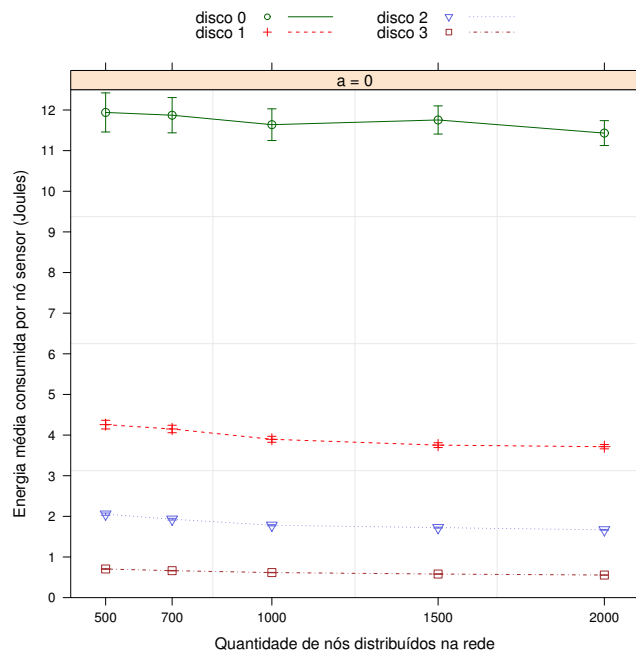
A Figura 22 – gráfico que confronta a energia média consumida pelos nós sensores, ao longo do tempo de simulação, com a densidade de nós presentes na região de monitoramento – mostra os resultados atingidos em nossas simulações.

O cenário utilizado é o caso em que a função de intensidade  $\kappa$ , do nosso modelo de distribuição  $I(n, g)$ , tem seu parâmetro de atratividade  $a$  configurado para o valor 0, ou seja, estamos tratando do caso em que os nós sensores são distribuídos de maneira aleatória e uniforme na região de interesse. Podemos concluir com a análise do gráfico que não existe influência marcante quando a densidade de nós sensores na rede aumenta, isto é, como previsto e confirmado em Li e Mohapatra (2007), a energia média consumida pelos nós sensores, em cada disco (representados pelas linhas, no gráfico), não tem alteração significativa. Portanto, o problema de *energy hole* não será resolvido, nem tão pouco minimizado, somente pelo fato de mais nós sensores serem alocados ao longo da região de monitoramento.

Além disso, podemos observar também que para esse cenário e para as características de densidade, de comunicação e de dispersão de nós utilizadas, comum em estudos de

RSSFs, o não balanceamento da energia consumida pelos nós sensores pertencentes a cada subregião delimitada pelos discos 0, 1, 2 e 3 é evidente. Isso pode ser melhor observado quando comparamos o consumo médio de energia dos nós alocados na região mais próxima da estação base, delimitada pelo disco 0, com o consumo médio de energia dos demais nós sensores da rede. O consumo é, em média, quase 3 vezes maior que o consumo de energia apresentado para os nós distribuídos na região delimitada pelo disco 1, em torno de 7 vezes maior quando a comparação é feita para os nós alocados na região demarcada pelo disco 2, e multiplicado por um fator de, aproximadamente, 28 na comparação com a energia consumida pelos nós alocados no disco 3, a região mais afastada da estação base.

Figura 22 – Impacto da densidade de nós no consumo de energia ao longo de uma RSSFs.



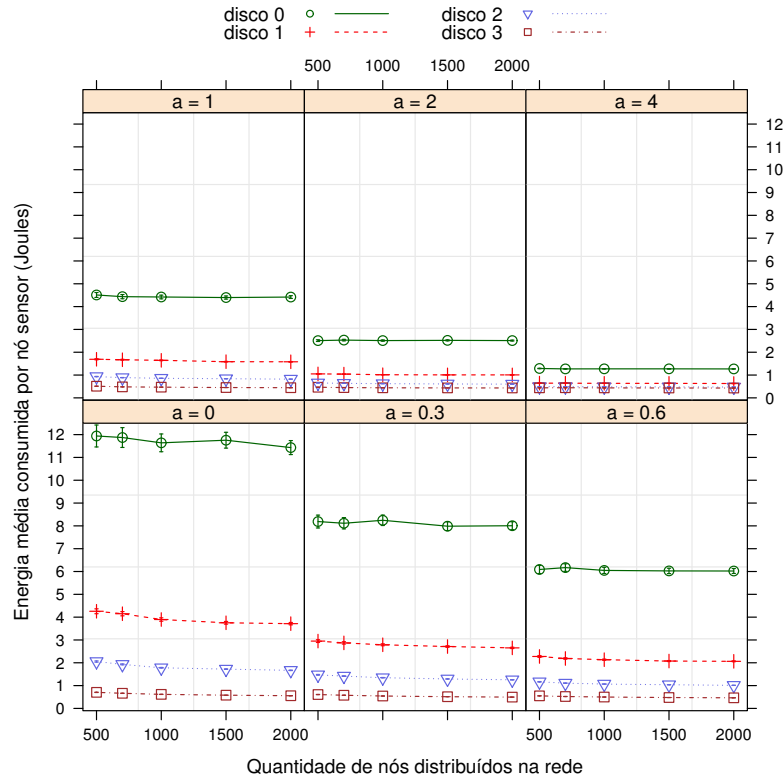
Fonte: Autor.

Diante dessa evidência e também pela hipótese levantada por Li e Mohapatra (2007), que em vez de aumentar a concentração de nós ao longo da área de monitoramento seria mais interessante avaliar o aumento da densidade de nós nas regiões mais próximas à estação base, construímos algumas simulações para analisar o mesmo confronto de informações da avaliação anterior, porém com o adicional proporcionado pelo modelo  $I$ .

Podemos observar os resultados obtidos na Figura 23, que apresenta seis gráficos, cada um representado os resultados de simulações obtidos para a variação do valor do parâmetro de atratividade  $a = \{0, 0.3, 0.6, 1, 2, 4\}$ . Para cada cenário, também foram utilizados os valores de 100 m para o diâmetro da região de monitoramento,  $r_c = 12,5$  m para o raio de comunicação dos nós sensores (também com o objetivo de dividir a região de monitoramento em quatro discos),  $n = \{500, 700, 1000, 1500, 2000\}$  para a quantidade

de nós sensores distribuídos e 2000 s para o tempo de simulação.

Figura 23 – Impacto da densidade de nós em regiões próximas à estação base no consumo de energia ao longo de uma RSSFs.



Fonte: Autor.

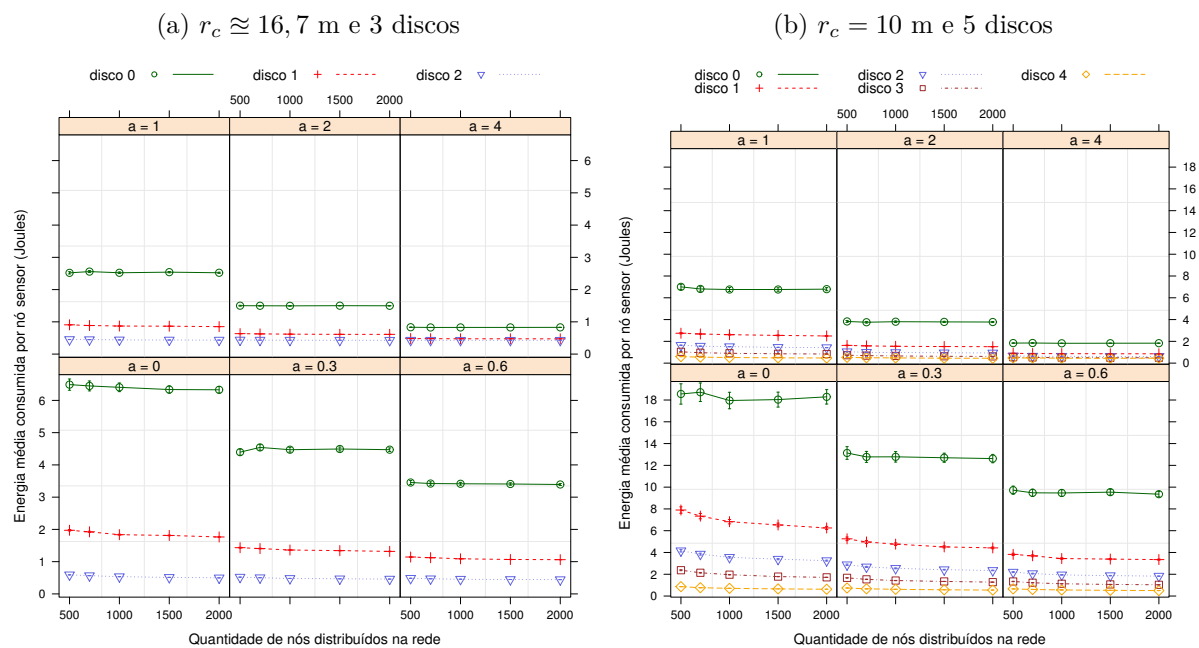
Apesar de observarmos o mesmo comportamento da Figura 22, que por sinal é o mesmo gráfico apresentado no retângulo indicado pelo valor  $a = 0$ , na barra superior de cada retângulo, os gráficos apresentam uma característica importante: quanto maior o valor do parâmetro de atratividade  $a$ , menor é a diferença entre o valor médio da energia gasta pelos nós em cada disco (indicados pelas linhas dos gráficos). Basta acompanharmos os gráficos na direção crescente dos valores do parâmetro de atratividade  $a$ . Comparando os gráficos  $a = 0$  e  $a = 0.3$ , por exemplo, já podemos ver que a energia média despendida pelos nós, no segundo gráfico, é menor para cada disco. No gráfico  $a = 0.6$ , a energia média consumida nos discos 0 e 1 é aproximadamente a metade da energia média despendida pelos nós sensores localizados nos mesmos discos para o caso apresentado pelo gráfico  $a = 0$ . Já para os gráficos  $a = 1$ ,  $a = 2$  e  $a = 4$ , as energias médias consumidas nas regiões demarcadas pelos discos 1, 2 e 3 são menores que a energia média despendida pelos nós alocados na região delimitada pelo disco 3, no caso em que  $a = 0$ .

Esses comportamentos têm um impacto significativo no tempo de vida de uma RSSFs, visto que os nós sensores “sobreviverão” uma quantidade maior de tempo. Esse impacto é ainda mais importante para os nós sensores distribuídos nas proximidades

da estação base, pois esses terão uma maior disponibilidade para realizar o papel de retransmissores de dados dos nós sensores localizados nos disco mais afastados.

Os gráficos da Figura 24 também podem ser interpretados da mesma maneira que os gráficos da Figura 23, porém aqui são avaliados diferentes valores de  $r_c$  e, conseqüentemente, diferentes quantidades de discos. Os gráficos da Figura 24a mostram a energia média consumida pelos nós sensores quando  $r_c \cong 16,7$  m e a região de monitoramento é dividida em 3 discos. Já os gráficos da Figura 24b apresentam a energia média despendida para  $r_c = 10$  m e para uma região de monitoramento particionada em 5 discos.

Figura 24 – Impacto da densidade de nós em regiões próximas à estação base no consumo de energia ao longo de uma RSSFs.



## 5.2 Cobertura

Conforme informamos na Seção 1.3, quando formulávamos a nossa proposta, uma das características do nosso estudo é que propomos um modelo de dispersão que aumenta a quantidade de nós sensores nas regiões mais próximas à estação base sem alterar a quantidade de nós distribuídos ao longo da rede, ou seja, dada uma quantidade de nós a ser distribuída, e como pretendemos oferecer uma alternativa aos modelos de dispersão de nós que organizam os nós sensores de maneira independente e uniforme, nós modelamos a organização dos nós de tal maneira que, na teoria, ao particionarmos a região de monitoramento em discos e ao propormos a utilização da função  $\kappa$  (apresentada na Seção 4.1.1.2) estamos atribuindo uma maior probabilidade dos nós serem depositados

nos discos mais próximos à estação base, com intensidade determinada pelo parâmetro de atratividade  $a$ .

Na prática, a alteração desse parâmetro pode ser interpretada de várias maneiras. Uma delas é que o modelo proporciona a subtração de alguns nós sensores das regiões mais afastadas da estação base e a posterior adição desses, de forma controlada, nas regiões mais próximas à estação base.

Diante dessa visão, sabemos que um dos fatores que contribuem para o comportamento apresentado nas Figuras 23 e 24 (na seção anterior - 5.1) está relacionado a essa subtração de nós nos discos que demarcam as regiões mais afastadas da estação base. Com isso, para valores maiores do parâmetro de atratividade  $a$  teremos cada vez menos nós sensores nessas regiões, fazendo com que:

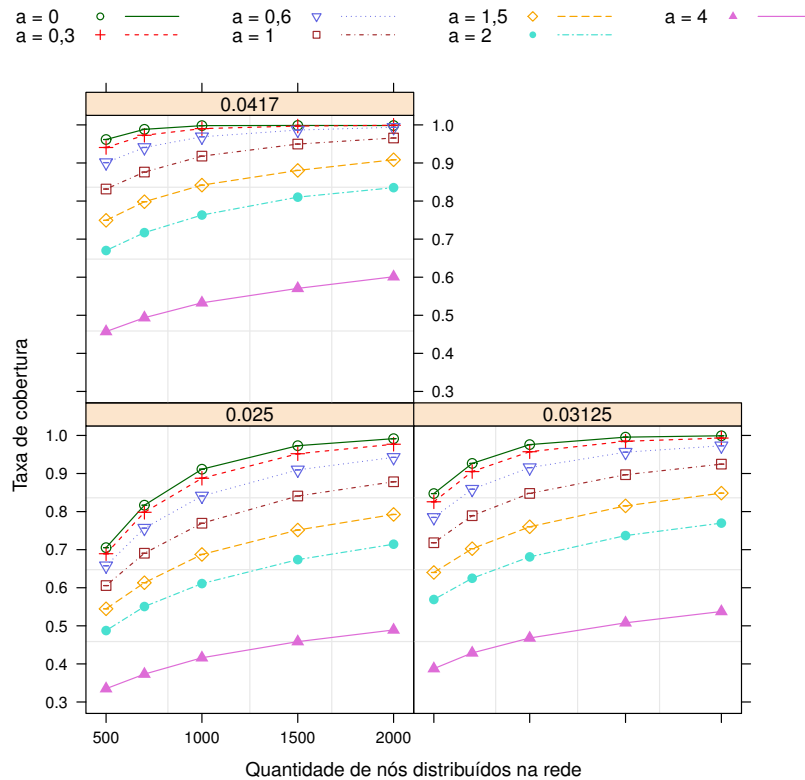
- a) cada vez menos dados sejam gerados nesses discos, porém com a preservação do gasto médio de energia originado pelas tarefas de sensoriamento;
- b) exista a diminuição da carga de mensagens a serem transmitidas e, consequentemente, diminuindo o gasto médio de energia promovido pela tarefa de retransmissão.

