

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Pedro Ivo Pereira Lancellotta

**Uma arquitetura diferenciada para apoio à tomada
de decisão baseada em semântica e redes complexas**

Juiz de Fora

2018

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Pedro Ivo Pereira Lancellotta

**Uma arquitetura diferenciada para apoio à tomada
de decisão baseada em semântica e redes complexas**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Victor Ströele de Andrade Menezes

Coorientador: Regina Maria Maciel Braga

Juiz de Fora

2018

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Lancellotta, Pedro Ivo Pereira.

Uma arquitetura diferenciada para apoio à tomada de decisão baseada em semântica e redes complexas / Pedro Ivo Pereira Lancellotta. -- 2018.

111 f. : il.

Orientador: Victor Ströele de Andrade Menezes

Coorientadora: Regina Maria Maciel Braga

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Ciência da Computação, 2018.

1. Inteligência de Negócios. 2. Ontologia. 3. Redes Complexas. 4. Visualização da Informação. I. Menezes, Victor Ströele de Andrade, orient. II. Braga, Regina Maria Maciel, coorient. III. Título.



Pedro Ivo Pereira Lancellotta

“Uma Arquitetura Diferenciada para Apoio à Tomada de Decisão baseada em Semântica e Redes Complexas”

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de Mestre.

Aprovada em 20 de setembro de 2018.

BANCA EXAMINADORA

Prof. Dr. Victor Ströele de Andrade Menezes – Orientador
Universidade Federal de Juiz de Fora

Prof^a. Dra. Regina Maria Maciel Braga Villela - Coorientadora
Universidade Federal de Juiz de Fora

Prof. Dr. Fernanda cláudia Alves Campos
Universidade Federal de Juiz de Fora

Prof. Dr. Cláudio Nápolis Costa
Empresa Brasileira de Pesquisa Agropecuária

*A Deus em primeiro lugar. A
minha família, amigos pelo apoio
incondicional.*

AGRADECIMENTOS

Primeiramente a Deus que permitiu a realização de todos meus objetivos, me dando força em todos os momentos da minha vida.

Aos meus pais e família, que sempre me deram apoio nessa jornada, sem o qual não conseguiria chegar aonde estou.

Aos meus amigos que sempre estiveram presentes, me dando força em momentos bons e ruins. Especialmente para meus amigos da Banda Olar, alguns amigos distantes e outros que sempre estiveram presentes. Um agradecimento especial àqueles que me acolheram e me ajudaram durante minha estadia em Juiz de Fora.

Agradeço a todos os professores do instituto de Ciência da Computação da Universidade Federal de Juiz de Fora e todos os meus supervisores de estágio por todas as lições, experiências e apoios transmitidos ao longo destes anos.

A todos que direta ou indiretamente fizeram parte da minha trajetória, o meu muito obrigado.

*Até mesmo a menor das
criaturas pode mudar o curso da
história.*

*J. R. R. Tolkien (Senhor dos
Anéis, O)*

RESUMO

O cenário de expansão da tecnologia da informação levou à necessidade de análise de grandes quantidades de dados para gerenciar informações para uma rápida tomada de decisão. A inteligência de negócios ou business intelligence, provê mecanismos para otimizar a eficiência da tomada de decisão nos mais diversos domínios de aplicação. Um dos mecanismos utilizados, a utilização de abstrações visuais, possibilita uma análise mais efetiva dos dados, de forma a tornar sua interpretação mais simples. A visualização da informação associada a soluções de processamento dos dados, como ontologias em conjunto a redes complexas, agregam informações semânticas e estruturais para os dados, constituindo uma poderosa ferramenta para tomada de decisão. Esta dissertação apresenta uma arquitetura que utiliza a interpretação semântica e estrutural como técnicas conjugadas para análise de dados no domínio de produção do leite, através de uma interface que suporta estratégias de visualização. A partir da arquitetura proposta, foi contruída uma plataforma de visualização de dados para tomada de decisão. A avaliação da plataforma foi feita através da aplicação de um estudo piloto e estudo de caso, ambos realizados com especialistas no contexto de pecuária leiteira utilizando dados de sistema de produção para gados leiteiros. Estes dados são provenientes do mundo real, utilizados de uma plataforma de gestão leiteira e análises de laboratórios de qualidade do leite. Os resultados demonstram a viabilidade da proposta.

Palavras-chave: Inteligência de Negócios, Ontologia, Redes Complexas, Visualização da Informação.

ABSTRACT

The expansion scenario in information technology has led to the need for the analysis of huge amounts of data, to manage information for quick decision making. The Business Intelligence provides mechanisms to optimize decision making for an application domain. One of the mechanisms used, called visual abstractions, allows a more efficient data analysis, making its interpretation simpler. Information Visualization associated with a data processing solution, such as the integration of complex networks, the aggregation of semantic information and data updating, are important tools for decision making. This dissertation has an architecture that uses the interpretation of the semantics and structural analysis as combined techniques to data analysis. From the proposed architecture, a data visualization platform was built to help in decision making. The evaluation of the platform was done through the application of a pilot study and a case study, both continued to be part of a research group data dairy production system. This data comes from the real world, used in a milk management platform and laboratory analyzes of milk quality. The results demonstrate the feasibility of the proposal.

Keywords: Business Intelligence, Complex Networks, Information Visualization, Ontology

LISTA DE FIGURAS

1.1	Estrutura da dissertação	22
2.1	Elementos do processo de BI (VUORI, 2007)	24
2.2	Exemplos de estratégias visuais (SCHOTS, 2014)	26
3.1	Núcleo da arquitetura, baseado no processo de BI	36
3.2	Conhecimento dos participantes em relação a gado leiteiro	40
3.3	Importância da visualização de dados no domínio de aplicação	40
3.4	Diferentes formas de visualização de dados no domínio de aplicação	41
3.5	Análise das relações entre indivíduos	41
3.6	Análise de indivíduos em destaque na rede de dados	41
3.7	Deficiências apontadas, separadas em grupos do processo de BI	43
3.8	Principais palavras citadas nas respostas dos participantes	44
3.9	Principais palavras citadas nos indicadores de vacas	45
3.10	Principais palavras citadas nos indicadores de touros	46
3.11	Fluxo do ciclo de vida de BI associado a soluções computacionais	49
3.12	Detalhamento da arquitetura proposta, expandindo o fluxo de ciclo de vida de BI	50
3.13	Relações da Ontologia de Domínio	52
3.14	Property Chain Gerencial	53
4.1	Dashboard para dados de produção de leite	63
4.2	Aplicação nas Redes Hierárquicas	64
4.3	Tela de filtro de dados na plataforma de visualização	68
4.4	Tabela de produção dos animais por rebanho	69
4.5	Tabela de produção de leite atendida por entidades gerenciais	70
4.6	Rede de vacas em destaque	71
4.7	Similaridade nas arestas entre vacas	72
4.8	Usabilidade da plataforma	75
4.9	Facilidade de acesso da plataforma	75

4.10 Suficiência da apresentação de visualizações durante a análise dos dados na	
plataforma	76
4.11 Aspectos positivos da plataforma de visualização	77
C.1 Quantidade de Artigos por Base de Dados	107
C.2 Relação entre Seleccionados e Escolhidos por Base de Dados	107
C.3 Relação de artigos por ano (antes do Quality Assessment)	108
C.4 Relação de artigos por ano (após o Quality Assessment)	108

LISTA DE TABELAS

2.1 Características de visualização	28
2.2 Tabela comparativa entre os artigos selecionados	33
3.1 Principais Deficiências Observadas	42
C.1 Elementos PICOC	102
C.2 Artigos encontrados nas bases de dados	103
C.3 Perguntas para avaliação da Qualidade	106
C.4 Condução da Revisão Sistemática	109

LISTA DE ABREVIATURAS E SIGLAS

BI Business Intelligence

MAPA Ministério da Agricultura, Pecuária e Abastecimento

RBQL Rede Brasileira de Laboratórios de Qualidade do Leite

LQL Laboratórios de Qualidade do Leite

CPL Cadeia Produtiva do Leite

IN Instrução Normativa

UFJF Universidade Federal de Juiz de Fora

SUMÁRIO

1 INTRODUÇÃO	15
1.1 MOTIVAÇÃO	16
1.2 PROBLEMA	17
1.3 QUESTÃO DE PESQUISA	18
1.4 ENFOQUE	18
1.5 OBJETIVO	20
1.6 METODOLOGIA	20
1.7 ORGANIZAÇÃO	21
2 FUNDAMENTAÇÃO TEÓRICA	23
2.1 BUSINESS INTELLIGENCE	23
2.1.1 Núcleo Genérico de Processo de BI	24
2.2 VISUALIZAÇÃO DE DADOS	25
2.2.1 Técnicas de Visualização de Dados	25
2.3 ARQUITETURA ORIENTADA A SERVIÇOS	27
2.4 ONTOLOGIA	28
2.5 REDES COMPLEXAS	29
2.6 TRABALHOS RELACIONADOS	30
2.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO	34
3 ARQUITETURA CONCEITUAL	35
3.1 ESTUDO EXPLORATÓRIO	37
3.1.1 Planejamento do Estudo Exploratório	37
3.1.2 Definição dos Participantes	38
3.1.3 Execução do Estudo Exploratório	38
3.2 ARQUITETURA	46
3.3 CAMADA DE ETL	50
3.4 CAMADA DE PROCESSAMENTO	51
3.4.1 Ontologia de Domínio	51
3.4.2 Desenvolvendo a Ontologia de Domínio	51

3.4.3	Redes Complexas de Indivíduos	54
3.4.3.1	Métricas de Redes Complexas	55
3.5	CAMADA DE VISUALIZAÇÃO	56
3.6	TECNOLOGIAS UTILIZADAS	57
3.7	CONSIDERAÇÕES FINAIS DO CAPÍTULO	57
4	AVALIAÇÃO DA PROPOSTA	59
4.1	DEFINIÇÃO	59
4.2	PLANEJAMENTO	60
4.2.1	Definição dos Participantes	61
4.3	CONDUZINDO O ESTUDO PILOTO	61
4.3.1	Cenário de Aplicação	62
4.3.2	Resultados Obtidos do Estudo Piloto	65
4.4	CONDUZINDO O ESTUDO DE CASO REGULAR	67
4.4.1	Execução do Estudo de Caso Regular	68
4.4.2	Resultados Obtidos do Estudo de Caso Regular	74
4.5	AMEAÇAS À VALIDADE	79
4.5.1	Validade Interna	79
4.5.2	Validade Externa	80
4.5.3	Validade do Constructo	80
4.5.4	Validade de Conclusão	80
4.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO	80
5	CONSIDERAÇÕES FINAIS	82
5.1	VISÃO GERAL	82
5.2	CONTRIBUIÇÕES	84
5.3	LIMITAÇÕES	85
5.4	TRABALHOS FUTUROS	85
	REFERÊNCIAS	87
	APÊNDICES	93
	Apêndice A QUESTIONÁRIO DE MÉTRICAS	94

Apêndice B QUESTIONÁRIO SOBRE USABILIDADE DA FERRAMENTA	98
Apêndice C MAPEAMENTO SISTEMÁTICO SOBRE VISUALIZAÇÃO DE INFORMAÇÃO	100
C.1 QUESTÕES DE PESQUISA	100
C.2 ESTRATÉGIA DE BUSCA	102
C.2.1 String de Busca	102
C.2.2 Seleção nas Bases	103
C.3 CRITÉRIO PARA SELEÇÃO DE ESTUDOS	103
C.3.1 Critérios de Exclusão	104
C.3.2 Critério de Inclusão	104
C.4 PROCESSO DE SELEÇÃO DOS ESTUDOS	105
C.5 AVALIAÇÃO DA QUALIDADE	106
C.6 EXTRAÇÃO DE DADOS E ESTRATÉGIA DE RESUMO	107
C.7 CONDUÇÃO DA REVISÃO SISTEMÁTICA	107
C.7.1 Resposta Q1: Técnicas de visualização utilizadas	108
C.7.2 Resposta Q2: Como técnicas de visualização influenciam a tomada de decisão	110
C.7.3 Resposta Q3: Qual a influência do domínio na escolha da estratégia de visualização	110
C.8 AMEAÇAS À VALIDADE	111
C.9 CONSIDERAÇÕES FINAIS	111

1 INTRODUÇÃO

Este capítulo tem como objetivo apresentar a motivação, problema, questão de pesquisa e o objetivo desta dissertação, assim como o contexto relacionado ao campo de estudo. A metodologia utilizada e a estrutura deste documento também são apresentados.

A tecnologia se mostra cada vez mais presente na rotina diária das empresas e indústrias, como dispositivos e ferramentas para manipulação de dados. A grande quantidade de dados gerada torna a análise manual uma atividade complexa a ser realizada. Esse mesmo efeito pode ser observado na agropecuária, onde o cenário de pressões econômicas e grande competitividade da pecuária leiteira tem induzido os produtores e instituições leiteiras a um aumento dos rebanhos em busca de melhoria da produção, da produtividade e da qualidade do leite, tornando importante a tomada rápida de decisões e gerenciamento de informações.

Em pesquisa relatada pelo MilkPoint¹, 78% dos produtores entrevistados apontaram o aspecto gerencial como o maior limitante à produção. Esta informação caracteriza a necessidade de capacitação e orientação do produtor para acesso e uso de processos gerenciais, eliminando restrições para melhoria da produtividade da produção e qualidade do leite.

Em 2002, foi criada pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) do Brasil, a Rede Brasileira de Laboratórios de Qualidade do Leite (RBQL), com a integração dos Laboratórios de Qualidade do Leite (LQLs) regionais. Estes laboratórios realizam análises de amostras de leite cru coletadas de animais em controle leiteiro e de tanques de resfriamento de produtores de leite e de estabelecimentos de laticínios do país, orientadas na instrução normativa (IN) n^o 62 do MAPA². Esta IN determina que as indústrias de laticínios devem realizar análises de amostras de leite dos rebanhos de produtores a fim de atingir as metas dos padrões de qualidade para o leite produzido no Brasil.

Assim, no contexto do manejo dos rebanhos leiteiros, o produtor de leite realiza os registros de desempenho zootécnico, isto é, das características de interesse econômico, sejam elas produtivas ou reprodutivas. Um exemplo dessa prática é a realização mensal do controle leiteiro, que consiste na pesagem do leite das vacas em lactação. Durante o

¹<https://www.milkpoint.com.br/>

²<http://www.apcbrh.com.br/files/IN62.pdf>

controle leiteiro, uma amostra do leite produzido é coletada de cada animal e encaminhada a um LQL para análise, cujos resultados caracterizam os indicadores de produção e qualidade de leite (gordura, proteína, lactose, sólidos totais). Estes controles geram grande volume de dados, necessitando, assim, de técnicas específicas para facilitar e, até mesmo, viabilizar sua interpretação.

Neste cenário, a análise se torna um processo não trivial de ser realizado, envolvendo um esforço cognitivo e, conseqüentemente, ocupando muito tempo por parte dos produtores e indústria leiteira. Há uma necessidade de informação em tempo real para suportar a habilidade de tomar decisões inteligentes de negócio rapidamente para que os produtores de leite se mantenham competitivos no mercado.

Portanto, é necessário que haja uma maneira eficiente de se armazenar, processar e interpretar estes dados complexos para que os produtores e instituições leiteiras sejam capazes de tomar decisões estratégicas em relação à produção do leite. Assim, esta dissertação propõe uma arquitetura que auxilie a Embrapa Gado de Leite (agentes da cadeia produtiva do leite) no processo de análise da produção e qualidade do leite de rebanhos leiteiros, utilizando ontologias e análise de dados, por meio de análise de redes complexas, para a extração de conhecimento implícito nos dados.

1.1 MOTIVAÇÃO

A competitividade de mercado atual, em conjunto com o grande volume de dados que precisa ser analisado para tomada de decisão, vem forçando as empresas a investirem em estratégias que auxiliem nesse processo. Neste contexto, a Inteligência de Negócios (Business Intelligence – BI) (CHEN et al., 2012) surge como um processo de coleta, organização e análise de informações que oferecem suporte a gestão de negócios em ambientes empresariais. O BI tem como objetivo principal o auxílio na tomada de decisões, reduzir custos, otimizar a eficiência do trabalho, contribuindo para a melhoria da gestão de processos (POPOVIĆ et al., 2012). Dentre as técnicas para apoio à tomada de decisão em BI, destaca-se o uso de componentes visuais, como Dashboards, gráficos e tabelas; bem como técnicas para extração de conhecimento e informações dos dados, como mineração de dados, aprendizado de máquinas e ontologias (RANJAN, 2008).

Diante do exposto, esta dissertação tem como motivação prover suporte computacional, por meio de uma arquitetura fundamentada nos conceitos de Inteligência de Negócios,

para que produtores e instituições leiteiras sejam capazes de analisar a produção do leite dos rebanhos.

Outra motivação deste trabalho é o emprego de técnicas de visualização de dados, ou visualização da informação, para auxiliar o processo cognitivo na interpretação de dados e descoberta de conhecimento (CARD et al., 2009) por parte das instituições leiteiras. Estas técnicas visam simplificar a maneira como estes dados são apresentados aos gerentes e tomadores de decisão, que podem observar tendências e informações implícitas em grandes volumes de dados.

Também motivadora é a criação de mecanismos que auxiliem na análise e interpretação de dados da produção leiteira através de técnicas de visualização de dados, como, por exemplo, o uso de componentes visuais de fácil interpretação por parte do usuário, consolidação de grande volume de dados, e redução da quantidade de informação exibida, para uma maior abstração destas informações, auxiliando na percepção do usuário (JONKER et al., 2013).

1.2 PROBLEMA

A diretriz do MAPA do Brasil (IN n^o 62 citada anteriormente), na qual há a necessidade de coleta periódica de amostras de leite, produzindo grande volume de dados a serem processados e armazenados, em conjunto com a necessidade da interpretação e análise desses dados para alcançar os padrões de produção e qualidade do leite produzido no Brasil, demonstra a necessidade de definição de uma arquitetura e o desenvolvimento de uma plataforma tecnológica, através da qual os produtores e as instituições leiteiras possam analisar os dados coletados e tomar decisões em relação à produção de leite dos rebanhos brasileiros.

Desta forma, o problema abordado por este estudo é: *como consolidar os dados oriundos de diferentes amostras de leite e prover uma análise visual e intuitiva, capaz de agregar valor para a tomada de decisão de produtores e instituições leiteiras em relação a produção do leite?*

1.3 QUESTÃO DE PESQUISA

A partir da motivação e problema apresentados, a questão de pesquisa deste trabalho é: *A concepção de uma arquitetura e o desenvolvimento de uma plataforma para visualização de dados sob a perspectiva da Inteligência de Negócios, proporciona um ferramental capaz de apresentar aos produtores e agentes da CPL conhecimento implícito nos dados, possibilitando a análise e tomada de decisão em relação a produção do leite?*

1.4 ENFOQUE

Para o desenvolvimento da proposta deste trabalho foram levantados, junto aos pesquisadores da Embrapa Gado de Leite, os requisitos funcionais e não-funcionais para o desenvolvimento da plataforma de visualização com foco na análise da produção do leite. Dessa forma, foi detectada a necessidade da análise semântica, com o intuito de prover conhecimento e uma visualização simplificada dos dados, bem como a necessidade de uma análise conjunta dos animais, para fornecer indicadores que os destaquem segundo suas produções leiteiras e possibilite uma fácil comparação entre eles. Assim, dois recursos computacionais principais foram selecionados para compor a solução proposta e suprir as necessidades identificadas: ontologias e redes complexas.

O uso de ontologia destaca-se como uma solução para a interpretação e análise semântica de dados, podendo ser utilizada em diversos domínios de aplicação (MIAH et al., 2007) (ZILLNER et al., 2008) (SILVA; CAVALCANTI, 2014) (LIU et al., 2016). A ontologia define relações explícitas e implícitas dos dados e é capaz de definir os conhecimentos referentes a um domínio, permitindo que este possa ser processado computacionalmente. Além disso, ela possibilita a extração de novas informações e atribuição de significados para os dados. O uso de ontologias agrega outros benefícios, tais como: maior integração dos dados, confiabilidade, portabilidade, reuso e manutenção (AKERMAN; TYREE, 2006). Considerando-se essas características, é esperado que o uso de uma ontologia do domínio de produção do leite possa auxiliar na análise e visualização dos dados.

As ontologias podem ser delineadas em forma de uma rede complexa, representando os indivíduos e suas associações e relações (HEIM et al., 2010). Ainda de acordo com (HEIM et al., 2010), estas redes permitem uma visualização mais efetiva das relações intrínsecas dos dados, observando as conexões entre indivíduos que compõem o domínio de aplicação.

As redes complexas são amplamente utilizadas com o objetivo de auxiliar na compreensão das formas de relacionamento entre diferentes entidades e identificação de elementos mais influentes em um conjunto de dados (BOCCALETTI et al., 2006).

Ao caracterizar a ontologia em forma de uma rede, é possível identificar como os dados estão associados, permitindo aplicar métricas de redes complexas, como análise de centralidade, grau dos nós, entre outras métricas próprias para identificar indivíduos em destaque no conjunto de dados observado. Neste sentido, visando suprir o requisito de comparação entre os animais considerando diferentes indicadores de produção do leite, foram utilizadas estratégias de análise de redes complexas, com o intuito de relacionar os animais e compará-los dinamicamente segundo os critérios definidos pelo especialista do domínio.

Diversos trabalhos utilizam ontologias e redes complexas individualmente para prover interpretações dos dados (QUEIROZ-SOUSA et al., 2013) (HEIM et al., 2010) (SILVA et al., 2017), porém poucos exploram ambas de maneira conjunta (FALCI et al., 2018) (LANCELLOTTA et al., 2018) (SAFAEIPOUR et al., 2017). Este trabalho explora a combinação de técnicas de ontologia e redes complexas de forma combinada, para complementar as informações de cada uma, de uma maneira automatizada ou semi-automatizada.

Ao combinar estas duas soluções, foram identificadas algumas vantagens que estas técnicas podem oferecer. As ontologias são capazes de oferecer benefícios para as redes complexas, dentre estes:

- (i) Abstração de nós da rede com a utilização de inferências;
- (ii) Agregação de valores semânticos para os dados da rede;
- (iii) Classificação dos indivíduos, identificando grupos semânticos que poderão ser analisados separadamente.

Por outro lado, a rede complexa também pode agregar benefícios para a ontologia, tais como:

- (i) Identificação de novos grupos de indivíduos, baseados em análises de entidades da ontologia, possibilitando a criação de outros grupos com características semelhantes, identificadas na rede;
- (ii) Identificação de novas relações entre os dados, automatizando o processo de criação de novas regras lógicas;
- (iii) Análise estrutural da ontologia e criação de novas propriedades dos dados, a partir

da utilização de métricas em redes complexas;

1.5 OBJETIVO

O principal objetivo desta dissertação é propor uma arquitetura sob a perspectiva da Inteligência de Negócios que define os componentes necessários para visualização da informação e apoio à tomada de decisão. A arquitetura utiliza ontologias e redes complexas de forma conjugada para obter conhecimento e prover a tomada de decisão. Como objetivos secundários destacam-se:

- (i) Desenvolver uma plataforma de visualização, baseada na arquitetura proposta, considerando a análise semântica e estrutural dos dados;
- (ii) Promover o uso de ontologias e redes complexas como técnicas conjugadas de forma a melhorar a interpretação de dados estratégicos;
- (iii) Desenvolver uma ontologia no domínio da produção do leite, relacionada a controle leiteiro;
- (iv) Usar componentes de visualização que possibilitem a interpretação e análise dos dados para tomada de decisão;
- (v) Avaliar a proposta e verificar a questão de pesquisa.

1.6 METODOLOGIA

A metodologia deste estudo foi realizada em 4 etapas principais: (i) revisão da literatura; (ii) definição da arquitetura sob a perspectiva da Inteligência de Negócios; (iii) implementação da plataforma para visualização dos dados de produção do leite que auxilie produtores e instituições leiteiras na tomada de decisão; e (iv) avaliação da proposta apresentada através de um estudo de viabilidade e um estudo de caso.

