

UNIVERSIDADE DE PASSO FUNDO
Programa de Pós-Graduação em
Computação Aplicada

Dissertação de Mestrado

ANÁLISE DE PERDA DE PACOTES DE UMA REDE LORAWAN

EMANUELE ARAUJO SOLAGNA



UNIVERSIDADE DE PASSO FUNDO
INSTITUTO DE CIÊNCIAS EXATAS E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

**ANÁLISE DE PERDA DE PACOTES DE UMA
REDE LORAWAN**

Emanuele Araujo Solagna

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Computação Aplicada na Universidade de Passo Fundo.

Orientador: Prof. Roberto dos Santos Rabello

Coorientador: Prof. Cristiano Roberto Cervi

Passo Fundo

2020

CIP – Catalogação na Publicação

S684a Solagna, Emanuele Araujo
Análise de perda de pacotes de uma rede LoRaWan
[recurso eletrônico] / Emanuele Araujo Solagna. – 2020.
4.8 MB ; PDF.

Orientador: Prof. Dr. Roberto dos Santos Rabello.
Coorientador: Prof. Dr. Cristiano Roberto Cervi.
Dissertação (Mestrado em Computação Aplicada) –
Universidade de Passo Fundo, 2020.

1. Redes de computadores. 2. Protocolo LoRaWAN.
3. Perda de pacotes. 4. Comutação por pacotes
(Transmissão de dados). I. Rabello, Roberto dos Santos,
orientador. II. Cervi, Cristiano Roberto, coorientador.
III. Título.


CDU: 004.41

Catalogação: Bibliotecária Jucelei Rodrigues Domingues - CRB 10/1569


ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO DO ACADÊMICO

EMANUELE ARAUJO SOLAGNA

Aos trinta e um dias do mês de março do ano de dois mil e vinte, às dezesseis horas, realizou-se, no Instituto de Ciências Exatas e Geociências prédio B5, da Universidade de Passo Fundo (UPF), não pública de defesa do Trabalho de Conclusão de Curso “Ferramenta de análise de perda de pacotes para uma rede Lorawan”, de autoria da **Emanuelle Araújo Solagna**, acadêmica do Curso de Mestrado em Computação Aplicada do Programa de Pós-Graduação em Computação Aplicada – PPGCA. Segundo as informações prestadas pelo Conselho de Pós-Graduação e constantes nos arquivos da Secretaria do PPGCA, a aluna preencheu os requisitos necessários para submeter seu trabalho à avaliação. A banca examinadora foi composta pelos doutores Roberto dos Santos Rabello, Cristiano Roberto Cervi, Carlos Amaral Hölbig, e Alexandre Tagliari Lazzaretti. Concluídos os trabalhos de apresentação e arguição, a banca examinadora considerou a candidata Aprovada. Foi concedido o prazo de até quarenta e cinco (45) dias, conforme Regimento do PPGCA, para a acadêmica apresentar ao Conselho de Pós-Graduação o trabalho em sua redação definitiva, a fim de que sejam feitos os encaminhamentos necessários à emissão do Diploma de Mestre em Computação Aplicada. Para constar, foi lavrada a presente ata, que vai assinada pelos membros da banca examinadora e pela Coordenação do PPGCA.




Prof. Dr. Roberto dos Santos Rabello – UPF
Presidente da Banca Examinadora
(Orientador)




Prof. Dr. Cristiano Roberto Cervi – UPF
(Coorientador)

Prof. Dr. Carlos Amaral Hölbig – UPF
(Avaliador Interno)



Prof. Dr. Alexandre Tagliari Lazzaretti – (IFSUL)
(Avaliador Externo)



Prof. Dr. Rafael Rieder
Coordenador do PPGCA

ANÁLISE DE PERDA DE PACOTES DE UMA REDE LORAWAN

RESUMO

O advento da IoT aliado ao surgimento de novas tecnologias, como por exemplo as redes LoRaWAN, possibilitaram capturar, processar e enviar informações sensoreadas por áreas extensas. No entanto para que seja possível transmitir dados por longas distâncias e ainda assim garantir uma comunicação eficiente e confiável, os dispositivos LoRa geralmente possuem um grande número de parâmetros de transmissão. Por exemplo, um dispositivo pode ser configurado para usar diferentes fatores de espalhamento, configurações de largura de banda, taxas de codificação e infinitas possibilidades de distância, resultando em diversas configurações possíveis e diferentes Taxas de Perda de Pacotes. É um desafio determinar qual desses parâmetros têm maior influência na Taxa de Perda de Pacotes da rede. Assim, este trabalho apresenta uma análise do impacto dos parâmetros de transmissão na Taxa de Perda de Pacotes, através da técnica de Análise de Correspondência Múltipla, aplicada utilizando Linguagem R. A rede submetida a avaliação está implantada na cidade de Passo Fundo - Rio Grande do Sul, a qual foi desenvolvida pela empresa Optim (residente UPF Parque). Os resultados mostraram que a distância de transmissão é o parâmetro que mais influencia na Perda de Pacotes. Também foi possível observar que os parâmetros de transmissão possuem baixa influência na perda, quando comparados com outros fatores como latência do gateway devido ao tipo de conexão utilizada por ele.

Palavras-Chave: Análise de Correspondência Múltipla, LoRa, Parâmetros de Transmissão.

PACKET LOSS ANALYSIS OF A LORAWAN NETWORK

ABSTRACT

The advent of IoT in addition to the emergence of new technologies, such as LoRaWAN networks, made it possible to capture, process and send sensed information over large areas. However, in order to be able to transmit data over long distances and still guarantee efficient and reliable communication, LoRa devices generally have a large number of transmission parameters. For example, a device can be configured to use different spreading factors, bandwidth settings, encoding rates and infinite distance possibilities, resulting in several possible configurations and different Packet Loss Rates. It is a challenge to determine which of these parameters has the greatest influence on the Packet Loss Rate of the network. Thus, this work presents an analysis of the impact of the transmission parameters on the Packet Loss Rate, through the Multiple Correspondence Analysis technique, applied using R Language. The network submitted to evaluation is implanted in the city of Passo Fundo - Rio Grande do Sul, which was developed by the company Optim (resident UPF Parque). The results showed that the transmission distance is the parameter that most influences Packet Loss. It was also possible to observe that the transmission parameters have a low influence on the loss, when compared to other factors such as gateway latency due to the type of connection used by it.

Keywords: LoRa, Multiple Correspondence Analysis. Transmission Parameters.

LISTA DE FIGURAS

Figura 1.	Dispositivos de IoT instalados em todo o mundo de 2015 a 2025 (bilhões). [9].	18
Figura 2.	Tecnologias de camada física [12]	19
Figura 3.	Estrutura da pilha de protocolos [14]	20
Figura 4.	Arquitetura de rede LoRaWAN. Adaptado de [18].	20
Figura 5.	Alcance vs Taxa de transmissão [21]	22
Figura 6.	Distribuição de canais para as faixas de frequência ISM EU 868MHz e US 915MHz	23
Figura 7.	Fluxo do algoritmo recomendado pela Semtech para o ajuste de ADR [25].	25
Figura 8.	Exemplo de pacote de dados transmitido pelos dispositivos.	26
Figura 9.	Interdisciplinaridade da Ciência de Dados a partir do diagrama de Venn. Adaptado de [27]	27
Figura 10.	Ciclo de Vida da Ciência de Dados. Adaptado de [26].	29
Figura 11.	Disposição dos dispositivos e gateways pela cidade de Passo Fundo.	31
Figura 12.	Estrutura do modelo MVC.	33
Figura 13.	Camadas funcionais da arquitetura da plataforma.	35
Figura 14.	Exemplo de pacote sensoriado, armazenado no servidor de rede e encaminhado para a aplicação de coleta.	37
Figura 15.	Componentes da camada de Fonte de dados.	38
Figura 16.	Componentes da camada de coleta de dados.	39
Figura 17.	Diagrama de classes dos dados armazenados.	40
Figura 18.	Componentes da camada de web.	41
Figura 19.	Forma de acesso a Plataforma Omni.	41
Figura 20.	Tela de visualização dos pacotes mais recentes coletados pela rede.	42
Figura 21.	Estrutura do módulo.	43
Figura 22.	Ciclo de vida dos dados no módulo de análise.	47
Figura 23.	Diagrama de Ishikawa utilizado para apresentar os parâmetros utilizados na avaliação de performance da rede em relação a perda de pacotes.	48
Figura 24.	Representação de uma matriz indicadora [44]	52

Figura 25.	Exemplo de nuvem de pontos com objetos classificados a partir variáveis.	55
Figura 26.	Resumo da Análise de Correspondência Múltipla [50].	58
Figura 27.	Dados antes da categorização.	59
Figura 28.	Dados de cada pacote já categorizados.	62
Figura 29.	Dados da medição da taxa de perda de pacotes por SF utilizado. Cada cor representa um valor de SF e as porcentagens indicam quantos pacotes perdidos utilizavam o SF indicado.	67
Figura 30.	Resultado da correlação entre variáveis utilizando Coeficiente de Pearson.	67
Figura 31.	Gráfico de correlação entre as variáveis.	68
Figura 32.	Gráfico de correlação entre as instâncias de cada variável.	69
Figura 33.	Nuvem de instancia das categorias, referente a correspondência do SNR com a distância e a perda de pacotes.	70
Figura 34.	Gráfico que pode ser visualizado na plataforma Omni, onde estão dispostas as informações de pacotes perdidos por dispositivo.	72
Figura 35.	Tela de apresentação dos dados referentes a avaliação da rede completa.	83
Figura 36.	Tela de apresentação dos dados referentes a avaliação da rede completa.	84
Figura 37.	Tela de apresentação dos dados referentes a avaliação da rede completa.	84

LISTA DE TABELAS

Tabela 1.	Exemplo de matriz indicadora utilizando as variáveis dos pacotes.	53
Tabela 2.	Tabela de valores de RSSI e SNR categorizados para o SF 7.	61
Tabela 3.	Tabela de valores de RSSI e SNR categorizados para o SF 8.	61
Tabela 4.	Tabela de valores de RSSI e SNR categorizados para o SF 9.	61
Tabela 5.	Tabela de valores de RSSI e SNR categorizados para o SF 10.	61
Tabela 6.	Tabela da distancia categorizada.	62
Tabela 7.	Pacotes disponíveis para calcular a MCA.	64
Tabela 8.	Valores encontrados para o total de pacotes perdidos, total de pacotes obtidos e taxa de perda de pacotes em porcentagem.	66
Tabela 9.	Taxa de perda de pacotes por dispositivo ordenada da maior (dispositivo 20), para a menor (dispositivo 19).	71
Tabela 10.	Configurações com maior porcentagem de sucesso na entrega de pacotes por dispositivo.	73

SUMÁRIO

1	INTRODUÇÃO	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	INTERNET DAS COISAS	17
2.2	INTERNET DAS COISAS NO ÂMBITO DE CIDADES INTELIGENTES.....	18
2.3	LORA E LORAWAN	19
2.3.1	LoRaWAN	19
2.3.2	LoRa	21
2.3.3	Parâmetros de Transmissão	21
2.3.4	Seleção dos parâmetros de Transmissão	24
2.3.5	Formato dos Pacotes	26
2.4	CIÊNCIA E ANÁLISE DE DADOS	27
2.5	A REDE LORAWAN DA OPTIM	30
2.6	A PLATAFORMA OMNI	31
2.6.1	Funcionamento	31
2.6.2	Arquitetura	33
2.6.3	Camada de Fonte de Dados	36
2.6.4	Camada de Coleta de Dados	38
2.6.5	Camada Web	40
2.6.6	Camada de Interação	41
2.6.7	Módulos da Plataforma Omni	41
2.6.7.1	Módulo de Análise de Perdas e Configuração de rede	42
2.7	TRABALHOS RELACIONADOS	43
3	MATERIAIS E MÉTODOS	47
3.1	TAXA DE PERDA DE PACOTES	49
3.2	MÉTODO DE ANÁLISE DE CORRESPONDÊNCIA	50
3.2.1	Análise de Correspondência Múltipla	51
3.2.2	Matriz indicadora	52
3.2.2.1	Nuvem de pontos	54
3.2.2.2	Considerações de dados para Nuvem de Categorias na Análise de correspondência múltipla	58

3.2.3	Correlação	60
3.3	AJUSTE E TRANSFORMAÇÃO DAS VARIÁVEIS	60
3.4	FERRAMENTAS DE ANÁLISE DOS DADOS	62
3.4.1	Análise de Correspondência múltipla e Correlação na Linguagem R ...	64
4	ANALISE E DISCUSSÃO DE RESULTADOS	65
4.1	AVALIAÇÕES DA REDE COMPLETA	65
4.1.1	Medição da Taxa de Perda	65
4.1.2	Taxa de Perda por SF	66
4.1.3	Influência dos parâmetros na Taxa de Perda	67
4.1.4	Avaliação da qualidade do sinal em relação a perda de pacotes	69
4.2	AVALIAÇÕES POR DISPOSITIVO	70
4.2.1	Medição da Taxa de Perda por dispositivo	70
4.2.2	Melhores configurações utilizadas por dispositivo	72
5	CONSIDERAÇÕES FINAIS	75
	REFERÊNCIAS	77
	APÊNDICE A – TELAS DO MÓDULO DE AVALIAÇÃO DE TAXA DE PERDA DE PACOTES E CONFIGURAÇÃO DA REDE	83

1. INTRODUÇÃO

Nos últimos anos, a Internet das Coisas tem impulsionado cada vez mais a experimentação e análise de novas infraestruturas de rede para comunicação de dados por longas distâncias [1], principalmente para uso nas chamadas Cidades Inteligentes. Neste contexto, as tecnologias e padrões de comunicação por radiofrequência têm sido amplamente utilizadas. Essas tecnologias são compostas geralmente por dispositivos que exercem o sensoriamento, a coleta e a transmissão dos dados. Como toda tecnologia de transmissão de dados, as tecnologias de radiofrequência estão sujeitas a condições que interferem em sua performance. E como são tecnologias relativamente novas, sua efetividade na entrega das informações ainda vem sendo explorada.

Dentre as infraestruturas que permitem atender as principais necessidades das Cidades Inteligentes, estão as redes LPWAN, um tipo de rede de baixa potência e área ampla. Esse tipo de rede é uma das principais vertentes tecnológicas que possibilitam o desenvolvimento de dispositivos de IoT. Dentre as LPWANs estão as redes que fazem uso da tecnologia LoRa. O Lora é uma tecnologia de camada física que utiliza uma técnica de modulação do sinal de rádio, que é o espalhamento espectral [2]. Esse tipo de modulação está relacionada a um parâmetro denominado fator de espalhamento (SF). Esse fator pode ser configurado com seis valores diferentes (SF7, SF8, SF9, SF10, SF11, SF12) [3]. Cada valor permite a transmissão e recepção de dados ao mesmo tempo e para cada um deles existe um compromisso entre a robustez à interferências e a taxa de transmissão. Por um lado, o aumento da taxa de transmissão acarreta no aumento da vazão dos dados, por outro, isso pode influenciar na perda de robustez, levando a interferências que podem induzir a perdas de pacotes.

A tecnologia é LoRa é de camada física e faz uso de um protocolo próprio, que define a arquitetura do sistema e os parâmetros de comunicação, esse protocolo é conhecido como LoRaWAN [4], e é definido como a camada lógica de rede. O LoRaWAN fornece um esquema chamado ADR (*Adaptive Data Rate*), que é usado para controlar a configuração dos parâmetros de transmissão do dispositivo para os gateways. Ao utilizar o ADR, os dispositivos permitem que o Servidor de Rede se encarregue de selecionar individualmente os parâmetros utilizados por cada dispositivo presente na rede, afim de minimizar a perda de pacotes e otimizar questões de consumo de energia. Portanto, em redes LoRaWAN, além do SF, existem outros parâmetros relacionados a integridade da transmissão dos dados, esses parâmetros são geralmente ajustados pelo servidor de rede e entre eles está a Largura de Banda e a Taxa de Código. Ademais, outro fator importante que pode interferir na transmissão é a distância de instalação dos dispositivos em relação aos Gateways.

Este trabalho está focado na avaliação de taxa de perda de pacotes em redes LoRaWAN, pois diferentes soluções têm envolvido o uso dessas redes, com o objetivo de cobrir grandes áreas geográficas em diferentes ambientes, como por exemplo, o projeto desenvolvido pela empresa Optim, que é uma empresa de Desenvolvimento de Hardware/-Firmware, situada no Parque Tecnológico da Universidade de Passo Fundo (UPF).

A empresa desenvolveu uma rede de porte industrial, que utiliza a tecnologia LoRaWAN para um de seus projetos. A rede encontra-se implantada na cidade de Passo fundo e possui aproximadamente 40 dispositivos em funcionamento. Como a rede é peça fundamental nos projetos desenvolvidos pela Optim, a empresa detinha de uma demanda por obter uma avaliação de efetividade relacionada a taxa de perda de pacotes de sua rede, a fim de certificar que a rede encontra-se com boa performance na entrega das informações. Assim, a partir deste trabalho, foi desenvolvida uma ferramenta que utiliza técnicas estatísticas de análise multivariada, com objetivo de obter informações sobre a relação da taxa de perda com os parâmetros que vêm sendo utilizado pela rede. A avaliação foi desenvolvida e integrada ao sistema interno da empresa, chamado de Plataforma Omni, e encontra-se disponível para que sejam feitas constantes verificações da taxa de perda em relação aos parâmetros utilizados.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais das áreas Internet das Coisas, LoRa, LoRaWAN e Ciência e Análise de Dados. Inicialmente é apresentado o termo Internet das Coisas, bem como suas definições e sua relação com o termo Cidades Inteligentes. Depois, são abordados conceitos e características de redes do tipo LoRaWAN. E finalmente, é apresentado o termo Ciência de dados, o qual foi utilizado para explorar informações pertinentes as redes LoRaWAN, por meio da análise dos pacotes de dados transmitidos por essas redes.

2.1 INTERNET DAS COISAS

A Internet das Coisas (IoT) é um termo relativamente novo, mas que vem sendo muito estudado. Foi inicialmente mencionado por Kevin Ashton em 1999. Ele usou o termo para vincular a ideia de identificação por radiofrequência (RFID) ao então novo tópico Internet [5]. Hoje, apesar de já estar estabelecida e em crescimento, a IoT ainda possui diferentes definições, que de modo geral apontam para um mesmo sentido, mas que se contrapõem entre si. Neste sentido, é possível notar divergências nas definições feitas por alguns autores.

Para Li et. al. [6] a Internet das Coisas é um conceito tecnológico, originário em conceitos de informação e tecnologia. O conceito básico da IoT é conectar as coisas, permitindo que essas “coisas” se comuniquem entre si e que as pessoas também se comuniquem com elas.

Como um conceito mais geral, a Internet das Coisas pode ser apresentada como uma ideia de que tudo possa estar conectado à internet, permitindo a coleta e o envio de dados por meio de sensores, processadores e dispositivos de comunicação no intuito de prover ações e tomadas de decisão mais eficientes [7].

Segundo a CompTIA, associação mundial de tecnologia [8], uma organização sem fins lucrativos, o número de "coisas"conectadas, desde computadores e monitores, até carros, deveria crescer a uma taxa anual de 23,1% entre os anos de 2014 e 2020, chegando a 50 bilhões de coisas no ano em questão. Já na projeção de Lucero [9], é esperado que em 2025 estejam instalados mais de 75 bilhões de dispositivos pelo mundo, como demonstrado na Figura 1.

De fato, essas afirmações levam em consideração a constante expansão do termo Cidades Inteligentes, o qual faz uso dos dispositivos de IoT nas mais diversas áreas.

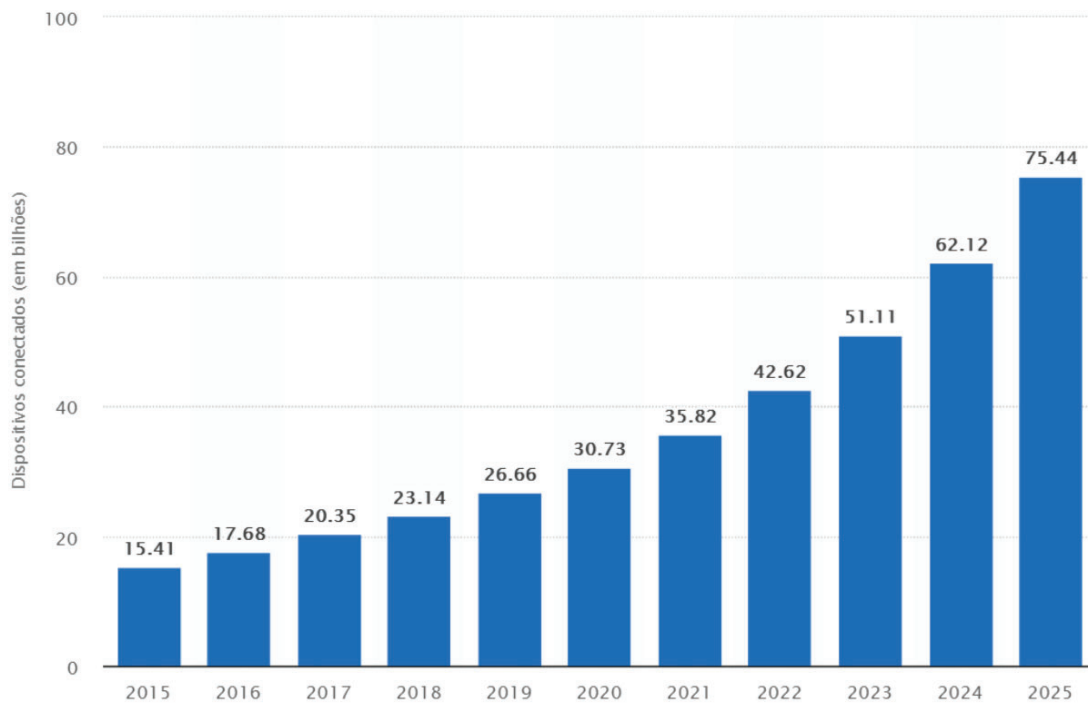


Figura 1. Dispositivos de IoT instalados em todo o mundo de 2015 a 2025 (bilhões). [9]

2.2 INTERNET DAS COISAS NO ÂMBITO DE CIDADES INTELIGENTES

A população humana nas cidades cresce exponencialmente devido à migração maciça e à urbanização, e a expectativa é a de que a população mundial vivendo em áreas urbanas deverá aumentar de 54% em 2014 para 66% até 2050, segundo a ONU [10]. Nesse panorama, tem-se a preocupação com o impacto que o aumento no número de pessoas tem no planejamento urbano das grandes metrópoles. Ao mesmo tempo, o advento de novas tecnologias da informação e comunicação (TICs), principalmente a IoT, permitiu a democratização da capacidade de produção dos cidadãos que passaram a ter poderes para participar da dinâmica de inovação de suas cidades [11]. E conceitos como cidades inteligentes estão sendo cada vez mais difundidos, desenvolvidos e postos em prática.

Exemplos de tecnologias incorporadas pela IoT em cenários de cidades inteligentes [4], podem ser visualizadas na Figura 2. A maioria delas são tecnologias baseadas em redes multi saltos, as quais não oferecem ampla cobertura. Por isso, principalmente para as aplicações que envolvem cobertura cidades, novas tecnologias de comunicação foram desenvolvidas com o intuito de suprir a necessidade de maior cobertura.

Soluções promissoras nesse cenário são as redes LoRaWAN, as quais são apresentadas a seguir.

Tecnologias	Alcance	Vida útil da bateria	Frequência	Sensibilidade	Taxa de Transmissão	Topologia	Soluções e Camadas Superiores
IEEE 802.15.4	70m	Anos	868.3 MHz; 915 MHz e 2.45 GHz	-90 dBm	250 Kbps	Malha	ZigBee, 6LoWPAN, Z-Wave
IEEE 802.15.1	70 m	Dias/Semanas	2.4 GHz	-90 dBm	1 Mbps	Ponto a Ponto/ Malha	BLE/CSRMesh
ISO 13157	10 cm	Meses/Anos	13.56 MHz	-90 dBm	424 Kbps	Ponto a Ponto	NFC
IEEE 802.11	30 m	Horas	2.4 e 5 GHz	-70 dBm	54 Mbps	Ponto Multi-Ponto	TCP/IP
Celular (GSM)	35 km	Dias	900 até 1900 MHz	-114 dBm	14.4 Kbps	Estrela	TCP/IP
LoRa	15km	Anos	169; 433; 915 e 868 MHz	-156 dBm	10 kbps	Estrela	LoRaWAN

Figura 2. Tecnologias de camada física [12]

2.3 LORA E LORAWAN

A tecnologia LoRa (Long-Range ou Longo Alcance) [3] é uma especificação proprietária, mantida pela Semtech [4], para redes LPWAN (Low Power Wide Area ou Longo alcance e baixa potência). O LoRa é uma tecnologia de camada física, que faz uso de uma técnica de modulação por espectro focada em dispositivos que seguem o paradigma da Internet das Coisas, ou seja, dispositivos presentes no cotidiano de seus usuários. Por poder ser utilizada em conjunto com diversos protocolos diferentes, é considerada uma tecnologia atrativa para o desenvolvimento de aplicações voltadas a IoT. Além disso, ela opera em frequências não licenciadas ISM (espectro de rádio reservado internacionalmente para fins industriais, científicos e médicos), reservadas internacionalmente para o desenvolvimento Industrial, científico e médico, com o intuito de reaproveitar a frequência em redes legadas de telefonia, entre outras [13].