Entretanto, como os nós subtraídos são alocados em posições cada vez mais próximas da estação base, para valores maiores do parâmetro de atratividade  $a$ , o modelo  $I$  produzirá uma concentração de nós cada vez maior nessas regiões. Podemos observar que:

- a) apesar da informação medida ser maior nessas regiões, o consumo médio de energia proveniente da atividade de sensoriamento também não sofre alteração;
- b) como há uma quantidade maior de nós disponíveis para promover a tarefa de retransmissão nos discos subsequentes, os nós mais afastados da estação base têm, em média, uma quantidade maior de vizinhos e a consequência direta disso é, também, a diminuição do gasto médio de energia com retransmissão em todos os discos.

Partindo dessas observações, mostramos que é possível atingir o objetivo de minimizar o problema do *energy hole* e, principalmente, minimizar a questão relacionada ao consumo não balanceado de energia. Entretanto, embora existam indícios fortes dos benefícios para o prolongamento do tempo de vida de uma RSSFs, nossa abordagem também tem uma influência em uma característica importante: a cobertura da região de monitoramento. Tomando como base as definições de raio de sensoriamento,  $r_s$ , abordadas na Seção 4.1.1.3 e entendendo como cobertura (ou área de cobertura do nós sensores) o somatório de todas as áreas de sensoriamento  $A_i$  delimitadas por cada nó sensor  $S_i$  presente na região de monitoramento, apresentamos os gráficos da Figura 25.

Um resultado bastante óbvio fica por conta do aumento de densidade ao longo da área de interesse. É fácil perceber que o aumento de densidade de nós sensores afeta

Figura 25 – Impacto do processo  $I$  na cobertura da região de monitoramento.

Fonte: Autor.

o grau de cobertura da região pois teremos, em média, uma quantidade maior de nós em qualquer porção (os discos que delimitam as regiões de salto de comunicação, por exemplo) da área de monitoramento. Para todas as experimentações de valores de  $r_s$  utilizados (representados pelos valores nas barras superiores dos três quadros), temos um comportamento similar para as curvas que representam as variações dos valores de intensidade do parâmetro  $a$ , isto é, quanto maior a densidade, maior a porção da região que é percebida conjuntamente pelos nós sensores.

Outro aspecto que merece destaque é a avaliação tomando como perspectiva o parâmetro de intensidade do processo de dispersão. Podemos observar que, independente da variação do valor de raio de sensoriamento  $r_s$ , o aumento do valor do parâmetro  $a$  reflete na diminuição da taxa de cobertura da rede.

Para explicar esse resultado, torna-se importante relembrar alguns cenários de aplicação do processo  $I$ , como os apresentados na Seção 3.3. Tomando como base os exemplos das Figuras 26a, 26b e 26c, percebemos que com a quantidade de nós sensores fixa, o acréscimo do valor do parâmetro  $a$  faz com que os nós sensores fiquem cada vez mais concentrados no centro da área de monitoramento, ou seja, próximos a estação base. O resultado é um crescimento da porção não percebida pelos nós sensores nas regiões de fronteira, o que naturalmente não é recuperado pelos nós sensores que estão aglomerados,



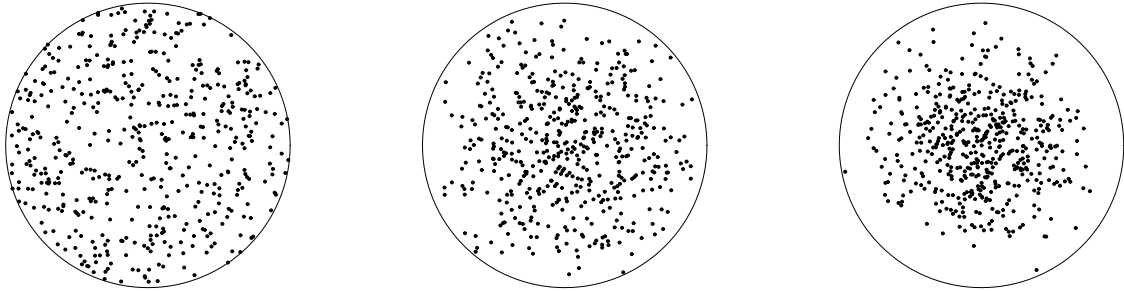
pois esses passam a medir regiões redundantes.

Figura 26 – Comportamento da dispersão de nós sensores quando aplicado o processo  $I$ , para o parâmetro de intensidade  $a = \{0, 1, 2\}$  e quantidade de nós  $n = 500$ .

(a)  $a = 0$  e  $n = 500$

(b)  $a = 1$  e  $n = 500$

(c)  $a = 2$  e  $n = 500$



Fonte: Autor.

Adicionalmente, se estabelecermos um limiar de, por exemplo, 70% para a taxa de cobertura, perceberemos que uma faixa de valores úteis para o parâmetro de intensidade é o intervalo  $0 \leq a \leq 1$ , principalmente com relação a um raio de sensoriamento com valor 0,3125, o qual corresponde ao caso em que a região de interesse está dividida em quatro discos, o nosso cenário base.

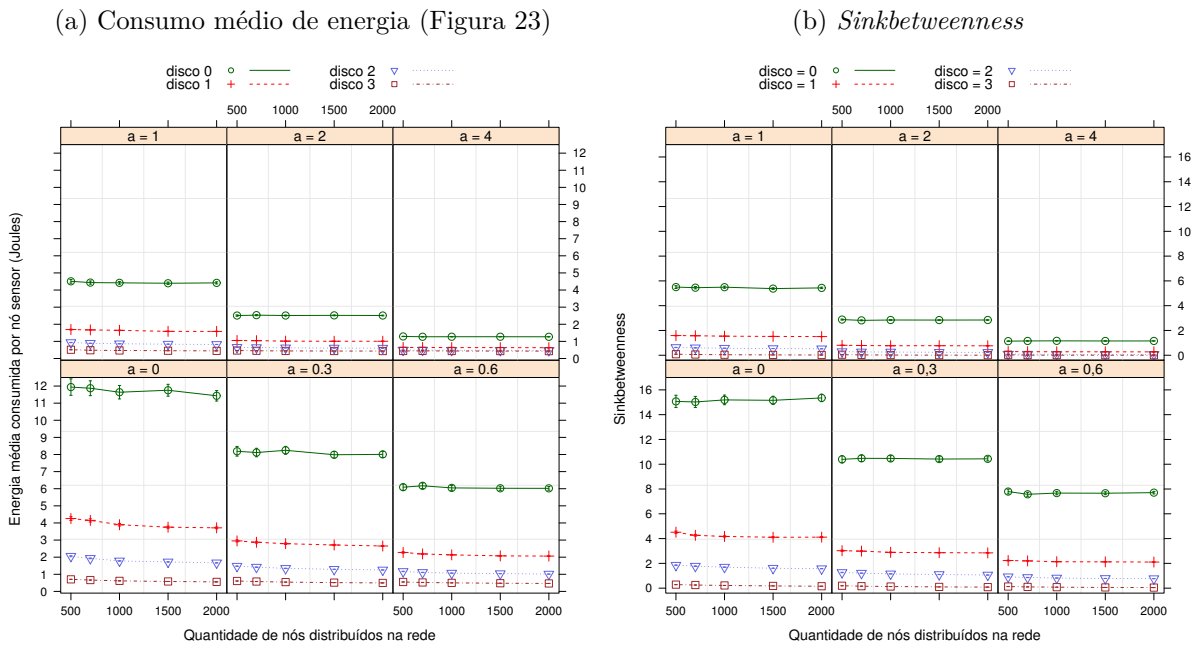
### 5.3 *Sinkbetweenness*

Antes de iniciarmos a análise comparativa entre Processo  $I$  e Processo  $M^2P^2$ , vejamos os gráficos exibidos na Figura 27.

Podemos perceber visualmente (independente do confronto de valores apresentados para o consumo médio de energia e para o *sinkbetweenness*, pois aqui a comparação faz sentido quando analisamos o comportamento das curvas dos gráficos, diante da alteração de valores para os fatores de intensidade  $a$ , de quantidade de nós  $n$  e de quantidade de saltos para a estação base) que, como já apontado por Ramos et al. (2011) no que diz respeito a forte correlação entre a medida de quantidade de mensagens transmitidas por um nó sensor e a medida de centralidade *sinkbetweenness*, existe também uma correlação entre as medidas de consumo médio de energia por nó sensor e em suas respectivas subregiões de distribuição (discos que representam os saltos – distâncias – até a estação base), e a medida de *sinkbetweenness*.

Confirmamos isso por meio dos gráficos exibidos nas Figuras 28 e 29. Para mais dois exemplos de configurações, com nós sensores configurados com raios de comunicação

Figura 27 – Avaliação do resultado da aplicação do processo I sob a ótica do consumo médio de energia e do *sinkbetweenness*.



Fonte: Autor.

$r_c = 10$  m e  $r_c \cong 16,7$  m, podemos ver o mesmo comportamento e a mesma indicação de correlação.

Portanto, assim como em Ramos et al. (2011), em que a medida de *sinkbetweenness* foi utilizado como forma de caracterizar a tarefa de retransmissão de um nó sensor, nós a utilizaremos com sentido similar, permitindo assim que nós possamos discutir alguns resultados da análise comparativa entre o processo  $I$  e o processo  $M^2P^2$ .

Na seção seguinte são abordados resultados e discussões relacionados a essa análise comparativa.

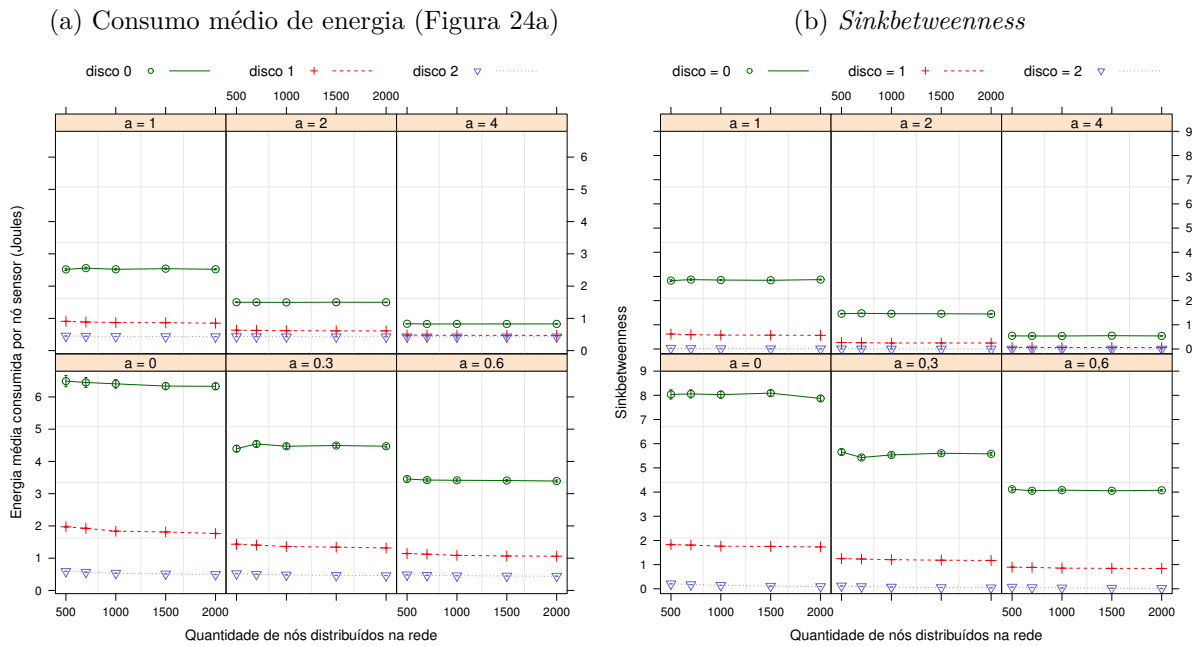
## 5.4 Análise comparativa entre Processo $I$ e Processo $M^2P^2$

### 5.4.1 *Sinkbetweenness*

Nessa seção, faremos uma análise comparativa entre as medidas de *Sinkbetweenness* entre os processo  $I$  e  $M^2P^2$ . Para isso, vejamos os gráficos presentes nas Figuras 30a (mesma que a Figura 27b) e 30b.