A revisão da literatura se deu pela condução de um estudo terciário para detectar as principais soluções em Inteligência de Negócios no domínio da agropecuária, visando encontrar estudos secundários que, por sua vez, podem ajudar a revelar o estado da arte da área, em particular na análise semântica para extração de conhecimento dos dados.

A definição da arquitetura se deu através dos requisitos funcionais e não-funcionais identificados pela condução de um estudo exploratório. Com os requisitos definidos a plataforma foi desenvolvida seguindo o modelo de desenvolvimento incremental, que com-

bina vantagens do modelo cascata, porém sendo aplicado de forma iterativa (PRESSMAN; MAXIM, 2016). O modelo é dividido em ciclos (incrementos) que contém as fases de comunicação, planejamento, modelagem (análise e projeto), construção (codificação e testes) e entrega.

O primeiro ciclo realizado gerou a primeira versão da plataforma, que foi avaliada através de um estudo de viabilidade com um especialista do domínio de produção de leite. Esta avaliação inicial teve o propósito de verificar a viabilidade técnica da proposta.

Posteriormente, novos ciclos de desenvolvimento foram realizados para ser possível inserir novos componentes na plataforma, e ainda promover melhorias de usabilidade e correção de defeitos identificados no estudo de viabilidade. Por fim, para avaliar se a proposta auxilia a tomada de decisão em um ambiente real, foi realizado um estudo de caso em uma instituição de pesquisa agropecuária envolvendo análise de dados de sistema de produção de leite.

1.7 ORGANIZAÇÃO

Este trabalho está dividido em 5 capítulos. O capítulo 2 aponta os principais fundamentos teóricos relacionados ao trabalho, explicando os principais conceitos necessários para o entendimento do mesmo, além de trabalhos no mesmo contexto. O capítulo 3 apresenta a arquitetura desenvolvida, assim como suas contribuições, funcionamento e principais conceitos. O capítulo 4 contém a avaliação desenvolvida com a utilização da arquitetura para observar sua viabilidade, destacando todo o processo de avaliação, assim como o seu planejamento e execução. Por fim, o capítulo 5 aponta as considerações finais, principais contribuições desta dissertação, assim como trabalhos futuros.

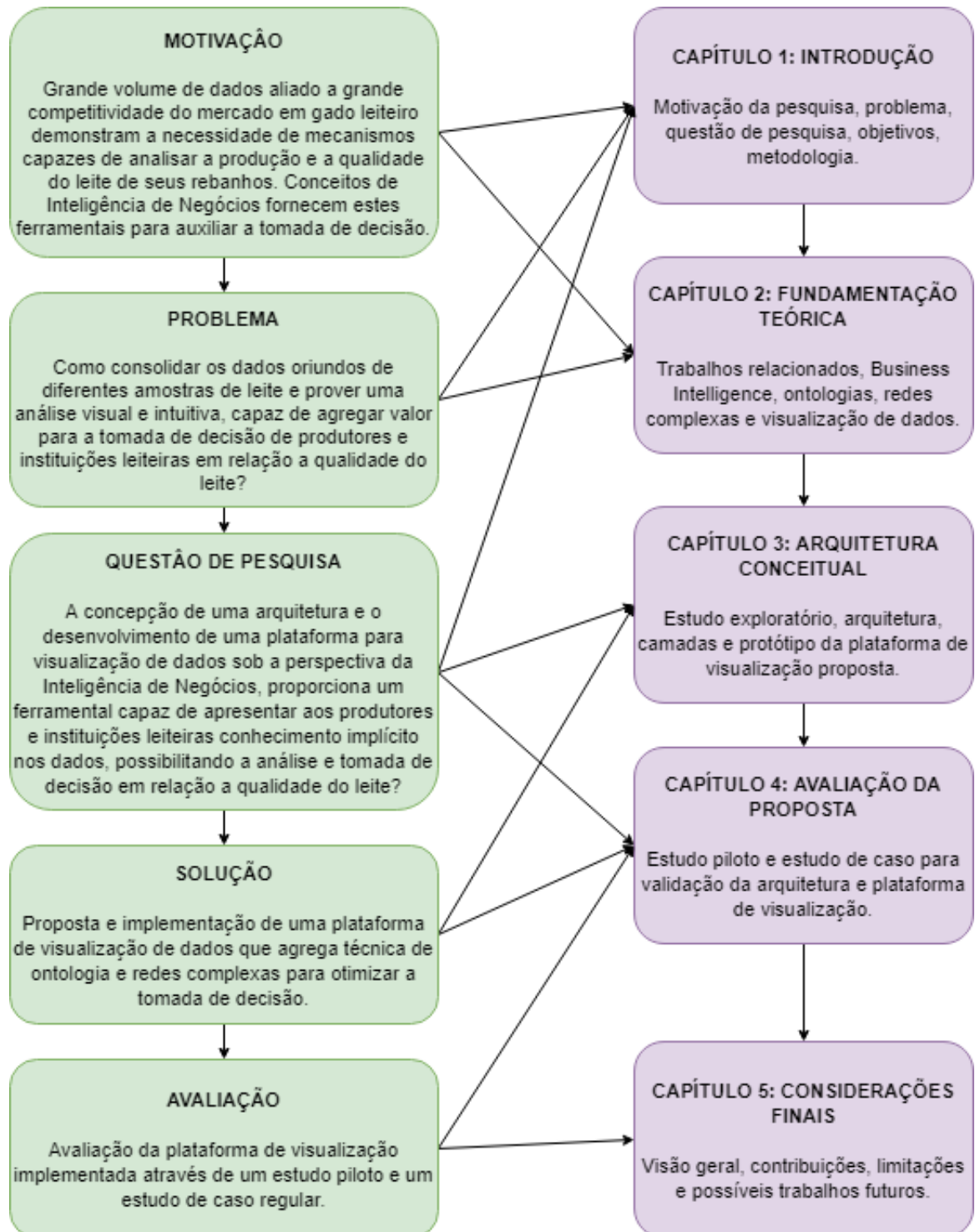


Figura 1.1: Estrutura da dissertação

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os principais conceitos para o entendimento desta dissertação, de forma a ilustrar como estes foram utilizados e, por fim, identificar trabalhos relacionados.

2.1 BUSINESS INTELLIGENCE

A inteligência de negócios, do inglês Business Intelligence ou simplesmente BI, provê mecanismos para otimizar a eficiência na tomada de decisão por parte de gestores de negócios ou empresas (LARSON; CHANG, 2016). A BI é utilizada de forma interdisciplinar, a fim de obter abstrações e informações em grandes volumes de dados independente do domínio de aplicação. Assim, aprimora o gerenciamento estratégico e fornece soluções para agilizar a tomada de decisão. Porém, a inteligência de negócios está constantemente em construção, com novas metodologias aplicadas em diferentes domínios, para fornecer informações mais rapidamente, que possam acompanhar o ritmo do ambiente de negócios onde há uma alta competitividade. Neste sentido, tomadas seguras de decisão realizadas com velocidade são fatores cruciais para manter a competitividade destas empresas no mercado.

A aplicação de BI fornece mecanismos para fomentar essa necessidade de informações rápidas no mercado, mas, com isso, surgem desafios, tais como identificar novos meios de representar os dados de uma maneira mais eficiente, identificar informações relevantes e apresentá-las efetivamente, possibilitando uma melhor tomada de decisão por parte dos gestores. A partir dos novos desafios e trabalhos gerados com a utilização de BI, esta área possui crescentes possibilidades de pesquisa (TRIEU, 2017).

Observando as principais lacunas encontradas no contexto de BI, uma das questões de pesquisa que necessita ser explorada expõe questionamentos sobre como a BI pode auxiliar na eficiência competitiva de uma empresa. Neste sentido, foi explorada esta questão com relação a utilização de ferramentas de visualização mais apropriadas, de forma a identificar se elas possibilitam uma melhoria na interpretação de grandes volumes de dados. Para isso foi realizado um mapeamento sistemático no Apêndice C que retrata um estudo sobre os principais meios de visualização e como sua aplicação pode influenciar na tomada de decisão.

2.1.1 Núcleo Genérico de Processo de BI

A BI provê mecanismos que apoiam a gestão de grandes conjuntos de dados, coletando e organizando, possibilitando a criação e visualização de indicadores que ajudam a mensurar valores destes dados e definir estratégias para tomada de decisão. Ao aplicar os processos de BI, as informações adquiridas por fontes heterogêneas são convertidas em informações úteis e que possibilitam uma melhor gestão das informações e dos dados em geral.

Um modelo genérico do ciclo de vida de um processo de BI pode ser definido como na Figura 2.1 (VUORI, 2007). A primeira etapa é a definição das informações necessárias para embasar os gerentes de negócio na tomada de decisão. Na segunda etapa, os dados devem ser coletados e organizados para que posteriormente, na terceira etapa, eles possam ser processados de forma a gerar conhecimento aos tomadores de decisão. Em seguida, na quarta etapa, a informação deve ser analisada para verificar o seu significado no contexto ao qual ela está inserida. Na quinta etapa essas informações são apresentadas aos tomadores de decisão que irão, na sexta etapa, utilizar as ferramentas disponíveis para analisar as informações consolidadas. Por fim, esses gestores apresentam seus feedbacks, que deverão ser incorporados como novas informações necessárias no processo.

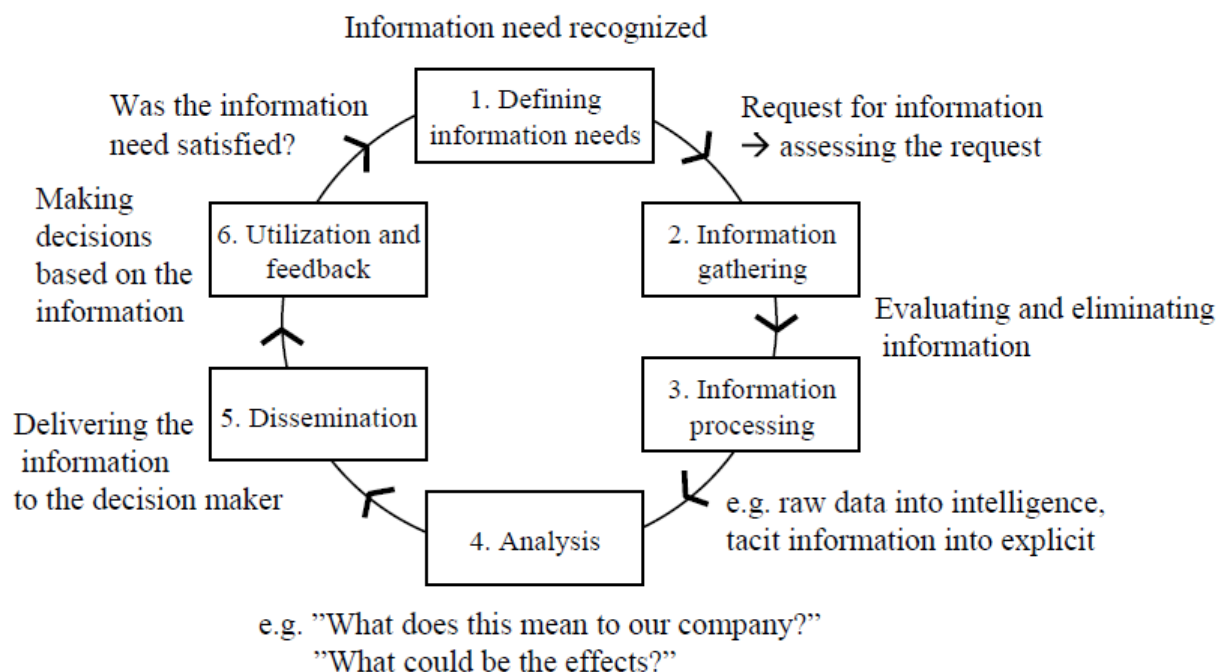


Figura 2.1: Elementos do processo de BI (VUORI, 2007)

Neste trabalho, esse ciclo foi utilizado para a definição do núcleo básico da arquitetura proposta no Capítulo 3.

2.2 VISUALIZAÇÃO DE DADOS

A visualização de dados desempenha um papel importante na compreensão e *awareness* de stakeholders em domínios de aplicação em que são fundamentais a análise de grandes volumes de dados. A escolha da abstração de visualização e técnicas para representar os dados, assim como técnicas de interação que são empregadas, depende fortemente da informação contextual (SCHOTS, 2014).

Neste contexto, a visualização de dados pode auxiliar no processo cognitivo, facilitando a interpretação de dados, pois a representação visual pode ser entendida de maneira mais simples do que um conjunto de textos ou números (RUSSELL et al., 2009). Existem diversos tipos de modelos visuais para análise, nos quais várias técnicas são utilizadas para conseguir fazer a representação destes dados. As visualizações têm grande influência na percepção humana, sendo que a abstração da informação pode ser mais facilmente interpretada ou, em alguns casos, podem ser inadequados. A utilização de elementos visuais de maneira errada podem prejudicar o entendimento das informações observadas pelo usuário (RIBEIRO et al., 2016).

Uma técnica de visualização precisa ser facilmente interpretável para o usuário, de forma que este consiga inferir as informações importantes de um conjunto de dados complexos, relacionadas a um domínio de aplicação. As representações visuais podem traduzir dados em uma forma visível que ajude o analista a perceber rapidamente aspectos destes dados (RIBEIRO et al., 2016). Organizando a natureza dos dados, abstrações visuais podem ser combinadas com representações técnicas para ressaltar aspectos relevantes de um determinado contexto. Metáforas visuais são uma abordagem que auxiliam na *awareness* (SCHOTS, 2014).

A partir disso, foram consolidadas técnicas de visualização de dados para realizar uma melhor abstração de dados. Com a identificação destas técnicas, atribuir visualizações que possibilitem uma melhor interpretação dos dados relacionados a produção do leite, domínio de aplicação estudado nesta dissertação.

2.2.1 Técnicas de Visualização de Dados

Um problema pode ser visualizado de diversas formas distintas. Ao escolher a melhor visualização para um objetivo, seus requisitos precisam ser expressados pela técnica uti-

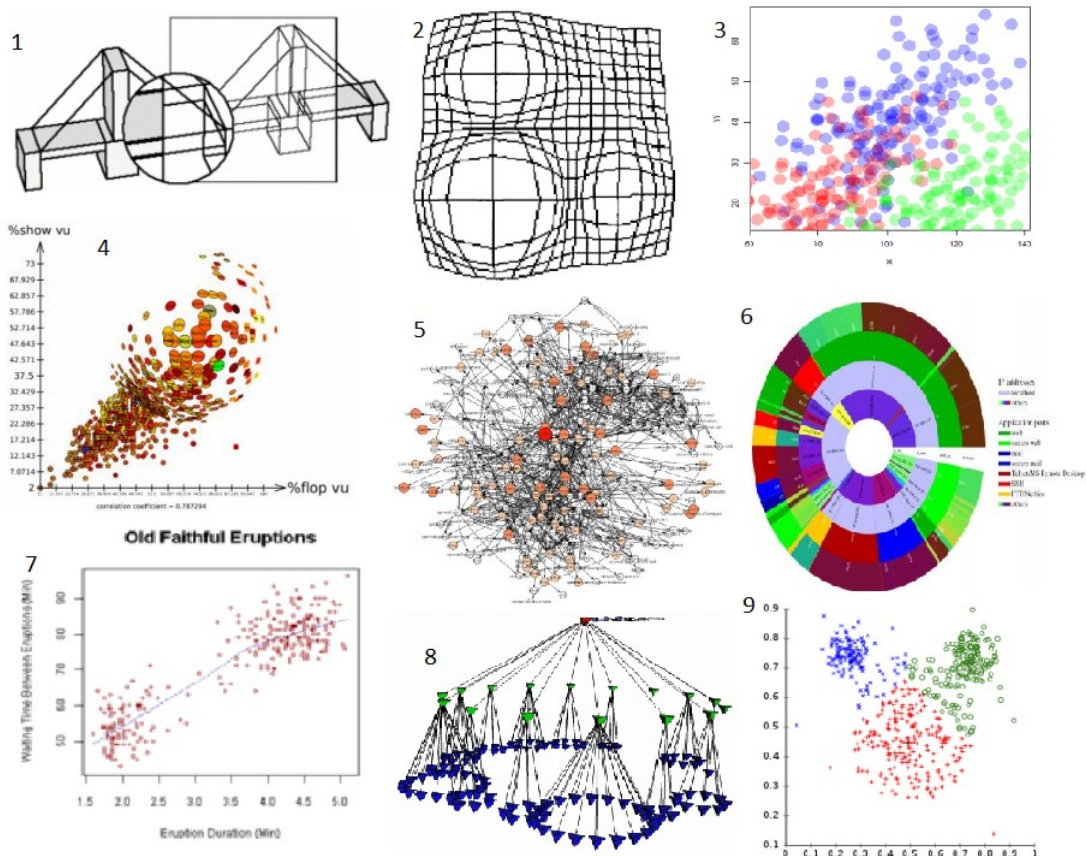


Figura 2.2: Exemplos de estratégias visuais (SCHOTS, 2014).

1- Magic Lens, 2- Polyfocal Display, 3- Transparency, 4- Fisheye View, 5- Overview, 6- Radial, 7- Scatter Plot, 8- Cone Tree, 9- Clustering

lizada. Desta forma, é possível categorizar um conjunto de técnicas e suas interações, de forma a organizá-las (SCHOTS, 2014). Esta categorização pode ser dividida em 3 formas:

- Visualização da Informação: Detalhamento das principais técnicas que permitem visualizar informação (gráficos, em geral). São apresentadas imagens para cada representação de visualização (Figura 2.2).

A imagem 1 é um exemplo de técnica de foco + contexto (focus + context), na qual há um foco de uma parte em especial dos dados onde se quer dar maior atenção. As imagens 2 e 4 representam gráficos de distorção (distortion - que também são gráficos de foco + contexto), na qual aumenta a região de interesse enquanto diminui as regiões ao redor. As imagens 3 e 7 são representações de sobreposição, utilizada quando a distribuição tende a ser esparsa com áreas de grande densidade. A imagem 5 representa tipos de visão geral (Overview), em que o objetivo principal é representar toda entidade de dados que a visualização da informação representa.

As imagens 6 e 8 ilustram representações visuais para análises hierárquicas, que são usadas para representar estruturas complexas e suas relações. Por fim, a imagem 9 tem como principal vantagem a divisão de subgrupos de acordo com sua similaridade em relação à uma medida.

- **Apresentação:** Representa as formas na qual as visualizações são apresentadas para o usuário, que podem ser sequenciais ou simultâneas. A distinção maior entre as formas é que a sequencial apresenta apenas uma visualização por vez em uma ordem, sendo interessante quando há limitações de tamanho para exibição da visualização. Na abordagem simultânea várias visualizações são apresentadas na tela ao mesmo tempo, proporcionando uma visão por várias perspectivas distintas.
- **Interação:** Representa as principais interações que o usuário pode ter com o conjunto de dados. Dentre estas interações, é importante ressaltar: filtragem, que tem como objetivo principal reduzir o espaço amostral, apenas enfatizando o que é relevante; Ordenação, que organiza os dados de acordo com algum critério; Rotação, onde o usuário pode interagir com os dados rotacionando a câmera da visão ou rotacionar o objeto que está sendo visualizado.

Como resultado do mapeamento sistemático (Apêndice [C](#)), os elementos de visualização identificados foram consolidados na Tabela [2.1](#). Pode ser observado que há técnicas de visualização que focam na interface com o usuário, buscando uma apresentação mais adequada da informação. Por outro lado, existem estratégias que focam na abstração dos dados, com o intuito de consolidá-los para gerar informação útil ao usuário final. Isso pode ser observado, principalmente, pelas características: Overview, Overview + Detail, Hierarchical Visualization e Details on Demand. Outro fator identificado na revisão da literatura é a importância do uso de componentes intuitivos para os usuários, independente do domínio de aplicação. Neste sentido, a proposta deste trabalho visa o uso de técnicas para a interpretação semântica e estrutural dos dados, apoiadas por componentes visuais intuitivos, possibilitando a tomada de decisão por parte dos usuários.

2.3 ARQUITETURA ORIENTADA A SERVIÇOS

A arquitetura orientada a serviços (*Service Oriented Architecture* - SOA) é um modelo organizado de forma lógica para estabelecer funções independentes, para otimizar a agili-

Tabela 2.1: Características de visualização

Característica Funcional	Definição	Exemplos
Focus + Context	Integra foco e contexto em uma única exibição.	Magic Lens; Distortion;
Distortion	Abrange técnicas de distorção espacial que tornam a interpretação de dados possíveis, mas preservando o contexto durante a exploração.	Fisheye View; Polyfocal Display;
Overview	Há uma visão geral do contexto dos dados. É uma representação de toda entidade de dados que a visualização pode representar.	Overview (graph network);
Overlap	Quando uma distribuição tende a ser esparsa, áreas de alta densidade são difíceis de se ver. Portanto, sobrepor itens se torna uma prática comum.	Scatter Plot; Parallel Coordinates; Transparency;
Hierarchical Visualization	Hierarquias são estruturas de informação geralmente usadas para representar estruturas complexas. Essa característica visa simplificar a representação visual dos dados através de camadas hierárquicas.	Treemap; Cone Tree; Radial; Clustering;
Layout	Apresenta a organização espacial, composição de múltiplos elementos na visualização.	3D;
Details on Demand	Aplicado quando um item ou subgrupo é selecionado para se capturar detalhes do mesmo.	Drill-Down; Labeling;
Binning	Esta abordagem supera o problema de sobreposição, especialmente quando o usuário afasta o zoom para observar os dados em escalas menores.	Maps;

dade e eficiência de organizações. Com o desacoplamento de responsabilidades, possibilita que funções sejam criadas em forma de serviço e os dados recebidos como resposta de sua execução sejam consumidos por um software (ERL, 2008).

Desta forma, é possível criar serviços que podem ser utilizados de forma distribuída, possibilitando uma integração de funcionalidades criadas em diversas plataformas. Os Serviços também podem interagir entre si, de forma a aumentar a gama de possibilidades de integração com serviços externos, tornando uma arquitetura importante para a utilização de dados provenientes de outras fontes e distribuídos geograficamente.

2.4 ONTOLOGIA

Ontologias podem ser consideradas como um sistema particular de categorias que é capaz de representar uma visão do mundo. O sentido mais utilizado no contexto de tecnologia, é o de que a ontologia se refere a um artefato de engenharia, constituído por vocabulários específicos que costumam descrever uma determinada realidade, além de possuir

um conjunto de hipóteses sobre o significado do vocabulário das palavras (GUARINO et al., 1998). Em seu lado metodológico, sua maior peculiaridade é a alta interdisciplinaridade, tornando possível analisar a estrutura de uma dada realidade em um alto nível de abstração e formulando um vocabulário rico e limpo.

Neste sentido, a ontologia é capaz de conseguir representar uma realidade. Assim, a interpretação e inferência de novos conhecimentos sobre um domínio específico de dados, pode constituir uma poderosa ferramenta de apoio ao processo cognitivo de análise semântica dos dados. Um dos usos fundamentais de uma ontologia é a conexão com um componente de banco de dados. O modelo conceitual resultante pode ser representado como uma ontologia processável pelo computador e mapeada de maneira concreta (GUARINO et al., 1998). Desta forma, o usuário pode fazer buscas na ontologia, sendo capaz de formular consultas no nível específico desejado. Uma das aplicações possíveis é a de identificar indivíduos semelhantes em bases heterogêneas, de forma a minimizar a perda de dados em fontes distintas.

Muitos dados disponíveis atualmente não apresentam aspectos semânticos, portanto, são interpretáveis apenas por humanos. Ao inserir o aspecto semântico, torna-se possível a manipulação destes dados através de máquinas (BERNERS-LEE et al., 2001). A ontologia consegue descobrir os significados de bancos de dados para quaisquer dados lidos. Neste sentido, a ontologia se torna uma poderosa ferramenta de interpretação, que pode descobrir novas informações a respeito de algum conjunto de dados.