No caso da LoRaWAN, está é uma tecnologia baseada em um modelo de camadas e implementa uma pilha de protocolos (Figura 3), que define a comunicação entre os componentes da arquitetura. Ou seja, a tecnologia LoRa é basicamente a parte física/eletrônica que permite a modulação e LoRaWAN é o protocolo padrão MAC (Media Access Control ou Controle de Acesso de Mídia) de controle de acesso, geralmente usado em redes LoRa. A associação da camada física LoRa com a camada lógica da rede constitui as redes LoRaWAN.

2.3.1 LoRaWAN

LoRaWAN é um protocolo do tipo MAC, suas especificações são abertas e padronizadas para LPWANs, sendo mantido pela LoRa Alliance [15], com objetivo de implementar os detalhes de funcionamento da rede LoRa [16].

As redes LoRaWAN são compostas por três dispositivos principais [17]:

l) Dispositivos Finais: podem ser sensores/atuadores, operados por baterias e conectados através do LoRa a um ou mais gateways;

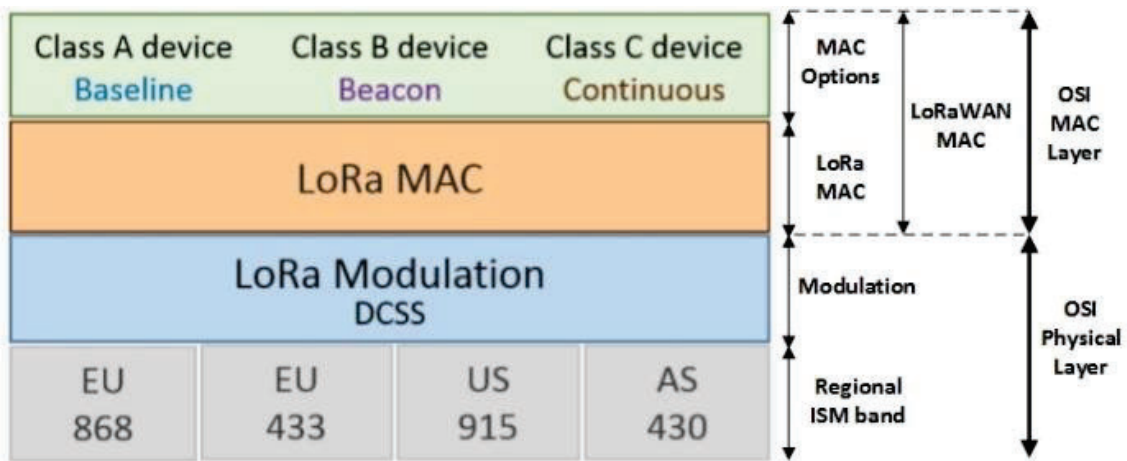


Figura 3. Estrutura da pilha de protocolos [14]

II) Gateways: concentradores que fazem o intermédio entre dispositivos finais e Servidor de Rede;

III) Servidor de Rede: controla toda a rede (gestão de recursos de rádio, admissão de novos dispositivos, segurança, disponibilização da informação, etc).

Os dispositivos transmitem as mensagens diretamente para um ou mais gateways (centralizadores), que encaminham mensagens de forma transparente por meio de uma conexão Protocolo de Internet (IP) para um servidor de rede. O servidor de rede remove mensagens duplicados e as encaminha para o servidor de aplicação.

A arquitetura desse tipo de rede, geralmente consiste na topologia de agrupamento em árvore (Figura 4).

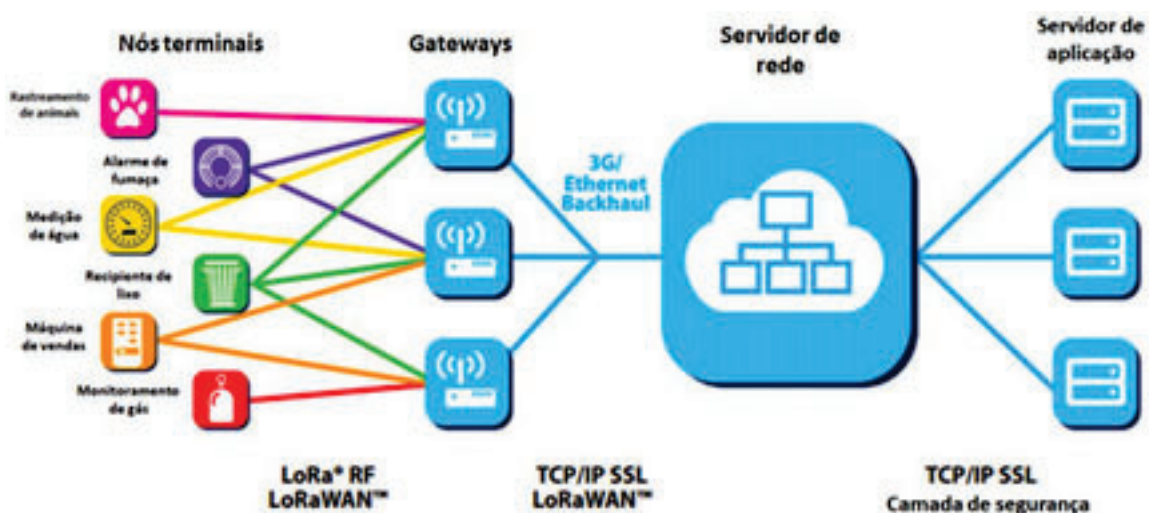


Figura 4. Arquitetura de rede LoRaWAN. Adaptado de [18].

A tecnologia LoRaWAN prevê o uso de três classes de dispositivos, cada uma associada a um modo de funcionamento diferente [19, 16]:

I) Dispositivos classe A: dispositivos que só podem receber dados dos gateways em janelas de tempos anteriormente determinadas e imediatamente após realizarem uma transmissão;

II) Dispositivos classe B: dispositivos sincronizados com o Servidor de Rede por meio de pacotes de beacon que são transmitidos pelos gateways, sua comunicação é bidirecional com janelas de recepção agendadas;

III) Dispositivos classe C: esta classe de dispositivos está sempre apta a receber dados do gateway.

Uma vez que é necessário consumir energia não só para transmitir os dados, mas também para receber mensagens,

As classes de operação se diferenciam pela diferença nos momentos de comunicação com os gateways e também pelo consumo de energia de cada uma. Quanto mais tempo disponível para receber e enviar informações, mais gasto energético.

2.3.2 LoRa

A camada física LoRa foi patenteada pela empresa Semtech em 2014 [4] e sua modulação é baseada na técnica Chirp Spread Spectrum (CSS). E com o uso da CSS o LoRa se torna robusto a interferências e ruídos, fazendo com que o receptor receba a mensagem correta mesmo com uma relação sinal-ruído baixa [19]. Esse tipo de modulação está relacionada a um parâmetro denominado fator de espalhamento (SF). Esse fator pode ser configurado com seis valores diferentes valores [3]. E para cada valor do SF, permite-se também que se tenha uma taxa de dados variável, possibilidade uma troca entre alcance, robustez, ou consumo de energia, mantendo uma largura de banda constante como mostrado na Figura 5 [17]. Além disso tecnologia LoRa permite que vários dispositivos possam transmitir simultaneamente no mesmo canal sem interferência pois os valores de SF são múltiplos e ortogonalmente diferentes [20].

A modulação LoRa permite que a sensibilidade do receptor atinja até -150 dBm, possibilitando comunicação a distâncias que podem chegar a mais de 5 quilômetros em ambientes urbanos e até 10 quilômetros em ambientes rurais [19]. Porém isto pode variar conforme os parâmetros de transmissão utilizados.

2.3.3 Parâmetros de Transmissão

Para que os dispositivos finais se comuniquem com os gateways, a tecnologia LoRa se baseia na adaptação da modulação CSS [22]. Para ajustar seu desempenho, um dispositivo LoRa pode ser configurado para usar diferentes Taxas de codificação (CR), que define a taxa de FEC (Correção antecipada de erros); Frequência (CF), que define a

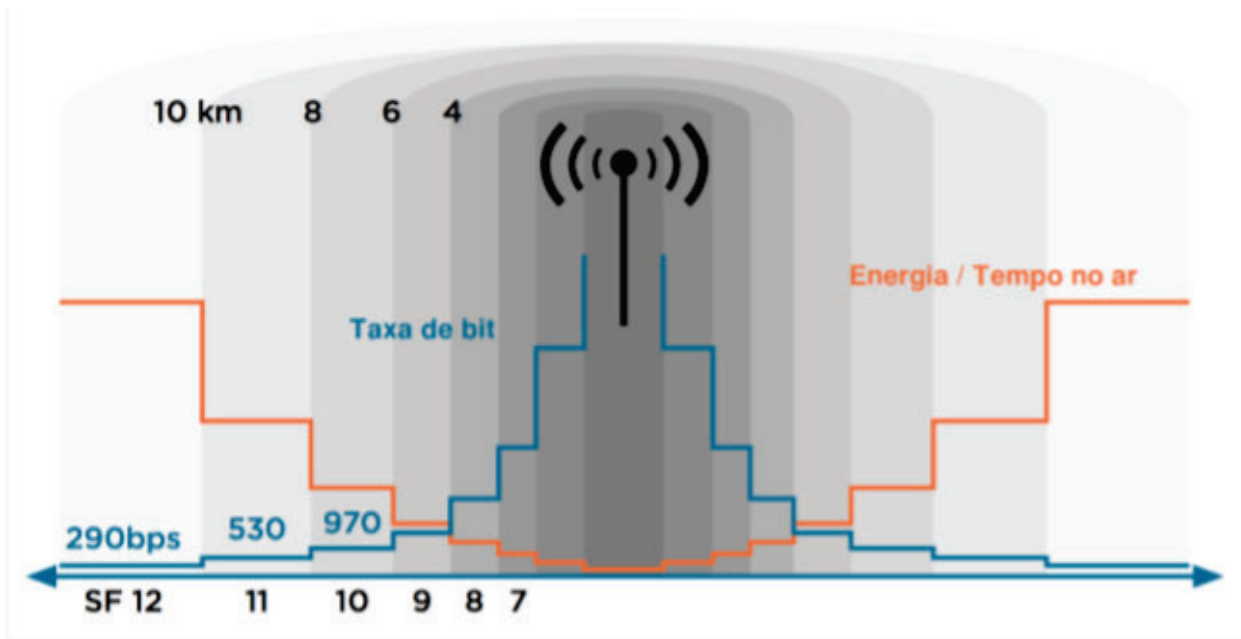


Figura 5. Alcance vs Taxa de transmissão [21]

frequência central para a banda de transmissão; Fator de espalhamento (SF), que define o espalhamento espectral; e a Largura de banda (BW) [2, 23].

a) Taxa de codificação (CR): CR é a taxa FEC usada pelo modem LoRa que oferece proteção contra surtos de interferência e pode ser configurada com os valores: 4/5, 4/6, 4/7 ou 4/8. Um CR mais alto oferece mais proteção, mas aumenta o tempo do pacote no ar [23]. A CR define a taxa de codificação T_c [4], como:

$$T_c = \frac{4}{4 + CR}, \text{ com } CR \in \{1, 2, 3, 4\}. \quad (1)$$

b) Frequência da portadora (CF): CF é a frequência central que pode ser programada dependendo do chip LoRa específico, esse intervalo pode ser limitado a 860MHz a 1020MHz. A frequência da portadora é definida de acordo com a região de operação dos equipamentos, ou seja, esse parâmetro não é em geral configurável, nem mesmo ajustável de acordo com aplicações. A Figura 6, especifica em formato de tabela a distribuição de bandas, sub-bandas, canais, divisão de sub-bandas e canal de uplink e downlink para as faixas de frequências mais estudadas na literatura: ISM EU 868MHz e US 915 MHz. Para o plano de frequências ISM do Brasil, especificamente é utilizada a faixa de 902-928 MHz.

c) Fator de espalhamento (SF): SF é - em resumo - a duração do *chirp*¹, ou seja O SF o que decide quantos chirps são enviados por segundo. A especificação LoRa define seis valores diferentes para o fator de espalhamento: SF7, SF8, SF9, SF10, SF11 e SF12 [4]. O SF está diretamente relacionado a relação sinal ruído (SNR), pois é devido ao uso da técnica de espalhamento pelo LoRa, onde grandes sequências de bits são codificadas

¹Um chirp é um sinal no qual a frequência aumenta ou diminui com o tempo.

Região	Faixa Freq. (MHz)	Canais	Sub-banda (MHz)	Divisão Sub-banda (kHz)	Canal Tx (kHz)	Canal Rx (kHz)
US	867-870	10	867,1-868,8 (SF7-SF12)	200	125	125
			868,3	—	250	250
			868,8 (FSK) (50 kbit/s)	—	—	—
			869,525 (SF9, RX2)	—	—	125
EU	902-928	64	902,3-914,9	200	125	—
		8	903,0-914,2 (SF7-SF12)	1600	125	—
			904,6 (SF8)	—	500	—
8	923,3-927,5 (SF7-SF12)	600	—	500		

Figura 6. Distribuição de canais para as faixas de frequência ISM EU 868MHz e US 915MHz

em um único símbolo, que é possível reduzir a relação sinal ruído a interferência de outras frequências nas transmissões de dados [19]. Quanto maior o valor do SF usado, há um aumento na SNR, a sensibilidade e o alcance; ou seja, o aumento da energia por bit torna maior o alcance entre transmissor e receptor, dependendo diretamente da potência de transmissão e da taxa de modulação.

As comunicações de rádio com SF diferentes são ortogonais, o que faz possível utilizar valores de SF diferentes em uma mesma rede [24]. A potência de transmissão é limitada pelas normas de regulação para bandas das frequências não licenciadas, enquanto a taxa de modulação é definida pelo SF. Também é definida uma taxa de transmissão R_b , teórica, definida em função do fator de espalhamento como:

$$R_b = SF \times \frac{T_c \times BW}{2^{SF}}, \text{ com } SF \in \{7, 8, 9, 10, 10, 11, 12\} \quad (2)$$

Então, o tempo necessário para enviar um símbolo LoRa é influenciado pelo tipo de modulação e o fator de espalhamento, em função da largura de banda do canal:

$$T_{sym} = \frac{2^{SF}}{BW}, \text{ com } BW \in \{125\text{kHz}, 250\text{kHz}, 500\text{kHz}\}. \quad (3)$$

d) Largura de banda (BW): BW é a largura das frequências na banda de transmissão. Um BW mais alto fornece uma taxa de dados mais alta (portanto, menor tempo no ar), mas uma sensibilidade mais baixa (devido à integração de ruído adicional). Um BW mais baixo fornece uma sensibilidade mais alta, mas com uma data mais baixa. O LBB mais baixo requer cristais mais precisos (menos ppm). Os dados são enviados a uma taxa de chip igual à largura de banda; uma largura de banda de 125 kHz corresponde a uma

taxa de chips de 125 kcps. Embora a largura de banda possa ser selecionada em uma faixa de 7,8kHz a 500kHz, uma rede LoRa típica opera em 500kHz, 250kHz ou 125kHz (respectivamente BW500, BW250 e BW125).

2.3.4 Seleção dos parâmetros de Transmissão

Um gateway LoRa é capaz de ouvir transmissões recebidas simultaneamente em todas as combinações de SF e BW, enquanto um dispositivo pode ouvir apenas uma combinação fixa de SF e BW por vez. Um dispositivo pode, portanto, transmitir aos gateways em qualquer combinação de parâmetros de transmissão sem antes concordar com a configuração. As mensagens transmitidas dos gateways para os dispositivos são transmitidas em um deslocamento configurável da taxa de dados de ligação ascendente na primeira janela de recebimento e, geralmente, com a configuração mais robusta (menor taxa de dados) na segunda janela de recebimento.

O LoRaWAN fornece um esquema chamado ADR (Adaptive Data Rate ou Taxa de Dados Adaptativa) que é usado para controlar a configuração dos parâmetros de transmissão para o uplink do dispositivo para o gateway. Um dispositivo tem a opção de selecionar sua taxa de dados e transmitir potência individualmente ou ter sua taxa de dados e transmitir potência controladas pelo servidor de rede usando o mecanismo ADR. O dispositivo indica que deseja usar ADR configurando o bitADR no cabeçalho do quadro. Quando o ADR é ativado, o servidor de rede comanda o LinkADRReq da camada MAC para controlar a taxa de dados e a potência de transmissão do dispositivo final. O comando LinkADRReq não permite selecionar nenhuma das possíveis configurações de transmissão do 6720 e fornece um subconjunto de apenas 8 configurações de taxa de dados e 6 configurações de potência de transmissão.

Embora o LoRaWAN especifique um esquema de sinalização de parâmetros de transmissão (através do comando LinkADRReq), ele não descreve como a sinalização deve ser usada. Não está descrito como o servidor de rede deve instruir os dispositivos sobre a adaptação da taxa: quando uma configuração deve ser alterada ou em que ordem as configurações devem ser alteradas. Porém sabe-se que sempre que o dispositivo transmite um pacote, ele utiliza as últimas configurações armazenadas (no caso a do pacote anterior), pois ele precisa utilizar essas configurações para enviar o bitADR e solicitar a mudança de configuração [25].

A maior parte dos servidores de rede não disponibilizam seus algoritmos de ADR, porém, existe um algoritmo recomendado pela Semtech para adaptação das taxas de dados [25]. O diagrama a seguir (Figura 7), descreve o fluxo de ADRs recomendado:

Primeiramente o algoritmo verifica se o bitADR está presente no pacote enviado, se a verificação não encontrar o bitADR então os últimos pacotes enviados são descarta-

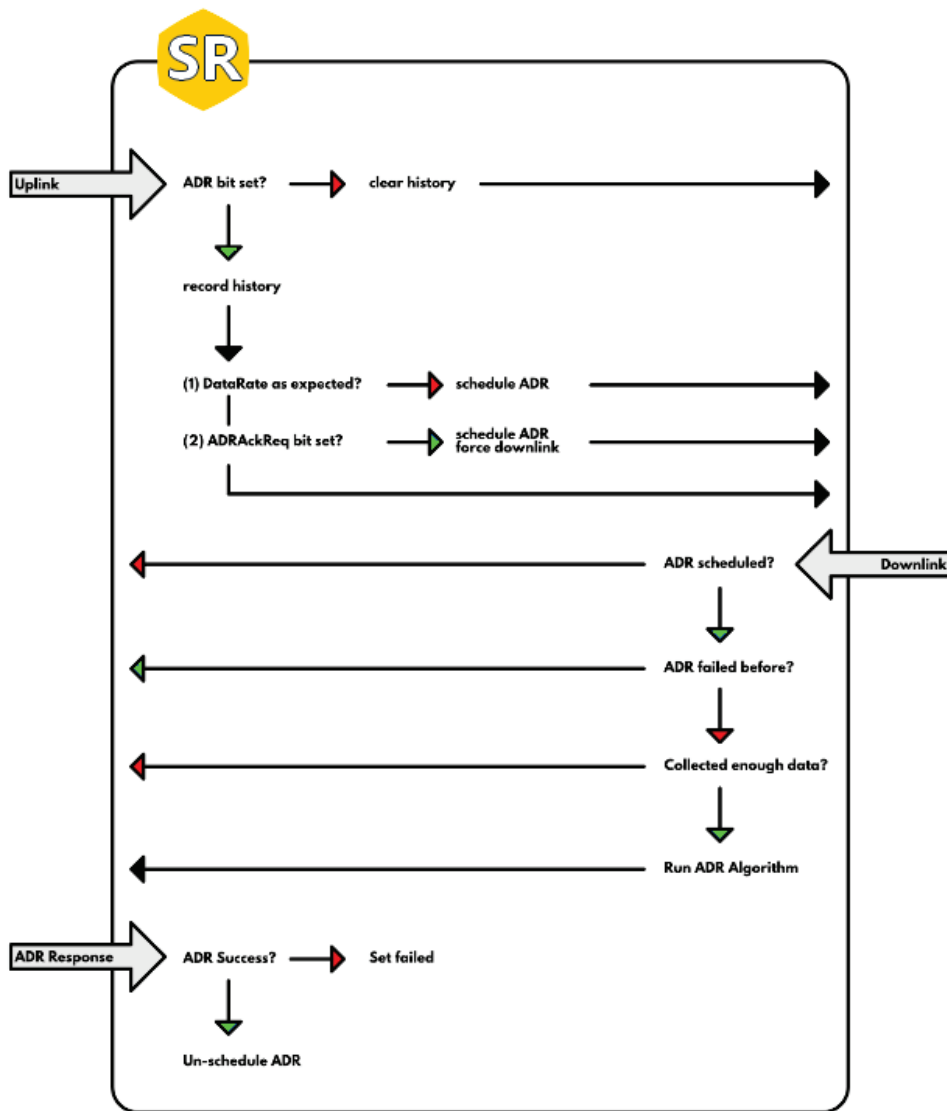


Figura 7. Fluxo do algoritmo recomendado pela Semtech para o ajuste de ADR [25].

dos, caso contrário o servidor recupera os últimos 20 pacotes enviados pelo dispositivo em questão, e avalia a taxa de dados desses últimos 20 pacotes. Se a taxa de dados não estiver bem ajustada e existirem anomalias (como perdas ou consumo excessivo de energia) no comportamento do dispositivo. O servidor verifica então se o ADR foi acionado e se alguma outra solicitação de ADR falhou anteriormente. Caso alguma das anteriores apresente falha, então ele tenta reenviar o ADR anterior. Caso não exista nenhum comando de ADR com falha anteriormente, o servidor roda o algoritmo de ADR, calcula os novos parâmetros e força o envio deles ao dispositivo. Se o comando de envio falhar e o dispositivo não responder, este é setado como falha. Do contrário, o servidor seta o envio como OK e o ADR pode ser "desprogramado". E assim sucessivamente, sempre ajustando a rede quando o Servidor de rede achar necessário, sem a necessidade de ajustes manuais.

2.3.5 Formato dos Pacotes

Conforme apresentado na Figura 8, a estrutura e os dados dos pacotes encaminhados dos dispositivos para o servidor de rede é a mesma de um documento JSON.

```
{
  "payloads_ul": {
    "id": 1544802401704,
    "deveui": "be7a040000000092",
    "timestamp": "2018-12-14T15:46:41.704Z",
    "devaddr": 271214779,
    "live": true,
    "dataFrame": "3039373b8a3039",
    "fcnt": 461,
    "session_id": "612b08bb-4e05-48c4-b7bc-ade080c23f83",
    "port": 12,
    "rssi": -87,
    "snr": 7.5,
    "freq": 916000000,
    "sf_used": 10,
    "dr_used": "SF10BW125",
    "cr_used": "4/5",
    "device_redundancy": 3,
    "time_on_air_ms": 370.688,
    "gtw_info": [
      {
        "gtw_id": "000000000b031938",
        "rssi": -87,
        "snr": 7.5
      }
    ],
    "decrypted": true
  }
}
```

Figura 8. Exemplo de pacote de dados transmitido pelos dispositivos.

Os pacotes são constituídos geralmente pelo dado sensorizados pelo dispositivo (dataFrame), juntamente com parâmetros de configuração utilizados para transmissão e informações sobre qualidade do sinal no qual o pacote foi transmitido (port, freq, sf_used, dr_used, cr_used, device_redundancy, decrypted, rssi, snr, time_on_air), dados do dispositivo que fez a transmissão (devAddr, deveui), dados sobre o gateway que encaminhou o pacote (GTW_info) e dados de controle (id, fcnt).

Como os pacotes possuem informações muito além de apenas os dados sensorizados, a partir deles é possível compreender questões sobre a qualidade da rede em diversos aspectos: qualidade de sinal, efetividade, entre outros.