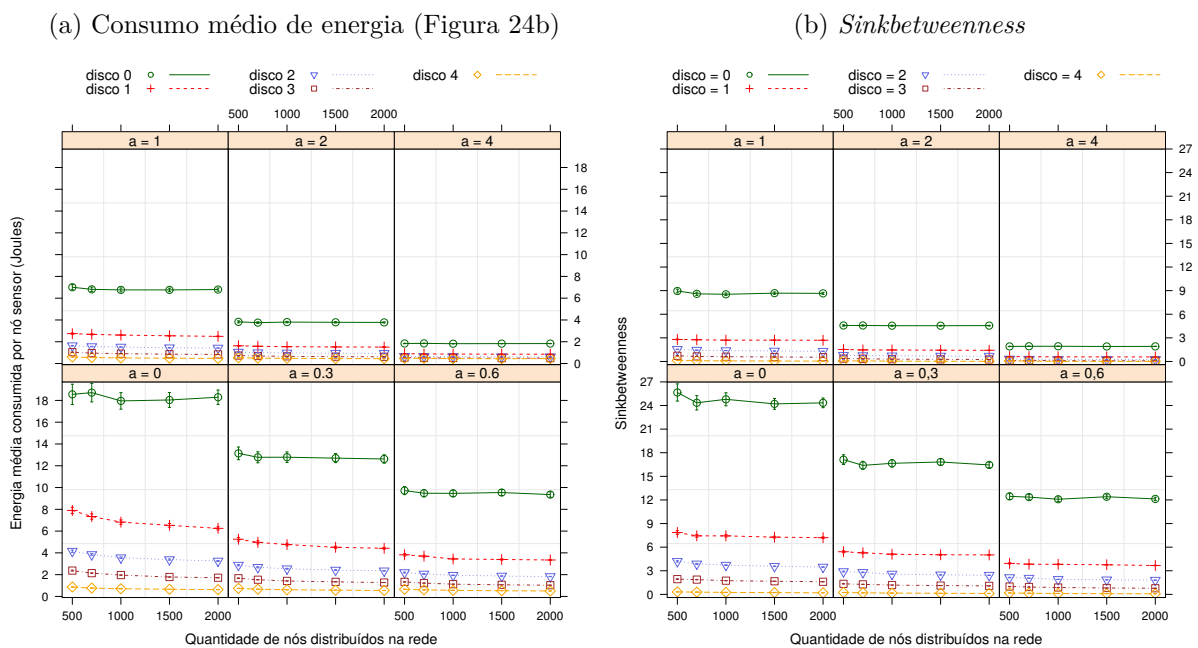
Primeiramente, é importante lembrar que os parâmetros de atratividade  $a$  têm influências diferentes para as respectivas funções de intensidade dos processos em questão. Dessa forma, os valores utilizados na construção das simulações são bem diferentes. Procuramos respeitar, para o processo  $M^2P^2$ , os valores apresentados pelo estudo de Ramos

Figura 28 – Avaliação do resultado da aplicação do processo  $I$  sob a ótica do consumo médio de energia e do *sinkbetweenness*, para uma região particionada em 3 discos de saltos de comunicação.



Fonte: Autor.

Figura 29 – Avaliação do resultado da aplicação do processo  $I$  sob a ótica do consumo médio de energia e do *sinkbetweenness*, para uma região particionada em 5 discos de saltos de comunicação.



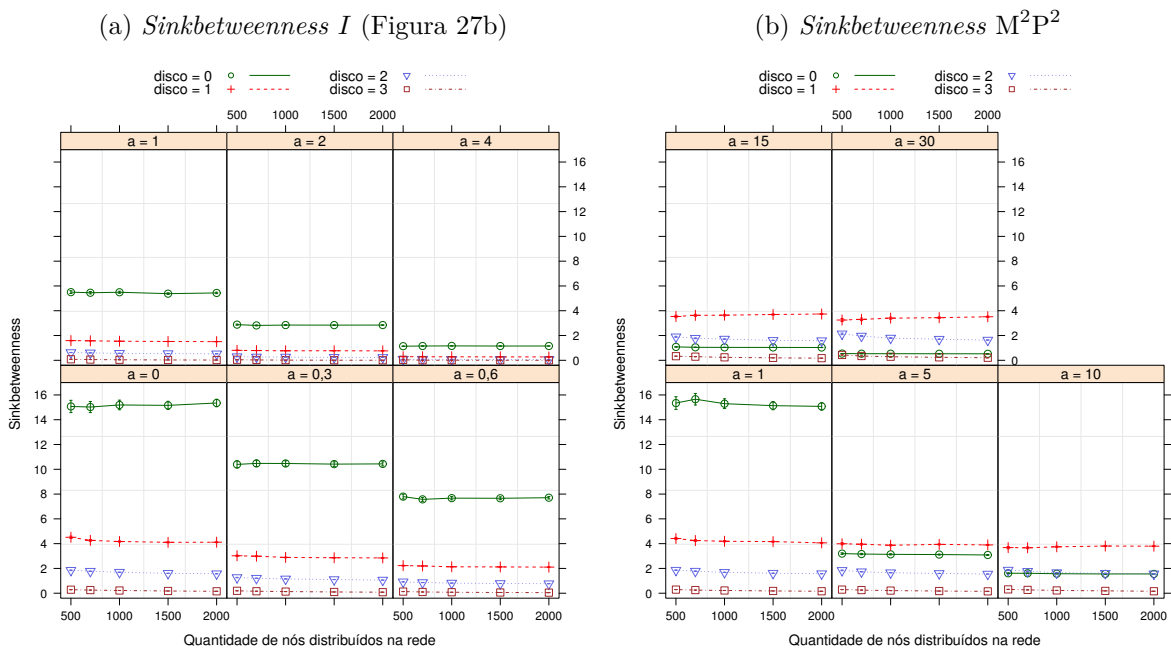
Fonte: Autor.

et al. (2011). Ao contrário do que possa parecer, somente dessa maneira podemos tornar os processos comparáveis entre si.

Diante de todos os fatos que já foram apresentados até o momento, neste trabalho, entendemos que um fator importante para a minimização do problema de *energy hole* é a aproximação da quantidade de energia média consumida (aqui falaremos em *sinkbetweenness*) entre as subregiões de comunicação entre os nós sensores espalhados ao longo de uma área de monitoramento, ou seja, o desejado balanceamento de carga. Percebemos, conforme explanado na Seção 5.1, que esse comportamento existe para o processo pontual não-homogêneo  $I$ . Podemos observar na Figura 30b que o objetivo também é alcançado pelo processo  $M^2P^2$ . De fato, os resultados representados por esses gráficos se tratam apenas de uma confirmação do que já havia sido bem apresentado em Ramos et al. (2011).

O que mais se destaca quando comparamos os resultados dos gráficos entre os processos, é que, aparentemente, o processo  $M^2P^2$  atinge esse objetivo de maneira mais rápida, já que no caso em que  $a = 5$  já podemos observar um valor médio de *sinkbetweenness* para a região mais próxima da estação base em um patamar bastante próximo dos valores médios de *sinkbetweenness* para as demais regiões.

Figura 30 – Análise comparativa do *sinkbetweenness* para o processo  $I$  e o processo  $M^2P^2$ .



Fonte: Autor.

Um comportamento que merece destaque também, ainda para o processo  $M^2P^2$ , é que a linha que representa a região a um salto da estação base começa a se posicionar entre as linhas que representam as outras regiões. Isso se explica devido ao fato do processo de adição de nós sensores só acontecer em uma região (que é justamente essa região a um salto da estação base, representada pelo disco 0). É importante perceber que as

outras regiões são afetadas com a subtração de nós sensores, mas os valores médios para o *sinkbetweenness* não acompanham um mesmo fator de subtração na mesma proporção (na verdade a diminuição do valor médio do *sinkbetweenness* nas demais regiões é quase imperceptível), deixando para a região mais próxima da estação base o trabalho de reduzir o consumo de energia da rede. Com isso, é provável que com o aumento do parâmetro de intensidade  $a$  o problema de *energy hole* seja transferido para as regiões seguintes. Isso significa que a estratégia utilizada pelo processo  $M^2P^2$  pode funcionar até um determinado valor de  $a$ , não tendo influência significativa para o tempo de vida da rede para valores acima desse limiar.

Se construirmos a mesma linha de raciocínio para a alternativa que propomos, o processo  $I$ , percebemos que a aproximação das linhas de valores médios para o *sinkbetweenness* acontece de maneira mais lenta que o processo  $M^2P^2$ . Na realidade, é provável que um comportamento similar ao apresentado nos gráficos da Figura 30b nunca chegue a acontecer para o processo  $I$ , por mais que a intensidade do processo seja elevada. Isso se deve ao comportamento apresentado justamente quando os valores do parâmetro de atratividade  $a$  são crescentes. É possível observar na Figura 30a que há uma diminuição significativa nos valores médios de *sinkbetweenness* para todas as subregiões de comunicação da área de monitoramento. Isso nos remete a uma possível conclusão de que, em relação ao tempo de vida da rede, o processo  $I$  pode ter uma influência mais positiva que o processo  $M^2P^2$ .

Resultados similares podem ser vistos para diferentes configurações de raio de comunicação  $r_c$ , nas Figuras 31, com  $r_c \cong 16,7$  m, e 32, com  $r_c = 10$  m.

Contudo, conforme apresentamos na Seção 5.2, o processo  $I$  também apresenta limitações para valores maiores do parâmetro de intensidade. Portanto, ainda é necessário realizar uma comparação entre os comportamentos dos dois processos do ponto de vista da cobertura total da região de dispersão dos nós.

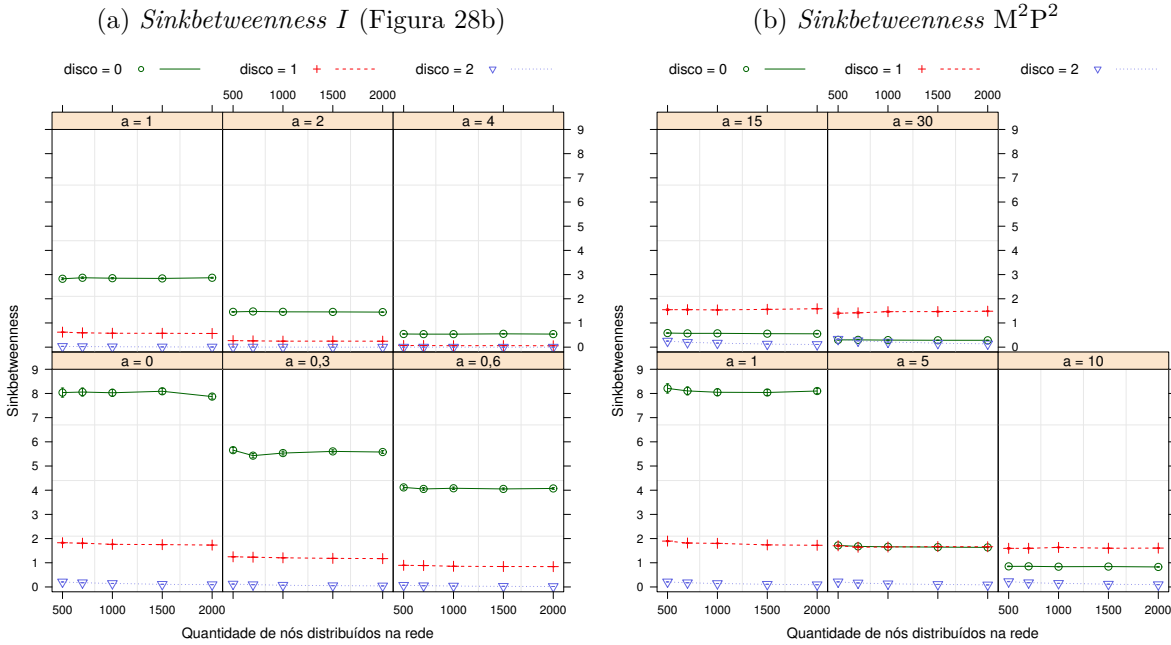
Vejamos essa análise na seção a seguir (5.4.2).

## 5.4.2 Cobertura

Para tratarmos da análise comparativa para os resultados de cobertura diante da variação da quantidade de nós e do valor do parâmetro de intensidade, e para alguns valores de raio de sensoriamento, vejamos primeiramente os gráficos da Figuras 33a e 33b, em que o raio de sensoriamento de cada nós sensor é  $r_s \cong 3,12$  m.

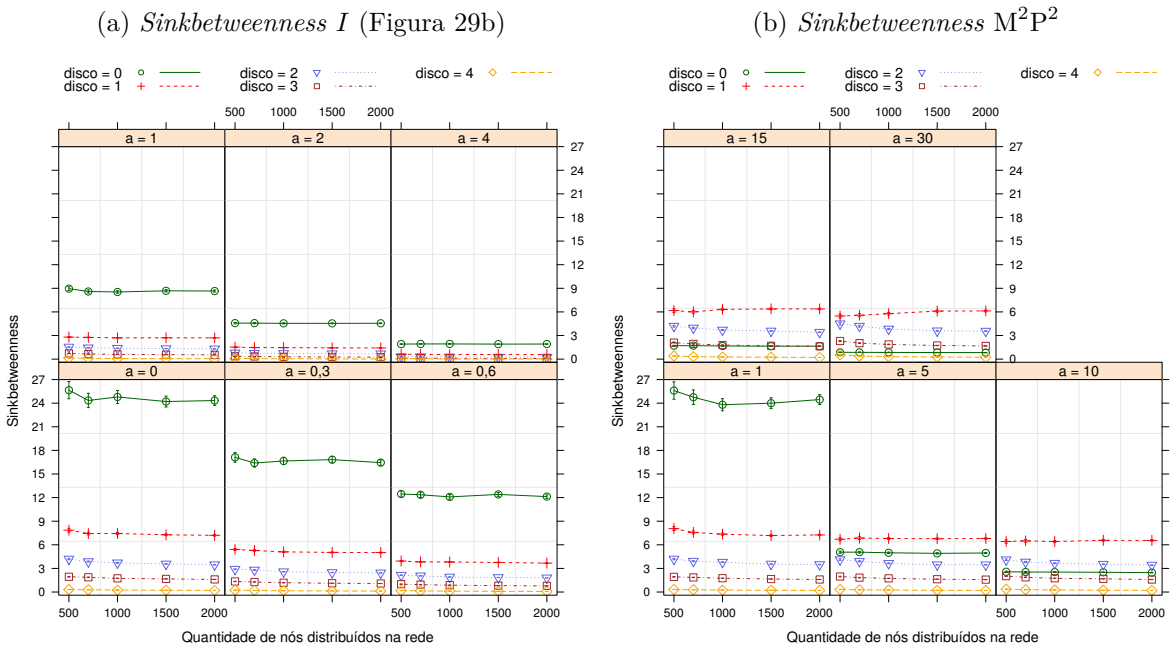
Tomando como base a variação da densidade de nós ao longo da área de monitoramento, trata-se de uma característica que afeta ambos os processos, em termos de cobertura. Entretanto, podemos perceber na Figura 33b que quanto maior a quantidade de nós sensores presente na região de monitoramento, ou seja, quanto mais densa for a rede, maior o impacto positivo para a cobertura da rede, no caso do processo  $M^2P^2$ . Isso

Figura 31 – Análise comparativa do *sinkbetweenness* para o processo  $I$  e o processo  $M^2P^2$ , para uma região particionada em 3 discos de saltos de comunicação.