2.5 REDES COMPLEXAS

As mais diversas áreas, como sociais, tecnológicas e biológicas possuem elementos e indivíduos que se relacionam através de associações explícitas, implícitas ou ocultas. Desta forma, estas redes possuem inúmeras conexões que podem não ser regulares. As redes podem ser consideradas complexas de acordo com suas características de relações entre os elementos que a compõem (KIM; WILHELM, 2008). Dentre estas características também podem ser ressaltadas a distribuição de grau nos nós desta rede e um alto coeficiente de clusterização.

As redes complexas são utilizadas em muitas áreas, assim como a biologia, que possui grandes quantidades de nós e estruturas topológicas complexas (LÜ et al., 2013). Com isto, são criadas novas funções e mecanismos com o objetivo de definir relações entre os

elementos desta rede, de forma a identificar padrões, reconhecer indivíduos mais relevantes e captar formas mais eficientes de interpretar estas redes. Em aplicações para outros domínios também são utilizadas métricas de redes complexas a fim de identificar elementos de destaque nestas rede, como por exemplo a utilização de métricas de centralidade para análise de ponto ótimo em circuitos elétricos (SALEH et al., 2018).

Duas métricas foram utilizadas nesta dissertação: *closeness centrality* (Equação 2.1) e *betweenness centrality* (Equação 2.2) (WASSERMAN; FAUST, 1994). A medida de *closeness centrality* representa a distância média dos caminhos mínimos de um nó até todos os outros nós da rede. Portanto, quanto mais central um nó é, mais próximo ele estará de outros nós da rede. Outra medida é *betweenness centrality*, que indica quantos caminhos mínimos passam através de um nó específico, permitindo identificar os nós mais centrais na rede e que possuem uma conexão forte com outros nós. As utilizações de ambas métricas permitem a definição de indivíduos relevantes dentro da rede, de forma a estabelecer grupos semânticos a partir da caracterização destes indivíduos.

$$C(x) = \frac{1}{\sum_y d(x, y)} \quad (2.1)$$

$$C_b(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(V)}{\sigma_{st}} \quad (2.2)$$

2.6 TRABALHOS RELACIONADOS

O principal objetivo desta seção é identificar trabalhos relacionados à proposta apresentada neste trabalho, assim como observar deficiências encontradas no contexto de visualização de dados. Foram revisados trabalhos que utilizam visualização da informação, assim como técnicas de BI para apoiar a tomada de decisão. Os trabalhos filtrados usam, principalmente, análises semânticas e de redes complexas como abordagem, contribuindo para a interpretação da informação e análise estratégica de conjuntos de dados de qualquer domínio.

Ao identificar os trabalhos relacionados, inicialmente foram analisados os trabalhos com aplicações de BI, além disso, como as abordagens de redes complexas e ontologias auxiliam na visualização da informação. Os principais desafios observados se referem ao grande volume de informação e mecanismos para otimizar a eficiência da tomada de

decisão em processos gerenciais. Analisar grandes volumes de dados é um desafio em diversas áreas de pesquisa. Algumas pesquisas focam em reduzir a quantidade de dados disponível para visualização (GU; WANG, 2011). Desta forma, a abstração da informação para o usuário é realizada com uma redução do espaço amostral, possibilitando análises em grupos menores, tornando menor o esforço de estabelecer relações entre grupos maiores e que apresentem uma topologia complexa.

Alguns estudos utilizam ontologias e / ou métricas de redes complexas para visualizar dados de forma mais eficiente e identificar elementos importantes para análises (QUEIROZ-SOUSA et al., 2013) (STRÖELE et al., 2017), e a criação de ferramentas de visualização que se comportam de forma automatizada ou semi-automatizada (SOBRAL et al., 2016). No entanto, poucos estudos usam estas técnicas de forma conjunta para interpretação dos dados (HEIM et al., 2010). Outros apenas usam diferentes formas de visualizar os dados (DÖRK et al., 2012) (RADICS et al., 2015) (SOKLAKOVA et al., 2016).

Neste sentido, foram examinados os principais benefícios e usos de abordagens relacionadas a BI em distintos domínios de aplicação, apoiando a tomada de decisão. Ontologias têm a habilidade de agregar informação semântica aos dados analisados, de forma a representar um domínio de aplicação, incorporando auxílios para o entendimento na especificação de um conjunto de dados (SUMALATHA et al., 2008). Uma utilização recorrente da ontologia é para caracterizar um domínio de aplicação de forma consistente. É possível representar atores que integram este domínio e suas inter-relações, criando um conceito consolidado e unificado de um domínio (JAYARAMAN et al., 2013). O uso de redes complexas permite uma investigação estrutural dos dados, desta forma, é possível identificar indivíduos que são mais influentes na rede assim como suas relações intrínsecas (LOPES et al., 2010). Ainda de acordo com (LOPES et al., 2010), para poder destacar indivíduos na rede, métricas de redes complexas são utilizadas, assim como grau dos nós, centralidade, entre outras. Estudos recentes fizeram esforços para identificar algoritmos e métricas para identificação de indivíduos relevantes e suas relações implícitas entre estes indivíduos em redes complexas, por exemplo, redes sociais (SILVA et al., 2017) (MUNIZ et al., 2017).

Com relação ao uso de BI, foram selecionados 4 artigos que utilizam abordagens de visualização em domínios específicos de aplicação. Todos os trabalhos apresentam relevância

na pesquisa em visualização de dados, sendo que alguns apresentam maiores contribuições no embasamento teórico para a recomendação da técnica para o gerente, tornando a experiência com a visualização algo mais simples e com grande poder de informação.

- **Artigo 1:** (THULKE et al., 2005)
- **Artigo 2:** (SWAYNE et al., 2003)
- **Artigo 3:** (RUSSELL et al., 2009)
- **Artigo 4:** (PARK; BASOLE, 2016)

Inicialmente, foi realizado um pequeno resumo sobre cada trabalho e, em seguida, criado um quadro comparativo entre os mesmos. (THULKE et al., 2005) chegou à conclusão que algumas abordagens tradicionais que utilizavam apenas duas dimensões de dados poderiam não ser suficientes para representar toda a informação necessária. Com isso, a necessidade de se utilizar ou criar técnicas que consigam ter a capacidade de visualizar em 3 ou mais dimensões para que se consiga analisar corretamente. Ao fim, (THULKE et al., 2005) recomenda a utilização da técnica desenvolvida para casos de análise de dados espaço-temporais.

(SWAYNE et al., 2003) apresenta uma ferramenta com possibilidade de integração com outras ferramentas, além da capacidade de utilização da importação dos dados com a utilização da linguagem XML. Além disso, é possível modelar representações de redes complexas assim como grafos, porém para uma melhor conclusão em relação à relevância da ferramenta seria melhor utilizá-la antes.

(RUSSELL et al., 2009) vê o potencial das análises visuais em melhorar o processo de gerenciamento de construções, há pouca literatura com este mesmo foco de pesquisa. Princípios gerais para guiar a utilização do design desenvolvido foram identificados no contexto de gerenciamento de projetos de construção. Assim como (THULKE et al., 2005), identificou a importância de representações gráficas que explorem mais de duas dimensões, facilitando a análise, além da interação dos usuários com a questão visual.

(PARK; BASOLE, 2016) observa indícios da viabilidade do uso de sua técnica desenvolvida, e apresenta alguns pontos em que se poderia melhorar, descrevendo algumas das limitações observadas. Também menciona que as competências individuais podem impactar na performance da interpretação da técnica e, consideram ampliar esse estudo para novos aprendizados.

Em resumo, os trabalhos apresentados chegam à conclusão que a utilização de meios visuais ajudam a facilitar a tomada de decisão por parte dos gerentes e tomadores de decisão. A escolha da técnica de visualização é um fator muito importante neste processo, podendo comprometer a interpretação correta dos dados. É preciso fazer um estudo mais aprofundado para avaliar a utilização de diferentes técnicas de acordo com o tipo dos dados ou o contexto. À partir destes estudos, pode-se observar uma área com muitos desafios de pesquisa a serem explorados ainda, com trabalhos futuros e contribuições. Com isto, algumas das questões de pesquisa podem ser modificadas ou incluídas novas questões a fim de explorar melhor este domínio. Por fim, é apresentada uma tabela comparativa que resume os aspectos encontrados nos trabalhos analisados (2.2). Esta tabela foi feita para uma análise resumida dos pontos mais relevantes encontrados nos artigos lidos. Com isso, é possível observar quais artigos tiveram uma maior contribuição para o embasamento teórico no domínio de estudo.

Tabela 2.2: Tabela comparativa entre os artigos selecionados

	Artigo 1	Artigo 2	Artigo 3	Artigo 4
Desenvolve uma nova técnica de visualização	x		x	x
Fundamenta a escolha da técnica de visualização	x		x	x
Realiza experimento empírico para validação			x	x
Leva em conta o contexto do usuário	x		x	x
Leva em conta o dispositivo do usuário	x			
Leva em conta o contexto dos dados	x		x	x
Utiliza técnicas de visualização para representação gráfica de dados	x	x	x	x

Em relação à esta dissertação, observamos o uso de BI em diferentes domínio de aplicação para auxiliar a tomada de decisão. De forma que para realizar uma análise mais eficiente, é necessário realizar uma análise minuciosa do contexto de aplicação, além de ressaltar a importância de realizar experimentos empíricos das visualizações criadas, obtendo feedbacks dos usuários, a fim de aprimorar as estratégias visuais utilizadas para sumarização. Para uma melhor fundamentação das visualizações, é necessário que se observe juntamente com os usuários quais as principais necessidades do domínio de aplicação.

Não foram encontrados trabalhos específicos de produção de leite com aplicações de BI utilizando redes complexas e ontologias, desta forma, foi identificada esta lacuna a partir dos trabalhos selecionados.

2.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo foram analisados os conceitos sobre ontologias, redes complexas e arquitetura orientada a serviços, utilizados nesta dissertação. Além disso, foram apresentados trabalhos relacionados com as principais aplicações associadas a arquitetura proposta, descrita no próximo capítulo.

3 ARQUITETURA CONCEITUAL

Este Capítulo apresenta o fluxo básico de soluções baseadas em BI, através do qual a arquitetura proposta se baseia e, além disso, as soluções computacionais que suportam o desenvolvimento da plataforma de visualização para tomada de decisão são descritas. Para tal, foi conduzido um estudo exploratório com especialistas do domínio, através do qual foram detectados os requisitos necessários para esta proposta, permitindo a definição de uma arquitetura sob a perspectiva de Inteligência de Negócios para o domínio de produção de leite.

Revisitando o problema a ser tratado neste estudo, são apresentados, neste capítulo, uma arquitetura e o desenvolvimento de uma plataforma para visualização de dados sob a perspectiva da BI, através da qual produtores e instituições leiteiras sejam capazes de tomarem decisões estratégicas visando a produção do leite. Baseado na revisão da literatura, identificamos a escassez de ferramentas que auxiliem produtores a tomarem decisões efetivas no contexto de gado leiteiro. Algumas ferramentas identificadas em trabalhos relacionados não possuem ou não oferecem em sua totalidade mecanismos para integração de dados de fontes distintas, visualização da informação, entre outros recursos, dificultando a agregação de informações provenientes de fontes heterogêneas e a tomada de decisão a partir das informações agregadas. Além disso, há um problema na confiabilidade das informações adquiridas, de forma que outras ferramentas exibem informações incompletas ou limitadas, inviabilizando a tomada de decisão efetiva a fim de otimizar as decisões estratégicas da cadeia produtiva do leite.

A partir do ciclo de vida analisado na seção [2.1.1](#), foi identificado que o núcleo da arquitetura proposta para esta dissertação necessita contemplar os elementos básicos do processo de BI ressaltados. Com isto, foi representado o fluxo desta arquitetura na Figura [3.1](#), na qual são definidas quatro camadas base que contemplam os processos padrão observados em demais processos de BI da literatura, assim como a transição de informação entre as camadas.

A Camada de ETL é responsável por extrair os dados de diferentes fontes de dados, consolidando-os em um repositório central. Esta camada contempla o passo 2 do fluxo de BI da Figura [2.1](#). No domínio agropecuário, identificamos que as bases de dados, onde são

encontradas as informações dos animais, estão disponíveis em diferentes locais e possuem contextos de aplicação diferentes, como análises laboratoriais, bases genéticas, informações de fazendas de produtores, planilhas eletrônicas, entre outras. Desta forma, esta camada deve integrar as informações oriundas de diferentes fontes de dados e consolidá-las em um repositório único.

A Camada de Processamento contempla os passos 3 e 4 do processo básico de BI. Nela as informações são organizadas e analisadas por especialistas do domínio no intuito de verificar se elas agregam valor considerando as necessidades definidas inicialmente.

A Camada de Visualização contém os componentes de visualização necessários para apoiar os gestores na tomada de decisão (passo 5 da Figura 2.1). Considerando o mapeamento sistemático realizado, essa camada é crucial para que os usuários consigam analisar grandes volumes de dados. É nessa camada que os componentes devem ser identificados para que as informações sejam apresentadas de formas variadas e complementares, apoiando os gestores na tomada de decisão.

A Camada Decision Making foi projetada para atender ao sexto passo do processo de BI. É importante que a solução proposta permita que os gestores interajam com a plataforma e visualizem os dados sob diferentes perspectivas. Neste sentido, esta camada é responsável por apresentar as informações utilizando diferentes componentes de visualização, permitindo que os gestores analisem os dados sob óticas diferentes.

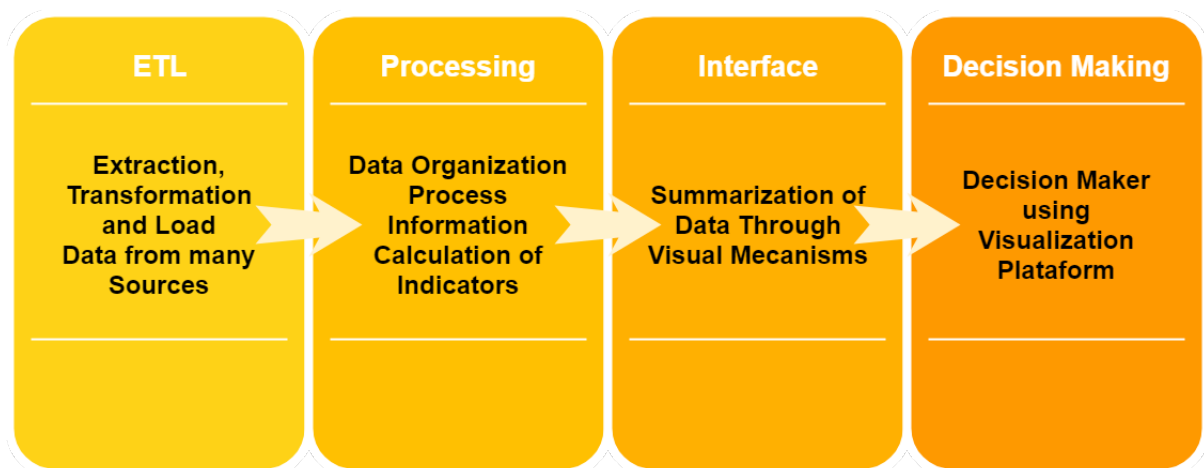


Figura 3.1: Núcleo da arquitetura, baseado no processo de BI

Com base no modelo de BI e na definição do núcleo genérico da arquitetura proposta, foi conduzido um estudo exploratório para identificar as necessidades dos gestores da Embrapa Gado de Leite, no domínio de produção do leite. Esse estudo está descrito na

próxima seção e, com o seu resultado, a arquitetura conceitual foi concebida.

3.1 ESTUDO EXPLORATÓRIO

Com a definição do processo de BI, visto na seção ??, é definido o fluxo básico, e camadas para a construção da arquitetura, é necessário identificar os requisitos funcionais e não funcionais para levantar as principais necessidades da CPL e, com isso, explicitar as principais necessidades de visualização e atributos que contém importância econômica relacionados aos atores que compõem o domínio de aplicação de produção do leite. Neste sentido, foi necessária a execução de um estudo exploratório a fim de identificar os principais requisitos.

3.1.1 Planejamento do Estudo Exploratório

Para execução do estudo exploratório necessitou-se realizar uma coleta de dados com auxílio de um questionário, seguindo os seguintes passos: (i) definição da relação do questionário com a teoria da pesquisa; (ii) projeto do questionário; (iii) definição do procedimento de aplicação; (iv) realização de um piloto; (v) coleta dos dados; (vi) análise dos dados e (vii) apresentação dos resultados (DRESCH et al., 2015).

A definição da relação do questionário com a teoria da presente pesquisa é baseada na identificação das necessidades dos pesquisadores da Embrapa Gado de Leite no contexto da produção do leite, para que sejam definidos os requisitos para o desenvolvimento da plataforma de visualização para tomada de decisão.

O projeto do questionário foi realizado contextualizando a pesquisa e definindo o uso de perguntas fechadas em escala de Likert, variando de 1 a 5 (ALLEN; SEAMAN, 2007). O questionário teve o propósito de avaliar os elementos da arquitetura e os requisitos para o desenvolvimento da plataforma. O objetivo principal do questionário foi verificar, juntos aos interessados, as informações necessárias para a tomada de decisão, no intuito de identificar as soluções computacionais que pudessem atender a essas necessidades.

Considerando a escala adotada, foi definido um critério de eliminação de componentes de acordo com as respostas obtidas. Não foram incorporados na arquitetura requisitos que obtiveram um valor médio de resposta inferior a 3 de acordo com escala adotada, ou seja, os elementos considerados não necessários pela maioria, não foram considerados no

desenvolvimento da plataforma.

Para o levantamento de requisitos, a principal necessidade é levantar todos os aspectos informativos que são aderentes ao domínio de aplicação. Para que, assim, as etapas que definem as especificidades do domínio sejam desenvolvidas de maneira aderente às principais necessidades. Desta forma, é fundamental que pesquisadores e usuários (como produtores, instituições leiteiras) possam fornecer quais as informações estratégicas imprescindíveis para a tomada de decisão. O questionário contido no Apêndice [A](#) foi utilizado para obter respostas relacionadas a estas necessidades. O questionário está disponível online e pode ser encontrado em um link da internet [1](#). As respostas obtidas foram compiladas e, com isso, foram identificados os principais elementos necessários para a definição da arquitetura e desenvolvimento da plataforma de visualização para tomada de decisão.

3.1.2 Definição dos Participantes

O estudo exploratório tem como objetivo levantar as principais necessidades estratégicas que a plataforma de visualização deve anteder, desta forma, os participantes foram selecionados de forma mais abrangente, ou seja, foram selecionados possíveis usuários da plataforma, para identificar as necessidades de usuários que possam ter contato com produções leiteiras ou gado de leite em geral.

Desta forma, foram selecionadas 22 pessoas com experiência em visualização de informação e no contexto do domínio de aplicação. Estes participantes enunciaram os principais indicadores e métricas que foram utilizados posteriormente na criação da plataforma e exibição dos indicadores de animais. As informações sobre requisitos foram colhidas através da utilização dos formulários aplicados, observações e entrevistas informais.

3.1.3 Execução do Estudo Exploratório

Uma prévia com 2 participantes com perfis distintos, um pesquisador da UFJF e um pesquisador da Embrapa, foi realizada antes do envio para todos os participantes, com o intuito de avaliar o questionário. Através do *feedback* dos participantes foi possível aprimorar o questionário em relação à qualidade das perguntas e organização das informações. O questionário utilizado possui 15 questões, perguntando sobre os requisitos que devem

¹<https://goo.gl/forms/ApFS2Mh5h0oApWWi2>

compor a plataforma, além de questões acerca do perfil do participante e do nível de conhecimento acerca do domínio de aplicação que é gado leiteiro. O questionário utilizado está disponível no Apêndice [A](#), e ficou aberto para recebimento de respostas pelo período de 4 semanas.

O questionário de levantamento de requisitos foi subdividido em 4 etapas. Inicialmente uma contextualização sobre o questionário, fazendo uma apresentação sobre a pesquisa desenvolvida e algumas de suas características. Em seguida, uma seção para a caracterização dos participantes com perguntas simples para identificação pessoal e reconhecimento sobre o nível de conhecimento em relação ao domínio de aplicação. A terceira seção determina as principais necessidades em relação a visualização da informação no contexto de produção do leite. Por fim, a quarta seção busca levantar indicadores econômicos relacionados aos animais, ou seja, as medidas de desempenho e qualidade associadas a esses animais. Os resultados obtidos com a análise dos questionários e após aplicação do critério de eliminação são apresentados a seguir.

Foi recebido um total de 22 respostas dos participantes selecionados. Dentro do contexto de gado leiteiro, foram selecionados participantes de áreas distintas, tais como economia, transferência de tecnologia, geoprocessamento, fisiologia vegetal, engenharia de software, análise laboratorial, melhoramento animal, relacionamento com o cliente, sensoriamento remoto e socioeconomia.

A primeira pergunta faz menção ao conhecimento do participante em relação ao contexto de gado leiteiro, visando caracterizar, essencialmente, o seu perfil e conhecimento em relação à produção de leite em gado leiteiro. Pode ser observado, pela Figura [3.2](#) que a maioria dos participantes possuíam conhecimento sobre o domínio de gado leiteiro. Com isso, há uma aderência maior das respostas em relação às principais necessidades apontadas no domínio de aplicação.

Na segunda etapa do questionário, com relação à importância da criação da plataforma de visualização e algumas das características necessárias para exibir informações de forma a auxiliar os gestores, nota-se a carência de mecanismos que apoiem a tomada de decisão no contexto de gado leiteiro, reforçando a necessidade da definição da arquitetura e o desenvolvimento da plataforma neste domínio de aplicação.

Inicialmente, foram avaliados se a visualização de dados de fato é relevante para uma melhor tomada de decisão. Neste sentido, observa-se no gráfico da Figura [3.3](#) a importân-

Em uma escala de 1 a 5, como você avalia seu conhecimento sobre gado leiteiro?

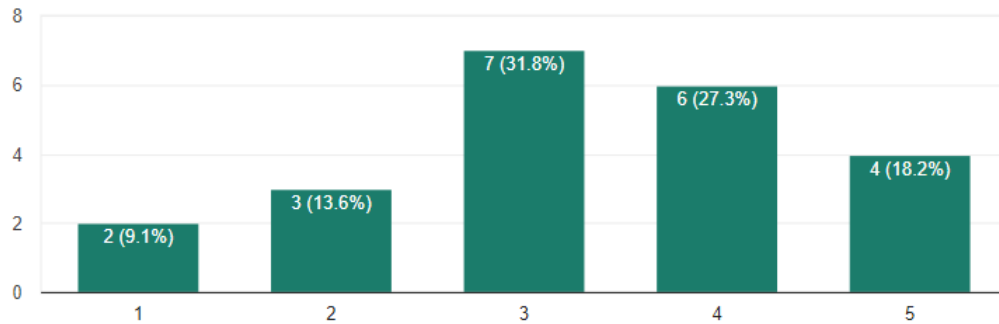


Figura 3.2: Conhecimento dos participantes em relação a gado leiteiro

cia da utilização de mecanismo visuais para a análise da quantidade de dados disponíveis no domínio trabalhado. Através do gráfico da Figura 3.4 pode ser observada a necessidade de apresentar diferentes representações, ao agregar diferentes artifícios de se analisar um mesmo conjunto de dados, de forma a esclarecer melhor suas informações implícitas.

Você julga importante a utilização de visualizações para análise de dados?