2.4 CIÊNCIA E ANÁLISE DE DADOS

A Ciência de Dados é um campo de atuação de competências interdisciplinares. Atribui-se à ela a extração de informação útil a partir de bases de dados complexas, dinâmicas, heterogêneas e distribuídas [26]. Conforme a Figura 9, para se atuar na Ciência de Dados, três domínios de conhecimento se inter-relacionam: Programação de Computadores; Estatística e Matemática; e Domínio do Conhecimento. Neste sentido, existem três pressupostos:

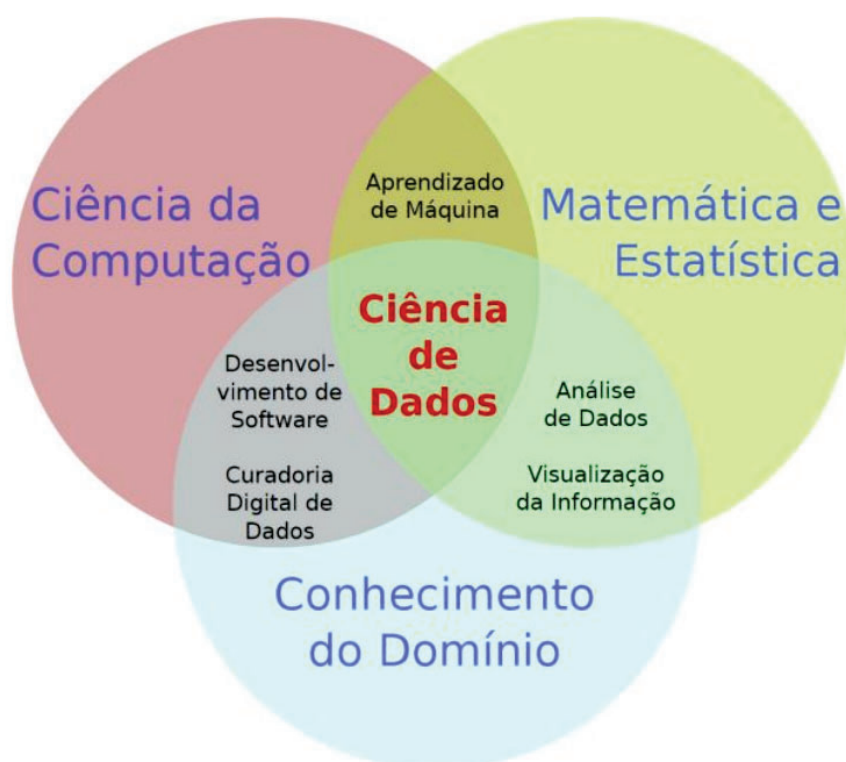


Figura 9. Interdisciplinaridade da Ciência de Dados a partir do diagrama de Venn. Adaptado de [27]

- Dentre as habilidades necessárias na Ciência de Dados, seus especialistas devem apresentar habilidades na área da Ciência da Computação, visto que primordialmente os dados são armazenados, manipulados e transmitidos por computadores. Neste contexto, os ambientes computacionais para o Desenvolvimento de Software são ferramentas essenciais para promover a Curadoria Digital e a implementar extração de conhecimento a partir dos dados, utilizando diversas técnicas computacionais. É imperativo saber utilizar essas tecnologias de modo a acessar e transformar os dados para abstrair e representar informação útil.
- O conhecimento sobre Matemática e Estatística também é necessário para a realização de atividades de Análise de Dados. Ou seja, os profissionais da Ciência de

Dados devem entender o funcionamento dos algoritmos de Aprendizado de Máquina, bem como, saber interpretar os resultados, estatisticamente. Interdisciplinarmente, a atividade de interpretação é facilitada pela Visualização da Informação, a qual privilegia a utilização elementos de representação gráfica da informação.

- Para o efetivo sucesso das soluções de Ciência de Dados, o Conhecimento do Domínio do problema deve ser disponível e amplamente utilizado no processo de tomada de decisão. Neste sentido, as soluções de Ciência de Dados são voltadas para a formulação de hipóteses e a aquisição de informação aderente como insumo no processo decisório.

Ressalta-se que em ecossistemas de Big Data (grandes quantidade de dados), o processo de tomada de decisão é guiado por dados [28]. Como pode ser percebido nos pressupostos relatados, geralmente, tal processo emprega soluções computacionais baseadas em algoritmos de aprendizado de Máquina à aquisição de informação relevante. Conceitualmente, o Aprendizado de Máquina é uma subárea da Inteligência Artificial que investiga a captura automatizada de modelos de abstração de informação a partir registros contidos em (volumosas) bases de dados [29]. Em outras palavras, o Aprendizado de Máquina aplica métodos computacionais e/ou estatísticos para a extração automatizada de informação útil a partir de dados históricos.

Para Scala [26], independentemente de método computacional de Aprendizado de Máquina utilizado, sete passos podem ser executados iterativamente em soluções de Ciência de Dados (Figura 10):

- Obtenção de Dados. Preconiza a realização de tarefas de avaliação e seleção de dados primários e seus metadados a partir: do processamento de arquivos de texto; do monitoramento de uma rede de sensores; de consultas a bases de dados de sistemas legados; de dados oriundos da web; dentre outros.

- Ingestão de Dados. Trata da transformação e carga dos dados primários de fontes diferentes e formatos diversificados em uma base de dados centralizada. Esta etapa implica em organizar e representar os dados de modo a inserir os recursos pré-processados em um repositório de dados principal, mitigando os esforços futuros da geração de informação relevante.

- Exploração de Dados. Privilegia a execução de estudos preliminares para estabelecer as conjecturas iniciais acerca dos dados disponibilizados em relação à informação requisitada. Neste sentido, esta atividade é importante para o estabelecimento do fluxo de trabalho, definindo o roteiro de como relacionar os dados primários à informação relevante.

- Definição dos Parâmetros. Passo intimamente ligado as escolhas necessárias para o emprego do(s) algoritmo(s) de Aprendizado de Máquina. Nesta atividade, por exemplo: I) converte-se os dados de entrada conforme os requisitos de manipulação do algoritmo de aprendizado; II) transforma-se os dados de saída de modo a refletir uma saída legível



Figura 10. Ciclo de Vida da Ciência de Dados. Adaptado de [26].

aos seres humanos; III) estabelece-se os intervalos dos parâmetros de entrada a serem considerados; IV) define-se os critérios de parada do algoritmo de aprendizado; v) o nível de confiabilidade exigido da resposta gerada; dentre outros.

- **Implementação do Modelo.** Primeiramente, utiliza-se os algoritmos de Aprendizado de Máquina para estabelecer modelos a partir dos dados de entrada e saída. Iterativamente, isso envolve o emprego de estratégias de treinamento e de testes dos algoritmos para a definição dos parâmetros mais adequados dentre aqueles avaliados. Como resultado, deve-se abstrair um modelo que estatisticamente melhor represente as características dos dados utilizados.

- **Utilização do Modelo.** Uma vez estabelecido um modelo, pode-se utilizá-lo para inferir informações sobre dados em um ambiente de produção. Isso confirmará o poder de generalização do modelo em gerar informação relevante perante situações do mundo real. Uma vez confirmado o poder de generalização, o modelo poderá ser empregado em Tarefas Intensivas em Conhecimento.

- **Tomada de Decisão.** Nas Tarefas Intensivas em Conhecimento, mediante a combinação do resultado gerado pelo modelo na análise dos dados com seu conhecimento especializado, o gestor ampara suas decisões tomadas. Uma parte fundamental nesta etapa envolve a customização da apresentação de dados e da visualização da informa-

ção através de relatórios e gráficos, respectivamente. Isso torna os *insights*² mais claros e convincentes, auxiliando as atividades cognitivas dos tomadores de decisão.

Considerando as atividades relatadas, assume-se que a geração de informação útil a partir de dados brutos normalmente é um processo iterativo e que pode ser aplicada a avaliação de redes LoRaWAN, pois é justamente a partir dos dados gerados pela rede que se espera obter informações sobre sua performance e efetividade. Neste sentido, os atores envolvidos formulam premissas iniciais a respeito dos parâmetros que serão avaliados e, gradualmente, esses parâmetros são refinados ao adicionar novas dimensões de dados. Em outras palavras, diante do volumoso conjunto de dados que foi avaliado, iterativamente, encontraram-se diversas formas de representação que são úteis acerca da rede. Subliminamente, isso evidencia o processo de evolução dados -> informação -> conhecimento. Ou seja, os dados são transformados em informações, que por sua vez, são agrupadas em padrões apresentados aos usuários para avaliação e descoberta de novas informações sobre a rede em questão, o que auxilia na tomada de decisão e por consequência faz com que se possa obter performance e eficiências cada vez mais otimizadas.

2.5 A REDE LORAWAN DA OPTIM

A rede Optim, faz uso da tecnologia de modulação LoRa e do protocolo LoRaWAN. Esta rede foi desenvolvida pela Optim Tecnologia, empresa situada no Parque Científico e Tecnológico UPF Planalto Médio (UPF Parque) e encontra-se implantada na cidade de Passo Fundo, no norte do estado do Rio Grande do Sul. A rede atua no sensoriamento de dados pertinentes a um projeto desenvolvido pela empresa, o qual teve início no ano de 2018 e que vem ainda sendo constantemente aprimorado conforme novas necessidades. No período selecionado para o desenvolvimento desse projeto, a rede era composta por aproximadamente 40 dispositivos finais classe A, com a tecnologia LoRa® [3] embarcada, cinco gateways LoRaWAN, sendo dois deles conectados a Internet através de rede cabeada e outro três com SIMcards de dados, e o servidor de rede Orbiwan [30]. Neste período, a maior parte dos dispositivos enviou os dados sensoriados de 1 em 1 hora, salvo algumas exceções onde alguns dispositivos comunicavam de 2 em 2 horas ou de 15 em 15 minutos. Essas exceções existiram pelo fato de alguns dispositivos necessitarem de ajustes relacionados aos valores sensoriados em si. Como nesse trabalho foi ignorado o conteúdo dos dados sensoriados, por não ser o foco da avaliação, em nada essas questões interferiram no resultado final.

Além da infraestrutura citada anteriormente, em sua configuração está definido o uso do padrão frequências australiano (AU915MHz), o qual foi recentemente regulamentado pela ANATEL (Administração Nacional de Telecomunicações) no ato 14448, definindo o plano de frequência para a América Latina sendo o padrão Australiano de 923MHz

²Entendimento de uma causa e efeito específicos dentro de um contexto específico.

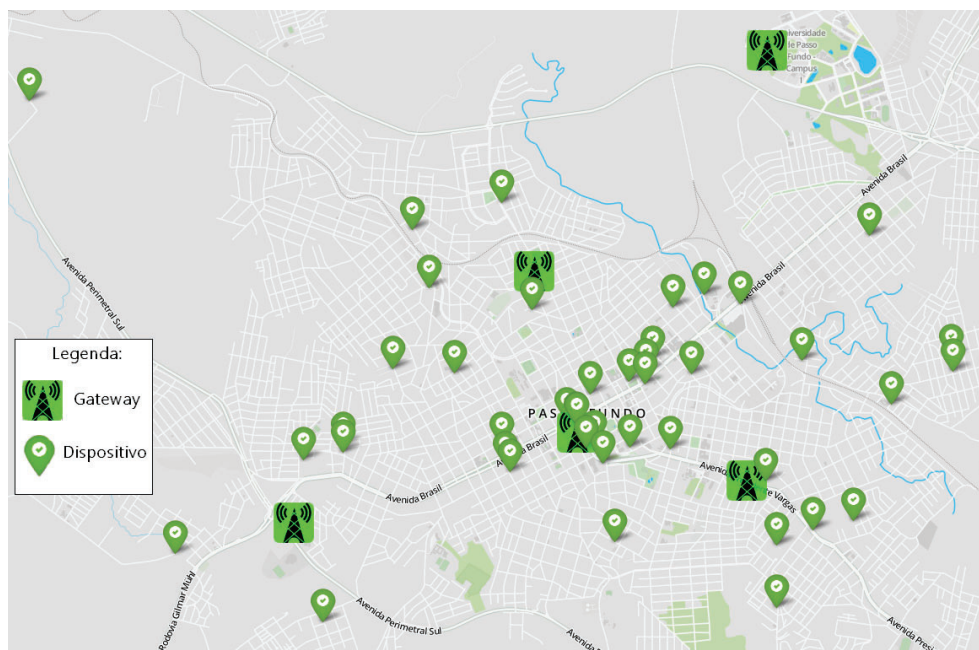


Figura 11. Disposição dos dispositivos e gateways pela cidade de Passo Fundo.

(915MHz a 928MHz) [31], e o uso do ADR, que faz com que a rede auto gerencie seus parâmetros de transmissão para cada dispositivo sempre que necessário

2.6 A PLATAFORMA OMNI

A plataforma Omni foi desenvolvida como sendo um sistema para coleta, processamento e visualização de dados relevantes acerca da rede Optim. A seguir é detalhado o funcionamento da Omni através de sua arquitetura e requisitos atendidos por ela, bem como as tecnologias utilizadas no seu desenvolvimento.

2.6.1 Funcionamento

O funcionamento da plataforma é descrito através de seus requisitos, eles são as especificações dos serviços que a Omni fornece, bem como suas restrições operacionais. Eles podem ser classificados em requisitos funcionais e não funcionais [32], e descrevem de forma clara o funcionamento do software:

- Requisitos funcionais: descrevem explicitamente o que o sistema deve e o que não deve fazer, as funções detalhadas do sistema, suas entradas e saídas, exceções, etc.
- Requisitos não funcionais: definem as propriedades e restrições do sistema como segurança, desempenho, padrões do cliente, legislação, espaço, entre outros.

Os requisitos funcionais, previamente levantados, apresentam a plataforma como um sistema computacional para auxílio na visualização dos dados transmitidos em tempo real pelos dispositivos presentes na rede Optim, além de ter a capacidade de armazenar informações pertinentes a rede e seus componentes, bem como auxiliar no envio de informações e configurações remotas para os dispositivos, o que é chamado de *downlink*³. Quanto aos requisitos não funcionais associados ao software em questão, destaca-se que há possibilidade de uma grande quantidade de dispositivos virem a enviar dados para a plataforma, visto que é de interesse da Optim que sua rede funcione em grande escala, e a necessidade de resposta no menor tempo possível. Sendo assim, a plataforma faz uso de tecnologias e padrões de design condizentes com as demandas identificadas. Além de uma solicitação em especial, a qual foi feita por parte da empresa, que a plataforma fosse desenvolvida na linguagem Java, sendo assim, a plataforma Omni teve seu desenvolvimento totalmente orientado à objetos. E neste sentido, ela foi desenvolvida utilizando tecnologias já amplamente adotadas, muito utilizadas para desenvolvimento de aplicações Web de médio e grande porte, como é o caso do Java Server Faces (JSF).

Conforme descreve Deitel & Deitel [33], Java Server Faces é um "*framework*"⁴ de aplicativo Web que simplifica o design da interface com o usuário de um aplicativo e separa ainda mais a apresentação de um aplicativo Web da sua lógica de negócio". Por sua vez, na visão da Oracle Corporation [34], JSF refere-se a um conjunto de APIs para apresentação de componentes visuais e controle dos seus respectivos estados, eventos, validação de entrada do usuário, definição de navegação, internacionalização e acessibilidade. Juntas, estas APIs constituem um arcabouço arquitetural que permite uma clara separação entre a lógica da aplicação e a sua apresentação. Em termos de padrões de design, o JSF acomoda a implementação do padrão Modelo-Visão-Controlador (Figura12), comumente conhecido pelo acrônimo MVC. Neste padrão arquitetural há forte apelo pela separação entre as camadas de uma aplicação, em especial, na separação dos modelos de dados criados para representar a percepção do mundo real da sua apresentação na interface por meio de componentes gráficos [34]. Assim, para atingir o adequado desacoplamento entre modelo e apresentação, o padrão MVC orienta o desenvolvimento da aplicação em três camadas distintas, camada de modelo, de visão e de controle.

Quanto ao alinhamento com as necessidades tecnológicas da plataforma Omni, o conjunto de características arquiteturais do JSF é adequado aos requisitos elencados para o seu desenvolvimento. O primeiro ponto positivo refere-se à incorporação do JSF ao Java EE. Essa característica permite que o código desenvolvido possa ser executado em qualquer container Java EE do mercado, como Glassfish, JBoss Application Server, entre outros. Além disso, ao executar em um container Java EE, a aplicação automaticamente herda todos os benefícios associados, como escalabilidade, redundância e alta disponibi-

³Dados enviados do servidor para o dispositivo.

⁴Abstração que une códigos comuns entre vários projetos de software provendo uma funcionalidade genérica.

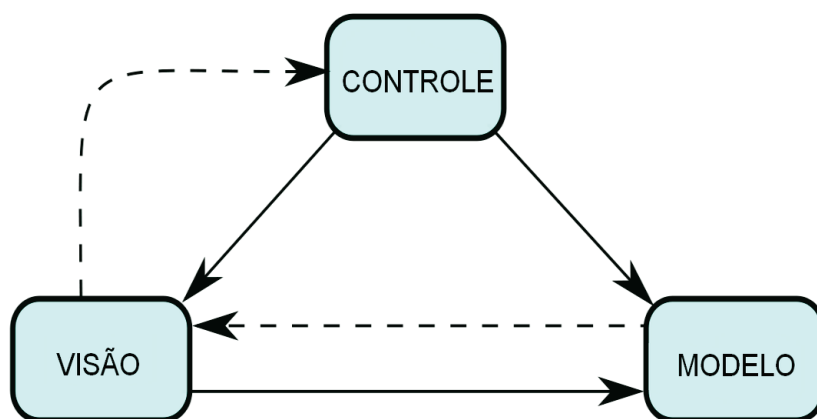


Figura 12. Estrutura do modelo MVC.

dade, por exemplo. Outra característica do JSF adequada ao desenvolvimento da Camada Web relaciona-se à alta propensão para modificações que a camada apresenta. Como era de conhecimento prévio, portanto, que novas funcionalidades poderiam ser agregadas à Omni no decorrer do tempo, isso tornou a preocupação com aspectos de qualidade algo constante durante seu desenvolvimento. Neste sentido, o uso do JSF e, por conseguinte, de um modelo arquitetural MVC, reduziu consideravelmente o acoplamento de código, visto que os modelos de dados ficaram separados dos aspectos relacionados a sua apresentação. Além disso, o desenvolvimento baseado em componentes, aliado às diversas APIs de programação providas pelo JSF tornaram o processo de desenvolvimento menos oneroso em termos de esforço.

Já com relação ao tipo de banco de dados utilizado, por enquanto, a plataforma faz uso de um banco de dados do tipo objeto-relacional, no caso o PostgreSQL. Isso pois foi determinado que também seria utilizado o Hibernate, um framework para o mapeamento objeto-relacional escrito na linguagem Java, em conjunto com a Java Persistence API que define um meio de mapeamento objeto-relacional para objetos Java simples e comuns. Porém, como já se espera que haja a necessidade por um banco de dados escalável, há grandes chances de no futuro migrar o sistema para a utilização de um banco de dados não-relacional como o MongoDB.

Apresentados os aspectos tecnológicos envolvidos, o passo seguinte refere-se ao entendimento da arquitetura da plataforma e da concepção do processo de desenvolvimento de seus módulos.

2.6.2 Arquitetura

Para alcançar seus objetivos, a arquitetura da plataforma Omni foi projetada em quatro camadas funcionais de operação (Figura 13). A primeira camada, identificada como camada de Fontes de Dados refere-se a rede em si, pois ela é exclusivamente a única

fonte que fornece o agregado de todos os dados sensoriados pelos dispositivos finais para a plataforma. A segunda camada, nomeada de Coleta e Processamento, consolida os processos de extração, "limpeza", carga e transformação dos dados enviados através da rede. A terceira camada, Camada Web, se refere a mineração e extração de conhecimento a partir dos dados coletados, com a utilização de técnicas e linguagens de ciência de dados. A última camada, definida como camada de Interação, descreve basicamente os usuários da plataforma que podem interagir diretamente com a Camada Web, ou seja a Optim.

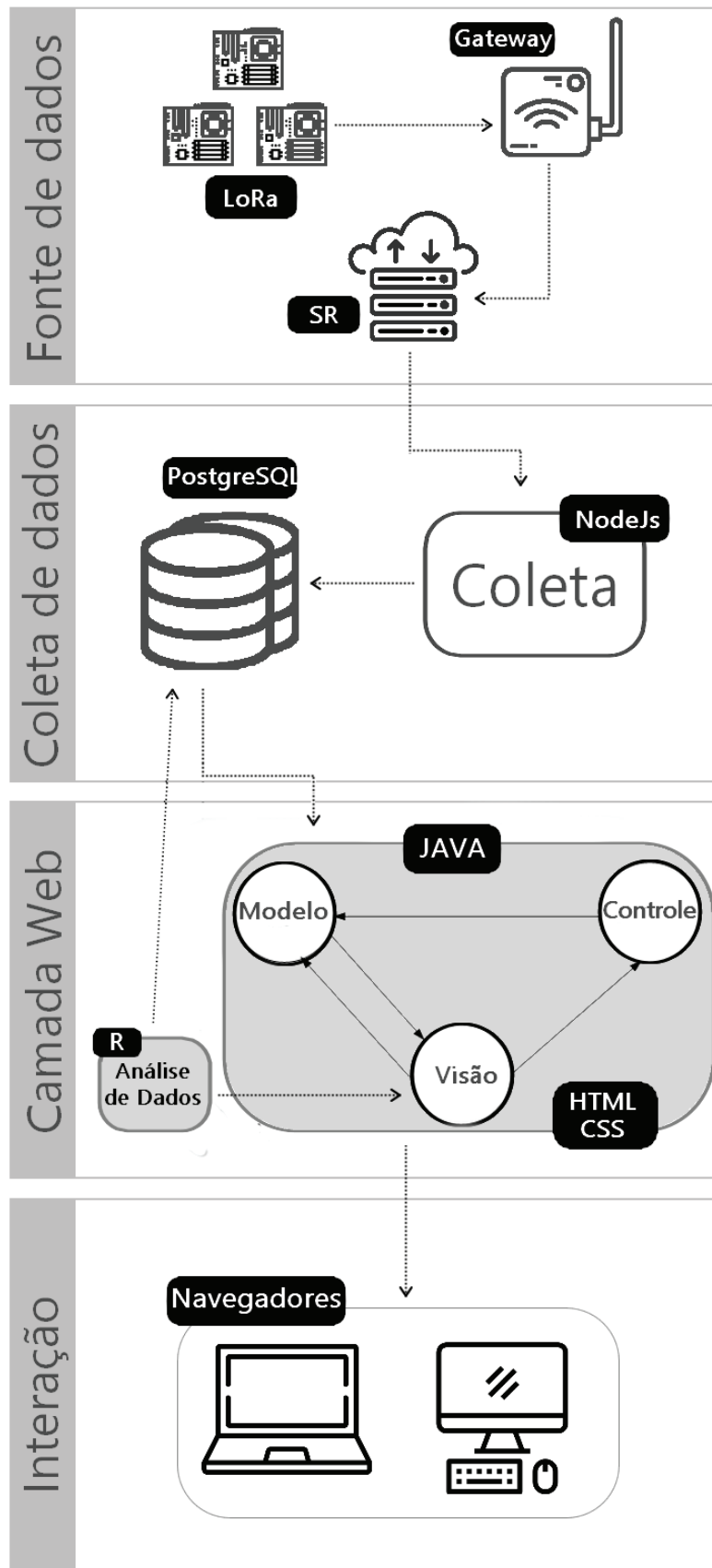


Figura 13. Camadas funcionais da arquitetura da plataforma.

Através da Figura 13 é possível apresentar o diagrama funcional de alto nível da arquitetura proposta, nela estão presentes todas as camadas da plataforma e seus compo-

mentes, bem como as tecnologias utilizadas no seu desenvolvimento. Detalhes sobre cada uma das camadas funcionais são descritos na sequência.

2.6.3 Camada de Fonte de Dados

A Fonte de Dados é a primeira camada da plataforma Omni, esta camada representa a própria rede Optim. Sendo assim, a escolha das tecnologias aplicadas a ela ficaram a cargo da empresa. Todas as tecnologias foram definidas em função das necessidades e particularidades do ambiente no qual a rede foi instalada, mantendo o requisito principal de se ter uma rede LoRa, que implementa o protocolo LoRaWAN.

O funcionamento da camada de Fonte de Dados se dá a partir dos dispositivos finais que coletam os dados por meio de sensores e os enviam como sinais para serem "lidos" por um dos gateways. Os gateways apenas encaminham os pacotes que receberam para o servidor de rede. É a partir da *API*⁵ do servidor de rede que os dados se tornam disponíveis.

Com relação aos dados, considera-se que os sensores coletam informações em diversos pontos diferentes na cidade de Passo Fundo, geralmente a cada 1 hora. Os dados sensorizados são chamados de *payload*, ou seja, a parte principal dos dados transmitidos, da qual se excluem as informações para facilitar a entrega, como cabeçalhos e metadados. Juntamente com os dados coletados pelos sensores, o pacote possui também dados de transmissão como o SF, RSSI, frequência e o SNR. Todos os dados dizem respeito a situação de configuração e intensidade de sinal em que cada pacote foi transmitido, ou seja, são dados referentes ao funcionamento da rede no momento em que o pacote foi recebido.