Fonte: Autor.

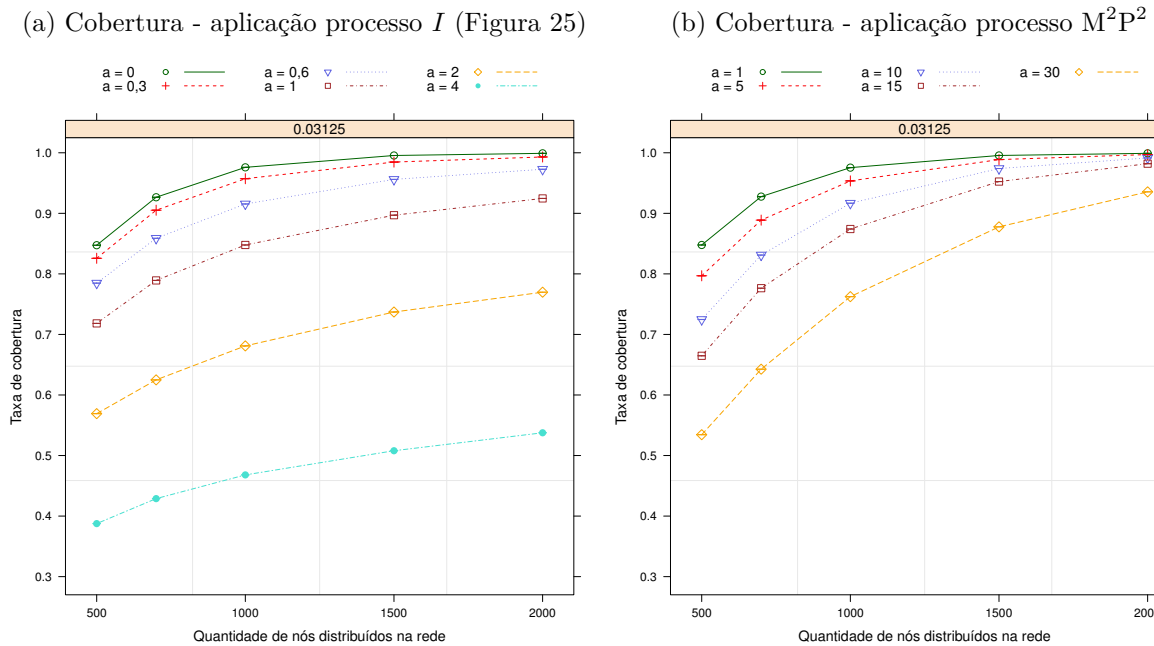
Figura 32 – Análise comparativa do *sinkbetweenness* para o processo  $I$  e o processo  $M^2P^2$ , para uma região particionada em 5 discos de saltos de comunicação.



Fonte: Autor.

acontece devido a, já falada, característica do processo de "subtração" dos nós sensores em todas as regiões, com exceção da região representada pelo disco 0, a qual "receberá"

Figura 33 – Análise comparativa para o impacto do processo  $I$  e do processo  $M^2P^2$  na cobertura da região de monitoramento.



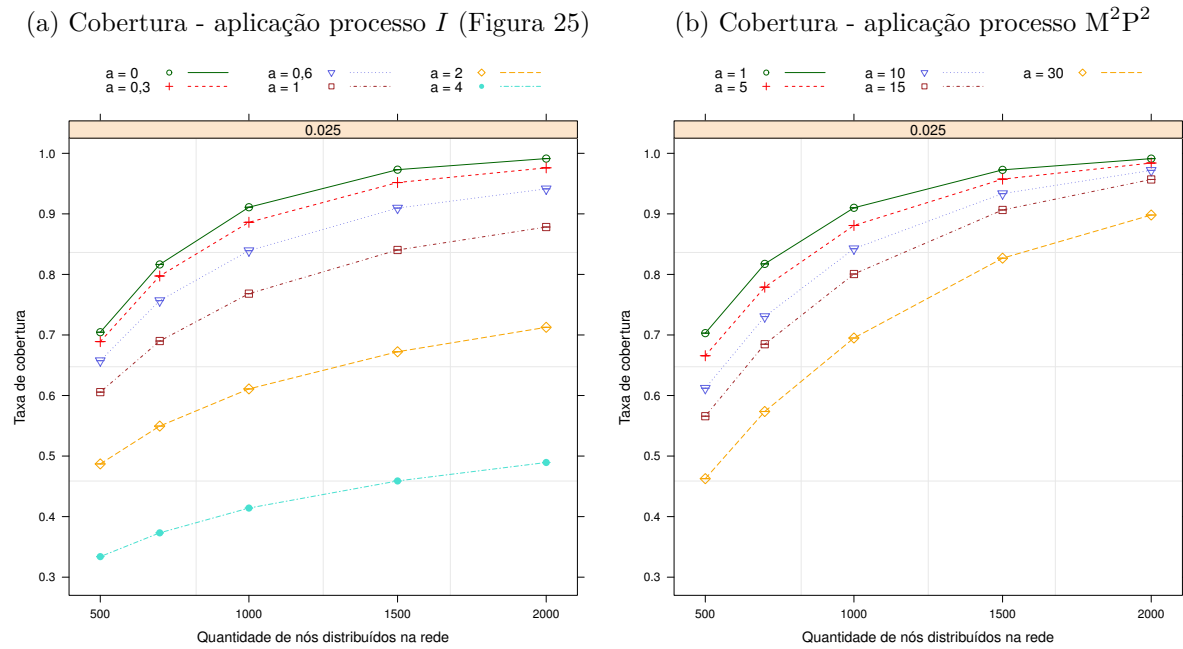
Fonte: Autor.

esses nós sensores "subtraídos". Quanto mais nós sensores estiverem presentes na rede, mais áreas cobertas existirão. Vejamos, por exemplo, os casos em que a quantidade de nós  $n = 1000$  e  $n = 2000$ . Para a primeira quantidade, as taxas médias de cobertura entre os casos em que  $a = 1$  (aleatório uniforme) e  $a = 30$  são, aproximadamente, 0,85 e 0,53, ou seja, um diferença de 0,32. Já para  $n = 2000$ , temos uma diferença de, aproximadamente, 0,08.

Esse é um comportamento bem diferente do apresentado na Figura 33a, para o processo  $I$ . Podemos observar que não existe uma influência significativa do acréscimo de nós ao longo da rede, para a taxa de cobertura. Tomemos como exemplo, assim como no caso anterior, os mesmos valores de  $n = 1000$  e  $n = 2000$ . Quando  $m = 1000$  vemos uma diferença de, aproximadamente, 0,46 para os casos de  $a = 0$  (aleatório uniforme) e  $a = 4$ . Já para o segundo caso, percebemos uma diferença de, aproximadamente, 0,47, para os mesmos valores do parâmetro de intensidade  $a$ . Esse comportamento pode ser explicado pelo fato dos nós sensores serem distribuídos cada vez mais (têm uma probabilidade maior de serem posicionados) próximos da estação base. Com isso, conforme já explicamos, as regiões mais afastadas da estação base ficarão com cada vez menos nós sensores, o que deixará essa região (representada pelo disco 4, nas figuras) cada vez menos percebida (sensoriada).

Esse comportamento nos remete ao que mais se destaca nessa análise comparativa: a influência do parâmetro de atratividade. É notória a percepção do melhor comportamento

Figura 34 – Análise comparativa para o impacto do processo  $I$  e do processo  $M^2P^2$  na cobertura da região de monitoramento, para um raio de sensoriamento  $r_s = 2,5m$ .



Fonte: Autor.

do processo  $M^2P^2$  em relação ao processo  $I$ , em termos gerais. Entretanto, diante de tudo que já relatamos até aqui, principalmente neste Capítulo e com as análises que foram feitas, é preciso lembrar que se tratam de parâmetros de intensidade com comportamentos e significados diferentes. Vejamos, por exemplo, que em termos gerais o processo  $M^2P^2$  apresenta melhores taxas de cobertura a medida que o valor do parâmetro de intensidade cresce, porém até os valor de  $a = 1$ , para o processo  $I$ , e  $a = 10$ , para o processo  $M^2P^2$ , vemos taxas de cobertura bastante aproximadas.

Finalmente, quanto aos diferentes valores de raio de sensoriamento  $r_s = 2,5m$  e  $r_s \approx 4,17$ , precisamos fazer uma análise das Figuras 34 e 35.

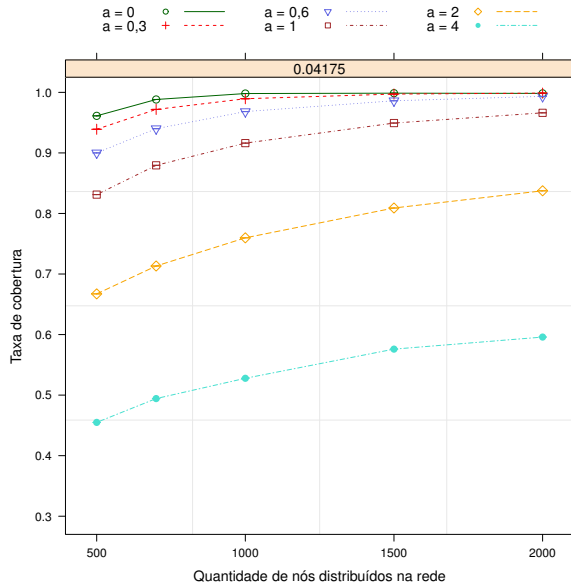
Podemos perceber que há bastante similaridade ao que foi descrito anteriormente, para o caso em que  $r_s \approx 3,12m$ , inclusive para o intervalo do parâmetro de intensidade  $a \leq 1$ , para o processo  $I$ , e  $a \leq 10$ , para o processo  $M^2P^2$ .

A exceção fica por conta das taxas médias de cobertura serem, em geral, menores quando o raio de sensoriamento é menor ( $r_s = 2,5m$ ) e maiores quando o raio de sensoriamento é maior ( $r_s \approx 4,17m$ ), o que, de certa maneira, já era um resultado esperado, visto que com um raio de sensoriamento menor, a capacidade (área) de percepção individual de cada nós sensor  $S_i$  presente na rede também será menor, aumentando consideravelmente a probabilidade de serem apresentadas áreas não percebidas por nenhum nó sensor.

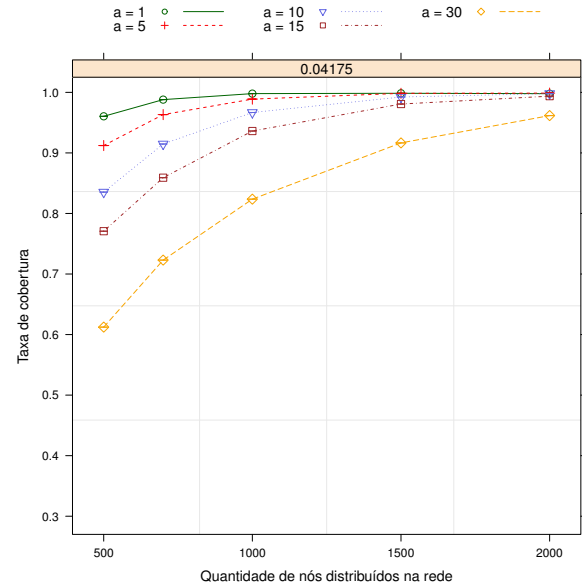


Figura 35 – Análise comparativa para o impacto do processo  $I$  e do processo  $M^2P^2$  na cobertura da região de monitoramento, para um raio de sensoriamento  $r_s \cong 4,17m$ .

(a) Cobertura - aplicação processo  $I$  (Figura 25)



(b) Cobertura - aplicação processo  $M^2P^2$



Fonte: Autor.



## 6 CONCLUSÃO

Dentre as motivações que nos levaram a discussões deste trabalho, buscamos destacar as RSSFs e o crescimento da importância desta ferramenta de aquisição de informações. Entendemos que essa característica é o ponto principal para a geração de interesse de áreas comerciais e acadêmicas.

Dando ênfase ao interesse acadêmico, na Seção 2 abordamos as características inerentes às RSSFs e como essas trazem diversos desafios. Buscamos concentrar nossa discussão da problemática de energia, tomando como base o problema do *energy hole*. Diante dessa abordagem, procuramos estabelecer um vínculo com uma das principais fontes de contribuição para o surgimento desse problemas: a localização dos nós.

Na Seção 3, propusemos um modelo de dispersão de nós capaz de traduzir a nossa discussão, além de permitir a avaliação do consumo de energia da rede e de contribuir, de fato, para um maior balanceamento de carga ao longo da região de monitoramento (com consequência direta no aumento do tempo de vida da rede).

Esse fato é abordado utilizando um conjunto de técnicas, das quais merecem destaques as encontradas nos trabalhos de Li e Mohapatra (2007), Alencar-Neto (2007), Frery et al. (2010), Ramos et al. (2011).