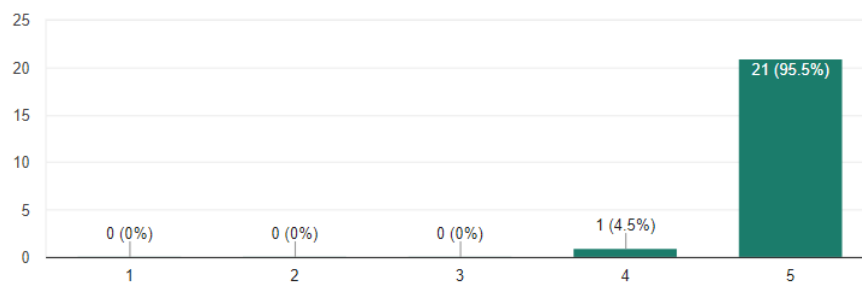


Figura 3.3: Importância da visualização de dados no domínio de aplicação

Além disso, foi explorada a necessidade de identificar a relevância dos indivíduos que se destacam dentro de um grupo pré-definido, assim como suas relações com outros indivíduos do domínio. Neste caso, retratando animais que podem se destacar de forma positiva ou negativa em relação aos outros animais. Nota-se que maioria dos participantes avaliou como um fator importante para ser incluso na plataforma (Figuras 3.5 e 3.6).

Foi solicitado que os participantes destacassem, caso pudessem identificar, a principal deficiência na análise e visualização de dados em ferramentas de apoio a informação para

Você acredita que a utilização de diferentes formas de visualização de dados (tabelas, gráficos, redes, etc.) pode facilitar a descoberta de novas informações?

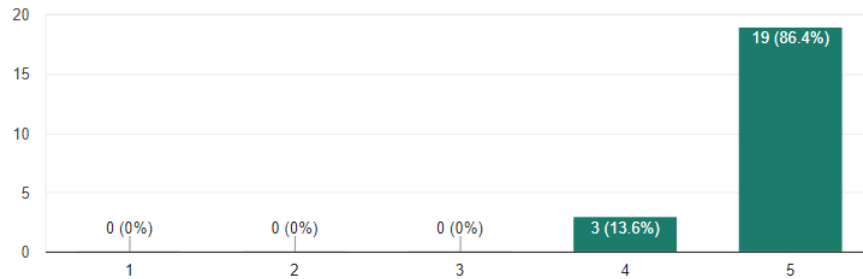


Figura 3.4: Diferentes formas de visualização de dados no domínio de aplicação

Você gostaria de ter uma visualização que permitisse observar as relações entre os indivíduos analisados (Por exemplo, relações de familiaridade entre os animais, produtores relacionados aos animais, entre outras relações pertinentes ao domínio)?

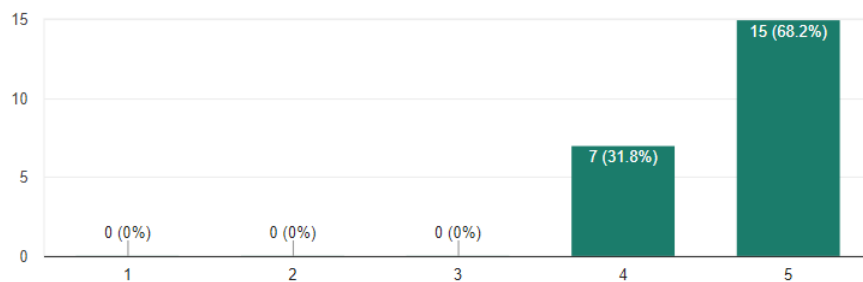


Figura 3.5: Análise das relações entre indivíduos

Você julga importante identificar indivíduos que se destacam (bons ou ruins - em termos de produção ou algum outro atributo), como vacas e touros com mais destaque em um grupo (um rebanho, animais de um estado ou outros critérios de agrupamento) para observá-los melhor?

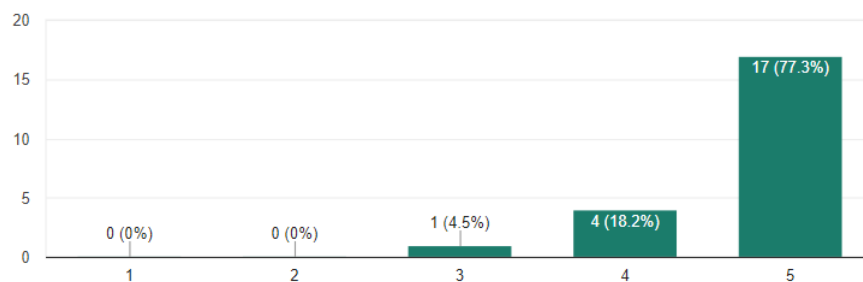


Figura 3.6: Análise de indivíduos em destaque na rede de dados

Tabela 3.1: Principais Deficiências Observadas

	Principais Deficiências
1	Falta de qualidade e clareza da visualização;
2	Visões univariadas sobre a produção;
3	Ausência de filtros e agrupamentos das informações
4	Escassez de recursos visuais para produção do gado leiteiro;
5	Falta de clareza ou imprecisão na construção das diferenças entre grupos;
6	Necessidade de opções para importação/exportação de dados;
7	Excesso de informações sem uma compilação prévia;
8	Necessidade de alternativas confiáveis para subsidiar as decisões;
9	Informações sobre o manejo que os animais são submetidos;
10	Teor de gordura, proteína e sólidos totais;
11	Conexão entre os dados;
12	Decisões de gestão do sistema produtivo;
13	Imprecisão no levantamento da informação primária;
14	Capacidade produtiva de um animal e a qualidade do leite;
15	A parte geográfica.

gado leiteiro. Esta questão era aberta, para definir o posicionamento dos participantes em relação às ferramentas já existentes e em quais pontos estas ferramentas são precárias, na questão de fornecer uma melhor informação aos usuários. Os principais pontos citados estão descritos na Tabela 3.1 (respostas retratadas como “não sei” ou “não pretendo opinar” foram removidas, para uma maior clareza das informações nesta tabela).

Revisitando a Figura 3.1, associamos as principais deficiências observadas em relação aos processos fundamentais de BI e obtivemos a Figura 3.7, onde analisamos as principais interseções entre as deficiências observadas e os processos de BI. É notório a intensidade de deficiências com relação ao processamento, importação e visualização dos dados. Grande parte das deficiências apontadas envolvem mais de um grupo, indicando que as relações entre os processos possuem muitas interligações e, caso um processo seja falho os outros provavelmente irão sofrer as consequências de um requisito mal definido. Assim, as etapas de BI precisam ser definidas de forma consolidada, para que todas as deficiências possam ser resolvidas corretamente.

Em relação às respostas dadas pelos participantes, observamos respostas variadas. Para conseguir filtrar as principais informações contidas nestas respostas, foi utilizada uma visualização chamada *Wordcloud*, que viabiliza o processamento de linguagem natural. A *WordCloud* é utilizada em domínios para análise de textos, como por exemplo, a mineração de opiniões (WU et al., 2010).

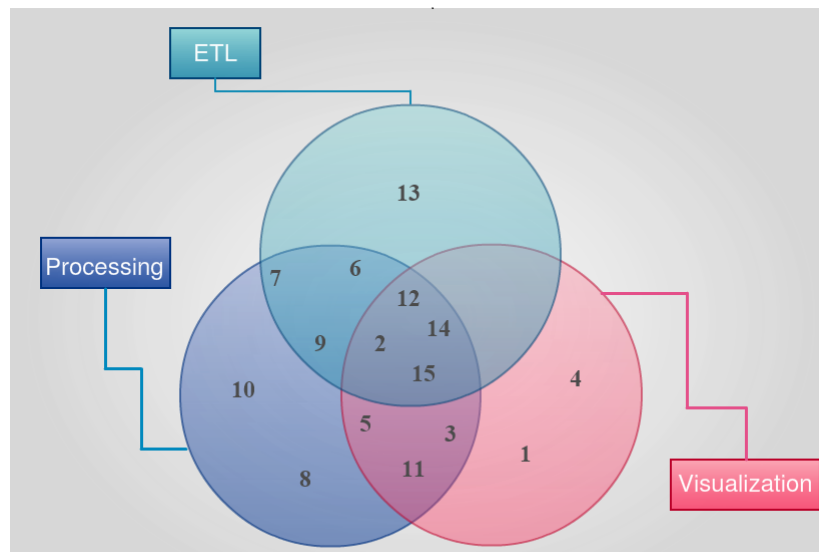


Figura 3.7: Deficiências apontadas, separadas em grupos do processo de BI.

Uma utilização popular deste tipo de visualização é a sumarização de textos, constituindo uma poderosa ferramenta para análise para o aprendizado de tópicos implícitos em grandes volumes de texto (HEIMERL et al., 2014). Esta visualização destaca as principais palavras contidas nas respostas recebidas, evidenciando quais foram mais citadas e reiterando os principais elementos contidos em respostas discursivas pelos participantes. As palavras com maior incidência, indicam uma maior preocupação por parte dos participantes, sugerindo um tópico de relevância para análise.

Desta forma, as palavras mais ressaltadas pelos participantes podem ser analisadas na Figura 3.8, onde identificamos que há uma falta de informação e necessidade de maior clareza em relação aos dados de gados leiteiros. Visto estas necessidades, a plataforma de visualização visa, essencialmente, preencher tais lacunas, a fim de minimizar estes problemas durante a tomada de decisão.

O objetivo da terceira etapa do questionário foi identificar as necessidades de informações para a tomada de decisão no domínio da produção do leite. Neste sentido, após respondidas as principais carências com relação a visualização da informação no contexto de aplicação, a última etapa é o levantamento dos principais indicadores e métricas que os participantes gostariam de poder analisar em uma ferramenta de visualização. Nesta etapa, o escopo foi restrito apenas aos animais (touros e vacas) por haver diversos indicadores e medidas que necessitam ser explorados. A seguir são apresentadas as perguntas e as respostas dadas a cada uma delas.

utilização de Dashboards, que permite representar através de um grupo visualizações um mesmo conjunto de informações ou de forma a complementá-las.

Com relação aos principais indicadores de eficiência dos animais, obtidos através das respostas dos participantes, notamos a necessidade de incluir soluções computacionais para viabilizar o processamento destes indicadores automaticamente, ou seja, agregar mais valor aos dados importados. Com isso, o uso de ontologias para representar o domínio de dados de maneira consolidada foi avaliado como uma solução viável e robusta. Desta forma, a ontologia pode representar os principais indicadores relacionados a cada indivíduo inerente ao domínio de aplicação e este consegue ser processado computacionalmente e, assim a eficiência de cada indivíduo pode ser processada sem necessidade de um conhecimento prévio sobre aquele tipo de dado, a ontologia agrega este valor semântico aos dados automaticamente.

Além disso, foi observada uma deficiência durante a execução do estudo exploratório. Esta deficiência consiste na existência de bases heterogêneas de dados e estes normalmente não possuem identificações confiáveis sobre os animais, ou seja, cada base pode identificar animais de maneiras diferentes, porém representam o mesmo indivíduo. Outro benefício que a ontologia agrega neste sentido, pois ao consolidar os dados na ontologia, é possível realizar inferências identificando indivíduos semelhantes e agregando mais informação aos mesmos.

Em seguida, os participantes ressaltaram a necessidade de observar relações entre indivíduos e quais destes possuem destaque (positivo ou negativo) a partir dos indicadores de eficiência destes em relação a um grupo observado, sendo que os critérios de agrupamento podem ser diversos, como por exemplo, animais por estado, rebanho, país, entre outros. Não foi possível solucionar este problema apenas com o uso da ontologia e, portanto, outra solução precisou ser identificada. Dentre as soluções, foram utilizadas redes complexas, que possuem como característica a análise estrutural dos dados e suas principais relações, além de possuir indicadores que permitem identificar nós relevantes dentro desta rede. Desta forma, a aplicação de visualizações em rede e suas métricas permitem auxiliar o tomador de decisão.

Um dos problemas observados ao representar um grande conjunto de dados em redes complexas é o volume de dados exibido na tela. Para resolver este problema, a integração com ontologia auxilia a reduzir o conjunto de nós observado, de forma que serão utilizadas

as *property chains* dos nós em questão na rede, omitindo as classes intermediárias, apenas representando o domínio e imagem das *property chains*. Isto foi exemplificado, de maneira mais clara, durante a avaliação realizada no Capítulo 4.

Analisando os requisitos obtidos no estudo exploratório, foi identificada a carência de soluções para a análise de subgrupos dentro de uma seleção maior de indivíduos. Desta forma, uma necessidade é a criação de mecanismos de criação de subgrupos baseados em algum indicador de desempenho ou qualidade, de forma que estes subgrupos possam ser analisados por indicadores distintos posteriormente. Para solucionar este problema, após a análise de relevância de indivíduos proporcionada pela aplicação das métricas em redes complexas, é possível criar estes grupos semanticamente e adicioná-los a ontologia. Desta forma, as redes complexas proporcionam incrementos na ontologia, tornando mais aderente as necessidades dos pesquisadores, aperfeiçoando as visão do domínio de aplicação a partir dos novos grupos semânticos identificados e criados. Ao criar estes novos grupos, estes poderão ser analisados por outras medidas de eficiência de acordo com a necessidade do tomador de decisão.

A partir disso, notamos que a integração entre as abordagens de ontologia e redes complexas permitem prover diversos benefícios quando utilizadas de forma agregada. Neste sentido, uma arquitetura conceitual foi proposta, de forma a se auto-alimentar e produzir novas informações. Em relação a arquitetura se auto-alimentar, reforçamos a necessidade de criação de novos subgrupos não existentes inicialmente, desta forma, a arquitetura criará novas definições na ontologia de domínio, de forma a representar as novas informações obtidas. O objetivo geral da arquitetura é possibilitar a visualização de dados e a interpretação de fontes de dados do ponto de vista semântico e estrutural.

Consolidando as informações obtidas no estudo exploratório, relacionando ao núcleo do processo de BI, um modelo básico da arquitetura é determinado a fim de identificar as principais etapas (Figura 3.11). Além disso, são representadas quais as soluções computacionais imprescindíveis para a construção e funcionamento de cada etapa do processo.

O processo de ETL representado, utiliza principalmente web services, gerando uma maior modularidade a arquitetura, que pode ser reaproveitada por outros softwares, que visem a utilizar estas funções criadas. Por fim, os dados importados, tratados, são consolidados em uma ontologia de domínio, este processo está definido com mais detalhes na seção 3.3. Com relação a camada de processamento, são utilizadas inferências das

ontologias, assim como métricas de redes complexas. As ontologias e redes complexas se complementam, fornecendo benefícios mútuos, incrementando o domínio de aplicação e fornecendo novas informações. A camada de visualização utiliza essencialmente sumarizações das informações efetuadas na camada de processamento. As principais visualizações destacadas são os *Dashboards*, tabelas e redes hierárquicas. Por fim, é ilustrada a camada de tomada de decisão, onde os tomadores de decisão utilizarão a plataforma de visualização web para uma melhor gestão das decisões gerenciais. A escolha da plataforma de visualização possuir tecnologia web foi devido ao fato de muitos pesquisadores estarem distribuídos geograficamente e, desta forma, é possível que todos tenham acesso a estas informações.

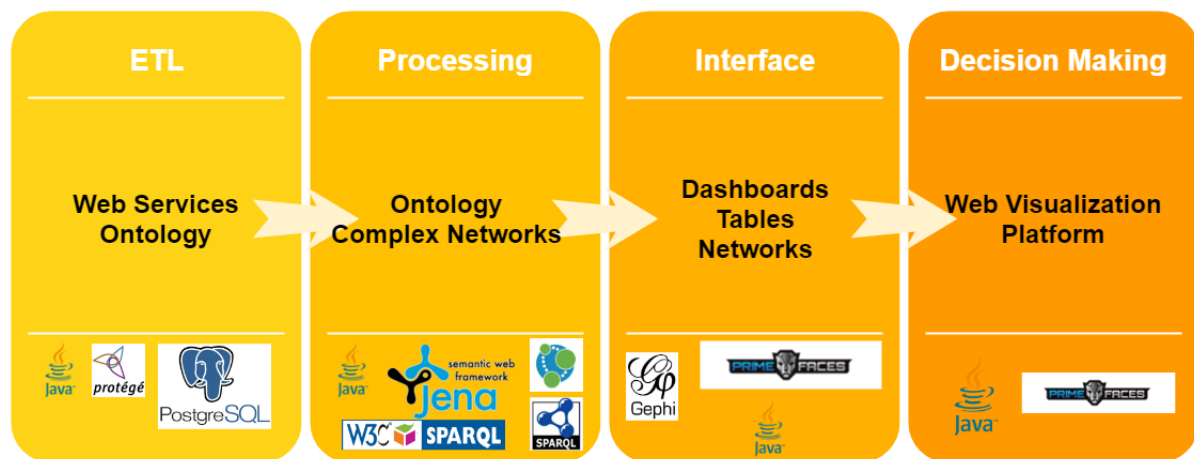


Figura 3.11: Fluxo do ciclo de vida de BI associado a soluções computacionais

Expandindo o modelo básico representado na Figura 3.11, foi criada a arquitetura detalhada, representando cada processo interno das camadas do núcleo de BI. O modelo conceitual da arquitetura é apresentado na Figura 3.12.

A operacionalidade da arquitetura ocorre de acordo com os seguintes passos. Inicialmente, (i) dados são coletados de um ou mais repositórios do mesmo domínio de aplicação, (ii) estes dados são organizados, traduzidos na camada de ETL e, então, importados na ontologia, em que algumas regras de domínio são definidas (iii), é traduzida como uma rede, mostrando o principal relacionamento entre os atores principais que constituem o domínio de aplicação (iv) finalmente, uma rede complexa é modelada com os indivíduos do domínio. Nesta rede, as relações podem ser similares às relações obtidas na ontologia, ou novas conexões entre os indivíduos são definidas considerando suas características intrínsecas. Desta forma, é feita uma análise semântica para abstrações de dados de forma

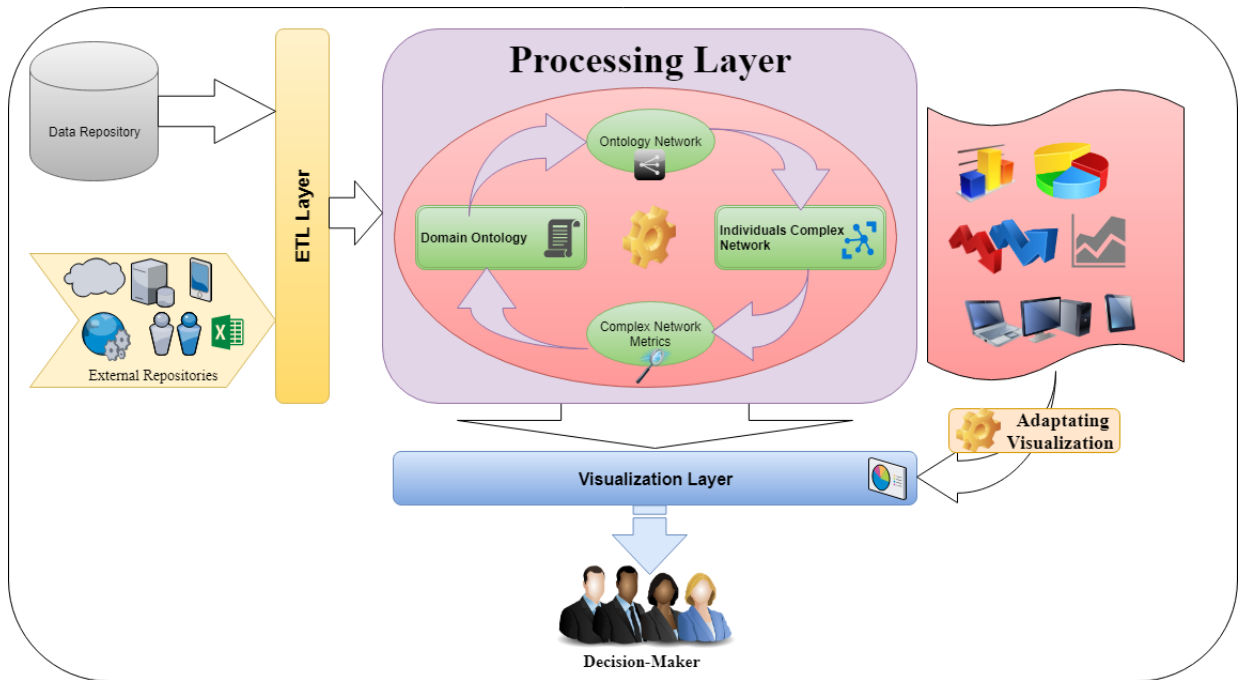


Figura 3.12: Detalhamento da arquitetura proposta, expandindo o fluxo de ciclo de vida de BI.

a sumarizar, reduzir o volume de dados, destacar dados mais relevantes e identificar novos agrupamentos de indivíduos e relações entre eles.

3.3 CAMADA DE ETL

A camada de ETL é responsável por extrair, transformar e carregar os dados para processamento futuro. Repositórios de dados podem vir de diversas fontes, mas todos se referem ao mesmo domínio de aplicação, ou pelo menos, possuem dados relacionados ao domínio de aplicação. As regras da ontologia serão responsáveis por relacionar indivíduos que possam ser iguais nestas fontes distintas, com a utilização das informações semânticas associadas, inclusive relacionar domínios semelhantes e que contenham informações sobre o mesmo indivíduo.

Neste sentido, a arquitetura importa os dados e carrega a ontologia com os indivíduos, de acordo com as regras de domínio. A principal funcionalidade da camada de ETL é a habilidade de interpretar dados de diferentes repositórios e consolidar em um conjunto de dados bem definido através da ontologia, obtendo todos os indivíduos e associando suas informações semânticas ao arquivo que é utilizado na camada de processamento.

3.4 CAMADA DE PROCESSAMENTO

O objetivo principal da camada de processamento é organizar e modelar os dados para que estes sejam enviados para a camada de visualização, permitindo uma análise mais clara dos mesmos, facilitando a tomada de decisão do usuário. Após receber os dados importados na camada de ETL, a ontologia de domínio é consolidada com todos os indivíduos do conjunto de dados.

3.4.1 Ontologia de Domínio

A ontologia é definida de forma a considerar as interações entre as entidades que estão presentes no conjunto de dados. *Property chains* são propriedades definidas através da sequência de propriedades de objeto, definindo inferências relacionadas aos dados. A ontologia tem o poder de tornar a arquitetura mais robusta e independente pois é capaz de interpretar os dados e inferir novos conhecimentos.

A partir desta camada semântica, a arquitetura cobre diferentes domínios de dados. Além disso, estas inferências auxiliam análises baseadas em redes complexas de forma a reduzir o número de nós que precisam ser representados através de técnicas de visualização, além de adicionar informação semântica aos nós e arestas.

3.4.2 Desenvolvendo a Ontologia de Domínio

Não foram identificados trabalhos na literatura que usem ontologias que representem o domínio de controle e produção do leite para gado leiteiro. Baseada no estudo exploratório, uma ontologia foi criada utilizando a ferramenta Protégé (protege.stanford.edu). A ontologia criada, tem como objetivo a representação do domínio de produção do leite, assim como os principais indivíduos que o integram. Desta forma, é possível agregar informações aos dados obtidos, provenientes de fontes distintas. A criação desta ontologia foi feita com a ajuda de especialistas na área de pesquisa em gado de leite para detalhar melhor o processo e suas necessidades.

As regras da ontologia representam as ligações semânticas entre os indivíduos, estabelecendo seus principais relacionamentos, assim como suas associações inferidas, ou seja, relações indiretas que podem ser identificadas através de processamentos feitos pela ontologia. A ontologia criada provê diversos benefícios para a plataforma de visualização, como

por exemplo, no processo inicial de integração dos dados. As principais fontes de dados dos animais no contexto de gado leiteiro estão contidas em repositórios heterogêneos, que não possuem integração entre os identificadores de animais, dificultando relacioná-los. A utilização da ontologia de domínio possui o poder de integrar estes dados de uma maneira mais concreta, possibilitando que dados de fontes genéticas e de dados laboratoriais de um mesmo animal possam ser integrados, maximizando a análise estratégica que pode ser feita para este.

Para isto, a ontologia de domínio conta com regras lógicas, que indicam quais os tipos de relações os indivíduos possuem, além das relações indiretas, chamadas cadeias de propriedades ou Property Chains. A ontologia criada conta com um total de 30 classes e 40 regras lógicas. A Figura 3.13 representa a ontologia de domínio criada e seus relacionamentos.