O conjunto de dados armazenado no servidor de rede que é enviado para a aplicação de coleta está em formato de documentos *JSON*. *JSON* é um acrônimo de "JavaScript Object Notation", um formato compacto de troca de dados simples e rápida entre sistemas, de padrão aberto independente de linguagem de programação derivado do JavaScript. Esse formato apresenta especificações peculiares que precisam ser considerados durante sua manipulação. Na imagem 14 está representada a estrutura e os dados de um pacote sensorizado.

A seguir estão listados cada um dos dados e o que cada um representa:

- *Deveui*: número hexadecimal que identifica o dispositivo o qual enviou o pacote;
- *DataFrame*: valor do consumo de água sensorizado a partir do hidrômetro instalado, o valor está criptografado em formato base64⁶;

⁵conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web

⁶Método para codificação de dados para transferência na Internet

```

{"payloads_ul":{
  "id":1544802401704,
  "deveui":"be7a040000000092",
  "timestamp":"2018-12-14T15:46:41.704Z",
  "devaddr":271214779,
  "live":true,
  "dataFrame":"3039373b8a3039",
  "fcnt":461,
  "session_id":"612b08bb-4e05-48c4-b7bc-ade080c23f83",
  "port":12,
  "rssi":-87,
  "snr":7.5,
  "freq":916000000,
  "sf_used":10,
  "dr_used":"SF10BW125",
  "cr_used":"4/5",
  "device_redundancy":3,
  "time_on_air_ms":370.688,
  "gtw_info":[{"gtw_id":"000000000b031938",
    "rssi":-87,
    "snr":7.5}],
  "decrypted":true}
}

```

Figura 14. Exemplo de pacote sensoriado, armazenado no servidor de rede e encaminhado para a aplicação de coleta.

- Port: valor decimal que representa a porta pela qual o valor foi recebido;
- Timestamp: data e hora em que o pacote foi recebido pelo servidor de rede;
- Fcnt: contador de frames que indica o número sequencial dos pacotes daquela sessão;
- RSSI : parâmetro que indica a força/intensidade do sinal recebido;
- SNR : valor que representa a relação entre o sinal e o ruído [35];
- Sf_used: valor do Spreading Factor utilizado no momento do recebimento do pacote.
- Id: identificador único (64-bits) do pacote;
- Live: indica se o pacote estava no armazenamento temporário ou se ele foi enviado em tempo real;
- Session_id: ID da sessão sob a qual o pacote foi recebido
- Decrypted: parâmetro que indica se o payload está criptografado ou não;

- Gtw_info: Informações sobre o gateway que enviou ou tentou enviar o pacote;
- Gtw_id: Valor hexadecimal, identificador do gateway;
- Gtw_gps: Valor inteiro o qual representa a localização do gateway;
- Gtw_rssi: parâmetro que indica a força/intensidade do sinal recebido;
- Gtw_snr: valor que representa a relação entre o sinal e o ruído do pacote enviado;

Na Figura 15, a camada Fonte de Dados é ilustrada conceitualmente. Por meio dela percebe-se o fluxo dos dados sensorizados até o armazenamento temporário no servidor de rede, para que possam ser acessados pela camada de Coleta de Dados.

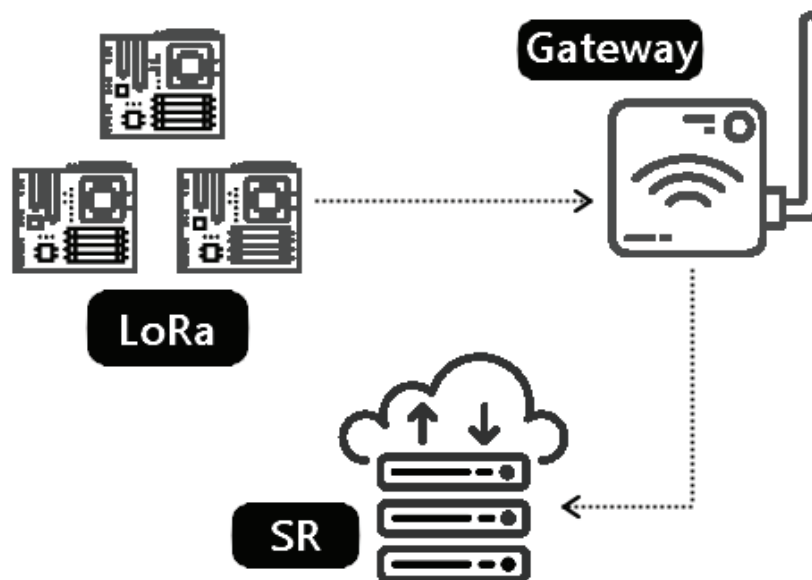


Figura 15. Componentes da camada de Fonte de dados.

2.6.4 Camada de Coleta de Dados

A camada de Coleta é caracterizada pelo uso de diversas tecnologias. A coleta dos dados ou pacotes é feita por uma aplicação escrita na linguagem Node.js (às vezes é também denominado como Node [36]), a qual funciona de forma independente a plataforma. A característica determinante para a escolha pela utilização dessa tecnologia foi a compatibilidade com a API do servidor de rede utilizado pela camada de Fonte de dados. E também pois Node.js, atende às demandas por meio do modelo de E/S orientado a eventos e nonblocking⁷ e JavaScript do lado do servidor [37, 38]. O mecanismo JavaScript V8 do Google Chrome é a base para o Node.js, pois o tempo de execução do JavaScript Node.js

⁷Capacidade de fazer operações de entrada e saída (acessar sistema de arquivos, banco de dados, rede, servidores, etc.) sem que a aplicação fique impedida de executar outras funções em paralelo.

é construído sobre ele [38]. Escrito em C++, a sua V8 é uma fonte aberta do Google como um mecanismo JavaScript de alto desempenho. Esta tecnologia é baseada em eventos e não em threads, então também é capaz de escalonar para milhões de conexões simultaneamente, enquanto usa um loop de eventos em uma única thread e não uma sobrecarga de várias threads. O Node.js processa suas operações de E/S de forma assíncrona assim como o AJAX [36].

O servidor de rede Orbiwan possui um modo de operação chamado Push-Mode Callback [30]. Esse modo de funcionamento consiste em disponibilizar sua API de modo a fazer o próprio servidor enviar os pacotes recebidos pelos dispositivos para a aplicação Node.js. No modo push, a aplicação cliente estabelece uma conexão TCP persistente com o servidor, via WebSocket. Quando um novo pacote chega, o servidor de rede armazena a mensagem temporariamente. Se a aplicação fez a requisição para iniciar o modo push e obteve o retorno, a API envia imediatamente o pacote para a aplicação com uma requisição POST [35]. Quando a aplicação terminar a solicitação de envio da mensagem (ou seja, retornar o código 200 ou 'OK' para o recebimento dos dados), o servidor de rede exclui automaticamente esse pacote de seu armazenamento temporário [30], e assim sucessivamente para cada pacote que for recebido pelo servidor. A Figura 16 representa a estrutura e componentes da camada de Coleta de dados.

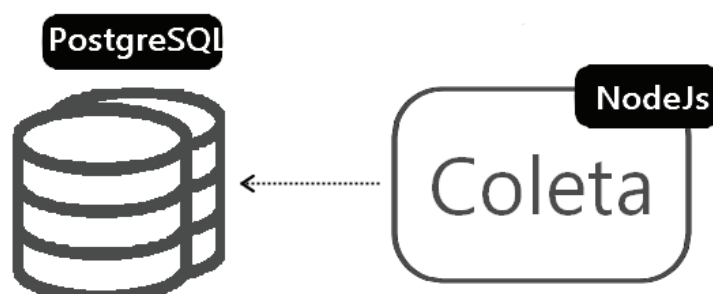


Figura 16. Componentes da camada de coleta de dados.

Os dados estão dispostos em 4 classes: Dispositivos, Gateways e Leituras, bem como o diagrama de classes apresentado na Figura 17. A classe Leituras representa as informações do pacote como um todo, as quais foram descritas na seção 2.6.3. A classe dispositivo representa cada um dos dispositivos finais da rede, os quais possuem uma localização fixa e dados de identificação, da mesma forma a classe Gateways. Essas classes representam também a estrutura em que os dados são persistidos no banco de dados. Ao chegar na aplicação Node.js, os dados do pacote são estruturados em uma query SQL e persistidos na base de dados PostgreSQL.

A tabela Leituras armazena todos os dados presentes no pacote, isso inclui os dados sensorizados e os dados referentes ao funcionamento da rede (RSSI, SNR, SF, etc.). Na tabela Gateways ficam armazenados os dados referentes ao cadastro manual de cada

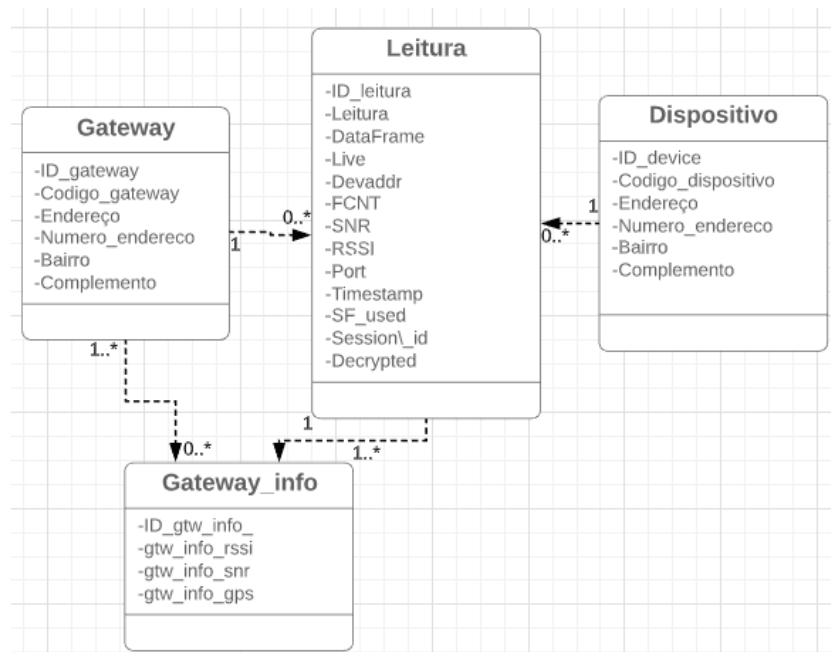


Figura 17. Diagrama de classes dos dados armazenados.

um dos 5 gateways, nela estão presentes os dados sobre localização e identificador único, entre outros. Da mesma forma a tabela Dispositivos armazena os dados (localização, identificador, etc..) de cada um dos 40 dispositivos. Já a tabela Gateway_info armazena dados sobre os Gateways que receberam determinado pacote e qual deles foi o responsável pelo envio ao servidor de rede. Uma vez persistidos, os dados ficam disponíveis para uso pela Camada Web.

2.6.5 Camada Web

A camada Web é responsável por toda a lógica de navegação e de interatividade idealizadas para a ferramenta. Várias tecnologias foram necessárias à construção da camada Web, destacando-se como principais os frameworks JSF (Java Server Faces) com o PrimeFaces e as tecnologias HTML5, Javascript, CSS3.

Esta camada foi desenvolvida seguindo o modelo MVC (Modelo, visão, controle), para facilitar a inclusão e exclusão de módulos. O modelo representa as informações (os dados) e as regras de negócio, a visão contém elemento de interface com o usuário, como textos, formulários, e o controle gerencia a comunicação entre o modelo e a visão. Então, basicamente, a visão foi posteriormente nomeada de Camada de Interação. A camada web está representada na Figura 18, onde é possível notar

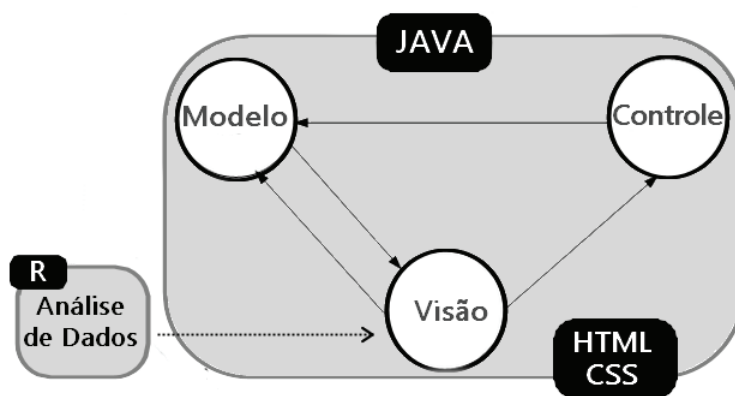


Figura 18. Componentes da camada de web.

2.6.6 Camada de Interação

Camada de Interação é onde fica disponibilizado o acesso a visualização dos dados e, como o nome diz, a interação com os componentes e módulos disponibilizados. Essa camada pode ser acessada via web Browser (navegador) e a partir dela é possível fazer uso de todos os módulos implementados na plataforma.



Figura 19. Forma de acesso a Plataforma Omni.

A Figura 20 mostra uma das telas da Plataforma, acessada pelo navegador Google Chrome.

2.6.7 Módulos da Plataforma Omni

A modularização é um mecanismo para melhorar a flexibilidade e compreensão de um sistema, que permite o encurtamento de tempo de desenvolvimento [39], pois possibilita decompor o sistema em "pedaços" menores, o que facilita a execução de testes e correções. Os módulos são partes da plataforma e representam as funcionalidades da Omni, eles são responsáveis por tarefas bem definidas e estão acoplados ao sistema, permitindo que a

Omni						
Início Leituras Automação Rede Downlink Usuários Sair Admin optim						
Intervalo: 01:59:57 Ordem: Data <input type="text"/> N° 100						
Leituras						
Data/Hora	Local	N° Device	Gateway	Leitura	RSSI	SNR
11/09/2019 09:53:15		Device 10	Gateway 4		-124 dBm	-4.5 dB
11/09/2019 09:51:52		Device 21	Gateway 2		-123 dBm	-6.5 dB
11/09/2019 09:47:48		Device 14	Gateway 2		-127 dBm	-12.8 dB
11/09/2019 09:46:52		Device 39	Gateway 2		-124 dBm	-4.2 dB
11/09/2019 09:46:09		Device 1	Gateway 1		-111 dBm	0.5 dB
11/09/2019 09:45:46		Device 305	Gateway 1		-82 dBm	10.2 dB
11/09/2019 09:45:31		Device 18	Gateway 2		-129 dBm	-7.5 dB
11/09/2019 09:45:27		Device 24	Gateway 1		-120 dBm	-1.0 dB
11/09/2019 09:44:41		Device 29	Gateway 2		-112 dBm	3.0 dB
11/09/2019 09:43:30		Device 16	Gateway 2		-109 dBm	2.2 dB
11/09/2019 09:42:41		Device 11	Gateway 2		-128 dBm	-7.8 dB
11/09/2019 09:41:07		Device 23	Gateway 2		-126 dBm	-7.0 dB
11/09/2019 09:40:46		Device 36	Gateway 1		-100 dBm	8.0 dB
11/09/2019 09:40:41		Device 28	Gateway 2		-129 dBm	-7.5 dB
11/09/2019 09:39:04		Device 22	Gateway 2		-127 dBm	-7.5 dB
11/09/2019 09:35:29		Device 20	Gateway 1		-77 dBm	10.0 dB

Figura 20. Tela de visualização dos pacotes mais recentes coletados pela rede.

plataforma execute as tarefas disponibilizada por seus módulos. Os módulos presentes na plataforma Omni são:

- **MÓDULO DE DISPOSITIVOS:** inclui dados de localização e configuração do dispositivos de rede;
- **MÓDULO DE GATEWAYS:** inclui dados de localização dos gateways;
- **MÓDULO DE DOWNLINK:** responsável por gerenciar e encaminhar comandos de configuração aos dispositivos da rede, bem como armazenar as configurações anteriormente utilizadas por cada dispositivo;
- **MÓDULO DE CONTROLE E NÍVEL DE ACESSO E PERMISSÕES:** responsável por gerenciar quais informações cada usuário têm acesso e também pelo cadastro de novos usuários;
- **MÓDULO DE ANÁLISE DE PERDAS E CONFIGURAÇÃO DE REDE:** responsável por disponibilizar diversas informações sobre a situação da rede em relação a perda de pacotes da rede.

2.6.7.1 Módulo de Análise de Perdas e Configuração de rede

Este módulo foi desenvolvido com intuito de extrair informações sobre os pacotes coletados, os quais são apresentadas na forma de *dashboards*⁸ (gráficos e listagens), na plataforma. Os códigos necessários para o funcionamento do módulo estão armazenados e

⁸Tipo de interface gráfica do usuário que geralmente fornece visualizações rápidas dos principais indicadores.

rodando em um Shiny Server, em uma máquina virtual mantida pela Amazon Web Services (AWS) e foi integrado a plataforma Omni via iFrame (trecho de código usado para incorporar outro documento no documento HTML atual). Uma visão geral do módulo é apresentada na Figura 21.

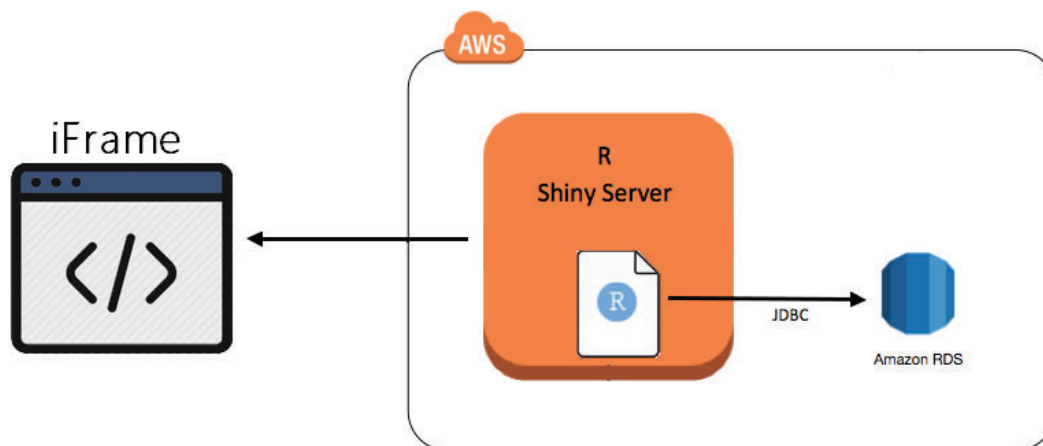


Figura 21. Estrutura do módulo.

2.7 TRABALHOS RELACIONADOS

Na literatura encontram-se alguns estudos referentes a avaliações de desempenho de redes LoRaWAN, muitos deles relacionados a protótipos ou a simulações.

No artigo “Caracterização de uma Rede Sem-fio de Baixa Potência e Longo Alcance para Internet das Coisas” [40], por exemplo, os autores fazem uma avaliação da tecnologia LoRa, analisando as métricas de desempenho de taxa de perda, a SNR, a potência do sinal e a vazão da rede em função dos diferentes fatores de espalhamento e diferentes distâncias entre os nós. Tais experimentos tiveram o intuito de compreender como as variações de distância estabelecem condições diferentes para o comportamento de uma rede LoRa e LoRaWAN. E da mesma forma que este trabalho, avaliam o desempenho do LoRa em função de diferentes configurações de modulação na camada física. Os experimentos realizados para a avaliação da tecnologia LoRa consistiram de medições práticas da taxa de perda, potência do sinal e relação sinal-ruído (SNR), em função de diferentes configurações do fator de espalhamento utilizado na modulação. O objetivo foi estabelecer parâmetros de funcionamento dos dispositivos usados e verificar a viabilidade de implementação para diferentes cenários da Internet das Coisas. O cenário utilizado para os experimentos e as medições para avaliação da tecnologia LoRa foi o campus da Ilha do Fundão da Universidade Federal do Rio de Janeiro (UFRJ). Dentro do campus foi estabelecido um ponto para a localização da unidade receptora (gateway) e cinco pontos para a localização das unidades transmissoras (dispositivos), de forma que, em diferentes momentos, as unidades transmissoras estivessem as distâncias de 0m, 500m, 1000m, 1500m e 2000m da unidade

receptora. Os resultados mostraram que o aumento do fator de 7 para 11 reduz a taxa de perda de 50% para 13%, para transmissões a distâncias de 2 km. Esse resultado os autores observaram que o SF 11 obteve melhor desempenho com relação a taxa de transmissão teórica.

Já no trabalho "LoRaWAN Performance Evaluation with Optimized Configuration"[41], os pesquisadores apresentam uma avaliação de desempenho de uma rede LoRaWAN, onde foram desenvolvidos três experimentos em cenário realista, e um estudo de escalabilidade realizado por simulação em computador. Os autores definiram como parâmetros, potência de transmissão, frequência, fator de espalhamento, largura de banda e taxa de codificação como os parâmetros a serem utilizados para a avaliação de desempenho da rede. A pesquisa consistiu no desenvolvimento de uma rede experimental, onde os autores desenvolveram os nós LoRa (dispositivos) e também o gateway. E como servidor de rede, utilizaram o The Things Network, um servidor de rede aberto mantido pela comunidade.

A partir do funcionamento dessa rede, foram avaliadas três métricas principais: I) desempenho da rede em termos de taxa de entrega de pacotes; II) consumo médio de energia por transmissão; III) energia média desperdiçada em colisão por dispositivos. A avaliação da rede foi feita usando simulação baseada em customização e se concentrou no desempenho geral da rede ao invés do desempenho por dispositivo. Os testes foram executados por 2 meses, com pacotes de 20 bytes sendo enviados de 10 em 10 minutos por dispositivo, onde cada um dos dispositivos foi submetido a 5 tipos de configurações nesse tempo, alterando sempre os parâmetros anteriormente citados.

Para fins de comparação com os experimentos, um caso padrão foi introduzido. O caso padrão é quando todos os dispositivos são configurados com a configuração comum LoRa, composto por: potência de transmissão de 14 dBm, frequência portadora de 915 MHz, fator de 12, largura de banda de 125 kHz e taxa de codificação de 4/5.

No primeiro experimento foi comparada a configuração que favorece robustez uma taxa de dados mais alta. A configuração mais lenta usou fator de espalhamento de 12, largura de banda de 125 kHz e taxa de codificação de 4/8 e a mais rápida usou fator de espalhamento 6, largura de banda de 500 kHz e taxa de codificação de 4/5. No segundo experimento foram incorporadas múltiplas randomizações de frequência que pode ser para minimizar a colisão. Nesta experiência, o dispositivo podia "escolher" aleatoriamente uma dentre três frequências para usar na transmissão do pacote. Para esse experimento, o caso mais lento do experimento 1 foi adotado.

Os três casos são comparados: o padrão, o mais lento e o mais lento com randomização de frequência. No terceiro experimento, foram comparadas duas otimizações simples. A primeira otimização, Opt1, configura os parâmetros em cada dispositivo para minimizar tempo de transmissão, que se baseia na sua distância até o gateway. A segunda otimização, Opt2, configura os parâmetros em cada nó para minimizar seu tempo de transmissão e consumo de energia. O experimento 3 demonstrou que ambos Opt1 e Opt2

obtem PDR significativamente mais alto até o ponto em que a diferença entre os casos quase imperceptível visualmente. Ao avaliarem os 3 experimentos juntamente com o caso padrão, a partir dos parâmetros definidos, os autores puderam evidenciar que o desempenho da rede LoRa é altamente dependente da configuração do fator de espalhamento, taxa de codificação e seleção correta da frequência.

No artigo "A Thorough Study of LoRaWAN Performance Under Different Parameter Settings"[42], os autores apresentaram uma análise sistemática dos impactos dos parâmetros ajustáveis do LoRaWAN em dois índices de desempenhos, nomeados UL-PDR e CPSR, que medem a eficácia da tecnologia na coleta de dados de dispositivos remotos. As métricas avaliadas foram desempenho em termos de taxa de entrega de pacotes, capacidade do sistema, eficiência e gasto de energia. Essas métricas foram avaliadas a partir da configuração do fator de espalhamento. A partir desse estudo, os autores observaram e explicaram o efeito de configurações de parâmetros diferentes em alguns cenários ilustrativos. A análise realizada foi baseada em simulação e revelou várias compensações, destacando algumas ineficiências no design do LoRaWAN padrão. Além disso, é demonstrado como ganhos significativos podem ser obtidos através da definição inteligente dos parâmetros do sistema e também evidenciado que a presença de confirmação das fontes de tráfego podem prejudicar significativamente o desempenho da rede.

Os trabalhos acima citados se diferem do trabalho desenvolvido, visto que este se trata da análise de uma rede implantada em um ambiente real e os demais trabalhos, em sua maioria, fazem análises de protótipos de redes LoRaWAN e/ou redes simuladas, outra diferença encontrada é a que nenhum dos trabalhos utiliza qualquer software ou linguagem estatística para avaliar as redes em questão, apenas são realizadas medições com parâmetros estáticos pré-definidos.