Diante desse contexto, apresentamos um experimento Monte Carlo na Seção 4 e discutimos alguns resultados na Seção 5.

Nossa percepção, no que diz respeito à análise comparativa entre os processos  $I$  e  $M^2P^2$ , é a dificuldade em afirmar qual é o melhor processo, tomando como referência o consumo de energia da rede. São abordagens diferentes que têm pontos fortes e fracos quando comparadas entre si. Se tomarmos como exemplo os resultados referentes à energia, veremos o processo  $M^2P^2$ , de maneira geral, convergindo mais rapidamente para o objetivo de redução no consumo de energia na região que abriga nós sensores a um salto de comunicação da estação base. Isso se deve ao fato de sua estratégia atuar especificamente nessa região, enquanto o processo  $I$  se utiliza de uma estratégia em que todas as regiões de comunicação são consideradas. Esse ponto reflete uma vantagem desse processo: há uma redução de energia em todas as regiões de comunicação, fazendo com que afirmação de redução de consumo médio de energia ao longo da rede seja válida.

Outro aspecto dessa análise comparativa está relacionada à questão da cobertura. No processo  $I$ , a estratégia está diretamente conectada à taxa de cobertura da região de monitoramento. É fácil perceber nos gráficos que para os casos em que o processo tem uma intensidade maior (a elevação do parâmetro de intensidade  $a$ ), a cobertura decresce em uma taxa mais rápida que o processo  $M^2P^2$ .

Acreditamos que algumas abordagens de trabalhos futuros que podem ser apontadas para reforçar os resultados alcançados no nosso trabalho são:

- a) Avaliação da efetividade dos dados sensoriados através do uso de técnicas de reconstrução de sinal, aproveitando o modelo de descrição de RSSFs apresentado por Alencar-Neto (2007), Frery et al. (2010);
- b) Análise da cobertura por região de comunicação (discos); nesta dissertação foi abordado apenas a cobertura geral da região de monitoramento;
- c) Uma RSSFs que contemple características de *duty cycle*, como discutido em Wu, Chen e Das (2008);
- d) Um diferente modelo de energia;
- e) Um modelo de roteamento alternativo.

## REFERÊNCIAS

- AKKAYA, K.; DEMIRBAS, M.; AYGUN, R. S. The impact of data aggregation on the performance of wireless sensor networks. *Wireless Communications & Mobile Computing*, v. 8, n. 2, p. 171–193, fevereiro 2008.
- AKYILDIZ, I. F. et al. A survey on sensor networks. *IEEE Communications Magazine*, v. 40, n. 8, p. 102–114, agosto 2002.
- AL-KARAKI, J. N.; KAMAL, A. E. Routing techniques in wireless sensor networks: A survey. *IEEE Wireless Communicatons*, v. 11, n. 6, p. 6–28, dezembro 2004.
- ALENCAR-NETO, J. *Estimação do Erro em Redes de Sensores sem Fios*. Dissertação (Mestrado em Modelagem Computacional de Conhecimento) — Universidade Federal de Alagoas, Maceió, AL, 2007.
- ALMIRON, M. *Estudos sobre a Conectividade em Redes de Sensores Sem Fios: Análise de Plataformas e Resultados de Percolação no Plano Contínuo*. Dissertação (Mestrado em Modelagem Computacional de Conhecimento) — Universidade Federal de Alagoas, Maceió, AL, 2009.
- ALMIRON, M.; ALMEIDA, E. S.; MIRANDA, M. The reliability of statistical functions in four software packages freely used in numerical computation. *Brazilian Journal of Probability and Statistics*, v. 23, n. 2, p. 107–119, outubro 2009.
- ALMIRON, M. G. et al. On the Numerical Accuracy of Spreadsheets. *Journal of Statistical Software*, v. 34, n. 4, p. 1–29, abril 2010.
- ALTMAN, M. A review of JMP 4.03 with special attention to its numerical accuracy. *American Statistician*, v. 56, n. 1, p. 72–75, fevereiro 2002.
- ALTMAN, M.; GILL, J.; MCDONALD, M. P. Accuracy: Tools for accurate and reliable statistical computing. *Journal of Statistical Software*, v. 21, n. 1, p. 1–30, julho 2007.
- ANASTASI, G. et al. Energy conservation in wireless sensor networks: A survey. *Ad Hoc Networks*, v. 7, n. 3, p. 537–568, maio 2009.
- ANDEL, T. R.; YASINSAC, A. On the credibility of manet simulations. *Computer*, v. 39, n. 7, p. 48–54, julho 2006.
- AURENHAMMER, F. Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Computing Surveys*, v. 23, n. 3, p. 345–405, setembro 1991.
- BADDELEY, A. Spatial point processes and their applications. In: BADDELEY, A. et al. (Ed.). *Stochastic Geometry*. [S.l.]: Springer, 2007, (Lecture Notes in Mathematics, v. 1892). p. 1–75.
- BUSTOS, O. H.; FRERY, A. C. Statistical functions and procedures in IDL 5.6 and 6.0. *Computational Statistics and Data Analysis*, v. 50, n. 2, p. 301–310, 2005.
- CANONICAL LTD. *Ubuntu Linux*. [S.l.], 2013. última consulta em outubro de 2013. Disponível em: <<http://www.ubuntu.com/>>.

CAVIN, D.; SASSON, Y.; SCHIPER, A. On the accuracy of MANet simulators. In: 2ND ACM INTERNATIONAL WORKSHOP ON PRINCIPLES OF MOBILE COMPUTING. Toulouse, France: ACM, 2002. p. 38–43.

CHANG, J. H.; TASSIULAS, L. Maximum lifetime routing in wireless sensor networks. *IEEE-ACM Transactions on Networking*, v. 12, n. 4, p. 609–619, agosto 2004.

CHONG, C. Y.; KUMAR, S. P. Sensor networks: evolution, opportunities, and challenges. *Proceedings of the IEEE*, v. 91, n. 8, p. 1247–1256, agosto 2003.

CULLER, D.; ESTIN, D.; SRIVASTAVA, M. Guest editors' introduction: Overview of sensor networks. *IEEE Computer Society*, v. 37, n. 8, p. 41–49, agosto 2004.

DARDARI, D. et al. Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks. *IEEE Transactions on Mobile Computing*, v. 6, n. 7, p. 790–802, julho 2007.

DEMIRKOL, I.; ERSOY, C.; ALAGOZ, F. MAC protocols for wireless sensor networks: a survey. *Communications Magazine, IEEE*, v. 44, n. 4, p. 115–121, abril 2006.

DISTRIBUTED COMPUTING GROUP. *Sinalgo - Simulator for Network Algorithms*. Zurich, Switzerland, 2007. última consulta em outubro de 2013. Disponível em: <<http://disco.ethz.ch/projects/sinalgo/>>.

FRERY, A. C. et al. Error estimation in wireless sensor networks. In: 23RD ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING. Fortaleza, CE, Brasil: ACM, 2008. p. 1923–1927.

FRERY, A. C. et al. Data driven performance evaluation of wireless sensor networks. *Sensors*, v. 10, n. 3, p. 2150–2168, março 2010.

HEINZELMAN, W.; CHANDRAKASAN, A.; BALAKRISHNAN, H. An application-specific protocol architecture for wireless microsensor networks. *IEEE Transaction on Wireless Communication*, v. 1, n. 4, p. 660–670, outubro 2002.

HOESEL, L. V. et al. Prolonging the lifetime of wireless sensor networks by cross-layer interaction. *IEEE Wireless Communications*, v. 11, n. 6, p. 78–86, dezembro 2004.

INTANAGONWIWAT, C. et al. Directed diffusion for wireless sensor networking. *IEEE-ACM Transactions on Networking*, v. 11, n. 1, p. 2–16, fevereiro 2003.

JAVA DEVELOPMENT TEAM. *Java*. Redwood Shores, CA, USA, 2013. última consulta em outubro de 2013. Disponível em: <<http://www.java.com/>>.

KEELING, K. B.; PAVUR, R. J. A comparative study of the reliability of nine statistical software packages. *Computational Statistics & Data Analysis*, v. 51, n. 8, p. 3811–3831, maio 2007.

KUMAR, S.; LAI, T. H.; BALOGH, J. On k-coverage in a mostly sleeping sensor network. *Wireless Networks*, v. 14, n. 3, p. 277–294, junho 2008.

LI, H. et al. Energy conservation in wireless sensor networks and connectivity of graphs. *Theoretical Computer Science*, v. 393, n. 1–3, p. 81–89, março 2008.

LI, J.; MOHAPATRA, P. Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. *Pervasive and Mobile Computing*, v. 3, n. 3, p. 233–254, junho 2007.

LIU, C. et al. Random coverage with guaranteed connectivity: joint scheduling for wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, v. 17, n. 6, p. 562–575, junho 2006.

LIU, S. C. A lifetime-extending deployment strategy for multi-hop wireless sensor networks. In: 4TH ANNUAL COMMUNICATION NETWORKS AND SERVICES RESEARCH CONFERENCE. Moncton, NB, Canada: IEEE, 2006. p. 53–60.

MARSAGLIA, G. *The Diehard Battery of Tests of Randomness*. 1998. última consulta em agosto de 2013. Disponível em: <<http://www.stat.fsu.edu/pub/diehard>>.

MATSUMOTO, M.; NISHIMURA, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, v. 8, n. 1, p. 3–30, janeiro 1998.

MCCULLOUGH, B. D. Assessing the reliability of statistical software: Part I. *American Statistician*, v. 52, n. 4, p. 358–366, novembro 1998.

MCCULLOUGH, B. D. Assessing the reliability of statistical software: Part II. *American Statistician*, v. 53, n. 2, p. 149–159, maio 1999.

MCCULLOUGH, B. D. Microsoft Excel's 'Not The Wichmann-Hill' random number generators. *Computational Statistics & Data Analysis*, v. 52, n. 10, p. 4587–4593, junho 2008.

MCCULLOUGH, B. D.; HEISER, D. A. On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics & Data Analysis*, v. 52, n. 10, p. 4570–4578, junho 2008.

MCCULLOUGH, B. D.; WILSON, B. On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics & Data Analysis*, v. 31, n. 1, p. 27–37, julho 1999.

MEDIDI, M.; ZHOU, Y. Extending lifetime with differential duty cycles in wireless sensor networks. In: IEEE GLOBAL TELECOMMUNICATIONS CONFERENCE. Washington, DC, USA: IEEE, 2007. p. 1033–1037.

NAKAMURA, E. F.; LOUREIRO, A. A. F.; FRERY, A. C. Information fusion for wireless sensor networks: Methods, models, and classifications. *ACM Computing Surveys*, v. 39, n. 3, p. artigo número 9, setembro 2007.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. *Statistical Reference Datasets*. 2003. última consulta em setembro de 2013. Disponível em: <<http://www.itl.nist.gov/div898/strd/index.html>>.

POTTIE, G.; KAISER, W. Wireless integrated network sensors. *Communication of the ACM*, v. 43, n. 5, p. 51–58, maio 2000.

- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. última consulta em outubro de 2013. Disponível em: <<http://www.R-project.org/>>.
- RAGHUNATHAN, V.; GANERIWAL, S.; SRIVASTAVA, M. Emerging techniques for long lived wireless sensor networks. *IEEE Communications Magazine*, v. 44, n. 4, p. 108–114, abril 2006.
- RAGHUNATHAN, V. et al. Energy-aware wireless microsensor networks. *IEEE Signal Processing Magazine*, v. 19, n. 2, p. 40–50, março 2002.
- RAMOS, H. S. et al. Topology-related modeling and characterization of wireless sensor networks. In: 8TH ACM SYMPOSIUM ON PERFORMANCE EVALUATION OF WIRELESS AD HOC, SENSOR, AND UBIQUITOUS NETWORKS. Miami, FL, USA: ACM, 2011. p. 33–40. ISBN 978-1-4503-0900-4.
- REIS, I. A. et al. Data-aware clustering for geosensor networks data collection. In: XIII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO. Florianópolis, SC, Brasil: INPE, 2007. p. 6059–6066.
- RIPLEY, B. D. *Stochastic Simulation*. New York, NY, USA: Wiley, 1987.
- RIPLEY, B. D. Thoughts on pseudorandom number generators. *Journal of Computational and Applied Mathematics*, v. 31, n. 1, p. 153–163, julho 1990.
- ROBERT, C. P.; CASELLA, G. *Monte Carlo Statistical Methods*. New York, NY, USA: Springer, 2000. (Springer Texts in Statistics).
- SADLER, B. M. Fundamentals of energy-constrained sensor network systems. *IEEE Aerospace and Electronic Systems Magazine*, v. 20, n. 8, p. 17–35, agosto 2005.
- SANTI, P. Topology control in wireless ad hoc and sensor networks. *ACM Computing Surveys*, v. 37, n. 2, p. 164–194, junho 2005.
- SIMON, S. D.; LESAGE, J. P. Assessing the accuracy of ANOVA calculations in statistical software. *Computational Statistics & Data Analysis*, v. 8, n. 3, p. 325–332, 1989.
- SOHRABI, K. et al. Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications*, v. 7, n. 5, p. 16–27, outubro 2000.
- THE ECLIPSE FOUNDATION. *Java*. Ottawa, ON, Canada, 2013. última consulta em outubro de 2013. Disponível em: <<http://www.eclipse.org/>>.
- TILAK, S.; ABU-GHAZALEH, N.; HEINZELMAN, W. A taxonomy of wireless microsensor network models. *ACM SIGMOBILE Mobile Computing and Communications Review*, v. 6, n. 2, p. 28–36, abril 2002.
- WICHMANN, B. A.; HILL, I. D. Algorithm AS 183: an efficient and portable pseudo-random number generator. *Applied Statistics*, v. 31, n. 2, p. 188–190, 1982.
- WU, X.; CHEN, G.; DAS, S. K. Avoiding Energy Holes in Wireless Sensor Networks with Nonuniform Node Distribution. *IEEE Transactions on Parallel and Distributed Systems*, v. 19, n. 5, p. 710–720, maio 2008.