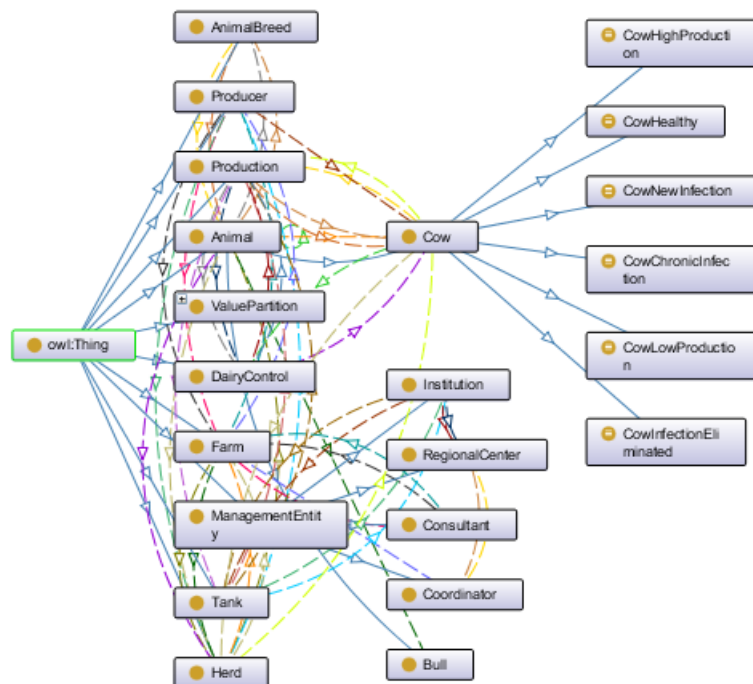


Figura 3.13: Relações da Ontologia de Domínio

Uma das regras utilizadas é detalhada na Figura 3.14. A property chain em questão ressalta as relações hierárquicas obtidas por uma instituição de leite. A instituição de leite é responsável pela análise da produção do leite de produtores de leite, porém não é possível se obter diretamente estes indicadores com os dados utilizados. Há uma cadeia hierárquica entre as entidades gerenciais, de forma que os consultores lidam diretamente

com o gerenciamento das fazendas e produtores que, por sua vez, são gerenciados por coordenadores pertencentes a um núcleo gerencial que envolvem instituição leiteira. Com um maior número de entidades gerenciando outras, torna-se complexo obter os principais indicadores de produção de leite para estas fazendas que são gerenciadas indiretamente pela instituição. Esta informação é de suma importância econômica, dado o fato de que decisões estratégicas podem ser tomadas para otimizar o custo e produção de leite nas fazendas problemáticas, assim como identificar as fazendas mais produtivas e buscar analisar as causas de seu sucesso.

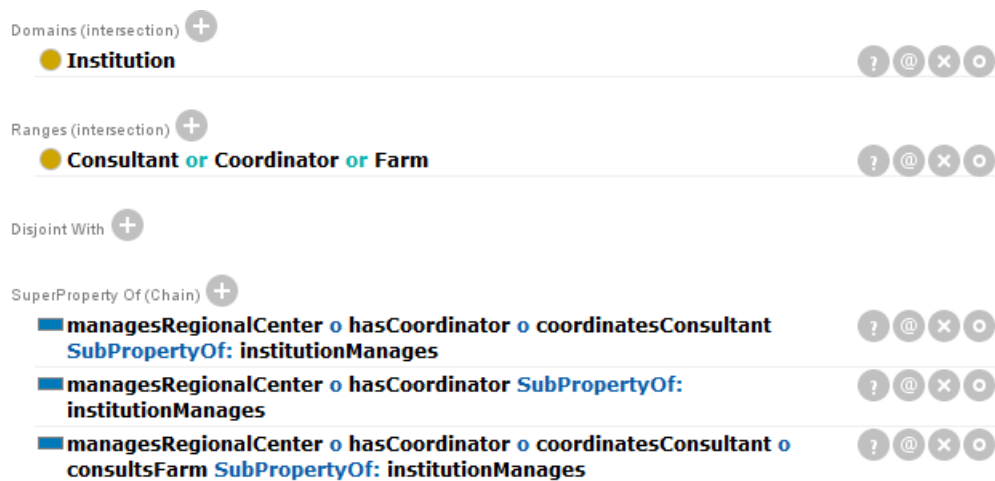


Figura 3.14: Property Chain Gerencial

Além disso, a ontologia também consegue realizar a inferência de informações semânticas agregadas aos indivíduos. Por exemplo, ao utilizar informações relacionadas a contagem de células somáticas de um animal, é possível inferir animais que estão saudáveis, com nova infecção, infecção eliminada ou com infecção crônica. Desta forma, é possível observar os dados associados a estes grupos semânticos de forma separada, de forma a identificar estes problemas antecipadamente ou verificá-los em tempo hábil de realizar decisões estratégicas. Uma agregação utilizada nesta dissertação é a de associar medidas de qualidade aos indivíduos da ontologia. A medida de qualidade determina a eficiência de uma classe, baseado nas informações semânticas previamente identificadas para esta classe. Por exemplo, uma classe de vaca pode ter como atributo de qualidade sua produção de leite, que representa sua eficiência. Uma classe pode ter um ou mais atributos de qualidade, a partir da necessidade.

Ao final do desenvolvimento da ontologia de domínio, a camada de ETL realiza a importação dos dados obtidos de fontes heterogêneas e os integra no arquivo de ontologia.

Este arquivo irá conter todos os indivíduos e suas relações, diretas e indiretas. Porém, mesmo com as informações semânticas agregadas a ontologia de domínio e os indivíduos, são necessários mecanismos automáticos que realizem o ranqueamento destes, de forma a avaliar quais são os principais elementos que compõem um conjunto, sendo que as redes complexas oferecem mecanismos para preencher essa lacuna na apresentação da informação pela plataforma de visualização.

Com a agregação destas informações semânticas aos dados importados e integrados, as suas relações permitem que sejam criadas as associações entre cada indivíduo, gerando redes complexas que indiquem as relações diretas entre os indivíduos que compõem a rede, assim como as relações ocultas entre estes.

A conexão entre os elementos que compõem a ontologia, naturalmente cria uma rede entre os atores inerentes ao domínio de dados observado. A modelagem e visualização desta rede permite algumas abordagens para análise de dados, assim como, (i) ilustra ao tomador de decisão os principais relacionamentos entre os elementos que compõem o domínio de dados, auxiliando a percepção e criação de novas relações através da observação direta e (ii) o uso de métricas (explicadas na seção [3.4.3.1](#)) para identificar elementos relevantes na ontologia.

3.4.3 Redes Complexas de Indivíduos

Nós na rede complexa representam indivíduos e as conexões entre eles representam indicadores de qualidade que podem ser definidos pelo usuário. Estes indicadores podem ser definidos em tempo de execução ou através de medidas de qualidade observadas na ontologia (observadas durante o estudo exploratório e detalhadas na próxima seção).

Esta camada utiliza dados inferidos da ontologia e reduz a quantidade de dados que é representado para o usuário. Como resultado, ela identifica indivíduos relevantes com métricas de redes complexas.

Considerando a análise semântica, os dados escolhidos pelo usuário são relacionados com indivíduos conectados através da ontologia, onde relacionamentos entre os dados e suas "regras lógicas" são analisados. Regras lógicas são responsáveis por sintetizar a informação, com isso, indivíduos se relacionam de maneira direta e indireta.

3.4.3.1 Métricas de Redes Complexas

Métricas de redes complexas são usadas para realizar a análise de nós que possuem destaque dentro de uma rede. A identificação de nós mais influentes é importante para uma análise mais específica de um elemento, no sentido de observar as causas de sua proeminência em relação aos outros indivíduos.

A rede complexa é definida como um grafo $G(V,E)$, com vértices V e arestas E . Os vértices deste grafo representam os indivíduos que pretendemos estabelecer as relações ocultas e identificar o seu grau de similaridade. As arestas entre estes grafos irão representar o grau de similaridade entre os vértices envolvidos.

Cada vértice deste grafo terá um indicador de qualidade estabelecida a critério do especialista de domínio. O indicador de qualidade define qual característica do vértice convém observar. Estabelecemos quão similares são as características entre dois indivíduos, definidas de acordo com as características que se deseja ressaltar no indivíduo. Com isto, a métrica é um cálculo que utiliza os atributos relacionados a esta característica, a partir das informações semânticas do mesmo.

A qualidade $Q(n_i)$ de um nó n_i é dada de acordo com a Equação 3.1, onde $C(n_x)$ representa o indicador de eficiência do nó. Este indicador de eficiência varia de indivíduo para indivíduo, de forma que a ontologia se encarrega de fornecer essa informação, que é definida em conjunto com o domínio de aplicação.

$$Q(v) = \frac{C(v)}{\Sigma C(v)} \quad (3.1)$$

Assim, a qualidade do indivíduo pode ser medido de acordo com sua característica (ou indicador) $C(v)$, comparada a esta característica de todos os outros indivíduos selecionados.

Ao definir os indicadores de qualidade observadas em cada tipo de nó, é possível estabelecer as relações de similaridade entre eles. A medida definida para as arestas neste trabalho é o módulo da diferença entre a qualidade de cada nó. Desta forma, uma aresta $E(i,j)$ é dada pela diferença entre as qualidades de v_i e v_j acrescida de uma unidade, para ajustar quando a similaridade entre dois nós for idêntica e o peso da aresta não ser zero.

$$E(i, j) = 1 + |Q(v_i) - Q(v_j)| \quad (3.2)$$

Normalizando as arestas, definimos a porcentagem de similaridade que um nó tem com seu vizinho, de forma que $0 < e(i,j) < 1$. Neste sentido, quanto maior for o valor da aresta, maior sua similaridade com o vizinho. As análises realizadas abordam todas as ligações entre os nós, desta forma a rede será completa. Neste trabalho serão destacadas as abordagens ponderadas para destacar os nós mais relevantes na rede, já que todas as arestas possuem pesos.

O objetivo desta seção é a definição de uma métrica para identificar os nós mais relevantes da rede e observar, de acordo com essa métrica, quais agrupamentos podem ser realizados, selecionando grupos dentre os nós em destaque. Para criar este agrupamento, utilizamos o *Boxplot*, pois devido a sua divisão em quartis, nos possibilita observar a distribuição estatística dos indivíduos, permitindo a criação de novos grupos semânticos na ontologia.

Ao identificar estes novos grupos semânticos, realizamos novas análises neles e assim podemos criar novos grupos para visualização.

3.5 CAMADA DE VISUALIZAÇÃO

A camada de processamento obtém os dados recebidos através dos serviços processados pelas ontologias e redes complexas e adaptá-los para as técnicas de visualização relevantes, levantados previamente através do estudo exploratório com especialistas de domínio. O principal objetivo é representar os dados da maneira mais clara possível para o tomador de decisão. Assim, o usuário pode otimizar seu processo cognitivo com as visões para obter um melhor entendimento sobre os dados. Após realizado o processamento dos dados obtidos, estes gráficos e informações são exibidos na camada de visualização, que por sua vez, tem o contato direto com o usuário. As principais definições desta etapa de adaptar a visualização são as criações de Dashboards, filtros em relação aos dados importados e construção das tabelas.

Com a interpretação dos dados pela ontologia e a análise estrutural das redes complexas, o usuário pode escolher os dados que são pertinentes para sua análise. Ele define medidas dos dados que pretende observar, assim como valores numéricos que representem produtividade, algum atributo que sugere contar ou qualquer outro atributo que seja mensurável.

Nesta camada é onde efetivamente o usuário visualiza as informações que irão apoiá-lo

na tomada de decisão. Os elementos gráficos, selecionados previamente, irão demonstrar os resultados e padrões da análise dos dados.

3.6 TECNOLOGIAS UTILIZADAS

Para a implementação da arquitetura foram utilizadas tecnologias atuais que possibilitam uma melhor manutenção do código, de forma a facilitar trabalhos futuros que visem uma complementação do mesmo. Para a execução da plataforma de visualização, foi utilizado um notebook com a seguinte configuração: processador Intel Core i5-4200 2.30 GHz, memória RAM de 6GB DDR3 e utilizando sistema operacional Windows 10 como arquitetura 64 bits. Inicialmente foi escolhida para o desenvolvimento uma arquitetura em camadas orientada a serviços de forma a otimizar a manutenção e portabilidade da plataforma de visualização.

Para o desenvolvimento da camada de ETL, foi criado código em Java para integrar bases de dados do PostgreSQL, planilhas eletrônicas e arquivos CSV. O código de importação foi desenvolvido através de serviços RESTful e a integração dos dados importados é feita em uma ontologia. A ontologia foi desenvolvida na ferramenta *Protegé*, assim como os códigos em *SPARQL* para realizar as consultas na mesma. Como principais bases de dados foram utilizados modelos relacionais no PostgreSQL, modelos de bancos em grafos utilizando Neo4j e algumas bibliotecas do mesmo para visualização e, por fim, a ontologia como repositório de dados.

A implementação das técnicas de visualização foi realizada utilizando um framework em Java chamado PrimeFaces², por englobar diversos mecanismos de visualização e ferramentas atualizados que contém componentes atrativos para possibilitar as tomadas de decisão mais efetivas. Além disso, foram utilizadas bibliotecas internas do Java para algumas representações gráficas e processamento das informações.

3.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou a arquitetura desenvolvida, que utiliza aspectos semânticos e estruturais para visualização de dados e tomada de decisão. Além disso, o capítulo ilustrou os benefícios da utilização das estratégias de forma combinada, assim como os principais

²<https://www.primefaces.org>

elementos para beneficiar a tomada de decisão pelo gerente de negócios. A arquitetura possui camadas com papéis bem definidos, de forma a se tornar mais modular, permitindo a inclusão de novos módulos e otimizações ao longo do tempo. Por fim, foi exibido um cenário real de aplicação da arquitetura, assim como a construção da plataforma de visualização e todos seus artefatos.

4 AVALIAÇÃO DA PROPOSTA

Este capítulo apresenta a avaliação da arquitetura proposta, que foi feita a partir da versão do plataforma de visualização para apoio a tomada de decisão. O objetivo desta avaliação é analisar a viabilidade da plataforma de BI e se esta auxilia na tomada de decisão de pesquisadores no domínio de produção do leite. Por fim, responder à questão de pesquisa considerando o uso da solução proposta em um contexto de mundo real, na Embrapa Gado de Leite.

O capítulo está descrito conforme a execução da avaliação, de forma a detalhar cada etapa realizada para a obtenção dos resultados.

4.1 DEFINIÇÃO

A tomada de decisão é um processo empírico e não há como gerar evidências concretas de que a mesma realmente resultou em boas ou más decisões estratégicas, desta forma, é preciso avaliar a viabilidade e aderência da plataforma para responder as perguntas de usuários que necessitam de realizar o processo cognitivo de descoberta e visualização da informação. Um estudo de caso é uma investigação empírica que se baseia na evidência de múltiplas fontes de evidência para investigar uma instância (ou um pequeno número de casos) de um fenômeno contemporâneo de um contexto atual, especialmente quando o fenômeno e o contexto não podem ser claramente observados (RUNESON et al., 2012). Neste sentido, adotamos o estudo de caso como melhor forma de avaliação, visto que a sua capacidade de interpretar resultados empíricos auxilia na resposta das questões de pesquisa propostas nesta dissertação.

Assim, o método de avaliação adotado nesta dissertação foi baseado em um estudo de caso conduzido no contexto de uma empresa de pesquisa agropecuária com dados reais extraídos da produção de gado leiteiro de rebanhos da Embrapa Gado de Leite. Além disso, identificar a viabilidade da plataforma de visualização que utiliza a arquitetura proposta para a tomada de decisão no contexto de gado leiteiro.

O escopo de avaliação foi definido baseado no método GQM (SOLINGEN et al., 2002) com o **propósito de caracterizar o efeito de** análises semânticas e estruturais, como abordagens de ontologia e redes complexas agregadas, **em** abordagens de visualização de

dados na tomada de decisão **do ponto de vista de** gerentes de negócio e instituições leiteiras **no contexto de** produção do leite”.

Desta forma, foi elaborada a questão de pesquisa primária:

- Q1. Como a utilização da arquitetura proposta pode auxiliar em uma melhor tomada de decisão?

A partir da questão de pesquisa primária, foram estabelecidas questões secundárias:

- Q1.1 Como a ontologia pode agregar novas informações para o domínio de aplicação?
- Q1.2 Como as redes complexas podem ressaltar indivíduos mais relevantes de um grupo selecionado?
- Q1.3 Como a agregação de técnicas de ontologia e redes complexas podem oferecer maior informação?
- Q1.4 Como a visualização pode otimizar as decisões estratégicas para tomada de decisão?

A partir das questões de pesquisa, foram definidas etapas que buscassem responder essas questões.

4.2 PLANEJAMENTO

Com o apoio dos dados levantados no estudo piloto, foi criada a ontologia de domínio, os serviços necessários para retorno de dados, processos de ETL e criada a plataforma de visualização para realização do estudo piloto. Esta etapa inicial possibilita reformular o planejamento de elementos da plataforma e aos tipos de serviços que criados para que sejam retornados os indicadores necessários e mais aderentes às necessidades das indústrias leiteiras para um melhor planejamento estratégico.

Para entender as principais necessidades e quais elementos da plataforma realmente são viáveis, foi realizado um estudo piloto na Embrapa Gado de Leite com alguns pesquisadores, de forma a refinar a plataforma de visualização. O estudo piloto proporcionou a observação da usabilidade da plataforma no ambiente de um pesquisador do contexto agropecuário que possui necessidades de tomar de decisão com os dados obtidos de análises de laboratórios e colhidas em fazendas e rebanhos experimentais. O estudo piloto

consistiu em observar as principais deficiências da plataforma e, através de um formulário de usabilidade e análises observacionais, detectar o que seria necessário para alterar na plataforma. Assim, descobrimos outras necessidades que poderiam ser atendidas com a utilização das técnicas propostas. A necessidade de integração entre outras bases de dados externas, novas aplicações comparativas entre os animais e, além disso, remover visualizações que não agregariam informação para a tomada de decisão.

Após analisar melhorias que deveriam ser feitas na plataforma, obtidas no estudo piloto, realizamos as melhorias e aplicamos o estudo de caso na Embrapa Gado de Leite, com novos participantes, pesquisadores da empresa altamente capacitados utilizando a plataforma. Para coletar evidências e responder à questão de pesquisa proposta, observações diretas, formulários e entrevistas foram utilizadas como fontes de dados.

4.2.1 Definição dos Participantes

A definição dos participantes se divide em duas etapas: estudo piloto e estudo de caso. No estudo piloto, a plataforma foi apresentada para um conjunto limitado de pesquisadores, de forma que estes pudessem opinar e indicar as principais deficiências na plataforma e o que poderia ser alterado, possibilitando uma experiência mais efetiva no quesito de uma tomada de decisão mais consistente. Com a participação de pesquisadores com grande conhecimento no domínio de aplicação, as modificações realizadas na plataforma foram otimizadas, possibilitando uma maior aderência às principais necessidades dos produtores de leite, instituições leiteiras e pesquisadores do domínio estudado.

Para o estudo de caso foram selecionados pesquisadores da Embrapa Gado de Leite, que convivem diariamente com dados e análises de gados leiteiros, possuindo uma vasta experiência com relação as necessidades do domínio de aplicação para uma tomada de decisão mais efetiva. Como a plataforma propõe visualizações da informação no contexto agropecuário, com foco na produção do leite, foram selecionados pesquisadores com conhecimento aderente a gado leiteiro e/ou experiência em produção de instituições leiteiras.

4.3 CONDUZINDO O ESTUDO PILOTO

Após concluir a implementação inicial da plataforma, esta foi apresentada na Embrapa, para que alguns pesquisadores pudessem opinar sobre as funcionalidades e usabilidade da

mesma, propiciando a inclusão de melhorias relacionadas à tomada de decisão e aos principais elementos de visualização e indicadores para uma melhor gestão e decisão estratégica por parte dos usuários.

Inicialmente, foram explicadas algumas noções sobre a arquitetura, a ontologia de domínio e sobre as redes complexas. Desta forma, ilustrando como essas metodologias foram aplicadas na plataforma. Após essa etapa, foram mostradas as principais funcionalidades da plataforma em um ambiente controlado, demonstrando como os dados são explorados na plataforma e quais as decisões estratégicas podem ser realizadas ao utilizá-la. Enquanto os pesquisadores observavam a utilização, fizeram algumas observações, que foram devidamente anotadas para melhorias posteriores.

4.3.1 Cenário de Aplicação

Considerando o domínio agrapecuário, particularmente da produção de leite, duas análises principais são realizadas pelos pesquisadores: gerenciais (entidades) e de produção (animais). Inicialmente, uma análise comparativa foi feita nas tabelas de dados gerenciais. Os especialistas de domínio observaram as relações hierárquicas entre algumas entidades (Centro Regional, Coordenadores e Consultores) e, sem o uso de ontologias, apenas inter-relações entre estes indivíduos foram identificadas. Devido ao grande volume de dados, é custoso identificar as relações indiretas entre estes dados, de forma a permitir que os especialistas observem as associações implícitas entre as entidades hierárquicas. Neste sentido, é muito complexo realizar a junção destas informações relacionadas as entidades e seus relacionamentos com outros dados, visto a quantidade de dados heterogêneos encontrados em bases agropecuárias referentes a gado leiteiro.

A utilização da ontologia preenche esta lacuna, de acordo com as inferências feitas, é possível agregar valores semânticos às informações. Além de possibilitar a identificação de dados semelhantes em repositórios distintos e organizá-los de forma a complementar as informações, tanto no aspecto gerencial quanto no produtivo.

Neste sentido, foram demonstradas tabelas e redes entre os indivíduos, obtidas através das regras lógicas da ontologia, para ilustrar as relações implícitas e explícitas entre os dados importados. Neste caso, os valores de produção de todas as propriedades atendidas por entidades gerenciais foram associados. Como resultado, foi possível fazer observações que eram anteriormente complexas, assim como identificar a produção de todas as

propriedades atendidas por determinados núcleos regionais. Além disso, também existem Dashboards que representam a produção de animais e indicadores relacionados aos rebanhos e seus animais, como ilustrado na Figura 4.1

Dairy Cattle Production and Managerial Data Visualization

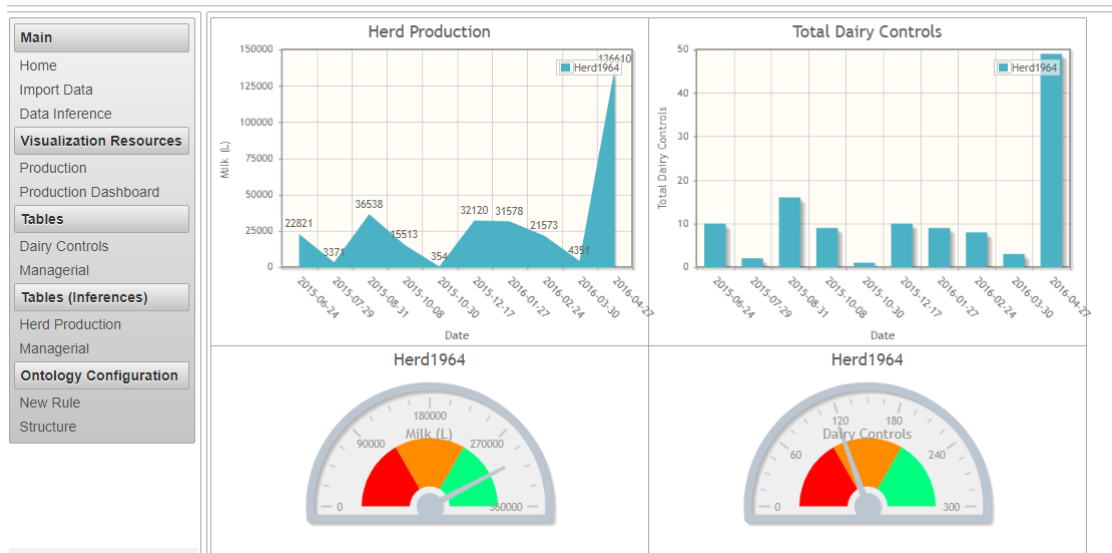


Figura 4.1: Dashboard para dados de produção de leite

Para a definição de níveis de abstração e análise estrutural, foram definidas redes complexas considerando a análise de relações ocultas entre os indivíduos, neste caso, baseada em sua similaridade e identificação de indivíduos de destaque em relação à rede como um todo. A rede gerencial busca visualizar a média, totalizadores e outras agregações de valores relacionadas às medidas de controles leiteiros de animais em um ou mais rebanhos de fazendas. Neste cenário, foram considerados valores de soma da produção de leite, por ser um indicador relacionado diretamente a produção e que revela bons valores econômicos para animais e fazendas. Estes valores foram utilizados como peso das arestas na rede gerencial.