A seguir são detalhados os Materiais e Métodos utilizados para o desenvolvimento do módulo responsável pela avaliação de desempenho da rede.

3. MATERIAIS E MÉTODOS

Este capítulo apresenta o desenvolvimento do módulo de Análise de Dados e Configuração de Rede, que está presente na plataforma Omni. Todas as informações são apresentadas pelo módulo por meio de tabelas e gráficos, basicamente como um *dashboard*.

O módulo funciona a parte da plataforma, coleta os dados via conexão feita JDBC diretamente entre ele e o Banco de dados construído pela camada de Coleta de Dados. Ele está incorporado a Omni através de um iFrame inserido na camada Web da plataforma, mais precisamente na parte de visão. Seu processo de desenvolvimento é demonstrado a partir da Figura 22:



Figura 22. Ciclo de vida dos dados no módulo de análise.

Primeiramente foram selecionados os dados para o experimento, que consistem nos pacotes sensoriados pela rede. A limpeza e processamento dos dados foi feita visando excluir dados que não fossem utilizados posteriormente. Além dos dados de transmissão e qualidade do sinal, os pacotes continham dados sobre o projeto desenvolvido pela Optim, dados estes que estão sendo utilizado pela empresa, mas que não são levados em consi-

deração por este projeto. Portanto estes e alguns outros dados foram descartados a fim de minimizar o tamanho do montante de dados a serem explorados.

Os parâmetros foram definidos a partir dos mais utilizados em avaliações de redes LoRaWAN presentes na literatura, mas também foram selecionados a partir dos objetivos do trabalho. Na parte de definição de técnicas, foi cogitada a utilização de outras técnicas como Regressão Logística e Regressão Linear Múltipla, porém, a Análise de Correspondência Múltipla (MCA) e a Correlação se adequaram melhor ao que se desejava analisar a partir dos dados, pois seu uso permite explorar relações entre linhas e colunas em uma tabela de classificação de diversos fatores.

Após a implementação do código em R, os gráficos e tabelas gerados foram utilizados para criar um dashboard onde são apresentadas as informações obtidas. Hoje o módulo opera disponibilizando uma análise da taxa de perda de pacotes por período, porém para exemplificar seu funcionamento e desenvolver este trabalho, foram realizados experimentos em um período pré-determinado. Os experimentos realizados consistem em processamento e mineração de aproximadamente 74 mil pacotes coletados pela rede entre 27 de abril de 2019 e 20 de Agosto de 2019 (116 dias). Esses pacotes contém o valor de cada um dos parâmetros utilizados na transmissão única de cada pacote sensorado. Assim, este capítulo descreve além da avaliação e experimentação, o projeto de software do módulo de análise de dados desenvolvido para compor a plataforma Omni. O objetivo é permitir que a Optim possa, a partir do módulo, constantemente avaliar a situação da rede visando obter maior efetividade e por consequência uma menor taxa de perda de pacotes.

É possível exemplificar a metodologia utilizada por meio de um diagrama de Ishikawa (Figura 23), o qual consiste na identificação de causas para o evento determinado, que nesse caso é a perda de pacotes.

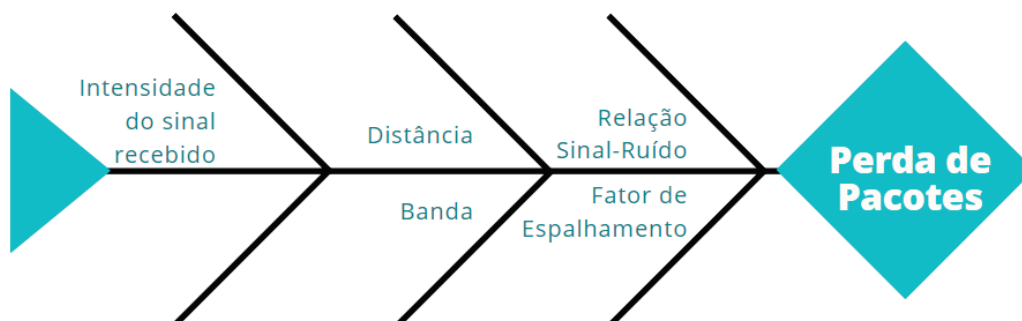


Figura 23. Diagrama de Ishikawa utilizado para apresentar os parâmetros utilizados na avaliação de performance da rede em relação a perda de pacotes.

3.1 TAXA DE PERDA DE PACOTES

Para avaliar a efetividade da rede, foi utilizada a Taxa de Perda de Pacotes (TPP) como métrica principal. Portanto todas as informações presente no módulo de Análise de Dados e Configuração de Rede está diretamente relacionada a TPP.

O protocolo LoRaWAN permite o envio de dois tipos de pacotes de informação - tanto de uplinks (dispositivo -> servidor de rede) como de downlinks (servidor de rede -> dispositivo) - os pacotes confirmados e não-confirmados. A diferença entre os dois é que o do tipo confirmado exige uma resposta de confirmação do recebimento por parte do destinatário da mensagem, enquanto o tipo não-confirmado não exige.

Na rede Optim, todos os pacotes enviados são do tipo confirmado, permitindo uma confiabilidade maior na comunicação. Os dispositivo foram configurados para 10 tentativas de envio de cada pacote, o que quer dizer que um dispositivo envia o pacote e cessa as tentativas de envio quando da confirmação de recebimento ou quando atingir um número máximo de tentativas. Além disso, uma função muito importante presente na comunicação LoRaWAN é o envio do FCNT (contador de frames) de cada pacote no preâmbulo da mensagem. Isso significa que o próprio protocolo controla os números de envios bem-sucedidos sem necessitar a intervenção do *firmware*⁹ do dispositivo ou mesmo do gerente da rede e sem gastar mais energia por isso.

No caso da rede Optim, a TPP está diretamente relacionada ao parâmetro FCNT (padrão do protocolo LoRaWAN), o qual é o contador de frames de uma sessão de envio de pacotes a qual cada dispositivo está submetido. O contador parte de 0 a N, sendo incrementado em 1 cada vez que um pacote é encaminhado para o servidor de rede e o servidor confirma o recebimento do mesmo. Sempre que o dispositivo envia um pacote para o SR, o mesmo incrementa o FCNT Down e o devolve ao dispositivo como resposta, se o FCNT Down atual for maior que o FCNT Down recebido pelo dispositivo anteriormente, então quer dizer que o dispositivo deve incrementar o FCNT UP. fazendo com que seja criada a sequência de FCNT, e sempre que um pacote é perdido existe uma lacuna entre o FCNT do pacote recebido e o anterior. Por exemplo, se o pacote atual possui FCNT = 7, o pacote anterior a este deve possuir o FCNT = 6, e caso isso não ocorra admite-se então que o pacote anterior foi perdido.

Assim, sabendo o número do FCNT de cada pacote, pode-se realizar o cálculo do percentual de perda de pacote, expresso pela equação 4

$$\% \text{ de perda de pacote} = \frac{\text{Contagem de pacotes perdidos} \times 100}{\text{total de pacotes}} \quad (4)$$

⁹Firmware é uma classe específica de software de computador que fornece controle de baixo nível para o hardware específico do dispositivo.

Neste ponto, ressalta-se que o módulo não visa encontrar os motivos pelas perdas de pacotes na rede, e sim avaliar qual a influência de cada um dos parâmetros a serem avaliados em relação a essas perdas. As informações apresentadas pelo módulo que estão relacionadas a TPP foram separadas em dois grupos: o primeiro grupo de informações da rede por dispositivo e o segundo de informações da rede completa (todos os dispositivos). Foi necessário fazer essa divisão em grupos, pois algumas métricas como a avaliação das configurações precisam ser aplicadas a cada dispositivo, visto que cada uma delas se encontra em condições diferentes uns dos outros.

No grupo 1 estão:

- Medições práticas de taxa de perda por SF: avaliação da TPP para cada os valores de SF utilizado.
- Avaliação da qualidade do sinal: utilização do Método de Análise de Correspondência para avaliar a potência do sinal, o qual pode influenciar na TPP;
- Avaliação de efetividade a partir da taxa de perda: medições da TPP da rede como um todo;
- Influência dos parâmetros selecionados na taxa de perda: avaliação dos parâmetros de transmissão na TPP por dispositivo;

Já o grupo 2 é composto por:

- Medições práticas de taxa de perda: avaliação de efetividade na entrega dos pacotes de cada um dos dispositivos;
- Avaliação das configurações utilizadas por dispositivo: avaliação de todas as combinações de parâmetros possíveis disponíveis para cada dispositivo.

3.2 MÉTODO DE ANÁLISE DE CORRESPONDÊNCIA

A Análise de Correspondência Simples (ACS) é uma técnica multivariada para análise exploratória de dados categorizados em tabelas de contingência de até duas variáveis. A ACS considera medidas de correspondência entre as variáveis de linhas e colunas [43], sendo possível utilizar seus procedimentos numéricos e algébricos independente do contexto. Assim sendo, este método se adéqua para o uso nesse trabalho, para avaliação dos parâmetros de transmissão de uma rede LoRaWAN, pois permite a identificação de diversos fatores pertinentes ao fenômeno de estudo (nesse caso a perda de pacotes).

A ACS permite investigar as relações entre as duas variáveis, por meio da associação entre suas respectivas categorias e possui múltiplas características que a distingue de

outras técnicas de análise. Uma delas é sua natureza multivariada, que permite investigar relações que não são facilmente percebidas a partir da comparação de variáveis. Porém, a AC possui algumas exigências: os dados devem estar dispostos em tabelas retangulares e devem obrigatoriamente ser positivos.

Para a análise de duas variáveis de entrada, é utilizada a AC Simples. Porém, quando a tabela possui mais variáveis, existe uma forma mais geral da AC chamada Análise de Correspondência Múltipla (ACM), a qual foi utilizada nesse trabalho, onde o objeto de estudo possui múltiplas variáveis.

3.2.1 Análise de Correspondência Múltipla

A análise de correspondência múltipla é considerada uma extensão da ACS quando se deseja envolver mais de duas variáveis na análise e sua dimensionalidade está ligada às categorias de cada variável. A aplicação desse método é feita em tabelas multidimensionais, onde as linhas representam os objetos observados e as colunas as diferentes categorias das variáveis estudadas. A representação gráfica pode ser feita tanto para os indivíduos (podendo ser utilizado em Sistemas de Recomendação para analisar similaridade de perfis), quanto para as variáveis, para as categorias (que é o caso desse trabalho) e para as categorias e indivíduos [44].

A ACM permite a visualizar relações que normalmente não seriam apontadas por simples comparações de variáveis par a par. E também tem a capacidade de ordenar e incorporar um alto número de indicadores categóricos, fazendo possível com que haja uma redução da dimensionalidade das entradas [45].

Um dos objetivos da ACM é classificar os dados em diferentes tipos a partir do conceito de semelhança. Assim, nesse caso os pacotes são considerados similares quando apresentam características em comum, ou seja, possuem uma representação próxima no espaço euclidiano gerado pela ACM [45].

Essencialmente, enquanto na ACS a tabela de contingência é usada, na ACM a matriz de dados pode ser representada por uma estrutura chamada de matriz indicadora z . Segundo [46], também é possível representar a matriz de dados através da matriz de Burt, $B = Z'Z$, que consiste em transformar a matriz retangular em uma matriz quadrada formada por tabelas de contingência bidimensionais onde são realizados todos os cruzamentos possíveis entre as variáveis envolvidas. Com relação a matriz indicadora, as linhas representam os objetos (ou no caso pacotes) e nas colunas as categorias de cada uma das variáveis. Cada linha contém todos os códigos correspondentes às categorias atribuídas a um pacote para cada uma das variáveis observadas. As categorias devem ser mutuamente exclusivas e exaustivas, ou seja, cada pacote deve possuir uma e somente uma categoria para cada variável. Ambas as possibilidades, matriz de Burt e matriz indi-

cadras são equivalentes, pois os gráficos resultantes são análogos e apenas divergem na escala das variáveis [46]. A análise de correspondência múltipla utilizada nesse trabalho foi desenvolvida utilizando matriz indicadora.

3.2.2 Matriz indicadora

Na matriz indicadora os elementos são categorizados e codificados como variáveis *dummy*¹⁰ na matriz indicadora, ou seja, 1 para a categoria escolhida como resposta de uma variável e 0 para as demais categorias da mesma variável, conforme ilustrado pela Figura 24.

Para explicar a ACM, foram adotadas as notações, conceitos e explicações dos mesmos por baseada em [44]:

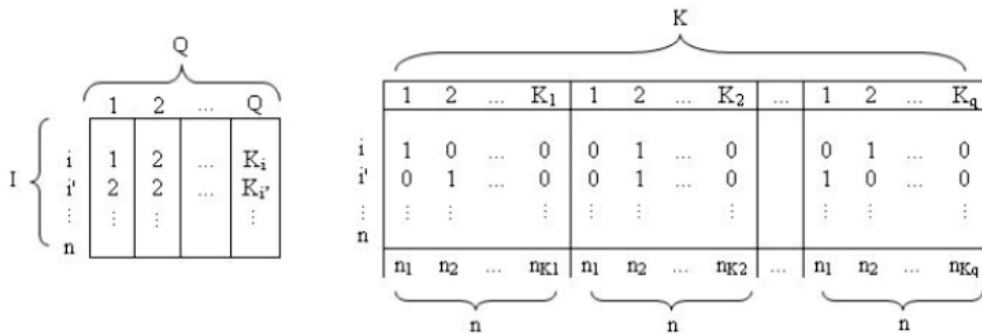


Figura 24. Representação de uma matriz indicadora [44]

Onde i representa o grupo de n objetos, Q o grupo de variáveis, K_q o número de categorias da questão q , K o número de total de categorias e K_i representa o grupo das Q categorias do objeto i .

Na Tabela 1 está representado um exemplo de matriz indicadora, onde são considerados 6 pacotes e três variáveis - SF, SNR e PERDA (V_1, V_2 e V_3) - portanto $Q = 3$, sendo três categorias na primeira variável ($V_1 = V_1 (7), V_2 (8), V_3 (9)$), portanto $KV_1 = 3$, três na segunda variável ($V_2 = (V_4 (B), V_5 (N), V_6 (R))$, logo ($KV_2 = 3$), e duas categorias na terceira variável ($V_3 = v_7 (S), v_8 (N)$), portanto $KV_3 = 2$. P tem-se um total de $k = 8$ categorias. Utilizando a matriz indicadora (Tabela 1), pode-se alcançar a resposta de todas as observações, por exemplo, a observação 1 tem as seguintes respostas: $V_1 = KV_3$; $V_2 = v_4$; $V_3 = v_7$, e assim sucessivamente são obtidas as respostas de todas as observações [47].

Seja $i = 1, 2, 3, \dots, n$ e $k = 1, 2, 3, \dots, k$ a linha marginal da matriz indicadora z , é a soma das observações na k -ésima coluna, que representa o número de objetos que possuem a categoria k , como apresentado na Equação 5:

¹⁰variáveis binárias (0 ou 1) criadas para representar uma variável com duas ou mais categorias.

Pacotes	Variáveis e suas instâncias								Total Linha
	SF (V_1)			SNR (V_2)			PERDA (V_3)		
	7 (V_1)	8 (V_2)	9 (V_3)	B (V_4)	N (V_5)	R (V_6)	S (V_7)	N (V_8)	
1	0	0	1	1	0	0	1	0	3
2	1	0	0	0	1	0	1	0	3
3	1	0	0	0	1	0	1	0	3
4	0	1	0	0	0	1	0	1	3
5	0	0	1	0	1	0	0	1	3
6	0	0	1	0	0	1	1	0	3
Total	2	1	3	1	3	2	4	2	18

Tabela 1. Exemplo de matriz indicadora utilizando as variáveis dos pacotes.

$$Z_{+k} = \sum_{i=1}^n Z_{ik} \quad (5)$$

A coluna marginal de z é a soma das observações da i -ésima linha, que representa o número de variáveis, sendo $z_{i+} = Q \forall i = 1, 2, 3, \dots, n$. O total geral representa o número de objetos multiplicado pelo número de variáveis [47]. Respectivamente ambas as equações são dadas por:

$$Z_{i+} = \sum_{k=1}^K Z_{ik} \quad (6)$$

$$Z_{++} = \sum_{i=1}^n Z_{i+} = \sum_{k=1}^K Z_{+k} = nQ \quad (7)$$

O vetor de frequências relativas marginais é denominado como o vetor contendo todas as frequências totais, podendo ser de linha ou de coluna. Esses vetores são conhecidos como massas e servem para normalizar as contribuições das linhas ou colunas, respectivamente, em função da distância euclidiana ponderada. E a massa de um elemento $i \in I$ é o quociente do total da i -ésima linha pelo total geral [45]:

$$r_i = \frac{z_{i+}}{z_{++}} = \frac{Q}{nQ} = \frac{1}{n} \quad (8)$$

Dependendo da quantidade de objetos, sendo a mesma para todo $i = (1, 2, 3, \dots, n)$. A massa de um elemento $k \in K$ é o quociente do total da k -ésima coluna pelo total geral, que corresponde ao quociente do número de objetos que utilizaram a k -ésima categoria pelo número total de objetos multiplicado pelo número de variáveis [44]:

$$c_k = \frac{z_{+k}}{z_{++}} = \frac{n_k}{nQ} \quad (9)$$

Um perfil do vetor de linha i que tem k elementos e um perfil do vetor de categoria k que tem i elementos:

$$r_i^k = \frac{z_{ik}}{z_{i+}} = \frac{z_{ik}}{Q} \quad (10)$$

$$c_I^k = \frac{z_{ik}}{z_{+k}} = \frac{z_{ik}}{nk} \quad (11)$$

Segundo Roux e Rouanet [44], as colunas podem ser consideradas como uma primeira quantificação das categorias das variáveis qualitativas uma vez que, os valores são pesos relativos de cada categoria dentro da sua respectiva variável. Assim a marginal linha corresponde às massas dos elementos $k \in K$, e o perfil marginal da coluna corresponde às massas dos elementos $i \in I$.

A ACM é geralmente baseada na observação de "nuvens" de pontos, as quais são definidas como um conjunto finito de pontos em um espaço [44]. A partir do procedimento da ACM é possível gerar dois tipos de nuvens de pontos. A primeira referente às linhas, chamada nuvem dos i pontos dos objetos, e a segunda referente às colunas, nuvem dos K pontos das categorias. Esse trabalho utiliza ambas para demonstrar a relação entre as variáveis selecionadas.

3.2.2.1 Nuvem de pontos

A nuvem de pontos dos objetos é estabelecida a partir do conjunto de cada linha, cada um associado à sua própria massa. Do mesmo modo, a nuvem de pontos das categorias é criada a partir do conjunto de cada coluna [48]. A nuvem de pontos sempre é projetada em um plano que possa preservar ao máximo a distância entre eles, mantendo maior parte da informação de forma original. O ponto médio G da nuvens de pontos, se resume no somatório de todas as distâncias entre um ponto P qualquer e todos os outros pontos da nuvem e dividido pelo total de pontos. Ou seja G pode ser definido como as médias das coordenadas dos pontos, onde o ponto médio não depende da escolha do ponto P . Um exemplo de nuvem de pontos é dado a partir da Figura 25.

Segundo Roux e Rouanet [44] a distância entre dois objetos é encontrada pelas diferentes escolhas de categoria para cada variável q denotada por $d_q(i, i')$, se os dois objetos utilizaram a mesma categoria tem-se que $d_q(i, i') = 0$. Ou seja, a distância entre dois objetos pode ser calculada somente quando ambos os pontos possuem diferentes categorias de uma mesma variável. Quando é usada a mesma categoria para ambos, os dois pontos coincidem na nuvem de pontos, e a distância entre eles é nula. Já a distância ao quadrado entre dois objetos que escolheram categorias diferentes, é dada pela Equação 12:

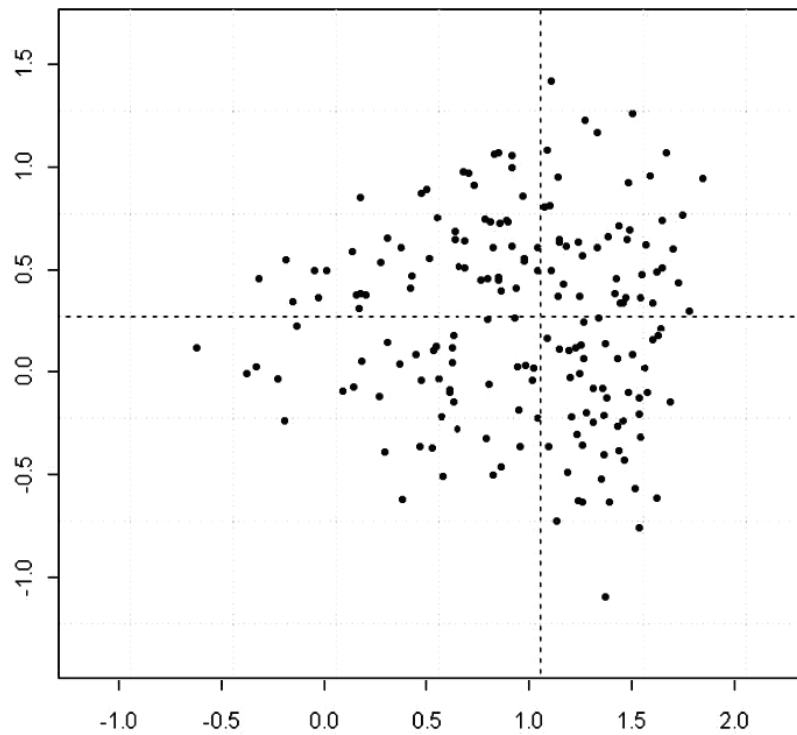


Figura 25. Exemplo de nuvem de pontos com objetos classificados a partir variáveis.

$$d_2^q(i, i') = \frac{1}{fk} + \frac{1}{fk'} \quad (12)$$

Onde fk representa a frequência relativa de objetos na categoria k , ou seja, $fk = \frac{n_k}{n}$, sendo n_k o número de objetos na categoria k e n o número total de objetos.

Para Roux e Rouanet [44], quanto menor a frequência das diferentes categorias, maior é a distância entre objetos. Sendo assim, um ponto qualquer M^i ficará localizado longe do centro da nuvem de pontos. O quadrado médio da distância total entre i e i' é definido pela Equação 13:

$$d^2(i, i') = \frac{1}{Q} \sum_{q=1}^Q d_q^2(i, i'). \quad (13)$$

A linhas marginais, ou seja, o peso de um ponto M^k referente a uma categoria k é n_k , e a soma dos pesos para cada categoria de uma dada variável é n e para todas, o total de categorias é nQ . A distância ao quadrado entre duas categorias é dada pela Equação 14:

$$d^2(M^k, M^{k'}) = \frac{n_k + n_{k'} - 2n_{kk'}}{n_k n_{k'} / n}, \quad (14)$$

sendo:

- M^k com $k = 1, 2, 3, \dots, K$ os pontos da nuvem de categorias;
- n_k o peso do ponto M_k (número de objetos que utilizaram a categoria k);
- n'_k o peso do ponto M'_k (número de objetos que utilizaram a categoria k);
- $nk k'$ o número de objetos que utilizaram ambas categorias k e k'

Tem-se que $n_k k' = 0$ quando k e k' são duas diferentes categorias de uma mesma variável. Segundo Roux e Rouanet[44] a distância entre M_k e $M_{k'}$ será menor quando o quanto mais categorias k e k' forem escolhidas pelo mesmo objeto, e, quanto menor a frequência da categoria k , mais o ponto M_k se distancia do centro da nuvem.

A Inércia é a medida de dispersão da nuvem de pontos em relação a um ponto qualquer da nuvem de pontos, ela é calculada pela média dos quadrados das distâncias entre os pontos de uma nuvem e um ponto qualquer, menos o quadrado da distância entre o ponto médio e o ponto escolhido.

A nuvem de categorias tem a mesma inércia da nuvem de indivíduos e a frequência de objetos em uma certa categoria influencia em sua contribuição, ou seja, categorias menos frequentes contribuem mais para a inércia global e quanto mais categorias uma variável possui, mais esta contribui para a variância da nuvem [49]. E a nuvem de pontos também é projetada em planos que possuam a capacidade de manter a maior parte da informação original. A projeção de uma nuvem corresponde à projeção ortogonal de seus pontos, portanto, a inércia total de uma nuvem ortogonalmente projetada é sempre menor ou igual á inércia total da nuvem inicial [44].