XIE, X. F.; ZHANG, W. J.; YANG, Z. L. Social cognitive optimization for nonlinear programming problems. In: FIRST INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS. Beijing, China: IEEE, 2002. v. 2, p. 779–783.

YE, W.; HEIDEMANN, J.; ESTRIN, D. Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE/ACM Transactions on Networking*, v. 12, n. 3, p. 493–506, junho 2004.

YICK, J.; MUKHERJEE, B.; GHOSAL, D. Wireless sensor network survey. *Computer Networks*, v. 52, n. 12, p. 2292–2330, agosto 2008.

YOON, S.; SHAHABI, C. The Clustered AGgregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks. *ACM Transactions on Sensor Networks*, ACM, New York, NY, USA, v. 3, n. 1, março 2007.

YOUNIS, M.; AKKAYA, K. Strategies and techniques for node placement in wireless sensor networks: A survey. *Ad Hoc Networks*, v. 6, n. 4, p. 621–655, junho 2008.

ZHANG, W. J.; XIE, X. F. Depso: Hybrid particle swarm with differential evolution operator. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. Washington, DC, USA: IEEE, 2003. v. 4, p. 3816–3821.

ZHANG, X.; WU, Z. D. The balance of routing energy consumption in wireless sensor networks. *Journal of Parallel and Distributed Computing*, v. 71, n. 7, p. 1024–1033, julho 2011.



## APÊNDICE A – AMBIENTE DE SIMULAÇÃO: UMA AVALIAÇÃO DA PRECISÃO NUMÉRICA

### A.1 Metodologia de avaliação da precisão numérica

A avaliação da precisão numérica foi realizada do ponto de vista do usuário, ou seja, independente da maneira como a função estatística foi implementada na plataforma (muitas vezes o código fonte não está disponível para avaliação).

Assim, a avaliação desses resultados consiste em submeter as funções implementadas nas plataformas a diferentes amostras. Depois 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 a seguir. Essa função é uma adaptação proposta por Almiron (2009) e baseada na medida proposta por McCullough (1998).

**Definição 12 (Matching Digits)** *A quantidade aproximada de dígitos significativos que coincidem entre dois valores é determinada por:*

$$MD(x, c) = \begin{cases} -\log_{10} \frac{|x-c|}{|c|} & \text{se } c \neq 0, \\ -\log_{10} |x| & \text{caso contrario.} \end{cases} \quad (19)$$

Nesse caso,  $MD(x, c)$  indica a quantidade aproximada de dígitos significativos que foram estimados corretamente, onde  $x$  é o valor estimado e  $c$  o correto.

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.

Com respeito as funções estatísticas avaliadas, foram eleitas para análise funções simples como a média, o desvio padrão, o coeficiente de autocorrelação e o estatístico  $F$  do procedimento de análise de variância ANOVA, e procedimentos estatísticos mais complexos como regressão não linear e geração de números aleatórios (também foram analisados procedimentos para regressão linear e funções para cálculos de quantis nas caudas de distribuições importantes, mas não abordaremos os resultados nesse trabalho).

## A.2 Resultados da avaliação

No restante desta seção, apresentamos os resultados da avaliação da precisão numérica obtidos através da metodologia descrita na seção anterior. Antecipamos que para maiores detalhes das funções avaliadas (nomeclaturas) e como torná-las comparáveis, das versões das plataformas e do ambiente computacional utilizado, sugerimos a leitura mais aprofundada dos trabalhos aqui destacados (ALMIRON; ALMEIDA; MIRANDA, 2009; ALMIRON et al., 2010).

Algumas das ferramentas podem, eventualmente, estar ausentes em algumas das tabelas apresentadas. Isso acontece quando não há uma função específica implementada para realizar esse cálculo nessa ferramenta. Desta maneira, mantemos a fidelidade da análise à plataforma.

Nas amostras de referência existem diferentes níveis de complexidade: baixa (b), média (m), e alta (a). Em cada amostra nas tabelas é indicado (entre parênteses e pelas iniciais de cada nível) o nível de dificuldade. Para facilitar a visualização, quando há diferença dentre as ferramentas, realçamos com **negrito** os melhores resultados e com **sublinhado** os piores valores de precisão.

Dando início a apresentação dos resultados, começamos com os estatísticos de resumo univariados. As Tabelas 1, 2 e 3, 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 1 – *MD* para a média amostral.

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

Fonte: Autor.

No cálculo da média (ver Tabela 1), todas as funções apresentam resultados bons. Notemos que *Oleo* teve alguns problemas com as amostras *NumAcc1* e *NumAcc3* e por consequência foi a plataforma 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 2 –  $MD$  para o desvio padrão amostral.

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

Fonte: Autor.

No que diz respeito ao cálculo do desvio padrão, a Tabela 2 mostra que os piores resultados foram aqueles obtidos com as plataformas Excel 2008, Oleo, e Ox. O símbolo ‘–’, na coluna correspondente a plataforma 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 plataforma que não teve problemas foi R, sendo desta vez a única que obteve a máxima precisão em todas as amostras.

Tabela 3 –  $MD$  para o coeficiente de autocorrelação amostral.

Amostra	Calc	Excel	Gnumeric	NeoOffice	Octave	Ox	Python	R
Lew (b)	5	5	5	5	5	14	<u>2</u>	<b>15</b>
Lottery (b)	3	3	3	3	4	14	<u>2</u>	<b>15</b>
Mavro (b)	4	4	4	4	4	<b>13</b>	<u>1</u>	<b>13</b>
Michelso (b)	6	6	6	6	8	<b>13</b>	<u>3</u>	<b>13</b>
NumAcc1 (b)	15	15	15	15	<u>0</u>	15	–	15
PiDigits (b)	5	5	5	5	<u>3</u>	<b>15</b>	<u>3</u>	13
NumAcc2 (m)	14/ <b>15</b>	11	<b>15</b>	11/ <b>15</b>	7	<b>15</b>	<u>3</u>	14
NumAcc3 (m)	12	12/ <u>1</u>	12	12	<u>3</u>	12	–	<b>14</b>
NumAcc4 (a)	11	11/–	11	11	3	11	<u>2</u>	<b>14</b>

Fonte: Autor.

Concluindo com as funções para o cálculo de estatísticos de resumo univariados, na Tabela 3 encontram-se os resultados para o cálculo do coeficiente de autocorrelação de primeira ordem. Claramente podemos perceber grande diferença por parte de Ox e R em

relação ao resto das plataformas para todas as amostras. Apesar de todas as plataformas apresentarem resultados bons para as amostras NumAcc1, NumAcc2, NumAcc3 e NumAcc4, nenhuma plataforma, com exceção de Ox e R, obteve desempenho satisfatório de uma forma geral. Os resultados apresentados duplamente (com o símbolo “/”) divergem devido as versões diferentes da mesma plataforma ou por estarem sendo avaliadas em diferentes plataformas de sistemas operacionais.

Apresentando agora a avaliação para o estatístico  $F$  do procedimento de análise de variância 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 e Lesage (1989). Na tabela 4, 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.

Tabela 4 –  $MD$  para o estatístico  $F$  de ANOVA.

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

Fonte: Autor.

Outro aspecto importante da avaliação de plataformas de análise de dados foi abordado em Almiron et al. (2010). O procedimento estatístico de regressão não linear foi avaliado em planilhas eletrônicas.

Em Almiron et al. (2010) a ideia foi analisar 27 amostras, transformando cada regressão em um problema de otimização não linear, em que o objetivo principal é minimizar o residual da soma dos quadros. Para isso, foram utilizadas as ferramentas de **Solver**.

Apesar de algumas planilhas eletrônicas não apresentarem nativamente essa ferramenta, foi possível fazer a avaliação através de extensões disponibilizadas nos sites oficiais dos fabricantes (casos do Calc 3.0.1 e NeoOffice 3.0). Essas extensões disponibilizam

dois algoritmos evolucionários não determinísticos para avaliação de problemas de regressão não linear, chamados DEPS (*Differential Evolution Particle Swarm*), uma abordagem multiagente baseada no paradigma de *Particle Swarm Optimization* (ZHANG; XIE, 2003), e SCO (*Social Cognitive Optimization*), baseado em mecanismos de inteligência social e aprendizado humano (XIE; ZHANG; YANG, 2002). Outro caso especial é o da plataforma Excel 2007, a qual também disponibiliza dois métodos para resolver problemas de regressão não linear, chamados Newton e Conjugado.

Além dos detalhes mais técnicos, é importante frisar que a metodologia de avaliação parte da visão do usuário. No caso da análise de regressão não linear essa metodologia ganha ainda mais força. Como McCullough (1998) comenta, as ferramentas de regressão não linear possuem diversos parâmetros de configuração. O ajuste dessas configurações podem resultar em soluções melhores. A configuração de alguns desses parâmetros faz parte da abordagem de avaliação de Almiron et al. (2010). Na apresentação dos resultados da avaliação, iremos nos referir como “configuração padrão” para os casos em que não houve alteração nos valores dos parâmetros do **Solver** e como “configuração personalizada” para os casos em que houve alteração.

Além disso, as amostras fornecidas pelo *NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY* apresentam valores certificados partindo de dois pontos de início (*starting points*): o primeiro apresenta um valor distante da solução final e representa a situação em que o usuário não tem, *a priori*, a ideia de qual é a solução ou, pelo menos, a ideia de para onde a solução deveria convergir; já o segundo apresenta um valor mais próximo da solução final. Como esse fator também faz parte da abordagem do trabalho de Almiron et al. (2010), também apresentaremos os resultados obtidos na avaliação e, a partir desse ponto, iremos nos referir a esses pontos de início por *Start 1*, para o ponto com valor mais distante da solução, e *Start 2*, para o ponto com valor mais próximo da solução.

Mais detalhes sobre as amostras utilizadas e das ferramentas de *Solver* analisadas, sugerimos a leitura mais aprofundada dos trabalhos de McCullough (1998) e Almiron et al. (2010) e a apreciação dos arquivos providos pelo *NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY*.

Dando início a apresentação dos resultados dessa avaliação, chamamos a atenção para os resultados mostrados nas Tabelas 5 e 6, em que foram utilizados os pontos de início *Start 1* e *Start 2*, respectivamente, e as configurações padrões das ferramentas de **Solver**. Podemos observar que para 16 (na Tabela 5) e 18 (na Tabela 6) das 27 amostras analisadas, foram alcançados pelo menos quatro dígitos significativos corretos. De acordo com McCullough (1998), esses valores de *MD* são resultados razoáveis para regressão não linear. De uma maneira geral, com o ponto de início *Start 2* os resultados apresentados são melhores que com *Start 1*, como era esperado.

Tabela 5 – *MD* para regressão não linear, iniciando em *Start 1* e configuração padrão.

Amostras	Calc 2.4.1 (U)	Calc 3.0.1 -DEPS (M)	Calc 3.0.1 -SCO (M)	Calc 3.0.1 -DEPS (U)	Calc 3.0.1 -SCO (U)	Calc 3.0.1 -DEPS (W)	Calc 3.0.1 -SCO (W)	Excel 2007 -Conjugate	Excel 2007 -Newton	NeoOffice 2.2.5	NeoOffice 3.0 -DEPS	NeoOffice 3.0 -SCO
Misra1a (b)	0	0	0	0	0	0	0	-	1.6	-	0	0
Chwirut2 (b)	-	5.6	4.4	4.8	4.9	6.1	5.0	0	4.3	0	5.4	4.8
Chwirut1 (b)	-	5.1	5.1	5.1	4.8	4.9	4.5	1.7	4.0	0	5.5	4.4
Lanczos3 (b)	-	0	-	0	-	0	-	-	-	-	-	-
Gauss1 (b)	0	5.7	6.1	5.8	5.9	6.0	5.7	1.0	4.8	0	6.1	5.8
Gauss2 (b)	0	5.8	-	6.1	-	5.9	5.3	1.3	4.6	0	5.9	-
DanWood (b)	1.1	4.3	4.1	4.7	5.0	4.6	4.2	4.0	4.7	1.0	4.2	4.8
Misra1b (b)	0	4.8	4.9	5.7	4.8	5.3	5.3	0	0	0	5.1	5.6
Kirby2 (m)	0	4.6	0	4.7	0	4.4	0	0	0	0	4.7	0
Hahn1 (m)	-	4.1	-	3.6	-	3.0	-	-	-	-	3.1	-
Nelson (m)	-	-	-	-	-	-	-	-	-	-	-	-
MGH17 (m)	-	-	-	0	-	-	-	-	-	-	-	-
Lanczos1 (m)	-	0	-	-	-	0	-	-	-	-	-	-
Lanczos2 (m)	-	-	-	-	-	-	-	-	-	-	0	-
Gauss3 (m)	0	5.9	-	6.0	-	6.1	3.0	0	4.3	0	5.4	0
Misra1c (m)	-	5.4	5.3	5.5	4.9	5.3	5.2	0	2.5	-	4.8	4.9
Misra1d (m)	0	6.0	4.5	5.9	4.6	4.7	5.4	0	1.0	-	5.0	4.6
Rozsman1 (m)	-	2.2	1.2	1.5	1.3	2.1	1.2	-	-	0	2.0	2.0
ENSO (m)	-	-	3.0	-	-	-	-	2.1	3.4	-	4.2	3.5
MGH09 (a)	-	-	-	-	-	-	-	-	-	-	-	-
Thurber (a)	0	5.8	0	5.8	0	6.1	0	-	1.3	0	6.6	0
BoxBOD (a)	-	-	-	-	-	-	-	-	-	-	-	-
Rat42 (a)	-	5.6	5.7	5.6	5.2	5.2	5.9	0	5.9	-	5.7	5.8
MGH10 (a)	-	6.5	-	6.7	-	6.1	-	-	-	-	6.2	-
Eckerle4 (a)	-	3.8	3.3	3.8	3.7	3.9	4.6	-	-	-	3.4	4.0
Rat43 (a)	-	-	-	-	-	-	-	-	-	-	-	-
Bennett5 (a)	-	0	0	2.8	0	0	0	0	0	0	0	0

Fonte: Autor.