Com ajuda desta rede, o especialista foi capaz de analisar quais fazendas produziam mais leite para um núcleo regional escolhido por ele de acordo com suas necessidades para tomada de decisão. No contexto da produção do leite, e a ontologia usada, foram duas principais redes para o sistema: gerencial e de produção de leite, para similaridade entre a produção dos animais. Foram realizadas as análises baseadas nestas duas redes de produção. Para a rede gerencial, foram definidas as relações hierárquicas entre os indivíduos e as representações simplificadas entre os mesmos (relações implícitas gerando

um novo grafo entre as entidades).

Na Figura 4.2, são ilustrados estes elementos: núcleos regionais (em roxo), coordenadores (em rosa), consultores (em amarelo) e, finalmente, fazendas (em cinza). É possível observar que com o aumento do número de instituições, a visualização da rede é comprometida devido ao volume de dados apresentados. Com a ontologia podem ser feitos mecanismos que reduzam o volume destas redes.

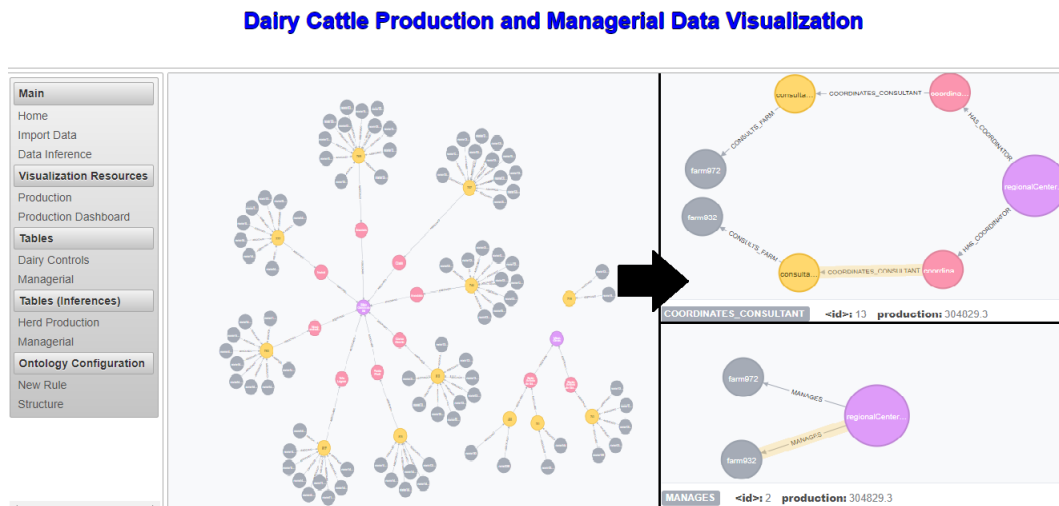


Figura 4.2: Aplicação nas Redes Hierárquicas

Na rede gerencial, algumas métricas foram definidas para destacar nós mais relevantes e com maior produção de leite. Foi observado que nós com maior *closeness centrality* são relacionados com Fazendas com uma mais entidades. Assim, comparando entidades de mesmo tipo (Coordenador, Consultor, Núcleos Regionais e Instituições) aquelas que são mais influentes na rede foram identificadas e apresentadas ao especialista. Isso é devido ao fato de lidarem com uma maior produção de leite. A métrica de *betweenness centrality* foi utilizada para identificar nós com maior influência na entidade e / ou propriedade gerencial. Assim, ao mensurar os caminhos mínimos entre os núcleos regionais e fazendas, foi possível identificar quais entidades estão presentes nestes caminhos, mostrando que a entidade gerencia uma maior gama de outras entidades e propriedades.

Na Figura 4.2, é possível observar a ação da ontologia na abstração dos dados apresentados na rede gerencial. Com esta abstração, o gerente é capaz de visualizar medidas identificadas com a análise dos totais das redes complexas. Neste caso, as arestas entre os nós (um consultor e um coordenador) foram selecionados, e o total de produção das fazendas consultadas por ele foi de 304829.3 litros de leite.

Assim, é possível observar através do fluxo que passa pelas arestas quais coordenadores ou entidades influenciam mais, considerando a produtividade de um certo núcleo regional ou alguma outra entidade desejada como elemento central nesta rede. A redução da quantidade de nós apresentados na tela (Figura 4.2) facilita a visualização pelos gerentes de negócio sem comprometer a informação contida na rede complexa.

A segunda métrica é para a rede de touros. Cada nó representa um touro, de forma semelhante aos filtros aplicados na rede de vacas. A diferença é que a qualidade do touro é medida de acordo com a quantidade de filhos $F(v)$ que este touro teve durante o período de tempo estabelecido. Desta forma, é possível medir a qualidade de um certo nó.

$$Q(v) = \frac{F(v)}{\Sigma F(v)} \quad (4.1)$$

A terceira métrica reflete uma análise de rede gerencial. Esta rede gerencial pode conter nós de tipos diferentes dentro da hierarquia determinada no gerenciamento. Neste caso, iremos observar as relações entre os nós de consultores. Os consultores atendem fazendas, desta forma, o indicador de qualidade das entidades é dada de acordo com a produção das propriedades gerenciadas $Pr(v)$ por estas entidades em relação a quantidade de propriedades atendidas $TotalP(v)$.

$$Q(v) = \frac{\Sigma Pr(v)}{TotalP(v)} \quad (4.2)$$

4.3.2 Resultados Obtidos do Estudo Piloto

As observações realizadas reforçaram os benefícios da utilização de ontologias e redes complexas na plataforma. Durante a visualização das inferências, foram realizadas perguntas referentes a possibilidades de integração entre outras bases de dados, de forma a obter informações de um mesmo animal, porém, em bases heterogêneas, ou até mesmo outros domínios, como por exemplo bases genéticas. A informação semântica agregada aos dados, permite que a arquitetura possa realizar esta leitura, de forma a integrar outras bases e conseguir inferir animais semelhantes dentro de uma estrutura de dados. As bases de dados utilizadas no estudo piloto foram limitadas, o que permitiu a integração de apenas poucos animais, porém isso permite uma maior flexibilidade da plataforma em integrar diversos domínios de aplicação.

Além disso, a visualização das redes não parecia muito clara ou interessante para os pesquisadores, portanto algumas visualizações de redes foram removidas da plataforma e uma delas alterada. Pois foi dito que a visualização destas redes para os problemas relacionadas a dados hierárquicos ainda era pertinente e, portanto, apenas buscamos melhorar como estes dados serão exibidos para o usuário, propiciando uma forma melhor de entender o que está contido nesta visualização. Os pesquisadores ressaltaram que a plataforma apresentava visualizações muito importantes e que há uma falta destas plataformas, relacionando informações de gado leiteiro, principalmente que possam integrar diversos domínios de dados e agregar informações semânticas aos mesmos.

Este estudo de caso piloto, demonstrou o uso da primeira versão do sistema de visualização e a tomada de decisão que cobre os aspectos das análises semânticas e estruturais dos dados observados. Todas ações foram tomadas no sistema, assim como as análises das métricas adotadas neste estudo de caso, foram feitas por um especialista do domínio agropecuário.

Além das observações diretas, foram conduzidas entrevistas e formulários de usabilidade da plataforma. Após o uso da plataforma desenvolvida, destaco 3 respostas dadas pelo especialista.

(i) A informação é apresentada claramente pelo sistema, apoiando o processo de tomada de decisão?

"Estes gráficos me possibilitaram a fazer uma análise imediata da produção de fazendas específicas comparando em um cenário nacional, apresentando informações que me apoiaram em um processo de tomada de decisão com produtores leiteiros."

(ii) Como você avalia o uso de diferentes recursos visuais (gráficos, tabelas e redes) para tomada de decisão?

"Os gráficos são importantes, mas precisamos analisá-los de maneira conjunta."

(iii) É possível analisar dados de diferentes instituições e sua contribuição na produção do gado nacional?

"Produção de leite é representada em uma forma de rede que nos permite identificar e comparar diferentes relações entre as entidade e sua produtividade. Isso é uma boa contribuição do sistema para análise de dados."

Baseado na primeira resposta, há a evidência de que a informação apresentada na visualização dos artefatos provê um melhor entendimento dos dados e suas características,

com isso, instituições leiteiras podem ajudar produtores leiteiros na administração de suas fazendas e rebanhos. De acordo com a segunda resposta, há evidência de que os componentes utilizados ajudaram no processo de tomada de decisão. Porém, ainda é preciso integrar estes componentes. A terceira resposta mostrou que o uso de análises semânticas e redes complexas como técnicas conjugadas podem ajudar no processo de tomada de decisão.

Para responder à questão de pesquisa, observamos (i) as interações dos especialistas com o sistema para obter informação; (ii) as análises feitas pelos mesmo no contexto de tomada de decisão para produção do leite; (iii) e suas respostas dadas na entrevista (formulário anexo) no estudo de casos. As respostas dos especialistas durante o estudo de casos apontaram para a viabilidade da proposta.

Considerando esses pontos, é possível afirmar que há evidências que a análise semântica e estrutural dos dados, apoiadas por técnicas de visualização são capazes de auxiliar a tomada de decisão, adicionando mais informações aos dados. Para isso, é feito um estudo de casos regular para uma melhor avaliação da proposta.

4.4 CONDUZINDO O ESTUDO DE CASO REGULAR

O estudo de caso regular visa responder às questões de pesquisa desta dissertação. Principalmente, sobre a capacidade de oferecer uma melhor tomada de decisão para o contexto de gado leiteiro a partir da análise de dados reais, utilizando a arquitetura proposta e com as estratégias de visualização definidas no levantamento de requisitos, assim como as entrevistas realizadas com os participantes das outras etapas da avaliação. Para a realização do estudo de caso, foram realizados convites para pesquisadores da Embrapa, de forma que estes viram uma apresentação inicial sobre a arquitetura e, em seguida, o seu funcionamento e principais telas para o apoio a tomada de decisão. Alguns dos participantes também fizeram parte do levantamento de requisitos para a plataforma. Outros não participaram das outras etapas, de forma a abranger uma maior gama de subáreas do contexto de gado leiteiro.

A sala de reuniões, onde a plataforma foi apresentada estava composta por 13 participantes, que analisaram o funcionamento da plataforma, assim como seus principais elementos de tomada de decisão. Foi realizada uma breve explicação sobre o contexto do trabalho e uma apresentação sobre os principais objetivos ao realizar a avaliação. Nesta

avaliação, foram utilizadas 3 fontes de coleta de dados: análise observacional, entrevistas e formulários para os participantes.

4.4.1 Execução do Estudo de Caso Regular

Neste estudo de caso foram realizadas melhorias com relação a plataforma e sua apresentação para os usuários. Algumas telas foram incluídas, outras telas mantidas por causarem um bom efeito visual e fazendo com que o usuário conseguisse tomar boas decisões. Uma das telas exibidas para os usuários como melhoria foi a tela de filtragem dos elementos visuais na tela observada na Figura 4.3. Esta tela possui como principal funcionalidade a redução dos elementos observados nas visualizações disponíveis. No caso, os filtros apresentados pertencem ao controle leiteiro das vacas onde podem ser controlados os seguintes atributos: qual a faixa de produção observada, ou seja, irão ser consideradas apenas as vacas que possuem controle leiteiro entre a quantia estipulada; O campo de percentual de gordura representa a gordura mínima a ser considerada para os componentes de leite das vacas; A caixa de seleção é pertinente ao estado de saúde dos animais, filtrando por animais saudáveis, com infecção crônica, doentes e infecção eliminada; Por fim, um componente de calendário que representa a partir de qual data os controles leiteiros devem ser examinados.

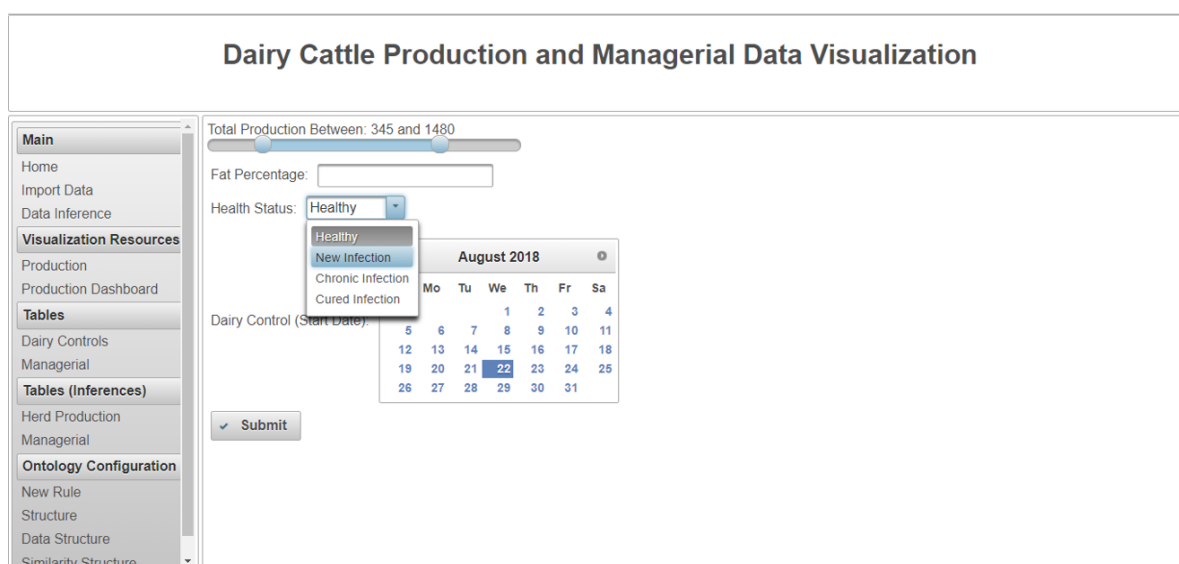


Figura 4.3: Tela de filtro de dados na plataforma de visualização

Além disso, foram executadas análises relacionadas com a inferência dos dados pela ontologia, sendo estas vista através de tabelas (além da exibição por *Dashboards*), visto

que tabelas também constituem uma poderosa ferramenta de análise dos indicadores. As Figuras 4.4 e 4.5 ilustram essas tabelas. Desta forma, a Figura 4.4 denota informações obtidas a partir de animais de rebanhos selecionados e comparando com outros rebanhos de interesse. Neste caso, observa-se que o rebanho do usuário (Herd1964) possui total de produção de leite acima do rebanho comparado (Herd315), destacando a superioridade deste indicador.

Herd Name	Total Production	Last Dairy Control
herd1964	304829.28	Wed Apr 27 00:00:00 BRT 2016
herd315	113857.81	Wed May 04 00:00:00 BRT 2016

Figura 4.4: Tabela de produção dos animais por rebanho

A Figura 4.5 representa uma outra informação aderente ao aspecto gerencial da informação. Determina algumas das entidades gerenciais e a produção atendida pelas entidades gerenciais com hierarquia menor. Desta forma, é visível que o Núcleo Gerencial (RegionalCenter195) atende o coordenador (Coordinator1283) que, por sua vez, coordena o consultor (Consultant109) e este presta consultoria para propriedades e estas propriedades produzem um total de 304829.3 litros de leite. Assim, é possível verificar as relações gerenciais entre todas estas entidades e identificar quais se destacam e verificar seus valores quantitativos referentes aos principais indicadores econômicos.

Durante a execução do estudo de caso, foi priorizada a análise de redes complexas, visto a dificuldade observada no estudo piloto. Algumas definições precisaram ser melhoradas para poder ressaltar os benefícios de algumas redes complexas para os participantes, de forma que as dificuldades técnicas com relação aos elementos computacionais dificultaram o entendimento das redes por parte dos participantes do estudo piloto. A identificação

Dairy Cattle Production and Managerial Data Visualization				
Main Home Import Data Data Inference Visualization Resources Production Production Dashboard Tables Dairy Controls Managerial Tables (Inferences) Herd Production Managerial Ontology Configuration New Rule Structure Data Structure Similarity Structure	Regional Center Name	Coordinator Name	Consultant Name	Served Property Production
	regionalCenter195	coordinator1283	consultant109	304829.3

Figura 4.5: Tabela de produção de leite atendida por entidades gerenciais

dos indivíduos relevantes ficou mais explícita na rede, a fim de destacar realmente dentro da rede quais animais possuem relevância maior de análise dentro do grupo identificado.

A arquitetura utiliza as redes complexas com dois principais objetivos: identificar os indivíduos mais relevantes dentro de um grupo semântico e observar relações ocultas entre indivíduos. No sentido de criar relações ocultas, são realizados processamentos baseados nas informações semânticas vindas da ontologia e aplicada estas informações na criação da rede, como associação de valores para as arestas e os vértices do grafo criado. Para a identificação de indivíduos mais relevantes, são utilizadas métricas de redes complexas de forma a obter indicadores que ressaltam a qualidade de indivíduos em relação aos seus vizinhos. Obtendo a proximidade destes indivíduos e utilizando algumas fórmulas, é possível identificar estas relações. Estas fórmulas e métricas serão detalhadas a seguir.

A análise feita nesta seção retrata um exemplo aplicado a rede de vacas. Ao obter as vacas de um rebanho, elas não têm nenhum relacionamento explícito indicado na ontologia ou pelo banco de dados. Porém, é importante que se possam criar recursos para comparar estes animais, visto que cada um pode possuir singularidades que os tornam melhores em relação aos demais. Para aplicar esta análise, são criadas relações de similaridade entre estes animais a partir de um atributo de qualidade, definido de acordo com o domínio de aplicação. A métrica de qualidade identifica a similaridade da produção de animais de um rebanho selecionado com 20 animais. Para observar os nós, inicialmente serão levantados

os dados desta rede, assim como a qualidade média dos nós da rede, a similaridade média entre todos os nós, a fim de observar a relevância dos nós e identificar como as métricas conseguem identificar os indivíduos relevantes dentro da rede apresentada. Na Figura 4.6, estão representados os nós baseados em seu grau ponderado, de forma que os maiores nós representam uma maior diferença em relação aos seus vizinhos, podendo ser de forma positiva ou negativa (qualidade alta ou baixa).

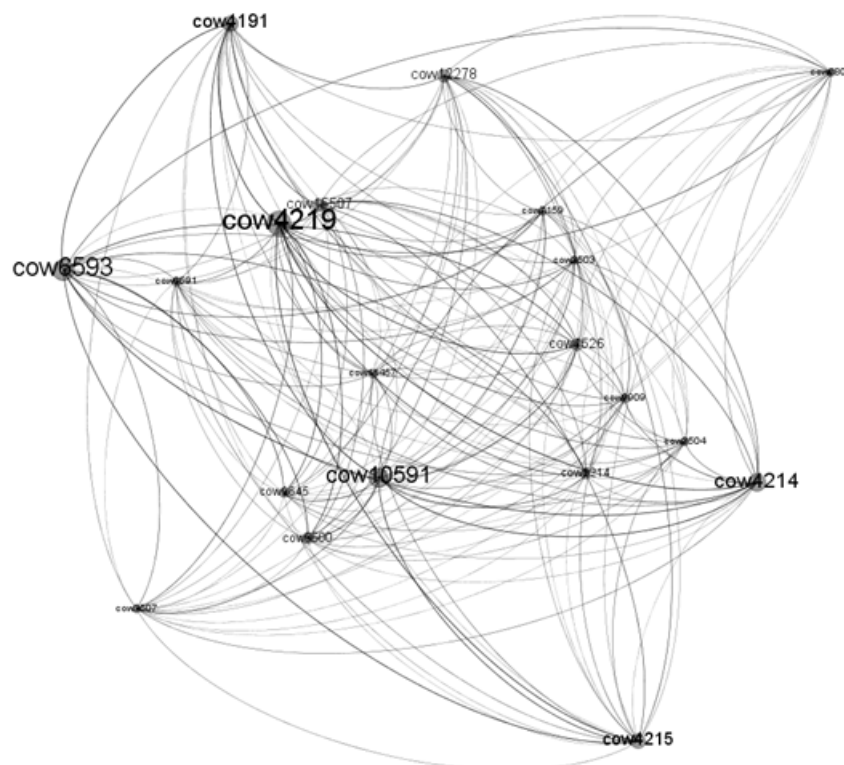


Figura 4.6: Rede de vacas em destaque

Os principais dados obtidos com relação a rede não-direcionada foram: total de nós: 20, arestas: 190, maior qualidade: 4898.72, menor qualidade: 217.07, média de qualidade: 2885.14, maior similaridade: 4898.72, menor similaridade: 217.07, média de similaridade: 2847.22.

Assim, observamos nesta rede todos os nós que não possuem uma relação direta ou indireta, mas sim uma relação oculta baseada na similaridade entre suas qualidades. Desta forma, a primeira análise realizada é o valor de qualidade de cada nó. Neste rebanho, obtemos que a maior qualidade é 4898.72 do animal cow4214 e a menor qualidade é 217.07, do animal cow4219. Desta forma, o objetivo ao analisar esta rede é, através de métricas em redes complexas, identificar indivíduos mais relevantes dentro da rede.