A nuvem de pontos é projetada sobre os chamados eixos principais, que corresponde a sua projeção em eixos arbitrários, através das distâncias entre os pontos. A inércia de cada eixo é restituída pelos autovalores λ determinados pela fatoração da matriz indicadora através do método de decomposição por valores singulares (DVS). A soma dos autovalores λ é igual a inércia da nuvem de pontos.

As coordenadas principais dos pontos definem a nuvem referida aos seus eixos principais. De acordo com Loesch [49], a coordenada principal de um ponto M^i da nuvem de objetos em relação ao eixo principal é denotada por y_i^i . Para a nuvem de categorias, as coordenadas são definidas da mesma forma, sendo a coordenada principal de M^k definida como y_i^k .

Para cada eixo principal tem-se que a média das coordenadas é nula e a variância é igual ao autovalor. Assim, tem-se as Equações 15 e 16:

$$\sum \frac{1}{n} y_i^i = 0; \sum \frac{1}{n} (y_i^i)^2 = \lambda l \quad (15)$$

$$\sum p_k y_i^k = 0; \sum p_k (y_i^k)^2 = \lambda l. \quad (16)$$

Como os eixos principais, as coordenadas são determinadas pela (DVS) da matriz indicadora.

A contribuição de um ponto a um determinado eixo consiste na importância desse ponto ao eixo, isto é, o quanto da inércia do eixo é devido ao ponto. Essa contribuição é calculada por meio da multiplicação do peso p e sua coordenada nesse eixo dividido pela inércia do eixo. Sendo y a coordenada relativa ao eixo de inércia λ , a contribuição do ponto a um eixo, segundo [44], é dada pela Equação 17:

$$Ctr = \frac{(py^2)}{\lambda}. \quad (17)$$

Essa contribuição indica a quantidade da inércia do eixo explicada por um ponto, logo, define-se a contribuição do ponto referente a nuvem de indivíduos M^i e a contribuição do ponto das nuvens de categorias (M^k), respectivamente por:

$$Ctr = \frac{\frac{1}{n}(y^i)^2}{\lambda} \quad (18)$$

$$Ctr_k = \frac{\frac{f_k}{Q}(y^k)^2}{\lambda} \quad (19)$$

A visualização das duas nuvens (objetos e categorias) de pontos é trabalhada separadamente. Para visualização das nuvens em um único espaço, utilizam-se as fórmulas de transição que tem como finalidade ligar as coordenadas principais referentes a nuvem de objetos y^i e as coordenadas principais referentes a nuvem de categorias y^k , o que permitiria analisar relações diferentes entre a nuvem de objetos e a nuvem de categorias. São elas:

$$y^i = \frac{1}{\sqrt{\lambda}} \sum_{k \in K_i} \frac{y^k}{Q}. \quad (20)$$

e

$$y^k = \frac{1}{\sqrt{\lambda}} \sum_{i \in I_k} \frac{y^i}{n_k}. \quad (21)$$

A primeira fórmula de transição permite calcular as coordenadas principais de um indivíduo através das categorias escolhidas por ele, e, com a segunda fórmula de transição, pode-se encontrar as coordenadas principais de uma categoria escolhida por um grupo de objetos. Na Figura 26 tem-se de forma sucinta o resumo da análise de correspondência simples e múltipla.

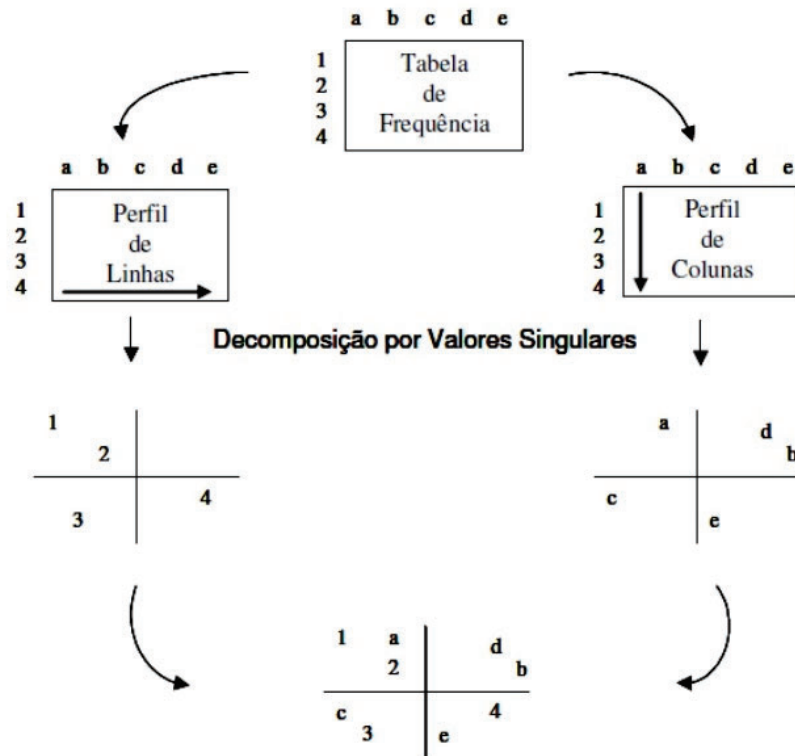


Figura 26. Resumo da Análise de Correspondência Múltipla [50]

3.2.2.2 Considerações de dados para Nuvem de Categorias na Análise de correspondência múltipla

Como o nome já diz, os dados utilizados pela nuvem de categorias são estritamente categóricos, ou seja, contêm um número finito de categorias ou grupos distintos, podendo não ter uma ordem lógica. Esses dados podem ser do tipo categóricos brutos, onde existem uma ou mais colunas de classificação com cada linha representando uma observação. Também podem representar categorias de forma numérica, de texto ou data/hora, ou até mesmo na forma de variáveis indicadoras, onde cada linha também representa uma observação e há uma coluna indicadora para cada nível de categoria. Os dados nesta forma consistem de 0s, exceto 1s que identificam as categorias às quais as observações pertencem.

As variáveis consideradas nesse trabalho foram:

- RSSI - representa a qualidade relativa de um sinal entre o gateway e o dispositivo.
- SNR - razão da potência do sinal (entre o gateway e o dispositivo) e a potência do ruído sobreposto ao sinal.
- Distancia - distância de transmissão do pacote, entre o gateway e o dispositivo.
- perda - variável que indica onde houveram perdas ou não.

- SF - define a sensibilidade e o alcance do sinal.

No caso deste trabalho, os dados utilizados em sua maioria eram dados contínuos, como o RSSI, o SNR e a distância. Outros como o SF já estavam dispostos como categorias. É possível demonstrar, por meio da Figura 27, os dados antes dos devidos ajustes e transformação necessários para o uso da MCA.

	numero_device	rssi	snr	sf_used	perda	distancia
1	14	-129	-9.8	10	0	4837.399
2	14	-133	-14.8	10	0	4837.399
3	14	-129	-9.8	10	0	2918.443
4	14	-132	-12.5	10	0	2918.443
5	14	-117	2.2	10	1	1168.290
6	14	-128	-6.8	10	0	2918.443
7	14	-129	-10.2	10	0	2918.443
8	14	-133	-13.0	10	0	2918.443
9	14	-133	-13.5	10	0	4837.399
10	14	-120	-10.0	10	0	2918.443
11	14	-124	-6.0	10	0	2918.443
12	14	-129	-10.8	10	0	2918.443
13	14	-127	-7.0	10	0	2918.443
14	14	-125	-7.2	10	0	2918.443
15	14	-129	-10.0	10	0	2918.443
16	14	-127	-7.5	10	0	2918.443
17	14	-129	-10.0	10	0	2918.443
18	14	-119	3.0	10	0	1168.290
19	14	-126	-6.5	10	0	2918.443
20	14	-124	-5.8	10	1	2918.443

Figura 27. Dados antes da categorização.

Nota-se também a presença da variável *perda*, a qual representa a contagem dos pacotes perdidos, onde 0 significa que não houve perda entre o pacote atual e o anterior e 1 representa que houve perda de pacote e a ausência das variáveis Largura de Banda (BW) e Taxa de código (CR). Ambas as variáveis foram descartadas das análises, pois possuem valores estáticos, BW = 125kHz e CR = 4/5, Não sendo necessário utilizá-las na avaliação das perdas, visto que seu valor não é alterado em nenhum momento.

3.2.3 Correlação

A análise de correlação busca aferir o grau de associação entre duas ou mais variáveis a partir de uma amostra selecionada. Para Triola [51], correlação é um instrumento utilizado para determinar a existência de significância na relação entre as variáveis a partir de parâmetros estatísticos. Há diferentes formas de correlação, porém neste trabalho será utilizada a correlação múltipla.

Tal técnica estatística é de natureza multivariada, assim como a MCA, e busca aferir a correlação entre mais de duas variáveis. De acordo com Penha et. al [52], a busca de associação entre variáveis é frequentemente um dos propósitos das pesquisas empíricas. A possível existência de relação entre variáveis orienta análises, conclusões e evidenciação de achados da investigação.

O grau de relação existente entre as variáveis estudadas a ferramenta estatística utilizada é o Coeficiente de Correlação de Pearson que consiste em um modelo para medir o nível de associação entre as variáveis, para isso o coeficiente pode assumir valores entre -1 ou $+1$, que em termos percentuais também pode ser representado entre -100% e $+100\%$, infere-se que quanto maior associação, representada pelo coeficiente “R”, ou seja, quanto mais próximo de $+1$ (induz que há correlação positiva perfeita entre as variáveis) ou -1 (infere-se que há uma correlação negativa perfeita entre as duas variáveis, ou seja, se uma aumenta a outra sempre diminui) [53].

O valor coeficiente de Pearson permite classificar a correlação, desse modo, Dancey & Reidy [54], classificaram os escores em relação ao valor atribuído a R, para $r = 0,10$ até $0,30$ apresenta-se uma correlação fraca; $r = 0,40$ até $0,6$ tem-se uma correlação moderada e para $r = 0,70$ até 1 há uma correlação forte). Correlações menores que isso nem mesmo são consideradas.

3.3 AJUSTE E TRANSFORMAÇÃO DAS VARIÁVEIS

Visto que podem ser implementadas em diferentes tipos de ambientes, com condições que também se diferem, as redes LoRaWAN em geral são como organismos únicos, os quais possuem suas particularidades em questão de funcionamento, alcance e configurações. Por isso não existe um consenso sobre a categorização dos parâmetros de RSSI e SNR entre bom ou ruim.

Portanto, para a categorização dos dados utilizados neste trabalho, assumiu-se o pior caso (dentro os dados coletados) para o valor de RSSI e SNR com cada SF utilizado, como péssimo/ruim e o melhor caso como ótimo/bom. A partir desses dois extremos, foi calculada a diferença entre eles e dividida por 5, para que fossem obtidas 5 categorias de ambos os parâmetros. As categorias definidas para cada nível dos parâmetros estão

dispostas nas tabelas a seguir (2, 3, 4, 5), sendo divididas por SF, pois sabe-se que o SF7 possui um alcance menor que o SF10, por exemplo, o que fez com que fosse necessário categorizar os dados conforme seu SF.

Fator de Espalhamento 7			
RSSI (dBm)	Categoria RSSI	SNR (dB)	Categoria SNR
-131 a -114	SINAL MUITO FRACO	-11 a -7	MUITO CORROMPIDO
-114 a -100	SINAL FRACO	-7 a -3	CORROMPIDO
-100 a -87	SINAL RAZOÁVEL	-3 a 3	RAZOAVELMENTE CORROMPIDO
-87 a -75	SINAL BOM	3 a 7	POUCO_CORROMPIDO
-75 a -60	SINAL ÓTIMO	7 a 11	NADA CORROMPIDO

Tabela 2. Tabela de valores de RSSI e SNR categorizados para o SF 7.

Fator de Espalhamento 8			
RSSI (dBm)	Categoria RSSI	SNR (dB)	Categoria SNR
-136 a -114	SINAL MUITO FRACO	-13 a -9	MUITO CORROMPIDO
-114 a -101	SINAL FRACO	-9 a -5	CORROMPIDO
-101 a -88	SINAL RAZOÁVEL	-5 a 5	RAZOAVELMENTE CORROMPIDO
-88 a -76	SINAL BOM	5 a 9	POUCO_CORROMPIDO
-76 a -60	SINAL ÓTIMO	9 a 13	NADA CORROMPIDO

Tabela 3. Tabela de valores de RSSI e SNR categorizados para o SF 8.

Fator de Espalhamento 9			
RSSI (dBm)	Categoria RSSI	SNR (dB)	Categoria SNR
-137 a -117	SINAL MUITO FRACO	-16 a -10	MUITO CORROMPIDO
-117 a -103	SINAL FRACO	-10 a -4	CORROMPIDO
-103 a -89	SINAL RAZOÁVEL	-4 a 4	RAZOAVELMENTE CORROMPIDO
-83 a -75	SINAL BOM	4 a 10	POUCO_CORROMPIDO
-75 a -60	SINAL ÓTIMO	10 a 16	NADA CORROMPIDO

Tabela 4. Tabela de valores de RSSI e SNR categorizados para o SF 9.

Fator de Espalhamento 10			
RSSI (dBm)	Categoria RSSI	SNR (dB)	Categoria SNR
-141 a -124	SINAL MUITO FRACO	-18 a -11	MUITO CORROMPIDO
-124 a -107	SINAL FRACO	-11 a -4	CORROMPIDO
-107 a -90	SINAL RAZOÁVEL	-4 a 4	RAZOAVELMENTE CORROMPIDO
-90 a -75	SINAL BOM	4 a 11	POUCO_CORROMPIDO
-75 a -60	SINAL ÓTIMO	11 a 18	NADA CORROMPIDO

Tabela 5. Tabela de valores de RSSI e SNR categorizados para o SF 10.

Já na categorização da variável *distância*, os dados foram categorizados em 7 categorias, tendo em vista que a menor distância de comunicação registrada foi de 15m e a maior distancia de comunicação foi de aproximadamente 7300m e que os dispositivos e

gateways são estáticos, o que faz com que esses números também se mantenham estáticos para cada dispositivo. O resultado da categorização dessa variável é apresentado a partir da Tabela 6;

Categorização de Distancias (metros)		
>=	<=	Categoria
1	1000	D1
1001	2000	D2
2001	3000	D3
3001	4000	D4
4001	5000	D5
5001	6000	D6
6001	>7000	D7

Tabela 6. Tabela da distancia categorizada.

Sendo assim, todos os dados utilizados foram categorizados para que fosse possível obter os resultados da Análise de Correspondência Múltipla, como pode ser visualizado na Figura 28.

	numero_device	nivel_sf	nivel_perda	nivel_snr	nivel_rssi	nivel_distancia
1	30	SF_8	NÃO_PERDA	SNR_NADA_CORROMPIDO	RSSI_SINAL_BOM	DISTANCIA_3001_4000
2	2	SF_9	NÃO_PERDA	SNR_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_1001_2000
3	25	SF_10	NÃO_PERDA	SNR_MUITO_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_1_1000
4	13	SF_10	NÃO_PERDA	SNR_RAZOAVELMENTE_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_1_1000
5	1	SF_10	NÃO_PERDA	SNR_RAZOAVELMENTE_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_4001_5000
6	32	SF_10	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_BOM	DISTANCIA_5001_6000
7	35	SF_9	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_2001_3000
8	22	SF_10	NÃO_PERDA	SNR_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_1001_2000
9	26	SF_8	NÃO_PERDA	SNR_RAZOAVELMENTE_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_1_1000
10	12	SF_10	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_RAZOAVEL	DISTANCIA_2001_3000
11	7	SF_9	NÃO_PERDA	SNR_RAZOAVELMENTE_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_2001_3000
12	27	SF_7	NÃO_PERDA	SNR_RAZOAVELMENTE_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_3001_4000
13	31	SF_10	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_1_1000
14	36	SF_9	NÃO_PERDA	SNR_NADA_CORROMPIDO	RSSI_SINAL_ÓTIMO	DISTANCIA_4001_5000
15	17	SF_9	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_1001_2000
16	38	SF_10	NÃO_PERDA	SNR_RAZOAVELMENTE_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_3001_4000
17	19	SF_10	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_RAZOAVEL	DISTANCIA_2001_3000
18	40	SF_7	NÃO_PERDA	SNR_POUCO_CORROMPIDO	RSSI_SINAL_FRACO	DISTANCIA_1_1000
19	34	SF_9	NÃO_PERDA	SNR_CORROMPIDO	RSSI_SINAL_MUITO_FRACO	DISTANCIA_3001_4000

Figura 28. Dados de cada pacote já categorizados.

3.4 FERRAMENTAS DE ANÁLISE DOS DADOS

O ambiente de desenvolvimento chamado de R é destinado a computação estatística e gráfica. Sendo uma ferramenta de código fonte aberto, com grande adesão na área

de análise e ciência de dados. O ambiente possui uma linguagem relativamente simples com uma gama de ferramentas disponibilizadas pelos próprios usuários, possibilitando a realização de diversas tarefas e podendo utilizá-lo nas mais variadas áreas do conhecimento [55]. Com o apoio de uma comunidade crescente de usuários de áreas interdisciplinares, as funcionalidades do R vêm se expandindo. A distribuição original do R possui alguns pacotes/pacotes ou bibliotecas, que são conjunto de funções e rotinas com diferentes utilidades. No repositório oficial do R, o CRAN (Comprehensive R Archive Network, em português "Rede abrangente de arquivamento R"), estão disponíveis diversos pacotes desenvolvidos pela própria comunidade que vão desde ferramentas para análise de dados obtidos da área de medicina e saúde, até conjuntos de funções para trabalhar com sistemas inteligentes que fazem uso de redes neurais, como é o caso do pacote "neuralnet", por exemplo. O download dos pacotes pode ser realizado no site oficial do CRAN ou instalados diretamente pelo console do R.

Entre as facilidades do ambiente R, podem-se destacar algumas que propiciaram sua crescente utilização na ciência de dados: conjunto de operadores para cálculos em *arrays*, em especial as matrizes; uma coleção coerente e integrada de ferramentas intermediárias para análise de dados; e uma linguagem de programação simples e eficaz, além de ampla variedade de técnicas estatística e gráficas, como modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação etc [55].

Assim, considerando o ambiente R uma ferramenta de apoio adequada aos estudos da performance de redes LoRaWAN, este trabalho tem como objetivo se utilizar dessa ferramenta para obter dados sobre o funcionamento da rede visando analisar a perda de pacotes por meio do uso de diversos pacotes disponibilizados.

A seguir é apresentada a lista de pacotes utilizados neste trabalho, com breve descrição:

- **tidyr**: Ferramentas para ajudar a criar dados organizados, onde cada coluna é uma variável, cada linha é uma observação e cada célula contém um único valor. Também inclui ferramentas para trabalhar com valores ausentes (implícitos e explícitos);
- **dplyr**: Uma ferramenta rápida e consistente para trabalhar com quadros de dados como objetos;
- **raster**: Leitura, escrita, manipulação, análise e modelagem de dados espaciais em grade. Utilizado para o cálculo de distancia entre Gateways e dispositivos a partir de suas coordenadas geográficas;
- **plotly**: Criação de gráficos interativos a partir de dados;
- **FactoMineR**: Métodos exploratórios de análise de dados para resumir, visualizar e descrever conjuntos de dados ;

- **factoextra**: Fornece funções para extrair e visualizar a saída de análises de dados multivariadas, incluindo 'PCA' (Análise de componentes principais), 'CA' (Análise de correspondência) e 'MCA' (Análise de correspondência múltipla);
- **RPostgreSQL**: Interface de banco de dados e *driver*¹¹ 'PostgreSQL' para 'R'. Este pacote fornece um driver compatível com 'Interface de banco de dados' DBI 'para que' R 'acesse os sistemas de banco de dados' PostgreSQL;
- **shiny**: Utilizado para criar aplicativos da web interativos com a R. A ligação "reativa" automática entre entradas e saídas e amplos *widgets*¹² pré-criados tornam possível criar aplicativos responsivos

Com a utilização dos pacotes acima listados, pode-se processar os dados anteriormente listados, por meio da técnica de análise de correspondência múltipla.

3.4.1 Análise de Correspondência múltipla e Correlação na Linguagem R

Várias funções de diferentes pacotes estão disponíveis no R para calcular a análise de correspondência múltipla. A tabela 7 lista algumas funções e seus pacotes, os quais podem ser utilizados para extrair e visualizar os resultados de uma MCA.

Função	Pacote
MCA()	FactoMineR
dudi.mca()	ade4
epMCA()	ExPosition

Tabela 7. Pacotes disponíveis para calcular a MCA.

O pacote utilizado nesse trabalho foi o FactoMinerR, um pacote R dedicado à análise de dados exploratória multivariada. Ele fornece ponto de vista geométricos da MCA, diversas saídas gráficas, ajuda a interpretar a descrição automática das dimensões, lida com valores ausentes através do parâmetro `missMDA` e possui uma interface em comum com o Shiny, que desenha gráficos interativos com o Factoshiny.

¹¹Interface que realiza a comunicação entre um banco de dados e uma linguagem de programação.

¹²Elemento de interação.

4. ANÁLISE E DISCUSSÃO DE RESULTADOS

Este capítulo apresenta os resultados obtidos do processamento e exploração dos dados coletados pela rede Optim, os quais foram tratados e submetidos a técnicas de análise de dados para a obtenção dos resultados observados.

Os resultados foram obtidos através da exploração de uma quantia significativa de pacotes sensoriados pela rede (aproximadamente 74 mil), os quais possuem dados referentes ao estado de configuração de cada dispositivo no momento do seu envio ao servidor de rede. Os resultados foram obtidos utilizando MCA e Correlação, implementadas em linguagem R e estão disponíveis para acesso pela empresa Optim, através do módulo de Análise de Taxa de Perda e Configuração de Rede presente na plataforma Omni, que fora desenvolvido a partir deste trabalho, para avaliar questões relacionadas a perda de pacotes na rede.

Os resultados observados a partir de cada métrica em seus devidos grupos, são apresentados nas seções a seguir.

4.1 AVALIAÇÕES DA REDE COMPLETA

Nas métricas de avaliação da rede completa, se faz necessário avaliar todos os dispositivos da rede em conjunto. A seguir são apresentadas as métricas: Efetividade em relação a Taxa de Perda, Perda de Pacotes por SF, Influência dos parâmetros na Taxa de Perda e Avaliação da qualidade do sinal em relação a perda de pacotes.

4.1.1 Medição da Taxa de Perda

Uma rede efetiva é aquela que faz a coleta e transmissão dos dados sem problemas aparentes ou perdas. Com relação a rede de Optim, acredita-se que o que indica se a rede pode ou não ser considerada efetiva, é a Taxa de Perda de Pacotes (TPP). Ou seja, quanto menor a TPP, mais a rede é considerada efetiva. Sendo assim, os experimentos realizados para a avaliação de efetividade da rede consistiram da medição da TPP da rede como um todo, independente do dispositivo.

$$\% \text{ de perda de pacote} = \left(1 - \frac{\text{Num. de frames recebidos}}{\text{Num. total de frames}}\right) \times 100\% \quad (22)$$

Como nem todos os dispositivos tinham tempos de comunicação iguais uns aos outros, a Equação 22 foi descartada para a avaliação de efetividade, pois com diferentes tempos de comunicação não foi possível estipular quantos dados deveriam ser sensoriados

no período selecionado. Porém, como todos os pacotes possuem o parâmetro FCNT (contador de frames por sessão), foi possível obter a contagem de pacotes perdidos e a partir dessa variável e através da Equação 23 foi obtida a TPP:

$$TPP = \frac{\text{pacotes perdidos} \times 100}{\text{total de pacotes}} \quad (23)$$

Para obter os resultados em porcentagem, sobre o total de pacotes perdidos, total de pacotes obtidos e taxa de perda de pacotes ([]) A partir da Equação 23 os resultados obtidos estão dispostos na Tabela 8 seguintes resultados foram obtidos:

Pacotes Perdidos	Total Pacotes	Taxa de Perda (%)
1714	73972	2.31

Tabela 8. Valores encontrados para o total de pacotes perdidos, total de pacotes obtidos e taxa de perda de pacotes em porcentagem.

4.1.2 Taxa de Perda por SF

Não há como indicar qual a configuração utilizada por um pacote perdido. Porém sabe-se que os dispositivos tendem a utilizar sempre a última configuração salva para se comunicar com o gateway, ou seja, a configuração do último pacote enviado. Assim é possível utilizar o pacote anteriormente enviado como sendo praticamente um espelho do pacote perdido. Quando o dispositivo vê a necessidade de ajuste de seus parâmetros, ele indica ao servidor de rede e solicita que o algoritmo de ADR reveja sua configuração para que seja realizada a otimização dos parâmetros. Então o ADR utiliza as 20 últimas configurações dos últimos 20 pacotes para indicar uma nova taxa de dados, e só a partir disso a configuração do dispositivo será modificada. Assim, é possível indicar, por exemplo, o SF que estava sendo utilizado pelo pacote perdido, pois tende a ter sido o mesmo do pacote anterior.