Para uma melhor visualização e interpretação dos resultados, podemos analisar os resultados apresentados do ponto de vista da complexidade das amostras utilizadas. Começando com as amostras com nível de dificuldade baixo, podemos observar que nos resultados obtidos com *Start 1* pelo menos um algoritmo apresenta bom resultado, exceto para as amostras **Misra1a** e **Lanczos3**, para as quais nenhum algoritmo conseguiu resultados confiáveis. Além disso, o algoritmo **Conjugado (Windows 2007)** apresentou sérias dificuldades em seus cálculos, só alcançando um bom resultado para a amostra **DanWood**. Entretanto, os piores resultados foram apresentados pelos algoritmos das plataformas **Calc 2.4.1** e **NeoOffice 2.2.5**, os quais não apresentaram nenhum dígito significativo correto



(com exceção dos cálculos para a amostra **DanWood** e mesmo assim o resultado não passou de 1 dígito significativo correto).

Dando destaque aos melhores resultados apresentados na Tabela 5, observamos que os algoritmos **DEPS** e **SCO** ficam com esse crédito. As exceções estão para os resultados obtidos pelo algoritmo **SCO**, quando utilizada a amostra **Gauss2**, nos sistemas operacionais **Ubuntu** e **Mac OS**.

Ainda a respeito das amostras de dificuldade baixa, agora com destaque aos resultados obtidos a partir do ponto de início *Start 2* (Tabela 6), todas as plataformas continuam tendo dificuldades para os cálculos envolvendo a amostra **Lanczos3**. Já com relação à amostra **Misra1a**, os resultados apresentados são melhores que os equivalentes para *Start 1*, entretanto **Calc 2.4.1**, **Conjugado** e **NeoOffice 2.2.5** ainda continuam apresentando os piores resultados. Novamente os melhores resultados ficam por conta dos algoritmos **DEPS** e **SCO**, porém não observamos melhoras significativas para o algoritmo **SCO** com relação à amostra **Gauss2**.

Partindo agora para a análise das amostras com nível de dificuldade médio, podemos encontrar vários problemas com os resultados obtidos a partir de *Start 1*. As amostras **Nelson**, **MGH17**, **Lanczos1**, **Lanczos2** e **Roszman1** representam um grande desafio para todos os algoritmos avaliados. Para o restante das amostras, o algoritmo **DEPS** conseguiu os melhores resultados, até mesmo para a amostra **ENSO**, para a qual **DEPS** é o único algoritmo a apresentar resultados para *MD* aceitáveis. Os algoritmos das plataformas **Calc 2.4.1**, **NeoOffice 2.2.5** e **Excel 2007** não apresentam resultados confiáveis para nenhuma amostra, exceto o algoritmo **Newton** que ainda consegue um bom resultado quando aplicado à amostra **Gauss3**.

Já com a utilização do ponto de início *Start 2*, a Tabela 6 mostra resultados melhores para as amostras **Nelson**, **MGH17** e **Roszman1**, entretanto essa melhora ainda não representa valores de *MD* aceitáveis. Adicionalmente, os resultados para as amostras **Lanczos1** e **Lanczos2** são ainda piores que os equivalentes obtidos para o ponto de início *Start 1*. Com relação ao algoritmo **DEPS**, que apresentou os melhores resultados para *Start 1*, devido a sua natureza não determinística não há evidências suficientes para declarar que existe uma melhora na utilização de *Start 2*. Contudo, esse ainda apresenta os melhores resultados. Como **DEPS**, o algoritmo **SCO** apresenta bons resultados, porém ainda falha quando aplicado às amostras **Kirby2**, **Hahn1** e **Gauss3**. Os algoritmos com pior desempenho para o ponto de início *Start 1* continuam com sem melhoras significativas quando aplicado *Start 2*.

Finalizando a avaliação dos resultados obtidos a partir de *Start 1* e *Start 2* com as configurações padrões das ferramentas de **Solver**, focamos na aplicação das amostras com nível de dificuldade alto. A partir da Tabela 5, podemos ver que, mais uma vez, algumas amostras se mostram bastante desafiadoras para os algoritmos, principalmente as amostras

Tabela 6 – *MD* para regressão não linear, iniciando em *Start 2* e configuração padrão.

Amostras	Calc 2.4.1 (U)	Calc 3.0.1 -DEPS (M)	Calc 3.0.1 -SCO (M)	Calc 3.0.1 -DEPS (U)	Calc 3.0.1 -SCO (U)	Calc 3.0.1 -DEPS (W)	Calc 3.0.1 -SCO (W)	Excel 2007 -Conjugate	Excel 2007 -Newton	NeoOffice 2.2.5	NeoOffice 3.0 -DEPS	NeoOffice 3.0 -SCO
Misra1a (b)	1.3	5.3	5.0	5.4	5.2	5.9	6.2	1.3	1.3	-	6.1	5.8
Chwirut2 (b)	0	5.8	5.1	5.1	5.1	5.6	5.1	1.0	4.7	1.3	5.3	4.8
Chwirut1 (b)	-	5.7	4.7	5.4	5.6	5.1	5.0	1.0	4.9	0	5.7	4.9
Lanczos3 (b)	-	-	-	-	-	-	-	-	-	-	-	-
Gauss1 (b)	1.1	6.0	5.7	5.8	5.5	5.8	6.0	1.4	4.6	1.1	6.1	6.0
Gauss2 (b)	0	5.9	-	5.9	0	5.7	0	0	4.5	0	5.9	0
DanWood (b)	3.4	4.1	4.0	4.8	5.1	4.7	4.6	5.1	4.8	1.9	4.7	4.3
Misra1b (b)	0	5.2	5.3	5.3	5.2	5.5	6.0	0	2.8	0	5.3	5.7
Kirby2 (m)	0	4.6	-	4.7	-	4.0	-	0	0	0	4.5	0
Hahn1 (m)	0	3.6	-	3.1	-	3.7	0	-	-	0	3.7	-
Nelson (m)	-	2.9	2.3	3.6	2.6	3.0	2.3	0	0	0	3.2	3.5
MGH17 (m)	0	1.2	0	1.3	0	1.1	0	0	0	0	1.2	0
Lanczos1 (m)	-	-	-	-	-	-	-	-	-	-	-	-
Lanczos2 (m)	-	-	-	-	-	-	-	-	-	-	-	-
Gauss3 (m)	0	6.0	2.9	5.8	-	5.7	0	0	4.4	0	5.8	0
Misra1c (m)	1.2	5.7	5.3	5.7	5.4	5.4	5.4	1.1	1.2	1.1	5.9	5.1
Misra1d (m)	0	5.0	4.7	5.5	5.4	5.5	5.5	1.5	1.5	1.5	5.5	6.0
Roszman1 (m)	0	1.9	0	1.7	2.2	1.7	0	0	0	0	2.5	1.2
ENSO (m)	-	1.2	1.2	1.2	1.2	1.2	-	2.5	3.4	-	1.2	-
MGH09 (a)	-	1.8	0	1.7	1.3	1.8	1.5	1.1	4.3	-	1.7	0
Thurber (a)	0	6.2	1.3	5.8	1.0	5.7	0	-	-	0	6.0	1.2
BoxBOD (a)	-	5.9	6.7	6.2	6.0	6.9	5.7	-	-	-	5.9	5.5
Rat42 (a)	-	5.9	5.0	5.4	5.1	5.7	5.6	1.2	5.2	-	5.7	5.5
MGH10 (a)	-	5.8	0	6.8	0	7.1	1.5	-	-	-	6.0	-
Eckerle4 (a)	0	4.0	3.8	3.7	3.7	3.8	3.4	5.1	4.7	0	3.8	3.7
Rat43 (a)	-	5.7	2.3	5.2	2.0	5.3	3.9	2.1	3.5	1.3	5.2	2.3
Bennett5 (a)	-	0	0	0	1.2	0	0	0	0	0	0	1.1

Fonte: Autor.

MGH09, BoxBOD, Rat43 e Bennet5. Praticamente todos os bons resultados são obtidos pelos algoritmos DEPS e SCO. Já com os resultados expostos na Tabela 6, podemos concluir que existe uma melhora significativa, principalmente com relação à amostra BoxBOD. O algoritmo Newton também apresenta melhoramentos relacionados às amostras MGH09 e Eckerle4.

Antes de darmos início as avaliações dos resultados expostos nas Tabelas 7 e 8, é importante esclarecer alguns pontos.

Cada **Solver** tem seus próprios parâmetros de acordo com o algoritmo usado. Dessa forma, Almiron et al. (2010) explicam que não é possível levantar um discussão a

respeito de um conjunto único de parâmetros a serem personalizados. Quando foi possível indentificar alguma semelhança, os parâmetros foram configurados de maneira equivalente. Adicionalmente, a velocidade de execução de um algoritmo, desde o início até o final de sua execução na busca de um resultado, não foi levado em consideração nessa avaliação. Apenas a precisão numérica continua sendo o ponto de estudo. Outro detalhe importante é que as plataformas Calc 2.4.1 e NeoOffice 2.2.5 não aparecem na análise porque não são passíveis de personalização. Para detalhes adicionais recomendamos novamente a leitura mais aprofundada de Almiron et al. (2010).

Dando início a avaliação dos resultados, podemos observar que, de uma maneira geral, para ambos os pontos de início *Start 1* e *Start 2* os algoritmos com configurações personalizadas produzem melhores resultados quando comparados aos algoritmos com configurações padrões. As Tabelas 7 e 8 mostram que para 20 e 23 (para *Start 1* e *Start 2*, respectivamente) das 27 amostras utilizadas no estudo o valor de *MD* alcança os quatro dígitos significativos corretos de precisão e, com isso, estão de acordo como critério estipulado em McCullough (1998).

Dividido novamente a análise pelo ponto de vista da complexidade das amostras, podemos observar que, para as amostras de dificuldade baixa, *Misra1a* e *Lanczos3* continuam sendo um desafio para os algoritmos. Além dessas duas amostras, os algoritmos apresentam dificuldades quando aplicados à amostra *Gauss2*. É importante destacar que apenas os algoritmos *DEPS* e *Newton* (Excel 2007) conseguem bons resultados para a amostra *Gauss2*. Adicionalmente, *Newton* é o único algoritmo que consegue bons resultados para a amostra *Misra1a*.

Observando os resultados obtidos com *Start 2*, vemos que melhores valores de *MD* são alcançados quando comparados aos valores obtidos com *Start 1*. Os problemas continuam somente para a amostra *Lanczos3*.

Novamente os melhores algoritmos são *Newton*, por apresentar os maiores valores de *MD*, e *DEPS*, por ser mais consistente e robusto.

A respeito das amostras com dificuldade média, podemos observar na Tabela 7 que há uma melhoria significativa de resultados se compararmos aos resultados do estudo com configurações padrões (ver Tabela 5), pois agora as dificuldades foram resumidas em três amostras (*MGH17*, *Lanczos1* e *Lanczos2*). Os melhores resultados novamente ficam por conta dos algoritmos *DEPS* e *Newton*.

Considerando os valores de *MD* apresentados na Tabela 8, ainda observamos problemas para as amostras *Lanczos1* e *Lanczos2*). Bons resultados continuam sendo obtidos pelos algoritmos *DEPS* e *Newton*). Ambos conseguem atingir bons valores de *MD* para 6 das 11 amostras de dificuldade média.

Finalizando a avaliação dos resultados obtidos a partir de *Start 1* e *Start 2* com

Tabela 7 – *MD* para regressão não linear, iniciando em *Start 1* e configuração personalizada.

Amostras	Calc 3.0.1 -DEPS (M)	Calc 3.0.1 -SCO (M)	Calc 3.0.1 -DEPS (U)	Calc 3.0.1 -SCO (U)	Calc 3.0.1 -DEPS (W)	Calc 3.0.1 -SCO (W)	Excel 2007 -Conjugate	Excel 2007 -Newton	NeoOffice 3.0 -DEPS	NeoOffice 3.0 -SCO
Misra1a (b)	0	0	0	0	0	0	-	4.5	0	0
Chwirut2 (b)	6.3	6.3	6.4	5.6	5.8	6.0	4.6	7.0	6.8	6.2
Chwirut1 (b)	6.6	6.3	6.4	6.8	6.6	7.3	4.4	7.1	6.8	6.2
Lanczos3 (b)	0	-	0	-	0	-	-	-	-	-
Gauss1 (b)	7.0	7.1	6.7	6.7	6.7	7.0	4.0	7.6	6.9	6.7
Gauss2 (b)	7.0	-	7.0	-	6.8	-	3.6	7.5	7.2	-
DanWood (b)	5.7	5.8	6.5	6.3	5.3	4.8	9.1	9.8	6.2	5.7
Misra1b (b)	6.1	5.6	5.9	6.3	6.5	5.8	0	5.6	6.5	5.7
Kirby2 (m)	5.2	0	5.3	0	5.3	0	0	0	5.1	0
Hahn1 (m)	4.7	0	4.8	-	5.0	-	-	-	4.7	-
Nelson (m)	-	-	-	-	-	-	0	4.8	-	-
MGH17 (m)	3.1	-	-	-	3.5	-	-	-	3.1	-
Lanczos1 (m)	0	-	0	-	0	-	-	-	0	-
Lanczos2 (m)	0	-	0	-	0	-	-	-	0	-
Gauss3 (m)	6.8	-	-	-	6.5	0	2.4	7.6	6.8	-
Misra1c (m)	5.9	6.0	8.0	6.0	7.0	5.9	0	2.9	5.9	6.0
Misra1d (m)	7.5	6.0	6.3	7.5	6.3	6.2	0	4.0	7.5	6.1
Rozzman1 (m)	3.2	1.2	2.6	1.3	2.7	1.5	-	9.1	3.2	9.0
ENSO (m)	-	-	-	-	-	-	5.8	7.3	-	-
MGH09 (a)	-	-	-	-	-	-	-	-	-	-
Thurber (a)	7.0	0	7.1	0	7.0	0	-	7.4	7.0	0
BoxBOD (a)	-	-	-	-	-	-	8.7	9.0	-	-
Rat42 (a)	6.2	6.7	6.2	6.2	6.5	6.1	4.8	8.4	6.2	6.0
MGH10 (a)	7.1	-	7.0	-	7.3	-	-	-	7.1	-
Eckerle4 (a)	5.0	4.8	4.9	5.1	4.7	4.9	-	-	5.0	4.8
Rat43 (a)	-	-	-	-	-	-	-	-	-	-
Bennett5 (a)	0	0	0	0	1.7	0	0	0	0	-

Fonte: Autor.

as configurações personalizadas das ferramentas de **Solver**, focamos na aplicação das amostras com nível de dificuldade alto. A partir da Tabela 7, podemos ver novamente que existem melhores resultados em comparação a abordagem com configurações padrões, porém para as amostras MGH09, Rat43 e Bennet5 os problemas ainda persistem. Assim como para as amostras de nível de dificuldade baixo e médio, os algoritmos DEPS e Newton apresentam bom desempenho.