Origem	Destino	Weight	Origem	Destino	Weight	Origem	Destino	Weight	Origem	Destino	Weight	Origem	Destino	Weight
14	5	4682.646	12	5	2357.726	18	5	1546.556	14	3	904.38	16	9	372.18
5	1	4422.386	14	8	2343.62	17	5	1531.876	16	4	889.11	6	4	369.85
14	11	4366.56	8	5	2340.026	17	6	1522.8	9	6	886.78	4	3	357.74
16	5	4310.636	14	12	2325.92	18	6	1508.12	17	13	860.51	2	0	349.49
14	7	4147.407	15	7	2324.737	16	2	1502.45	18	13	845.83	19	10	347.29
11	1	4106.3	14	13	2292.26	9	0	1479.76	16	10	828.98	19	2	328.18
16	11	3994.55	7	2	2273.947	16	15	1451.66	17	12	826.85	11	5	317.086
9	5	3939.456	17	3	2248.39	8	3	1440.24	18	12	812.17	10	3	297.61
7	1	3887.147	18	3	2233.71	12	3	1422.54	17	8	809.15	19	4	287.16
5	3	3779.266	14	0	2222.95	13	3	1388.88	19	9	804.09	19	15	277.39
16	7	3775.397	11	0	2144.61	6	1	1369.71	19	8	797.34	14	1	261.26
11	9	3623.37	8	1	2083.36	17	15	1329.1	18	8	794.47	6	2	245.49
10	5	3482.656	13	11	2075.3	3	0	1319.57	19	12	779.64	11	7	220.153
11	3	3463.18	12	1	2065.66	18	15	1314.42	19	13	745.98	15	6	194.7
5	4	3422.526	12	11	2041.64	19	1	1287.02	14	9	744.19	9	3	161.19
9	7	3404.217	13	1	2032.0	17	2	1278.31	6	3	726.59	8	0	121.67
7	3	3244.027	11	8	2023.94	18	2	1263.63	8	6	714.65	16	1	112.75
11	10	3186.57	16	8	1971.61	14	4	1261.12	12	6	696.95	12	0	103.97
17	14	3151.77	1	0	1962.69	16	6	1257.96	19	0	676.67	19	6	83.69
18	14	3137.09	16	12	1953.91	18	11	1230.47	10	2	674.47	13	0	70.31
19	5	3136.366	17	10	1951.78	17	11	1215.79	13	6	663.29	10	4	61.13
11	4	3106.44	18	10	1937.1	14	10	1200.99	3	1	644.12	13	8	52.36
6	5	3053.676	7	0	1925.457	19	16	1175.27	19	3	643.9	15	2	51.79
10	7	2947.417	16	13	1920.25	10	8	1143.63	15	10	623.68	13	12	34.66
17	1	2891.51	17	4	1891.65	9	2	1131.27	4	2	614.34	12	8	18.7
7	4	2887.287	18	4	1876.97	12	10	1125.93	6	0	593.98	18	17	15.68
18	1	2876.83	14	2	1874.46	13	10	1092.27	15	4	563.55			
15	5	2859.976	13	7	1856.147	8	4	1083.5	7	5	536.239			
19	11	2820.28	16	0	1850.94	15	9	1080.48	16	9	532.37			
5	2	2809.186	15	14	1823.67	12	4	1065.8	15	8	520.95			
17	16	2779.76	12	7	1822.487	13	4	1032.14	9	4	517.93			
18	16	2765.08	9	7	1804.767	10	0	1022.96	15	12	503.25			
11	6	2737.59	14	6	1629.97	18	7	1011.317	9	1	483.93			
19	7	2601.127	2	1	1614.2	4	1	1000.86	8	2	470.16			
15	11	2543.89	19	17	1605.49	17	7	996.637	15	13	469.59			
7	6	2518.437	9	8	1600.43	3	2	971.08	10	9	457.8			
11	2	2493.1	19	18	1590.81	4	0	962.83	12	2	452.46			
5	0	2460.696	12	9	1582.73	10	1	940.73	10	6	429.98			
17	9	2408.58	15	1	1563.41	17	0	929.82	13	2	418.8			
18	9	2393.9	13	9	1549.07	15	3	920.29	15	0	400.28			
13	5	2391.386	19	14	1547.28	18	0	915.14	18	14	373.01			

Figura 4.7: Similaridade nas arestas entre vacas

Neste trabalho iremos utilizar algumas características da rede e métricas, como grau ponderado para análise. Cada aresta da rede representa a disparidade entre as medidas de qualidade entre dois nós. Assim, obtemos que uma aresta com peso grande identifica que a medida de qualidade entre dois indivíduos é muito grande, ou seja, eles possuem uma diferença muito grande entre as qualidades. Neste sentido, identificamos que um nó com alto grau ponderado, representa um indivíduo com uma grande diferença de qualidade em relação aos seus vizinhos, enquanto um indivíduo com uma baixa média ponderada representa indivíduos que estão mais próximos da média de qualidade da rede. Desta forma, a métrica utilizada combina estas análises. Baseado no ranking de graus ponderados e juntamente com uma análise geral da rede, é possível identificar se um indivíduo é bom ou ruim para a rede.

Nesta análise de grau ponderado, conseguimos identificar os dois extremos de indivíduos, ou seja, aqueles que possuem uma medida de qualidade muito alta e muito baixas. Para identificar se ele é bom ou ruim, é necessário comparar sua medida de qualidade com a medida de toda a rede. Neste sentido, é projetada uma métrica para identificar a taxa do indivíduo em relação a rede. Esta métrica é definida como a qualidade do nó subtraída da média de qualidade da rede, desta forma, se o valor da taxa for negativo significa que o indivíduo analisado está com destaque negativo, caso contrário tem um destaque positivo dentro da rede. representa a taxa, enquanto indica a qualidade do nó x e, por último,

representa a qualidade média na rede.

$$T_x = Q_x - Q_e \quad (4.3)$$

Com esta métrica é possível projetar dentro da rede de animais os que se destacam em relação aos demais, observando se esse destaque é positivo ou negativo dentro da rede utilizando a taxa de qualidade em relação a rede. Essa estratégia de identificação de nós pode ser utilizada para quaisquer redes que contenham uma métrica de qualidade para o nó e que se queira realizar a similaridade entre os nós, de forma que é possível observar os nós mais próximos da média de qualidade da rede e os nós que destoam mais, como os nós com maior e menor qualidade.

Como exemplo iremos realizar os cálculos para alguns nós da rede e observar os resultados. O cálculo é feito para os 4 primeiros nós identificados no ranking de grau ponderado.

$$Q(5) = 217.074$$

$T(5) = 217.074 - 2885.14 = -2668.066$, ou seja, o valor é negativo muito distante da média de qualidade, o que indica que o nó está, realmente, muito abaixo da qualidade da rede.

$$Q(11) = 533.16$$

$T(11) = 533.16 - 2885.14 = -2351.98$, ou seja, como possui um valor negativo muito abaixo da média, também reflete um animal com pouca qualidade.

$$Q(7) = 752.313$$

$T(7) = 752.313 - 2885.14 = -2132.827$, se enquadra no mesmo grupo dos anteriores.

$$Q(14) = 4898.72$$

$T(14) = 4898.72 - 2885.14 = 2013.58$, ou seja, possui um valor alto, indicando que sua qualidade é acima da média.

Desta forma, podemos obter as seguintes conclusões analisando estes fatores: é possível identificar que os graus ponderados ressaltam os nós com maior disparidade dentro da rede, de forma que é possível selecionar os N primeiros nós desse ranking e agrupá-los de acordo com a sua taxa. Por exemplo, os indivíduos que possuem uma taxa acima de um determinado valor podem determinar um grupo de animais relevantes dentro da rede, por seu desempenho acima da média. Enquanto enquadrarmos um grupo de taxas negativas, ou seja, os animais que merecem mais atenção por sua baixa produtividade.

Ao mesmo tempo, utilizando um ranking inverso, observamos os animais que possuem um desempenho médio na rede, o que pode ser interessante caso o usuário busque os animais que mantêm um maior equilíbrio em seus indicadores de eficiência.

4.4.2 Resultados Obtidos do Estudo de Caso Regular

No estudo de caso regular, obtivemos as respostas dos participantes em relação aos principais aspectos a respeito da tomada de decisão proporcionada pela plataforma de visualização que utiliza a arquitetura proposta nesta dissertação. Os resultados foram baseados na percepção dos usuários em relação ao uso da plataforma e seus elementos visuais aplicados às análises dos dados no domínio de produção do leite. Os resultados apontam para a viabilidade da plataforma, respondendo as questões de pesquisa propostas.

As respostas das questões de pesquisa Q1.1, Q1.2 e Q1.3 foram observadas durante a etapa de entrevista dos usuários ao apresentar estas técnicas e como elas representam os dados relacionados aos animais. Estas técnicas não foram expostas no formulário, já que se tratam do domínio específico de computação, onde alguns dos usuários poderiam ter dificuldades técnicas de entender do que se tratavam estas questões. Desta forma, foi observado que as ontologias propiciam uma melhor percepção do domínio de aplicação, podendo adaptar a plataforma para novas bases de dados de acordo com a criação de uma nova ontologia de domínio. Além disso, a escolha de métricas baseadas em redes complexas para representar indivíduos relevantes assim como as associações hierárquicas entre atores, serviu ao propósito de análise por parte dos participantes, que observaram possíveis adaptações para outros contextos utilizando abordagens semelhantes. Por fim, a questão Q1.4 foi respondida através da observação direta com os usuários fazendo uso da plataforma, de forma que conseguiram identificar análises e tendências com relação aos dados apresentados nas visualizações presentes na plataforma. Tais análises foram citadas ao longo deste capítulo de avaliação.

A integração entre ambas técnicas possibilitou vantagens para os participantes, de forma que a agregação de informações semânticas aos dados de redes hierárquicas e a criação de novos grupos semânticos a partir das métricas de redes complexas, possibilitou diferentes análises por parte dos usuários, criando uma percepção mais efetiva de novos grupos além dos grupos expostos inicialmente.

Duas questões foram propostas para analisar se as informações na plataforma estavam

disponíveis de uma maneira clara, de forma que o usuário teve a possibilidade de encontrar todos os recursos necessários para o que pretendia. Nas Figuras 4.8 e 4.9, observamos que maioria dos usuários julgou a plataforma como boa usabilidade e facilidade para encontrar informações na mesma. A escala representada nos gráficos está disposta no formulário disponível no Apêndice B.

Como você avalia a usabilidade geral da ferramenta?

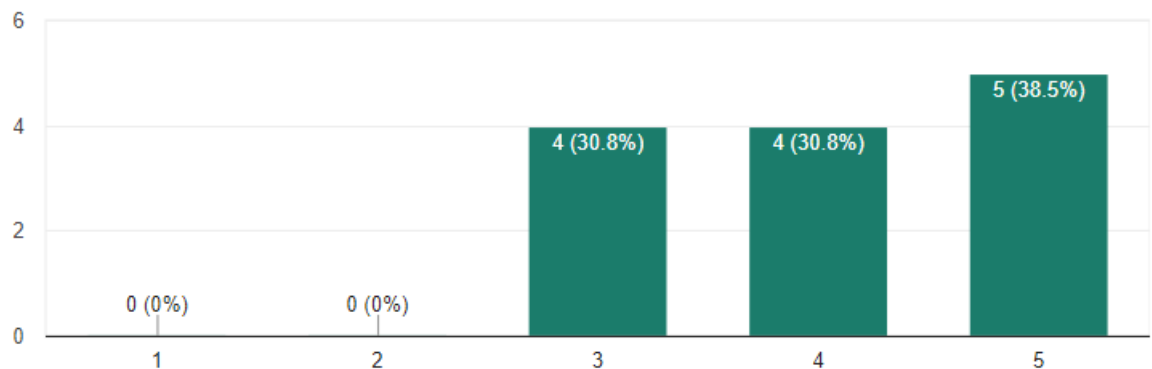


Figura 4.8: Usabilidade da plataforma

Teve facilidade para encontrar as informações buscadas nos menus da ferramenta?

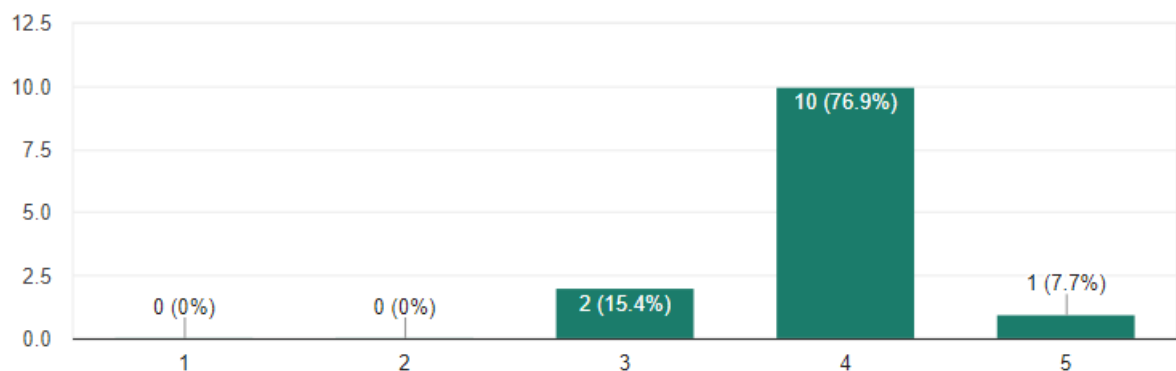


Figura 4.9: Facilidade de acesso da plataforma

Em seguida, buscamos identificar se as visualizações propostas possibilitaram fomentar as principais necessidades em obtenção de novos conhecimentos baseado nos dados analisados. A grande maioria dos participantes julgaram como suficientes as visualizações

apresentadas, observado na Figura 4.10 exceto pela possibilidade de incrementar as visualizações apresentadas com dados provenientes de outros domínios, conforme identificado na entrevista realizada com os usuários dentro do estudo de caso. A partir das observações destes gráficos, obtivemos indícios de que as visualizações apresentadas conseguem otimizar a tomada de decisão, oferecendo mecanismos que integrem volumes de dados extensos, fornecendo uma melhor percepção do domínio de dados. Com isso, a visualização associada às técnicas apresentadas, possibilita uma melhor seleção dos dados, indicando que a tomada de decisão se torna mais efetiva.

As visualizações apresentadas foram suficientes para conseguir analisar os dados e obter novos conhecimentos?

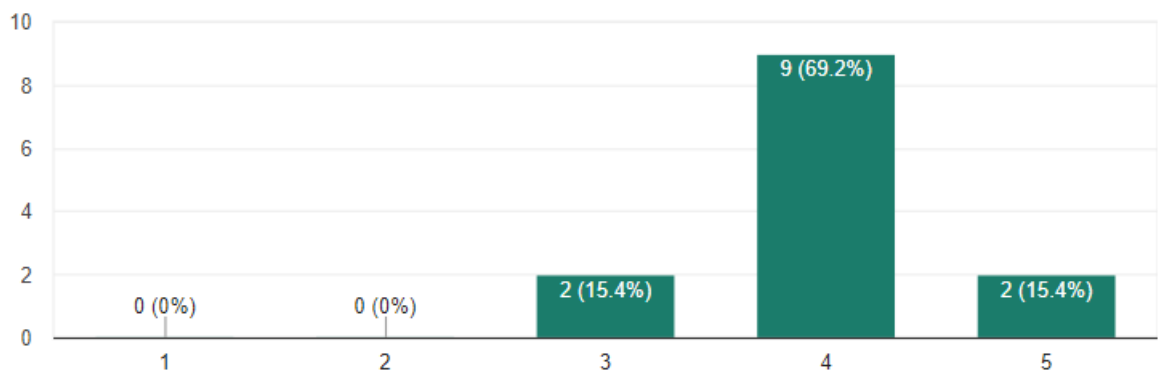


Figura 4.10: Suficiência da apresentação de visualizações durante a análise dos dados na plataforma

Por fim, pedimos para que os participantes ressaltassem os principais aspectos positivos na plataforma, de forma a observar quais elementos foram mais efetivos para a tomada de decisão. Podemos observar que os elementos mais relevantes são os gráficos com dados implícitos para identificar novos conhecimentos, os quais utilizam ontologias e regras de inferência para demonstrar os resultados e, juntamente com os gráficos, a identificação de indivíduos mais relevantes nos grupos e possibilidade de criar novos agrupamentos, possibilitado pelas análises de métricas em redes complexas. Desta forma, observamos como as técnicas abordadas na arquitetura proposta auxiliaram na tomada de decisão e planejamento estratégico.

Algumas perguntas discursivas foram realizadas com o intuito de perceber quais elementos os participantes gostariam de ressaltar sem haver nenhuma interferência de res-

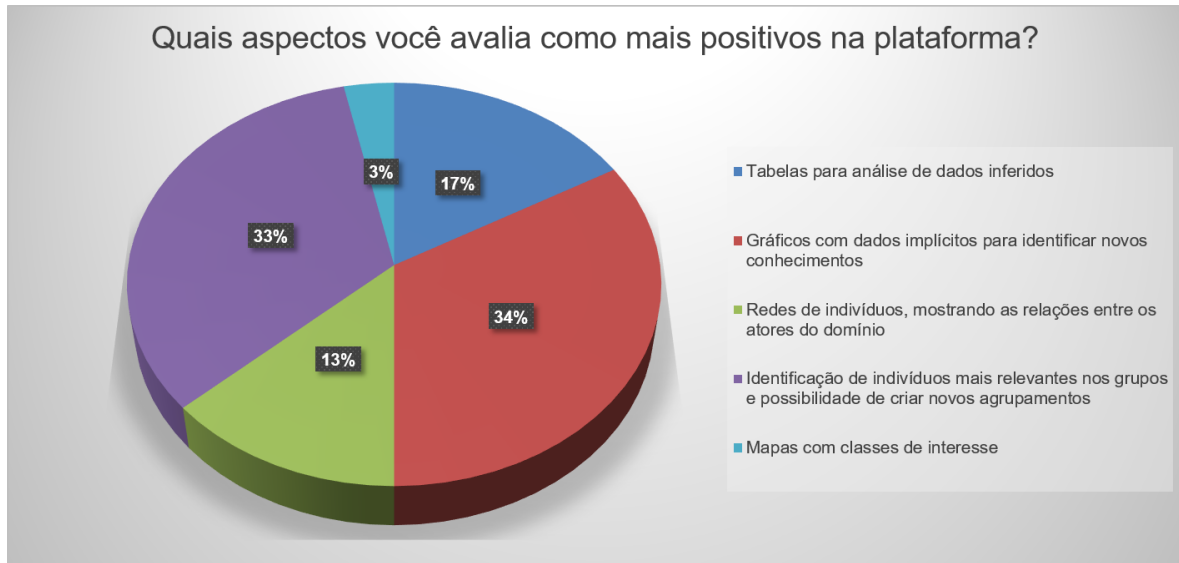


Figura 4.11: Aspectos positivos da plataforma de visualização

postas previamente criadas.

- (i) O que você alteraria ou adicionaria para melhorar a visualização?

Para isto, as principais respostas envolveram melhorar o design gráfico, maior clareza sobre as métricas apresentadas e textos explicativos sobre alguns gráficos. Em resumo, poderiam ser utilizados alguns outros recursos gráficos, como melhoria nos estilos de elementos visuais da página, ou seja, melhorias no front-end apresentado.

- (ii) O que você gostaria de ressaltar em relação aos itens avaliados?

Os participantes, ressaltaram os benefícios em realizar consultas dinâmicas, resultados com impacto nos processos estruturados, boa usabilidade, aspecto visual relevante para a aplicação dos dados e a indicação do posicionamento de indivíduos perante seu grupo homogêneo.

- (iii) Ocorreram quaisquer problemas técnicos durante a utilização da plataforma? Em caso positivo, citar quais foram.

A grande maioria dos participantes avaliou sem ocorrência de quaisquer problemas técnicos durante a utilização da plataforma, porém em um caso houve uma demora no processamento, ocasionada durante a atualização do banco de dados e realização da inferência na ontologia de domínio.

- (iv) Campo livre para sugestões e considerações.

Os participantes ressaltaram o interesse em aplicar a plataforma em outros domínios, para complementar as visualizações criadas nos dashboards e elementos visuais da plataforma. Além disso, citaram a relevância do problema estudado para a cadeia de produção do leite, de forma que a tomada de decisão apresenta aspectos gerenciais vitais para uma melhoria da gestão por parte de produtores de leite. Um participante considerou a melhoria na apresentação da plataforma, como melhorias na página inicial, de forma a apresentar melhor os elementos de busca.

Durante a etapa de entrevista, foi observado que os participantes se mostraram muito interessados com a criação da plataforma de visualização, expondo a escassez de ferramentas com visualizações confiáveis no contexto do domínio de aplicação. Houveram algumas dúvidas com relação a quais bases de dados estavam sendo exibidas na plataforma, de forma que cada participante citava as possibilidades de incluir bases de dados e visualizações específicas de sua área, como por exemplo, um participante que possuía maior experiência em geoprocessamento pensou na possibilidade de mecanismos visuais representando mapas, representando zonas de calor, ou seja, rebanhos e animais que possuem destaque dentro dos grupos analisados. As etapas de atualização do banco e processamento da máquina de inferência na ontologia, foram realizados de maneiras normal, sem muitas perguntas sobre o processo em si. Com relação as tabelas, foram elogiadas e cumpriram seu objetivo em conseguir representar análises numéricas, que também são necessárias no processo de tomada de decisão. Porém, indicaram a necessidade de exportar essa tabela no formato de alguma planilha eletrônica ou outros tipos de arquivo.

Com relação aos dashboards, que representam informações semânticas inferidas sobre os dados importados, os participantes ressaltaram a importância e relevância de se possibilitar a observação de um mesmo conjunto de dados de maneiras diferentes e, nesta etapa, muitos participantes fizeram observações imaginando adaptações das visualizações do dashboard para o contexto de aplicação que os mesmos trabalham, ressaltando várias métricas para indicadores de qualidade em contexto específicos. Uma das visualizações mais importante nesse processo de tomada de decisão foi o *gauge chart*, que representa o quão bom um elemento está em comparação com médias de grupos pré-definidos.

Por fim, foram realizadas perguntas durante a apresentação de forma a esclarecer melhor se os elementos visuais otimizaram a tomada de decisão e corresponderam às expectativas das informações que gostariam de encontrar, buscando responder à questão

de pesquisa proposta.

(i) A análise dos dados em diferentes perspectivas (utilização de dashboards) melhora o processo de tomada de decisão?

Os participantes sugeriram diversas aplicações para o dashboard proposto, de forma a aplicar o mesmo em diferentes domínios, visto que seu potencial de tomada de decisão é alto. Explorando algumas métricas e indicadores de qualidade do rebanho em comparação com agrupamentos determinados, pode otimizar a percepção dos dados analisados em relação à uma média geral.

(ii) Vocês acham que as informações são apresentadas de maneira clara na plataforma?

Alguns participantes encontraram alguns problemas com relação ao layout, de forma que a aplicação de novos estilos poderia tornar mais eficiente a apresentação dos dados. Apoiaram a utilização de novas bases de dados mais enxutas para exibição de alguns indicadores, que poderiam estar incompletos para animais utilizados na base de dados importada.

Por fim, ao responder as questões de pesquisa secundárias, observamos que a utilização da arquitetura proposta apresenta indícios de viabilidade para uma melhor tomada de decisão e gerenciamento de novas informações, respondendo a questão de pesquisa principal.

4.5 AMEAÇAS À VALIDADE

A condução do estudo de caso pode apresentar elementos subjetivos do pesquisador, de forma que estas podem apresentar ameaças a confiabilidade do estudo apresentado. Para avaliar as ameaças, são considerados quatro aspectos de qualidade: interna, externa, constructo e conclusão (YIN, 2015).

4.5.1 Validade Interna

Como ameaças de validade interna podemos destacar elementos de incompreensão de questões técnicas em relação a plataforma desenvolvida. Os principais elementos computacionais foram apresentados de forma mais técnicas, para minimizar esta validade, porém os participantes podem ter dificuldades de compreender os mesmos. Problemas relacionados a dificuldades na busca de informações podem influenciar na tomada de decisão por

parte dos participantes.

4.5.2 Validade Externa

A arquitetura proposta tem como fundamento a possibilidade de adaptação, de acordo com a ontologia de domínio criada e os web services construídos para fomentar a análise dos dados. Porém, neste trabalho a arquitetura foi aplicada apenas no contexto de produção do leite e apenas uma ontologia de domínio foi desenvolvida. Desta forma, a validade externa remete à limitação do contexto de aplicação.

4.5.3 Validade do Constructo

Os indicadores de qualidade aplicadas neste trabalho estão relacionadas ao levantamento de requisitos obtido no estudo exploratório inicial, desta forma, algumas das questões do formulário podem estar complicadas ou abrangentes demais para os participantes. Com isso, dificultando a percepção dos participantes e, conseqüentemente, uma ameaça para o levantamento dos requisitos do sistema. Além disso, a construção da ontologia de domínio também pode ser considerada uma ameaça a validade, já que seus principais aspectos estão relacionados aos requisitos levantados inicialmente.

4.5.4 Validade de Conclusão

Como ameaças a validade podemos citar o número limitado de participantes. Através desta avaliação não é possível generalizar os resultados e identificar os resultados em outras aplicações utilizando a mesma arquitetura desenvolvida.

4.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou a avaliação feita sobre a viabilidade da arquitetura proposta no Capítulo 3. De forma a apresentar as principais etapas para as conclusões apresentadas. A avaliação aponta para a viabilidade da arquitetura assim como uma melhor tomada de decisão baseada na agregação das técnicas de ontologia e redes complexas. O capítulo apresenta dados quantitativos obtidos a partir das respostas de formulários, assim como análises observacionais e entrevistas com os participantes.

Os principais elementos de visualização também foram representados no capítulo, observando os principais resultados obtidos a partir da análise dos mesmos.