Para obter o valor de taxa de perda de pacotes por SF, primeiramente foi considerado o fato da rede possuir dispositivos individuais, isto porque o intuito dessa medição era avaliar todos os SFs independentemente do dispositivo, podendo assim também validar a influência do ADR adaptativo na perda de pacotes.

Através da Figura 29 é possível notar que, o SF = 10 foi o valor utilizado por mais da metade dos 1714 pacotes perdidos. Isso se explica pelo fato do algoritmo de ADR geralmente utilizar o maior valor de SF para obter maior cobertura e "calibrar" os demais parâmetros, diminuindo este valor após o comportamento do dispositivo se "alinhar".

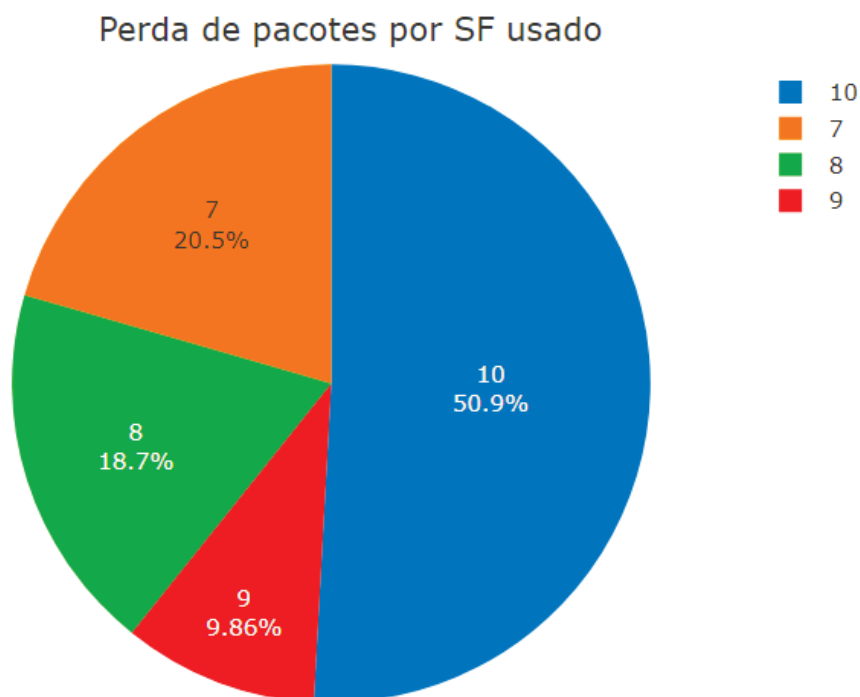


Figura 29. Dados da medição da taxa de perda de pacotes por SF utilizado. Cada cor representa um valor de SF e as porcentagens indicam quantos pacotes perdidos utilizavam o SF indicado.

4.1.3 Influência dos parâmetros na Taxa de Perda

A partir dos aproximadamente 74 mil pacotes coletados, foi possível correlacionar os parâmetros (RSSI, SNR, Distância e SF) com a perda de pacotes, utilizando correlação múltipla. O resultados obtidos por meio da correlação entre as variáveis estão dispostos em formato de tabela e podem ser visualizados Figura 30, onde é apresentado o coeficiente de correlação de Pearson de todas as variáveis.

	rsssi	snr	sf_used	numero_device	perda	distancia
rsssi	1.0000000	0.8827205	-0.0873943	-0.0537904	-0.0011481	-0.1800377
snr	0.8827205	1.0000000	-0.2286909	-0.0117521	-0.0173360	-0.2708259
sf_used	-0.0873943	-0.2286909	1.0000000	-0.0929844	0.0013193	0.1422916
perda	-0.0011481	-0.0173360	0.0013193	0.0012200	1.0000000	0.0166486
distancia	-0.1800377	-0.2708259	0.1422916	-0.0135260	0.0166486	1.0000000

Figura 30. Resultado da correlação entre variáveis utilizando Coeficiente de Pearson.

Conforme apresentado na Figura 30, a correlação das variáveis observadas em relação a perda é praticamente nula. As variáveis devem possuir um valor acima de 0,40 ou -0,40 como coeficiente para que sejam consideradas moderadamente correlacionadas através do coeficiente de Pearson. Níveis abaixo desses valores são considerados como

baixos ou nulos. Portanto, percebe-se que como nenhum dos coeficientes passa de 0,01 ou -0,01, essas variáveis tem pouquíssima influência na taxa de perda de pacotes.

Apesar dos resultados da correlação mostrarem que não há grande influência das variáveis observadas na perda de pacotes, foi também gerada a nuvem de variáveis, para que se possa entender graficamente a relação entre elas.

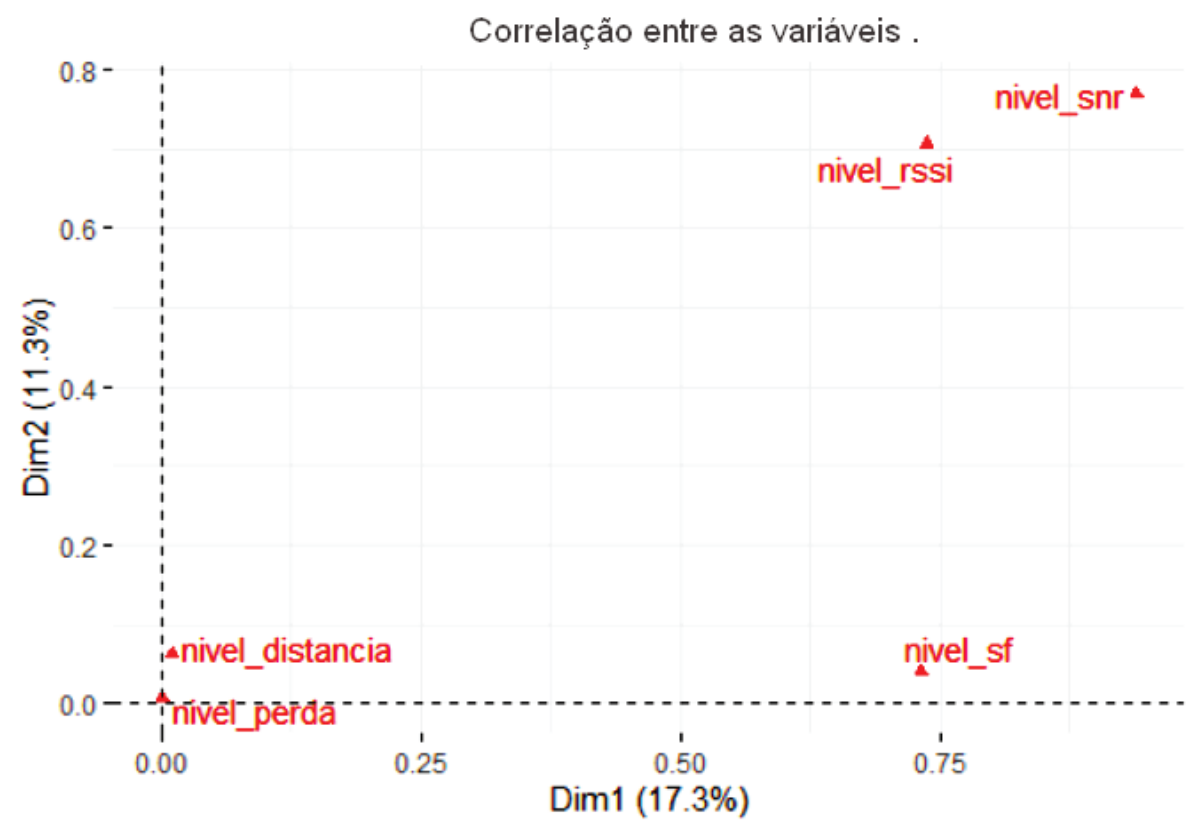


Figura 31. Gráfico de correlação entre as variáveis.

Através da nuvem de variáveis gerada pela ACM, foi possível identificar o parâmetro mais influente na perda de pacotes, dentre os parâmetros selecionados. Por meio da Figura 31 é apresentada essa nuvem, onde é possível notar que a distância de transmissão está mais próxima do ponto "nível_perda", o que indica que o mesmo possui maior influência na perda que os demais.

Através da Figura 32 é possível visualizar a nuvem de correlação entre todas as instâncias de cada uma variáveis utilizadas. Com a nuvem de instâncias foi possível verificar que a perda de pacotes está mais relacionada a níveis corrompidos de SNR (geralmente acima de -10dB), distâncias de transmissão do dispositivo a até Gateway acima de 4000m, ao SF10 (que foi o que obteve mais perdas), e a qualidade muito fraca do sinal.

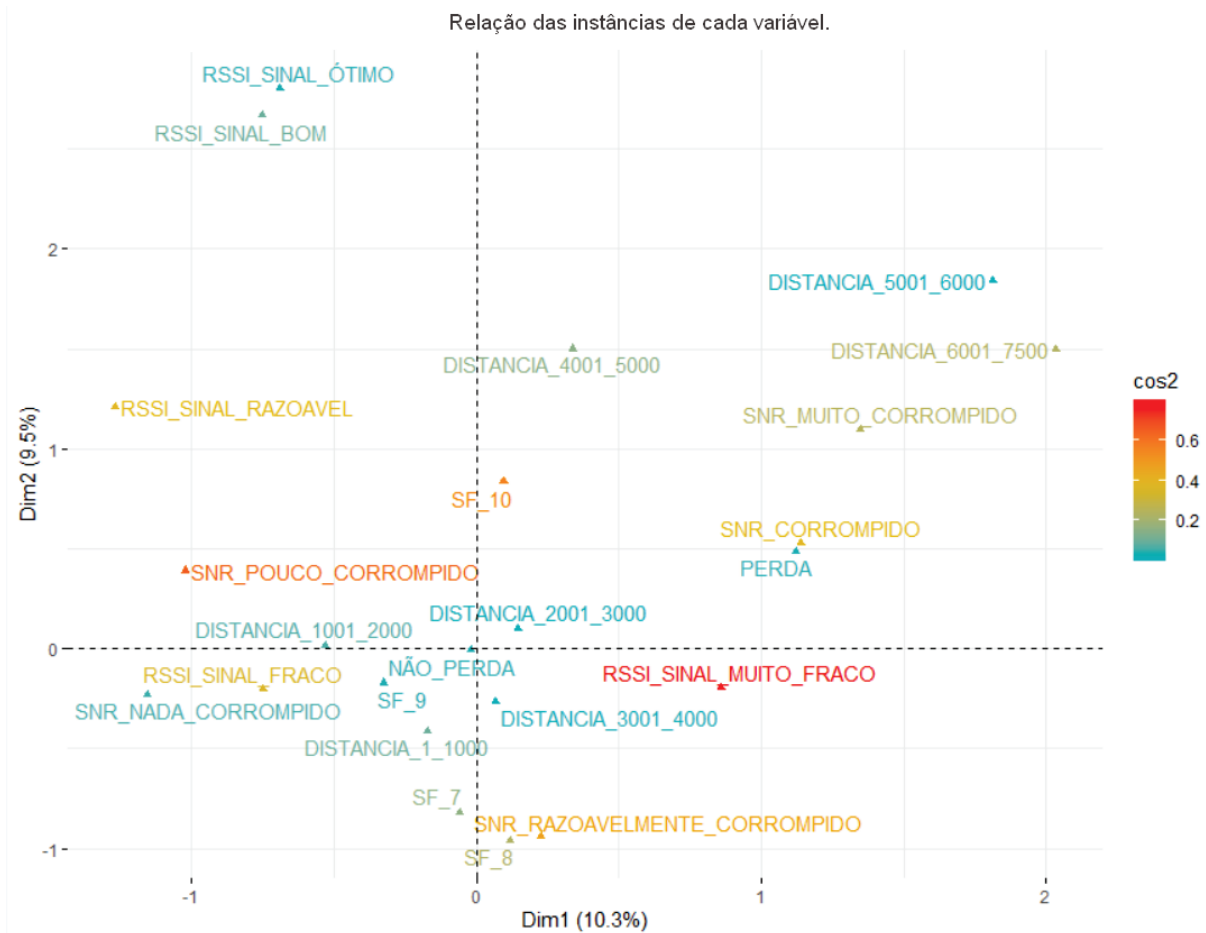


Figura 32. Gráfico de correlação entre as instâncias de cada variável.

4.1.4 Avaliação da qualidade do sinal em relação a perda de pacotes

O sinal da rede LoRaWAN pode ser avaliado relacionando o parâmetro de SNR com a distância do dispositivo. A Figura 33 apresenta a nuvem de categorias gerada por meio do Método de Análise de Correspondência Múltipla, correlacionando todas as categorias dessas duas variáveis com a perda de pacotes.

Através da Figura 33 é possível notar que quando a distância de comunicação está entre 1m e 3000m, tem-se um sinal pouco ou razoavelmente corrompido. Este cenário está mais relacionado a não perda de pacotes e tende a influenciar menos nas perdas.

Porém, é possível observar que a distâncias acima de 4000m, o nível de ruído no sinal tende a aumentar, e por consequência isto está mais relacionado a perda de perda de pacotes.

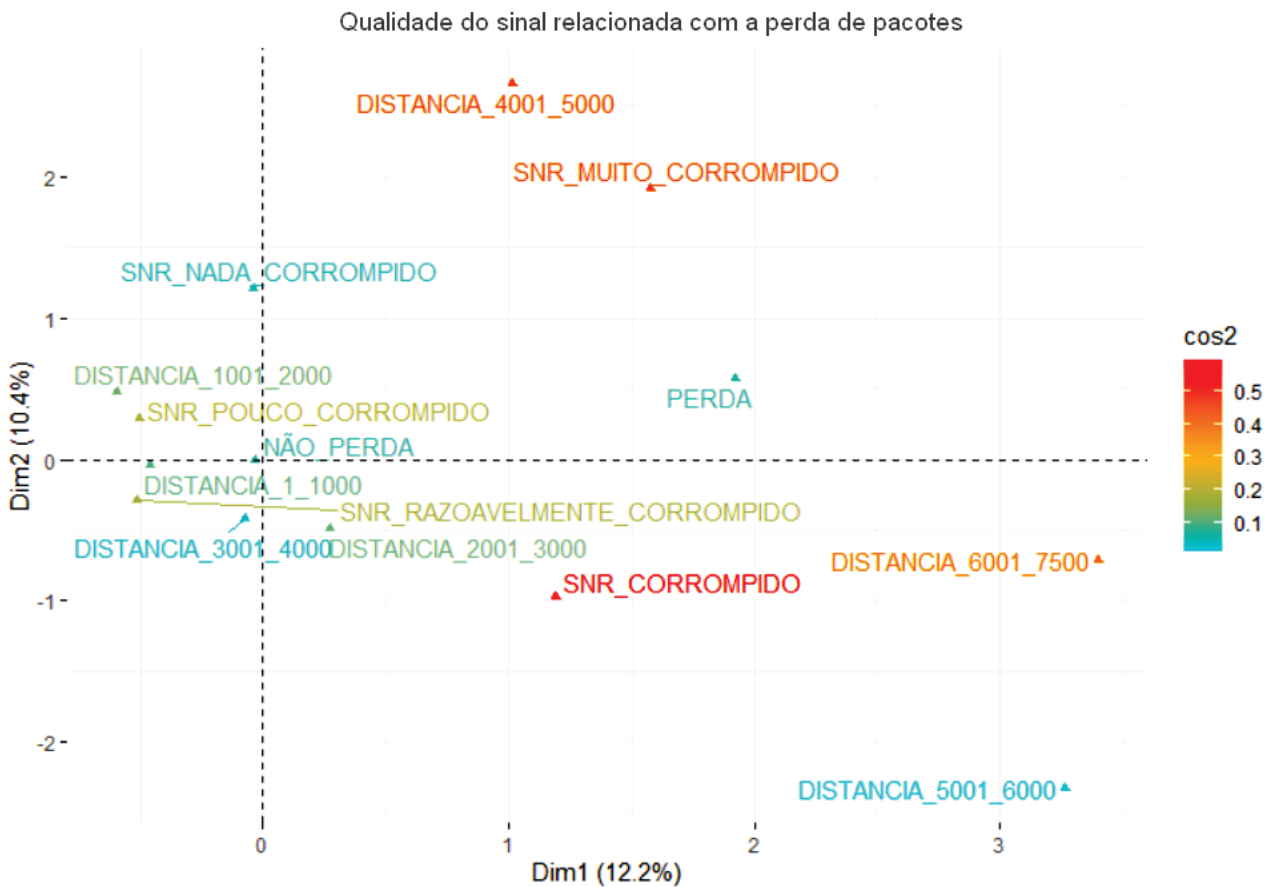


Figura 33. Nuvem de instancia das categorias, referente a correspondência do SNR com a distância e a perda de pacotes.

4.2 AVALIAÇÕES POR DISPOSITIVO

Em redes LoRaWAN, cada dispositivo é como um organismo único na rede, portanto cada um deles têm comportamentos que se divergem entre si, mesmo que sua configuração seja a mesma. Isso ocorre por diversas questões como relevo, distância do gateways até mesmo posição de instalação. Portanto, para algumas métrica é necessário que as avaliações sejam feitas também por dispositivo, para avaliar a situação de cada um dentro da rede.

4.2.1 Medição da Taxa de Perda por dispositivo

A taxa de perda de pacotes também foi calculada por dispositivo. As informações obtidas estão dispostas na Tabela 9. Além da taxa de perda de pacotes por dispositivo, através da tabela 9 também é possível notar a diferença no total de pacotes de cada um, isso se deve ao tempo de comunicação dos dispositivos ser diferente.

Na plataforma Omni, esses mesmos dados são apresentados em forma de gráfico, como demonstrado na Figura 35. A partir deles é possível identificar as diferenças de perda

Dispositivo	Pacotes Perdidos	Total Pacotes	Taxa de Perda (%)
20	199	2753	7.23
38	87	1259	6.91
22	76	1250	6.08
18	82	1457	5.63
25	81	1522	5.32
10	126	2505	5.02
24	55	1260	4.37
3	100	2924	3.42
30	40	1258	3.18
5	77	2560	3.01
11	75	2562	2.93
26	29	1215	2.39
21	26	1181	2.20
14	119	5646	2.11
2	58	2938	1.97
1	61	3281	1.86
23	45	2426	1.85
7	45	2719	1.66
39	19	1203	1.58
40	32	2057	1.56
28	22	1570	1.40
6	37	2724	1.36
31	15	1234	1.21
13	30	2591	1.16
37	14	1235	1.13
35	14	1250	1.12
36	13	1236	1.05
12	27	2599	1.04
34	10	1055	0.95
4	25	2737	0.91
17	11	1260	0.87
15	13	1575	0.83
29	9	1210	0.74
27	9	1215	0.74
8	18	2762	0.65
16	7	1215	0.58
32	5	1268	0.39
19	3	1260	0.24
Total	1714	73972	100%

Tabela 9. Taxa de perda de pacotes por dispositivo ordenada da maior (dispositivo 20), para a menor (dispositivo 19).

entre um dispositivo e outro e visualizar os dispositivos que estão tendo mais ou menos perdas.

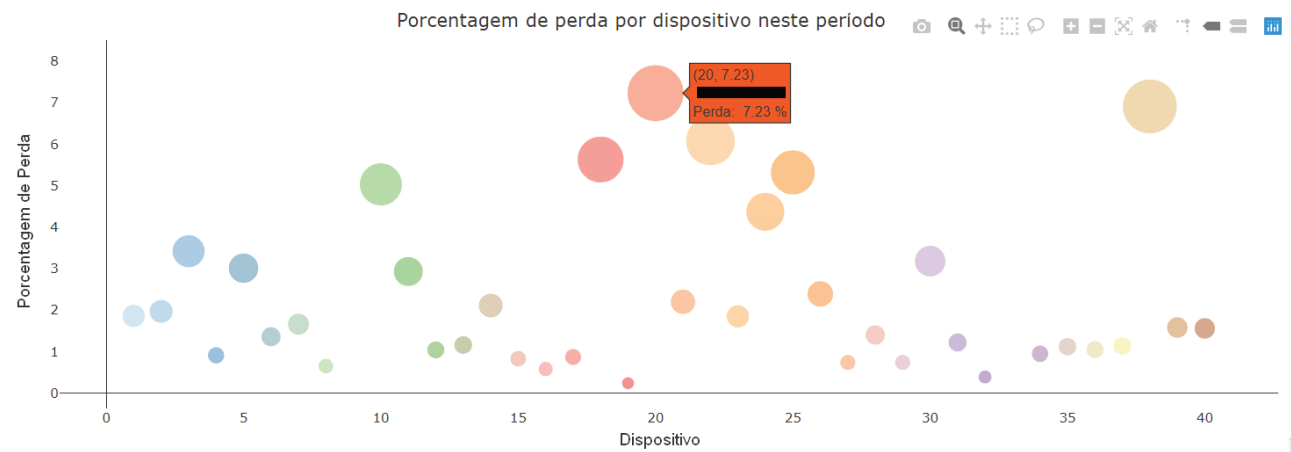


Figura 34. Gráfico que pode ser visualizado na plataforma Omni, onde estão dispostas as informações de pacotes perdidos por dispositivo.

4.2.2 Melhores configurações utilizadas por dispositivo

A seleção dos parâmetros de transmissão das redes LoRaWAN pode prejudicar o desempenho de comunicação e, por consequência, pode influenciar na taxa de perda de pacotes. Mais precisamente, a seleção afeta o alcance da transmissão e a resiliência à interferência.

A partir das de configurações utilizadas por cada dispositivo, foi possível indicar quais delas obtiveram menor taxa de perda de pacotes. Os dados observados estão dispostos na Tabela 10:

Como demonstrado na Tabela 10, o dispositivo 1 obteve maior sucesso na entrega dos pacotes ao utilizar a configuração $S_i = SF10, BW125, CR4/5, 2577m$. Já o dispositivo 7 obteve um maior sucesso com a configuração $S_i = SF7, BW125, CR4/5, 2756m$, uma diferença de distancia de apenas 200m de uma configuração para a outra, porém utilizando um valor SF mais baixo que o dispositivo anterior.

Nota-se também que o SF10 foi o que obteve o melhor desempenho em relação aos demais. Isso pode ser relacionado ao ADR, pois o algoritmo tende a sempre aumentar o SF quando solicitado, pois com um SF mais alto, maior a sensibilidade e o alcance do sinal.

Dispositivo	SF	Distancia (m)	BW	CR	Pacotes entregues
1	10	2577.72380	125	4/5	1240
2	7	1800.12653	125	4/5	1347
3	7	959.45883	125	4/5	1722
4	10	1331.61082	125	4/5	1571
5	10	2328.58679	125	4/5	1212
6	9	782.73883	125	4/5	888
7	7	2756.59605	125	4/5	1097
8	10	3219.39503	125	4/5	1229
9	8	18.38374	125	4/5	354
10	10	3389.40893	125	4/5	1744
11	8	659.66504	125	4/5	856
12	10	2015.85266	125	4/5	1366
13	10	15.43136	125	4/5	1358
14	10	2918.44266	125	4/5	2920
15	10	315.97029	125	4/5	1553
16	10	3139.07551	125	4/5	459
17	10	1924.58058	125	4/5	761
18	10	574.23298	125	4/5	1361
19	10	2533.18666	125	4/5	359
20	10	6347.75598	125	4/5	2250
21	9	2703.14628	125	4/5	814
22	10	1501.79304	125	4/5	785
23	8	2328.54649	125	4/5	735
24	8	3094.27730	125	4/5	329
25	10	467.43504	125	4/5	1424
26	7	260.69775	125	4/5	667
27	7	3464.69636	125	4/5	458
28	8	780.51965	125	4/5	654
29	7	1230.83246	125	4/5	432
30	7	3349.69953	125	4/5	471
31	8	390.35532	125	4/5	386
32	10	2356.05475	125	4/5	707
33	7	3780.85620	125	4/5	613
34	9	3791.06262	125	4/5	543
35	7	2346.04744	125	4/5	394
36	8	825.68536	125	4/5	406
37	10	116.75433	125	4/5	391
38	8	3496.10315	125	4/5	516

Tabela 10. Configurações com maior porcentagem de sucesso na entrega de pacotes por dispositivo.

5. CONSIDERAÇÕES FINAIS

O uso cada vez maior de infraestruturas para coleta e transmissão de dados tende a por cada vez mais em prática o termo Cidades Inteligentes. Através de redes como a LoRaWAN é possível obter os mais variados dados independente do contexto em que são utilizadas. Porém, sabe-se que mesmo sendo redes tidas como robustas, as redes LoRaWAN também estão sujeitas a interferências que levam a perda de informações. Diante disso, é necessário que se obtenha maior conhecimento sobre as possibilidades de configuração dessas redes e se essas configurações podem ou não influenciar nas perdas. E, se possível, definir técnicas, métricas, métodos e tecnologias para análise do comportamento dessas redes.