Um caso de melhoria interessante acontece com o ponto de início *Start 2*. O resultados apresentados na Tabela 8 mostram que os algoritmos só apresentam dificuldade quando aplicados à amostra Bennett5.

Tabela 8 – *MD* para regressão não linear, iniciando em *Start 2* e configuração personalizada.

Amostras	Calc 3.0.1 -DEPS (M)	Calc 3.0.1 -SCO (M)	Calc 3.0.1 -DEPS (U)	Calc 3.0.1 -SCO (U)	Calc 3.0.1 -DEPS (W)	Calc 3.0.1 -SCO (W)	Excel 2007 -Conjugate	Excel 2007 -Newton	NeoOffice 3.0 -DEPS	NeoOffice 3.0 -SCO
Misra1a (b)	6.0	6.0	6.1	5.9	6.5	6.3	1.3	4.3	6.7	6.6
Chwirut2 (b)	6.3	6.0	5.9	5.5	6.3	6.2	3.5	7.0	6.2	6.1
Chwirut1 (b)	6.2	6.4	7.1	6.3	6.9	6.6	4.9	7.0	6.2	6.4
Lanczos3 (b)	-	-	-	-	-	-	-	-	-	-
Gauss1 (b)	7.1	6.9	7.1	7.3	7.0	7.2	4.3	7.6	6.7	6.7
Gauss2 (b)	7.2	-	7.1	-	7.2	-	3.6	7.5	7.2	6.8
DanWood (b)	5.6	5.2	5.7	4.9	6.0	5.3	9.0	9.4	5.3	5.7
Misra1b (b)	5.9	6.2	6.9	6.6	7.0	5.8	0	3.4	6.1	6.3
Kirby2 (m)	5.8	0	5.6	0	5.6	0	0	0	4.9	0
Hahn1 (m)	4.4	0	4.2	-	4.7	-	-	-	4.4	-
Nelson (m)	5.0	4.0	4.3	4.0	4.2	4.2	0	4.8	4.1	3.7
MGH17 (m)	3.5	0	1.2	0	3.1	0	0	4.4	3.5	0
Lanczos1 (m)	-	-	-	-	-	-	-	-	-	-
Lanczos2 (m)	-	-	1.1	-	-	-	-	-	-	-
Gauss3 (m)	6.7	3.6	6.4	0	7.0	-	2.9	7.8	6.7	-
Misra1c (m)	7.0	6.3	6.0	7.3	6.2	6.2	1.2	3.1	7.0	6.5
Misra1d (m)	6.4	6.6	6.5	5.8	6.7	6.6	1.5	3.3	6.4	6.6
Roszman1 (m)	2.8	1.3	3.5	2.6	2.5	1.4	0	8.1	2.8	1.4
ENSO (m)	1.2	-	1.2	1.2	1.2	1.2	6.2	6.5	1.2	-
MGH09 (a)	2.7	2.3	2.5	2.3	2.6	2.4	3.4	5.0	2.7	2.0
Thurber (a)	6.5	0	6.8	0	6.8	0	0	7.2	6.5	0
BoxBOD (a)	7.4	7.4	7.1	7.1	7.3	8.0	-	-	7.4	6.9
Rat42 (a)	6.5	6.9	6.7	6.5	6.1	6.7	6.2	8.5	6.5	6.7
MGH10 (a)	7.9	-	7.2	-	7.3	-	-	1.4	7.9	-
Eckerle4 (a)	4.7	4.7	-	4.7	4.9	4.8	6.2	5.9	4.7	5.1
Rat43 (a)	6.4	4.2	7.1	4.3	7.1	4.6	3.7	8.9	6.4	4.9
Bennett5 (a)	1.2	0	1.4	0	0	0	0	0	1.2	1.5

Fonte: Autor.

Por último, mas não menos importante do que as outras avaliações já apresentadas até o momento, fazemos um comentário sobre a geração de números pseudoaleatórios. Conforme já afirmamos, nosso estudo terá como base o uso de simulações estocásticas com abordagem Monte Carlo. Devido a isso, plataformas de simulação e análise de dados que apresentem bons algoritmos de geração de números pseudoaleatórios são altamente desejadas.

Almiron et al. (2010) trazem como referência para determinação de um bom algoritmo os trabalhos de Ripley (1987) e Ripley (1990), em que podemos encontrar declarações de que um bom algoritmo deve prover uma sequência de números com as

seguintes propriedades:

- a) eles são semelhantes aos resultados de uma variável aleatória uniformemente distribuídas;
- b) vetores de dimensões moderadas dessas variáveis aleatórias são coletivamente independentes;
- c) possibilidade de especificar a semente em uma grande variedade de ambientes computacionais;
- d) velocidade;
- e) períodos longos.

Alguns testes têm sido propostos na literatura para avaliar essas propriedades e outros aspectos desses geradores. O padrão do *NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY*, com mais informações disponíveis em <http://csrc.nist.gov/groups/ST/toolkit/rng>, e o teste de *Diehard* de Marsaglia (1998), disponível em <http://i.cs.hku.hk/~diehard>, são exemplos de ferramentas disponíveis para esse tipo de avaliação.

Do ponto de vista do usuário, uma boa documentação pode ser suficiente para a escolha da melhor plataforma. Podemos ver em Almiron et al. (2010) que a maioria das plataformas de planilhas eletrônicas avaliadas não possuem documentação suficiente para determinar, por exemplo, quais algoritmos são utilizados em suas implementações. A exceção fica por conta das plataformas *Gnumeric*, que implementa o algoritmo Mersenne-Twister (MATSUMOTO; NISHIMURA, 1998), e *Excel*, cujo fabricante declara a implementação do algoritmo Wichmann e Hill (1982) (para as versões 2003 e 2007). Entretanto, podemos ver que ambas as plataformas não são as melhores escolhas quando se tratam de números aleatórios, pois *Gnumeric* implementa um ótimo algoritmo mas não possibilita uma maneira (pelo menos ao nível do usuário final) de configurar a semente e a declaração feita para o *Excel* não é confirmada em McCullough (2008).

Já para as plataformas estatísticas, Almiron (2009) resume que as plataformas analisadas implementam vários algoritmos de geração de números pseudoaleatórios, apresentando ótimas propriedades de velocidade, longo período e possibilidade de configuração de semente.

### A.3 Conclusões da avaliação

Com relação aos estatísticos de resumo univariados, Almiron, Almeida e Miranda (2009) e Almiron et al. (2010) concluem que todas as plataformas estatísticas e todas as planilhas eletrônicas envolvidas no estudo, quando aplicadas a resoluções de problemas envolvendo a média amostral, apresentam resultados satisfatórios. Os maiores problemas acontecem com a plataforma *Ox*, no estudo de desvio padrão amostral, e com as pla-

taformas `Octave`, `Python`, `Excel 2008` e `NeoOffice 2.2.5` no cálculo do coeficiente de autocorrelação amostral. Adicionalmente, a plataforma `Python` apresenta um problema extra no estudo relacionado ao estatístico  $F$  de ANOVA. O mesmo problema ocorre para a plataforma `Excel 2007`.

As plataformas que apresentam os melhores resultados para os estatísticos de resumo univariado são `Gnumeric` e `R`, sendo essa última a melhor escolha. Almiron, Almeida e Miranda (2009) ainda apresenta uma informação adicional com respeito plataforma `R`: a versão 2.6.2 introduz alguns melhoramentos com relação as implementações de funções estatísticas feitas em versões anteriores (KEELING; PAVUR, 2007).

Já para a avaliação do comportamento das plataformas diante de problemas de regressão não-linear, a abordagem abrangente de Almiron et al. (2010) (envolvendo quatro situações: uso das configurações padrões e configurações personalizadas dos algoritmos implementados nas planilhas eletrônicas com a possibilidade de dois pontos de início, denominados longe (*Start 1*) e perto (*Start 2*) do valor certificado) permitiu a observação de que existem problemas sérios com os algoritmos avaliados, embora em diversas situações alguns desses algoritmos tenham apresentado bons resultados.

Além dessa abordagem, Almiron et al. (2010) ainda adiciona uma pequena análise exploratória aos seus resultados: um terceiro ponto de início, usado em conjunto com a configuração personalizada de parâmetros. Esse terceiro ponto se trata do próprio valor certificado. Essa avaliação é extremamente útil para entender a habilidade do algoritmo de identificar quando alcançou a solução. Os algoritmos `Conjugado` e `Newton` (disponíveis para a plataforma `Excel`) reconhecem quando estão diante da solução ótima, exceto quando aplicados à amostra `Bennett5`. Já o algoritmo `DEPS` (disponíveis para as plataformas `Calc` e `NeoOffice`) apresentam falhas para 12, 13 e 9 amostras das 27 amostras avaliadas, em `Windows`, `Mac OS` e `Ubuntu`, respectivamente. Entretanto, é importante informar que apesar da falha na identificação da solução ótima, o algoritmo `DEPS` apresenta pelo menos valores de  $MD = 7$  para todas as amostras, exceto para as amostras `Rat43` e `Bennett5`.

Apesar dos bons resultados alcançados pelos algoritmos `DEPS` e `Newton`, esse último principalmente no cenário com pontos de início mais próximos aos valores certificados e com configurações personalizadas, o estudo de Almiron et al. (2010) conclui que a escolha de planilhas eletrônicas para estudos estatísticos sérios e mais elaborados de regressão não linear pode não ser o melhor caminho.

Ainda com respeito a escolha de plataformas de estudo estatístico de regressão não linear, embora não tenhamos apresentado no nosso trabalho resultados relacionados a plataformas estatísticas mais completas, mostramos aqui uma informação adicional extraída de Almiron, Almeida e Miranda (2009).

Na Tabela 9 são apresentados resultados para a avaliação de regressão linear feita

Tabela 9 –  $MD$  menores para o coeficiente  $\hat{\beta}$  e o desvio padrão residual RSD em R aplicando regressão não linear.

Amostras	$\hat{\beta}$	RSD	Amostras	$\hat{\beta}$	RSD	Amostras	$\hat{\beta}$	RSD
Misra1a (b)	6	6	Hahn1 (m)	6	6	ENSO (m)	4	5
Chwirut2 (b)	4	5	Nelson* (m)	5	5	MGH09* (a)	4	4
Chwirut1 (b)	5	6	MGH17* (m)	5	5	Thurber (a)	5	4
Lanczos3 (b)	6	4	Lanczos1 (m)	NA	NA	BoxBOD* (a)	5	5
Gauss1 (b)	6	6	Lanczos2 (m)	7	–	Rat42* (a)	7	6
Gauss2 (b)	6	6	Gauss3 (m)	6	5	MGH10* (a)	6	6
DanWood (b)	8	7	Misra1c (m)	8	6	Eckerle4* (a)	7	7
Misra1b (b)	6	6	Misra1d (m)	6	6	Rat43* (a)	5	5
Kirby2 (m)	6	6	Roszman1 (m)	5	6	Bennett5 (a)	5	4

Fonte: Autor.

para a plataforma R. Podemos observar que a plataforma R produz bons resultados para praticamente todas as amostras, exceto para as amostras **Lanczos1** e **Lanczos2**.

Tentando fazer uma equivalência com a abordagem realizada para a avaliação do comportamento das planilhas eletrônicas, todos os cálculos foram realizados com as configurações padrões da função `nls` (disponível através do pacote `stats`) e para a maior parte das amostras foi utilizado o ponto de início *Start 1* (para 8 amostras, destacadas pelo símbolo “\*” na Tabela 9, o cálculo foi realizado a partir de *Start 2*). Além das questões relacionadas às amostras, que são iguais as fornecidas pelo *NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY*, essa avaliação aborda os resultados obtidos para os cálculos do coeficiente de precisão  $\hat{\beta}$  e o desvio padrão residual RSD (ALMIRON; ALMEIDA; MIRANDA, 2009). Apesar dessa avaliação não ser tão abrangente quanto a realizada em Almiron et al. (2010), podemos ter uma valiosa ideia do comportamento da plataforma R diante de problemas de regressão não linear, confirmando que, assim como na análise dos estatísticos de resumo univariados, essa plataforma se mostra como a melhor escolha para análise de dados.

Já com respeito aos geradores de números pseudoaleatórios, além das conclusões já apresentadas para a plataforma de planilhas eletrônicas, as quais não são boas escolhas para esse objetivo, vale destacar que a plataforma R implementa, dentre outros algoritmos, o algoritmo Mersenne-Twister, um dos melhores (se não for o melhor) algoritmos disponível. Esse algoritmo é um gerador com dimensionalidade 623, um período de  $2^{19937}$ , um conjunto 624-dimensional de sementes inteiras de 32 *bits* e passa no teste *Diehard*.