Senso assim, através deste trabalho, foi possível constatar que a estatística multivariada pode ser aliada no processo de extração de conhecimento acerca dessas redes. Mostrando possuir métodos de avaliação eficazes para encontrar a relação entre os parâmetros de transmissão e a taxa de perda de pacotes. Através desses métodos, foi possível constatar que os parâmetros Fator de espalhamento, RSSI e SNR têm menos influência na taxa de perda de pacotes da rede do que a distância. E através dessa constatação, presume-se indiretamente que o algoritmo de ADR utilizado pelo servidor de rede Orbiwan cumpre seu papel na minimização das perdas e otimização da rede. Devido aos resultados das medições de taxa de perda apontarem que, com a utilização do mesmo, a rede possui taxas de perda de pacotes abaixo de 3%, o que valida o bom funcionamento do algoritmo. Mesmo com coeficientes de correlação muito baixos, o parâmetro que demonstrou ter uma maior influência na taxa de perda foi a distância de instalação entre o dispositivo e o Gateway.

Ao evidenciar que os parâmetros analisados tem influência mínima na perda de pacotes, outras questões foram levantadas. Essas questões dizem respeito a busca por quais outros fatores estariam influenciando nas perdas. Uma das possibilidades é a de que o uso de SIMCard e conexão 3G por alguns gateways possa influenciar na taxa de perda, pois sabe-se que esse tipo de conexão tende a ter uma latência relativamente elevada e instável, para este tipo de aplicação. Outra possibilidade é a de que o relevo e obstáculos como prédios possam também influenciar no sinal de radio.

Portanto, diante do exposto, entende-se que para a compreensão da taxa de perda de pacotes de redes LoRaWAN, se faz necessário a avaliação de outros parâmetros além dos parâmetros de transmissão, pois estes têm uma influência mínima na perda. Considera-se ainda que a partir dos parâmetros de transmissão é possível inferir outras questões relacionadas a redes LoRaWAN, e que o uso de ferramentas de estatística como a Linguagem

R podem possibilitar extrair inúmeras informações sobre a rede a partir de seus próprios dados.

Como trabalho futuro, é de interesse investigar a influência de outros fatores como relevo e latência do Gateway na taxa de perda, para evidenciar qual o principal fator para a ocorrência de perda de pacotes em redes LoRaWAN. Além disso, pretende-se integrar ao módulo outras funções como análise completa da qualidade do sinal, visto que é de suma importância a verificação de sinal em redes de radiofrequência.

Por fim, se almeja transformar o módulo de análise em um sistema em si. Sistema esse que possibilite avaliar a eficiência de qualquer rede LoRaWAN a partir de seus próprios dados, utilizando outras técnicas e métricas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] SCHAFFERS H.AND KOMNINOS, N.; PALLOT, M. e. a. Smart cities and the future internet: Towards cooperation frameworks for open innovation. *In: The Future Internet: Future Internet Assembly 2011: Achievements and Technological Promises, Springer Berlin Heidelberg*, n. 31, p. 431–446, 2011.
- [2] CORPORATION, S. *LoRa™ Modulation Basics AN1200.22*. [S.l.], 2015. Disponível em: <<https://www.semtech.com/uploads/documents/>>. Acesso em: Agosto de 2018.
- [3] LORA™-ALLIANCE. *LoRa™-Alliance Technology*. [S.l.]. Disponível em: <<https://www.lora-alliance.org/technology>>. Acesso em: Agosto de 2018.
- [4] SEMTECH. *AN120.22 LoRa Modulation Basics*. 2018. Disponível em: <<https://www.semtech.com/uploads/documents/an1200.22.pdf>>. Acesso em: Setembro de 2018.
- [5] ASHTON, K. That 'internet of things' thing. v. 22, p. 97–114, 2009.
- [6] LI, Y. et al. Towards a theoretical framework of strategic decision, supporting capability and information sharing under the context of internet of things. *Information Technology and Management*, v. 13, n. 4, p. 205–216, 2012.
- [7] EVANS, D. The internet of things: How the next evolution of the internet is changing everything. 2011.
- [8] 2015, C. *INTERNET OF THINGS INSIGHTS AND OPPORTUNITIES*. 2005. [Online; Acesso em: 21 outubro 2019.]. Disponível em: <<http://www.comptia.org/resources/internet-of-things-insights-and-opportunities>>.
- [9] LUCERO, S. *IoT Platforms: Enabling the Internet of Things*. *In: HIS Technology*. 2016. [Online; Acesso em: 22 outubro 2019.]. Disponível em: <<https://cdn.ihs.com/www/pdf/enabling-IOT.pdf>>.
- [10] ONU. *Perspectivas de urbanização mundial: a revisão de 2014, destaques*. [S.l.]: Departamento de Assuntos Econômicos e Sociais, 2014.
- [11] CAPDEVILA, I.; ZARLENGA, M. I. Smart city or smart citizens? the barcelona case. *Journal of Strategy and Management*, v. 8, n. 3, p. 266–282, 2015.
- [12] SARAIVA, L. da S. *Projeto de hardware e software para dispositivos finais em uma rede LoRaWAN*. 2017. Engenharia de Telecomunicações - Instituto Federal de Santa Catarina (IFSC), São José – SC.

- [13] SONG, Q.; NUAYMI, L.; LAGRANGE, X. Survey of radio resource management issues and proposals for energy-efficient cellular networks that will cover billions of machines. *EURASIP Journal on Wireless Communications and Networking*, v. 11, n. 1, p. 140, 2016.
- [14] OLIVEIRA, L. et al. Protocols for internet of things: A survey. *Future Internet 2019*, v. 11, n. 16, 2019.
- [15] LORATM-ALLIANCE. *LoRaWAN Specification V1.0.2*. [S.l.], 2016. Disponível em: <<https://www.lora-alliance.org/>>. Acesso em: Setembro de 2018.
- [16] DUCROT, N.; RAY, D.; SAADANI, A. *LoRa Device Developer Guide*, Orange. 2017. Disponível em: <<https://partner.orange.com/wpcontent/uploads/2017/11/LoRa-DeviceDeveloper-Guide-Orange.pdf>>. Acesso em: Agosto de 2018.
- [17] VANGELISTA, L.; ZANELLA, A.; ZORZI, M. Long-range iot technologies: The dawn of lora™. In: ATANASOVSKI, V.; LEON-GARCIA, A. (Ed.). *Future Access Enablers for Ubiquitous and Intelligent Infrastructures*. [S.l.]: Springer International Publishing, 2015. p. 51–58.
- [18] EMBARCADOS. *Introdução ao LPWAN (Low Power Wide Area Network)*. 2017. Disponível em: <<https://www.embarcados.com.br/introducao-ao-lpwan/>>.
- [19] CENTENARO, M. et al. Long-range communications in unlicensed bands: the rising stars in the iot and smart city scenarios. *IEEE Wireless Communications*, v. 23, n. 5, p. 60–67, Out 2016.
- [20] PETÄJÄJÄRVI, J. et al. Evaluation of lora lpwan technology for remote health and well-being monitoring. In: *2016 10th International Symposium on Medical Information and Communication Technology (ISMICT)*. [S.l.: s.n.], 2016. p. 1–5.
- [21] DEKKERS, P. *LoRa, the Internet of Things*. 2016. Disponível em: <<https://blog.surf.nl/en/lora-the-internet-of-things>>.
- [22] Reynders, B.; Pollin, S. Chirp spread spectrum as a modulation technique for long range communication. In: *2016 Symposium on Communications and Vehicular Technologies (SCVT)*. [S.l.: s.n.], 2016. p. 1–5. ISSN null.
- [23] BOR, M.; VIDLER, J.; ROEDIG, U. Lora for the internet of things. In: *Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks*. [S.l.]: Junction Publishing, 2016. p. 361–366.
- [24] Bor, M.; Roedig, U. Lora transmission parameter selection. In: *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. [S.l.: s.n.], 2017. p. 27–34.

- [25] THE Things Network. Acessado: 14 mai. 2017. Disponível em: <<https://www.thethingsnetwork.org/docs/>>. Acesso em: 14 mai. 2017.
- [26] BUGNION, P.; MANIVANNAN, A.; NICOLAS, P. R. *Scala: Guide for Data Science Professionals*. [S.l.]: Birmingham: Packt Publishing, 2017. 530 p.
- [27] CONWAY, D. *The Data Science Venn Diagram*). 2010. Disponível em: <<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>>. Acesso em: Agosto de 2018.
- [28] PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. *Big Data*, v. 1, 03 2013.
- [29] BLUM, A.; HOPCROFT, J.; KANNAN, R. *Foundations of Data Science (2018)*. 2018. [Online; acessado em 24-Outubro-2019]. Disponível em: <<https://www.cs.cornell.edu/jeh/book.pdf>>.
- [30] ORBIWISE. *Orbiwan Documentation*. 2019. Disponível em: <<https://www.orbiwise.com/solutions/orbiwan>>.
- [31] TELECOMUNICAÇÕES, A. N. de. *Plano de Atribuição, Destinação e Distribuição de Frequências no Brasil*. 2016. Disponível em: <<http://www.anatel.gov.br/institucional/acervo-documental>>.
- [32] SOMMERVILLE, I. *Engenharia de software*. PEARSON BRASIL, 2011. ISBN 9788579361081. Disponível em: <<https://books.google.com.br/books?id=H4u5ygAACAAJ>>.
- [33] DEITEL, H.; DEITEL, P.; PEARSON, E. *Java® Como Programar*. PEARSON BRASIL. ISBN 9788543004792. Disponível em: <<https://books.google.com.br/books?id=2gNRswEACAAJ>>.
- [34] CORPORATION, O. *Introduction to Javasever Faces*. 2005. Disponível em: <<https://www.oracle.com/technetwork/topics/index-090910.html>>.
- [35] THOMSEN, P. K. *Application - Data Access Interface specification*. Confidential, 2016. Version 018.
- [36] CHHETRI, N. *A Comparative Analysis of Node.js (Server-Side JavaScript)*. Dissertação (Mestrado) — Saint Cloud State University - Department of Computer Science and Information Technology, St. Cloud, Minnesota - USA, 2016. Disponível em: <<https://repository.stcloudstate.edu/cgi/viewcontent.cgi?referer>>.
- [37] MCCUNE, R. R. Node.js paradigms and benchmarks. *University of Notre Dame*, 2012.

- [38] JOYENT. *Home page of Node.js*. 2016. Disponível em: <<https://nodejs.org/en/>>. Acesso em: Abril de 2019.
- [39] PARNAS, D. L. On the criteria to be used in decomposing systems into modules. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 15, n. 12, p. 1053–1058, dez. 1972.
- [40] ORTIZ, F. M. et al. Caracterização de uma rede sem-fio de baixa potência e longo alcance para internet das coisas. In: *Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Porto Alegre, RS, Brasil: SBC, 2018. ISSN 2177-9384. Disponível em: <<https://sol.sbc.org.br/index.php/sbrc/article/view/2485>>.
- [41] HARWAHYU, R.; PRESEKAL, A.; SARI, R. Lorawan performance evaluation with optimized configuration. Department of Electrical Engineering, Universitas Indonesia, Indonesia, p. 51–68, 07 2018.
- [42] MAGRIN, D.; CAPUZZO, M.; ZANELLA, A. A thorough study of lorawan performance under different parameter settings. 06 2019.
- [43] GREENACRE, M. J. *Theory and applications of correspondence analysis*. London, UK: Academic Press, 1984.
- [44] ROUX, B. L.; ROUANET, H. Multiple correspondence analysis (mca). In: _____. [S.l.: s.n.], 2010.
- [45] BLASIUS, J. et al. Editorial: Special issue on correspondence analysis and related methods. *Comput. Stat. Data Anal.*, Elsevier Science Publishers B. V., NLD, v. 53, n. 8, p. 3103–3106, jun. 2009.
- [46] NAITO, S. D. N. P. *Análise de correspondências generalizada*. 2007.
- [47] ABDI, H.; VALENTIN, D. Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics*, 01 2007.
- [48] SOUZA, A. C. *Análise de Correspondência aplicada à ECINF: a diversidade do setor informal urbano no Brasil*. 2004.
- [49] LOESCH, C.; HOELTGEBAUM, M. *Métodos Estatísticos Multivariados*. São Paulo, Brasil: Saraiva, 2012.
- [50] CLAUSEN, S. E. *Applied Correspondence Analysis: An Introduction*. California: Sage Publications Inc, 1998.
- [51] TRIOLA, M. F. *Introdução à estatística*. Rio de Janeiro: LTC, 2005.

- [52] PENHA, N.; OLIVEIRA, S.; PEREIRA, D. F. Análise de correlação entre comportamentos de matrizes pesadas e variáveis ambientais. *Revista Brasileira de Engenharia de Biosistemas*, v. 1, p. 209, 11 2009.
- [53] MARTINS, G. de A.; DOMINGUES, O. *Estatística Geral e Aplicada*. [S.l.]: Atlas, 2017.
- [54] DANCEY, C. P.; REIDY, J. *Estatística sem Matemática para Psicologia*. [S.l.]: Pearson, 2013.
- [55] VENABLES, W. N. S. D. M. *An Introduction to R: A Programming Environment for Data Analysis and Graphics*. [S.l.]: R Core Team, 2014.

APÊNDICE A – TELAS DO MÓDULO DE AVALIAÇÃO DE TAXA DE PERDA DE PACOTES E CONFIGURAÇÃO DA REDE

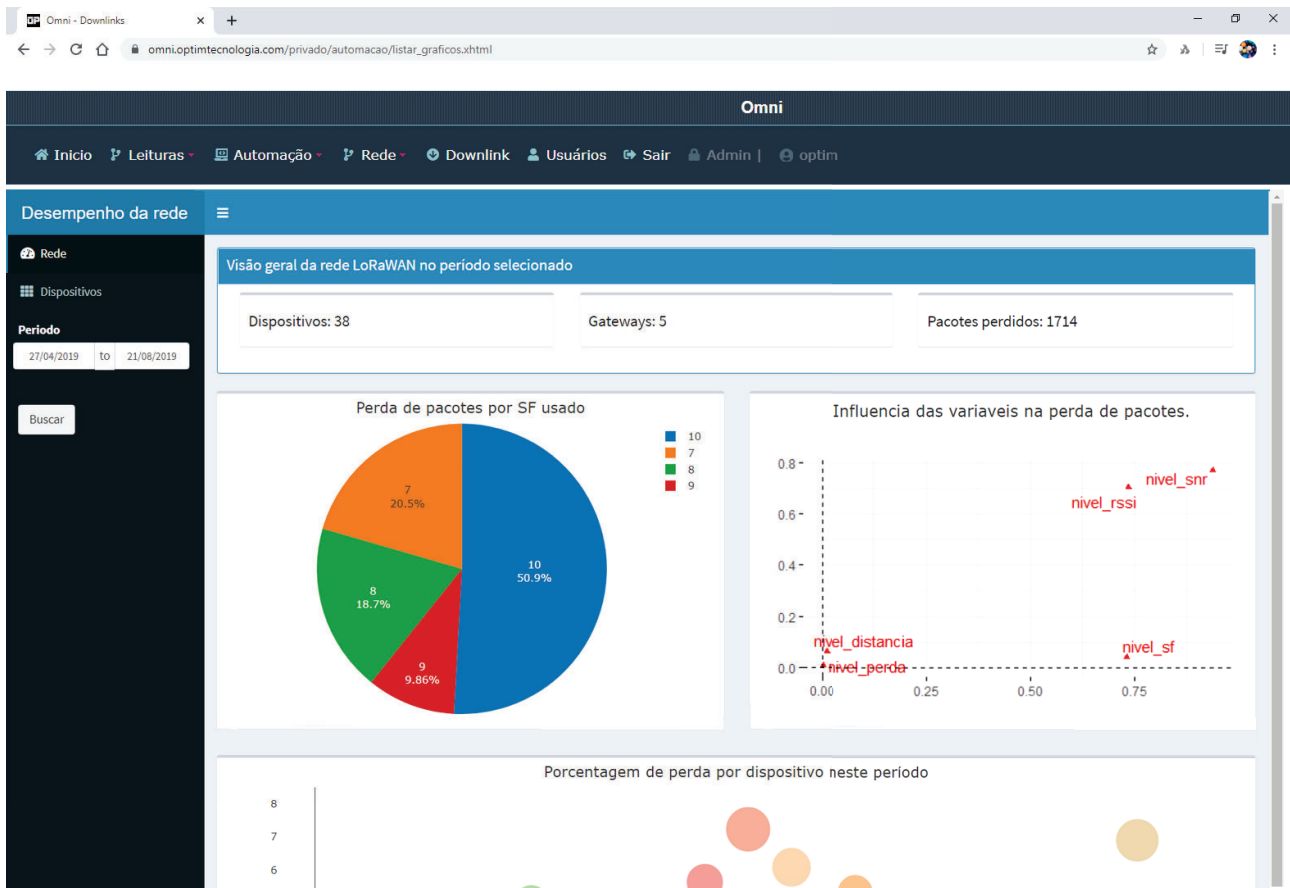


Figura 35. Tela de apresentação dos dados referentes a avaliação da rede completa.

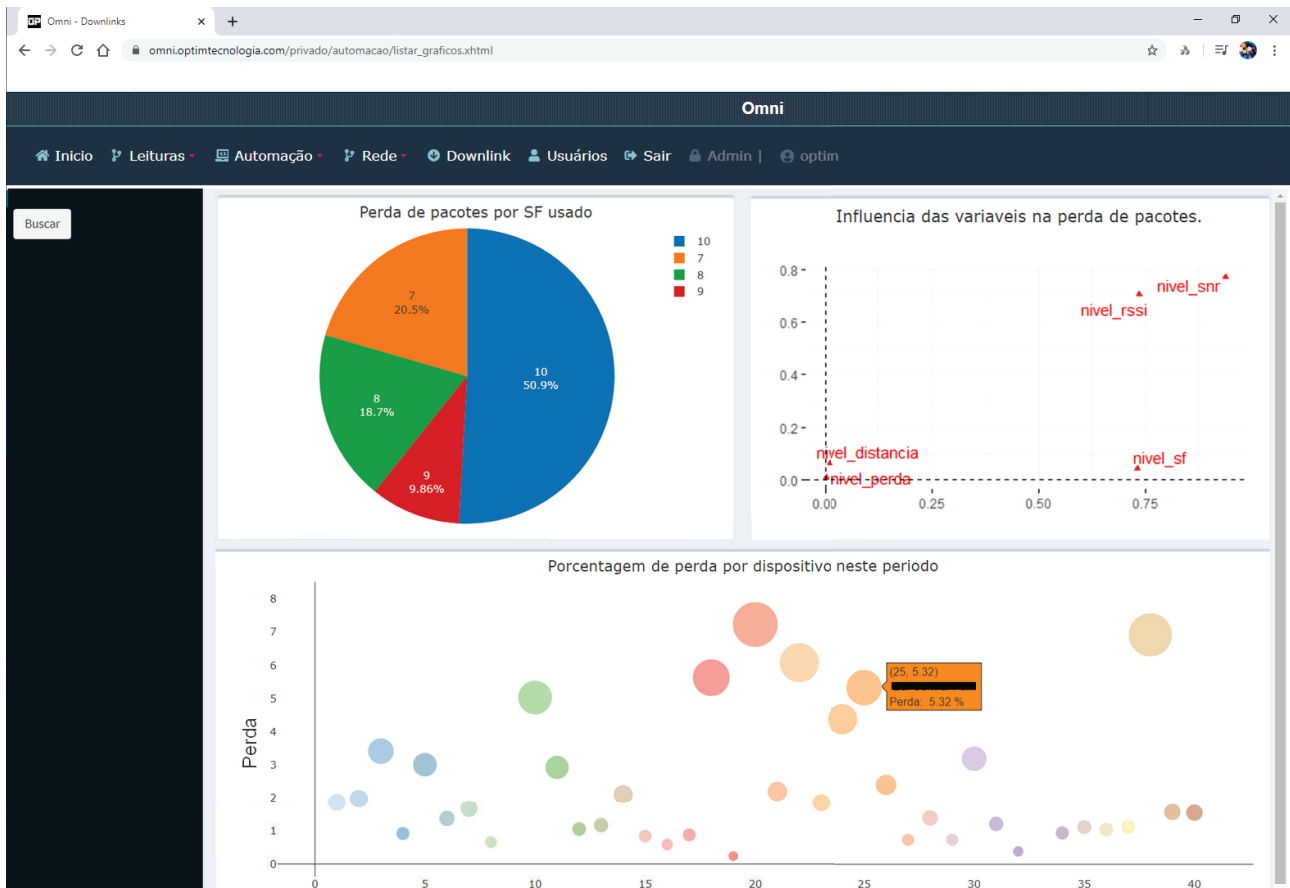


Figura 36. Tela de apresentação dos dados referentes a avaliação da rede completa.

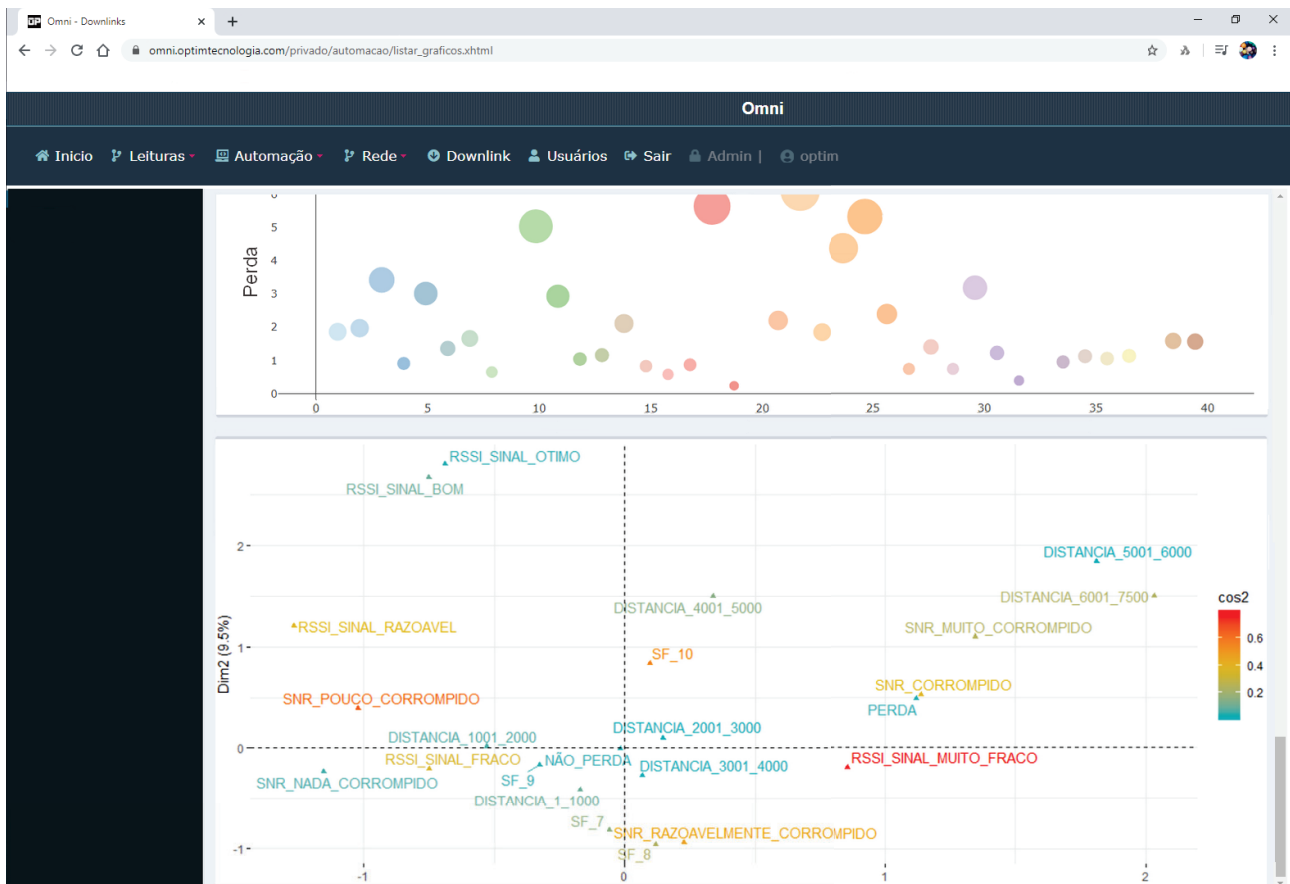


Figura 37. Tela de apresentação dos dados referentes a avaliação da rede completa.



UPF

UNIVERSIDADE
DE PASSO FUNDO

UPF Campus I - BR 285, São José
Passo Fundo - RS - CEP: 99052-900
(54) 3316 7000 - www.upf.br