5 CONSIDERAÇÕES FINAIS

Este capítulo tem como objetivo apresentar as contribuições desta dissertação em relação a arquitetura proposta, além de suas limitações e trabalhos futuros.

5.1 VISÃO GERAL

Esta dissertação propôs uma arquitetura de visualização de dados baseada nos processos associados a inteligência de negócios, que visa otimizar decisões, reduzir custos e dar suporte a gestão da informação. A arquitetura proposta utiliza técnicas de ontologias e redes complexas de forma agregada para aprimorar a obtenção de novas informações em bases de dados heterogêneas do contexto de produção do leite, no domínio agropecuário.

A arquitetura, apresentada no Capítulo 3, conseguiu consolidar dados oriundos de diferentes bases de dados relacionadas a amostras de leite, além disso uma avaliação foi realizada de forma a identificar se a arquitetura também foi capaz de agregar valor a tomada de decisão, reforçando o problema desta dissertação, tratado no Capítulo 1.

O Capítulo 4 apresentou a avaliação desta plataforma, realizada na Embrapa Gado de Leite com dados reais de rebanhos da Embrapa e análises de laboratório. Os dados foram consolidados e apresentados na plataforma de visualização, de forma que os pesquisadores avaliaram a tomada de decisão em relação aos dados utilizados. A avaliação contou com coletas de informação em forma de entrevistas, observação e formulários aplicados aos participantes.

Os resultados da avaliação apontaram para a viabilidade da proposta, que implica em indícios que a *concepção de uma arquitetura e o desenvolvimento de uma plataforma para visualização de dados sob a perspectiva da Inteligência de Negócios, proporciona um ferramental capaz de apresentar aos produtores e instituições leiteiras conhecimento implícito nos dados, possibilitando a análise e tomada de decisão em relação a produção do leite*, que era a questão de pesquisa desta dissertação.

Em relação aos objetivos propostos: *propor uma arquitetura sob a perspectiva da Inteligência de Negócios que define os componentes necessários para visualização da informação e apoio à tomada de decisão*, consideramos que foram atendidos. No que diz respeito aos objetivos secundários, é possível destacar alguns pontos relevantes:

- A utilização de ontologias proporcionou maneiras eficientes de agregar bases de dados heterogêneas, de forma a agregar valores semânticos aos dados. A possibilidade de relacionar indicadores de desempenho aos animais provê informações relevantes por parte do especialista, que entende quais os principais indicadores de eficiência para um indivíduo pertencente ao domínio de aplicação.
- Agregar técnicas de ontologia e redes complexas proporcionou benefícios em ambas as abordagens utilizadas na arquitetura proposta. A utilização de ontologias agregou informações semânticas às redes complexas, fornecendo soluções para reduzir a quantidade de nós através de property chains, omitindo nós intermediários e apresentando apenas o domínio e a imagem definidas nas property chains. Com relação aos benefícios da ontologia, possibilitou uma representação estrutural das principais classes que a compõem além de propiciar o reconhecimento dos indivíduos mais relevantes dentro dos conjuntos de dados.
- A criação de uma ontologia de domínio de produção do leite foi uma inovação no contexto de aplicação, já que durante a revisão da literatura não foram encontrados trabalhos com ontologias específicas relacionadas ao contexto abordado nesta dissertação. Com a criação e utilização desta ontologia, a representação do domínio de aplicação se tornou mais eficiente, provendo formas de representar as informações semânticas agregadas a todos os indivíduos presentes no domínio, assim como suas relações implícitas e explícitas.
- As visualizações empregadas na arquitetura possibilitaram boas abstrações da informação do grande conjunto de dados provenientes das fontes importadas (sistemas e análises laboratoriais). No estudo piloto foram identificadas que algumas das visões relacionadas às redes complexas precisavam de mais informações para que os usuários pudessem entender seu contexto real.
- A utilização das abstrações visuais atingiram o objetivo de auxiliar na tomada de decisão, de forma a conseguir representar de maneira simplificada informações complexas dos dados.

5.2 CONTRIBUIÇÕES

As principais contribuições realizadas nesta dissertação foram:

- Construção de uma arquitetura que permite integrar técnicas de ontologia e redes complexas que, agregadas, possibilitam um processamento mais efetivo dos dados para gerar informação gerencial;
- Desenvolvimento uma plataforma de visualização para apoio a tomada de decisão no contexto de gado leiteiro;
- Construção de uma ontologia de domínio, aplicada para o domínio de produção de leite;
- Desenvolvimento de um processo de ETL que importa dados proveniente de fontes heterogêneas, como dados de análises de laboratório e ferramentas de gestão de gado leiteiro;
- Desenvolvimento de serviços web para análise dos dados interpretados semanticamente, permitindo uma análise destes dados por parte de pesquisadores e agentes da CPL;
- Criação de visualizações que auxiliam na tomada de decisão por parte de gestores de negócio, gerando uma maior abstração dos dados, otimizando as decisões estratégicas;

A partir da avaliação realizada no capítulo 4, identificamos que os artefatos gerados auxiliam em uma decisão estratégica eficiente por parte dos gestores de negócio. Com isso, o desenvolvimento da plataforma permitiu que os pesquisadores participantes conseguissem analisar os dados provenientes das análises laboratoriais e de outras fontes heterogêneas, de uma forma consolidada e realizar tomadas de decisão com as informações obtidas.

A criação da ontologia de domínio também permitiu consolidar os principais elementos relacionados ao domínio de aplicação, de forma que todos os principais elementos que o constituem puderam ser representados na plataforma e realizadas análises nos mesmos.

5.3 LIMITAÇÕES

A arquitetura proposta foi aplicada para atender as necessidades do domínio de produção do leite, a partir de requisitos levantados por pesquisadores da Embrapa entre outros usuários. O número limitado de participantes para o levantamento de requisitos influencia nas informações disponibilizadas além dos indicadores de qualidade utilizadas para representar o desempenho de produção dos animais. Com isso, possivelmente se o levantamento de requisitos fosse mais abrangente, ou seja, disponibilizado para mais pesquisadores e produtores, novas métricas iriam surgir para a análise da eficiência dos animais e, conseqüentemente, novas visualizações seriam derivadas para suprir estas necessidades.

Os dados provenientes de análises laboratoriais e fazendas é crescente, desta forma o grande volume de dados gerado pode limitar a utilização da ontologia em relação a utilização de inferências sobre os dados. Assim, é preciso testar a aplicação da arquitetura com um maior conjunto de dados e observar o comportamento. Ainda com relação aos dados importados, também podem ser considerados como uma limitação, visto que dados provenientes de mais fontes geram outras informações e conseqüentemente análises diferentes, é válido analisar a aplicabilidade das mesmas na arquitetura.

Algumas das visualizações utilizadas estão limitadas ao framework utilizado (*PrimeFaces*) e bibliotecas internas do Java, é possível explorar outras ferramentas visuais, de forma a identificar outras abstrações visuais para representar outras perspectivas com relação aos dados importados.

5.4 TRABALHOS FUTUROS

Algumas das limitações apresentadas, podem ser sugeridas como trabalhos futuros, visto que a aplicabilidade da plataforma é muito extensa. Uma das principais sugestões de melhoria na plataforma é a aplicação da arquitetura utilizada em outros domínios de aplicação, assim como citado por participantes da avaliação realizada que gostariam de observar a análise dos dados no contexto de geoprocessamento, análises econômicas específicas, dentre outras aplicações. Além disso, também pode ser desenvolvidas outras ontologias de domínios fora de gado do leite e analisar o comportamento da arquitetura e como a visualização e os novos serviços criados irão interagir.

A ontologia de domínio criada ainda é muito simplória, de modo que esta pode ser

expandida para oferecer uma maior gama de informações inferidas com relação aos dados. Um levantamento de requisitos utilizando um maior número de participantes pode auxiliar no levantamento de novas necessidades e indicadores a serem aplicados na ontologia, construindo um domínio mais robusto para as aplicações de gados leiteiros.

As visualizações retratadas na plataforma constituem as principais necessidades levantadas através do estudo exploratório realizado, com isso, a exploração de novas abstrações da informação são bem vindas. Como a construção da plataforma foi realizada de forma incremental, é possível desenvolver novas funcionalidades e serviços, inclusive permitindo integrá-la com softwares externos, o que é um fator relevante durante o desenvolvimento, vindo a fornecer uma usabilidade mais efetiva da plataforma.

REFERÊNCIAS

- AKERMAN, A.; TYREE, J. Using ontology to support development of software architectures. **IBM Systems Journal**, IBM, v. 45, n. 4, p. 813–825, 2006.
- ALLEN, I. E.; SEAMAN, C. A. Likert scales and data analyses. **Quality progress**, v. 40, n. 7, p. 64–65, 2007.
- B., K.; S., C. Guidelines for performing systematic literature reviews in software engineering. In: **Technical report, Ver. 2.3 EBSE Technical Report**. EBSE, 2007.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. et al. The semantic web. **Scientific american**, New York, NY, USA:, v. 284, n. 5, p. 28–37, 2001.
- BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D.-U. Complex networks: Structure and dynamics. **Physics reports**, Elsevier, v. 424, n. 4-5, p. 175–308, 2006.
- CARD, S.; MACKINLAY, J.; SHNEIDERMAN, B. Information visualization. **Human-computer interaction: design issues, solutions, and applications**, CRC Press, v. 181, 2009.
- CHEN, H.; CHIANG, R. H.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. **MIS quarterly**, v. 36, n. 4, p. 1165–1188, 2012.
- CHENG, T.; TEIZER, J. Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. **Automation in Construction**, v. 34, p. 3 – 15, 2013. ISSN 0926-5805. Information Technologies in Safety Management.
- DRESCH, A.; LACERDA, D. P.; JÚNIOR, J. A. V. A. **Design science research: método de pesquisa para avanço da ciência e tecnologia**, 2015.
- DörK, M.; CARPENDALE, S.; WILLIAMSON, C. Visualizing explicit and implicit relations of complex information spaces. **Information Visualization**, v. 11, n. 1, p. 5–21, 2012.

- ERL, T. Soa: principles of service design. Prentice Hall Upper Saddle River, v. 1, 2008.
- FALCI, M.; BRAGA, R.; STROËLE, V.; DAVID, J. Software process improvement through the combination of data provenance, ontologies and complex networks. In: , 2018. v. 2, p. 61–70.
- GRIERSON, H.; CORNEY, J.; HATCHER, G. Using visual representations for the searching and browsing of large, complex, multimedia data sets. **International Journal of Information Management**, v. 35, n. 2, p. 244 – 252, 2015. ISSN 0268-4012.
- GU, Y.; WANG, C. Transgraph: Hierarchical exploration of transition relationships in time-varying volumetric data. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2015–2024, 2011.
- GUARINO, N. et al. Formal ontology and information systems. In: **Proceedings of FOIS**, 1998. v. 98, n. 1998, p. 81–97.
- HEIM, P.; LOHMANN, S.; STEGEMANN, T. Interactive relationship discovery via the semantic web. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6088 LNCS, n. PART 1, p. 303–317, 2010.
- HEIMERL, F.; LOHMANN, S.; LANGE, S.; ERTL, T. Word cloud explorer: Text analytics based on word clouds. In: **IEEE. System Sciences (HICSS), 2014 47th Hawaii International Conference on**, 2014. p. 1833–1842.
- JAYARAMAN, P.; GUNASEKERA, K.; BURSTEIN, F.; HAGHIGHI, P. D.; SOETIKNO, H.; ZASLAVSKY, A. An ontology-based framework for real-time collection and visualization of mobile field triage data in mass gatherings. In: , 2013. p. 146–155.
- JONKER, D.; LANGEVIN, S.; BOZOWSKY, N.; WRIGHT, W. Aperture: An open web 2.0 visualization framework. In: , 2013. p. 1485–1494.
- KIM, J.; WILHELM, T. What is a complex graph? **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 387, n. 11, p. 2637–2652, 2008.
- LANCELLOTTA, P.; STRÖELE, V.; BRAGA, R.; DAVID, J.; CAMPOS, F. Semantic analysis and complex networks as conjugated techniques supporting decision making. In: , 2018. v. 2, p. 195–202.

- LARSON, D.; CHANG, V. A review and future direction of agile, business intelligence, analytics and data science. **International Journal of Information Management**, v. 36, n. 5, p. 700 – 710, 2016. ISSN 0268-4012.
- LIU, Y.; WANG, G.; JIANG, Z.; QIAN, J.; CHEN, Y. An integrated decision support system for emergency evacuation management. In: IEEE. **Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on**, 2016. p. 000940–000946.
- LOPES, G. R.; MORO, M. M.; WIVES, L. K.; OLIVEIRA, J. P. M. de. Cooperative authorship social network. In: **AMW**, 2010.
- LÜ, J.; CHEN, G.; OGORZALEK, M. J.; TRAJKOVIĆ, L. Theory and applications of complex networks: Advances and challenges. In: IEEE. **Circuits and Systems (ISCAS), 2013 IEEE International Symposium on**, 2013. p. 2291–2294.
- MIAH, S.; GAMMACK, J.; KERR, D. Ontology development for context-sensitive decision support. In: , 2007. p. 475–478.
- MUNIZ, C.; CHOREN, R.; GOLDSCHMIDT, R. Using a time based relationship weighting criterion to improve link prediction in social networks. In: , 2017. v. 1, p. 73–79.
- PARK, H.; BASOLE, R. C. Bicentric diagrams: Design and applications of a graph-based relational set visualization technique. **Decision Support Systems**, v. 84, p. 64 – 77, 2016. ISSN 0167-9236.
- PARK, H.; BELLAMY, M. A.; BASOLE, R. C. Visual analytics for supply network management: System design and evaluation. **Decision Support Systems**, p. –, 2016. ISSN 0167-9236.
- POPOVIČ, A.; HACKNEY, R.; COELHO, P. S.; JAKLIČ, J. Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. **Decision Support Systems**, Elsevier, v. 54, n. 1, p. 729–739, 2012.
- PRESSMAN, R.; MAXIM, B. Engenharia de software-8^a edição. McGraw Hill Brasil, 2016.
- QUEIROZ-SOUSA, P. O.; SALGADO, A. C.; PIRES, C. E. A method for building personalized ontology summaries. **Journal of Information and Data Management**, v. 4, n. 3, p. 236, 2013.

- RADICS, P.; POLYS, N.; NEUMAN, S.; LUND, W. Osnap! introducing the open semantic network analysis platform. In: , 2015. v. 9397.
- RANJAN, J. Business justification with business intelligence. **VINE**, v. 38, n. 4, p. 461–475, 2008.
- RIBEIRO, F. C.; CAETANO, B. P.; PAULA, M. M. de; FERREIRA, G. X.; OLIVEIRA, R. S. de. Keep calm and visualize your data. **Tópicos em Sistemas de Informação: Minicursos SBSI 2016**, p. 31, 2016.
- RUNESON, P.; HOST, M.; RAINER, A.; REGNELL, B. **Case study research in software engineering: Guidelines and examples**, 2012.
- RUSSELL, A. D.; CHIU, C.-Y.; KORDE, T. Visual representation of construction management data. **Automation in Construction**, v. 18, n. 8, p. 1045 – 1062, 2009. ISSN 0926-5805.
- SAFAEIPOUR, H.; ZARANDI, M.; BASTANI, S. Crisp to fuzzy ontology conversion in the context of social networks: A new approach. In: , 2017.
- SALEH, M.; ESA, Y.; MOHAMED, A. Applications of complex network analysis in electric power systems. **Energies**, Multidisciplinary Digital Publishing Institute, v. 11, n. 6, p. 1381, 2018.
- SAVOSKA, S.; LOSKOVSKA, S.; DIMITROVSKI, I. Information visualization from the public utilities databases of local municipality for municipalities managers. In: , 2008. p. 237–242. ISSN 13301012.
- SCHOTS, M. On the use of visualization for supporting software reuse. In: ACM. **Companion Proceedings of the 36th International Conference on Software Engineering**, 2014. p. 694–697.
- SILVA, M. A. A. da; CAVALCANTI, M. C. Combining ontology modules for scientific text annotation. **Journal of Information and Data Management**, v. 5, n. 3, p. 238, 2014.
- SILVA, P.; DIAS, S.; BRANDÃO, W.; SONG, M.; ZÁRATE, L. Formal concept analysis applied to professional social networks analysis. In: **Proceedings of the 19th International Conference on Enterprise Information Systems**, 2017. v. 1, p. 123–134.

- SOBRAL, T.; COSTA, V.; BORGES, J.; FONTES, T.; GALVÃO, T. Obavum: An ontology-based approach to visualizing urban mobility data. In: , 2016.
- SOKLAKOVA, T.; ZIARMAND, A.; OSADCHYIEVA, S. Big data visualization in smart cyber university. In: , 2016.
- SOLINGEN, R. V.; BASILI, V.; CALDIERA, G.; ROMBACH, H. D. Goal question metric (gqm) approach. **Encyclopedia of software engineering**, Wiley Online Library, 2002.
- STRÖELE, V.; CAMPOS, F.; DAVID, J. M. N.; BRAGA, R.; ABDALLA, A.; LANCELLOTTA, P. I.; ZIMBRÃO, G.; SOUZA, J. Data abstraction and centrality measures to scientific social network analysis. In: IEEE. **Computer Supported Cooperative Work in Design (CSCWD), 2017 IEEE 21st International Conference on**, 2017. p. 281–286.
- SUMALATHA, M.; RAVI, A.; ARAVIND, M.; PRASANNA, N. Collective intelligence in distributed systems and semantic data visualization. In: , 2008. p. 292–297.
- SWAYNE, D. F.; LANG, D. T.; BUJA, A.; COOK, D. Ggobi: evolving from {XGobi} into an extensible framework for interactive data visualization. **Computational Statistics & Data Analysis**, v. 43, n. 4, p. 423 – 444, 2003. ISSN 0167-9473. Data Visualization. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167947302002864>>.
- THULKE, H.-H.; SELHORST, T.; MÜLLER, T. Pseudorabies virus infections in wild boar: data visualisation as an aid to understanding disease dynamics. **Preventive Veterinary Medicine**, v. 68, n. 1, p. 35 – 48, 2005. ISSN 0167-5877.
- TRIEU, V.-H. Getting value from business intelligence systems: A review and research agenda. **Decision Support Systems**, Elsevier, v. 93, p. 111–124, 2017.
- VUORI, V. Studies of business information needs: a systematic literature review. 05 2007.
- WASSERMAN, S.; FAUST, K. **Social network analysis: Methods and applications**, 1994.
- WU, Y.; WEI, F.; LIU, S.; AU, N.; CUI, W.; ZHOU, H.; QU, H. Opinionseer: interactive visualization of hotel customer feedback. **IEEE transactions on visualization and computer graphics**, IEEE, v. 16, n. 6, p. 1109–1118, 2010.

YIN, R. K. Qualitative research from start to finish. Guilford Publications, 2015.

ZILLNER, S.; HAUER, T.; ROGULIN, D.; TSYMBAL, A.; HUBER, M.; SOLOMONIDES, T. Semantic visualization of patient information. In: , 2008. p. 296–301.

APÊNDICES

A QUESTIONÁRIO DE MÉTRICAS

Apresentação do Questionário

Meu nome é Pedro Ivo Pereira Lancellotta e sou estudante do Programa de pós-graduação em Ciência da Computação pela Universidade Federal de Juiz de Fora. Por meio desta mensagem convido-o a colaborar em um estudo exploratório que objetiva levantar informações para compor métricas e melhorias em uma ferramenta de visualização e análise de dados. A ferramenta é aplicada ao contexto agropecuário, mais precisamente, na gestão da produção do leite de bovinos leiteiros. Os dados relacionados à pesquisa serão apresentados e discutidos na dissertação de mestrado. As informações obtidas de colaboradores, serão compiladas para se realizar melhorias na ferramenta de visualização.

O contexto da dissertação e da ferramenta: A identificação de similaridades e o agrupamento de indivíduos relevantes (definidos por atributos ou indicadores de qualidade) em um conjunto de dados é uma tarefa complexa de ser realizada. A qualidade do indivíduo pode ser entendida, de maneira simplificada, como uma fórmula que calcula o quão bom o indivíduo é. Os indivíduos podem possuir diversas associações diretas e indiretas entre si. As associações indicam comportamentos, padrões ou relações, etc. que ajudam a estabelecer um melhor entendimento sobre a estrutura e conteúdo da informação dos dados que subsidiam um estudo/tema de interesse. Outra possibilidade é o estudo de relações ocultas nos dados, ou seja, estabelecer relações que não podem ser percebidas direta ou indiretamente. Neste sentido, poderia ser possível identificar a similaridade entre indivíduos através de métricas definidas ou, a partir de um determinado ponto de vista, identificar como seria possível comparar dois indivíduos que, a princípio, não teriam associações em comum.

Antecipadamente, meus agradecimentos pela sua participação e valiosa colaboração.

Identificação do Participante

Nome do Participante

Área de atuação (Esta resposta é obrigatória apenas no sentido de caracterizar os participantes)

E-mail

Em uma escala de 1 a 5, como você avalia seu conhecimento sobre gado leiteiro?

Baixo (1) (2) (3) (4) (5) Alto

Questionário sobre a Ferramenta

A ferramenta de visualização emprega meios que consigam fornecer maiores informações (implícitas, explícitas e ocultas) sobre conjuntos de dados complexos. Neste sentido, busca-se nesta seção, identificar quais aspectos são mais relevantes para você no momento de analisar um conjunto de dados. Os campos "Outros" podem ser utilizados para descrever observações caso a resposta não seja direta.

1. Você julga importante a utilização de visualizações para análise de dados

Pouco (1) (2) (3) (4) (5) Muito

2. Você acredita que a utilização de diferentes formas de visualização de dados (tabelas, gráficos, redes, etc.) pode facilitar a descoberta de novas informações?

Pouco (1) (2) (3) (4) (5) Muito

3. Você gostaria de ter uma visualização que permitisse observar as relações entre os indivíduos analisados (Por exemplo, relações de familiaridade entre os animais, produtores

relacionados aos animais, entre outras relações pertinentes ao domínio)?

Pouco (1) (2) (3) (4) (5) Muito

4. Você julga importante identificar indivíduos que se destacam (bons ou ruins em termos de produção ou algum outro atributo), como vacas e touros com mais destaque em um grupo (um rebanho, animais de um estado ou outros critérios de agrupamento) para observá-los melhor?

Pouco (1) (2) (3) (4) (5) Muito

5. Se pudesse identificar alguma deficiência na análise e visualização de dados em ferramentas de apoio a informação para gados leiteiros, qual seria a principal deficiência?

6. Campo livre para observações de sugestões

Questionário sobre a Ferramenta

Esta parte do questionário é para a identificação de atributos relevantes ao fundamentar as métricas para análise dos dados.

1. Quais atributos de vacas você acredita serem relevantes para analisar a qualidade de produção de uma vaca para compará-las umas com as outras? () Produção de Leite

() Produção 305

() Percentual de Gordura

() Percentual de Proteína

() Percentual de Lactose

() Período de Lactação

() Outros

2. Se pudesse elaborar uma fórmula para qualidade de comparação entre vacas, como seria ou o que você julgaria essencial de ser considerado?

3. Quais atributos de touros você acredita serem relevantes para analisar a qualidade de um touro para compará-lo uns com os outros?

- Quantidade de Filhos
- Produção das Vacas Filhas
- Outros

4. Se pudesse elaborar uma fórmula para qualidade de comparação entre touros, como seria ou o que você julgaria essencial de ser considerado?

B QUESTIONÁRIO SOBRE USABILIDADE DA FERRAMENTA

Nome do Participante

E-mail

Como você avalia a usabilidade geral da ferramenta?

Ruim (1) (2) (3) (4) (5) Boa

Teve facilidade para encontrar as informações buscadas nos menus da ferramenta?

Pouca (1) (2) (3) (4) (5) Muita

As visualizações apresentadas foram suficientes para conseguir analisar os dados e obter novos conhecimentos?

Pouca (1) (2) (3) (4) (5) Muita

O que você alteraria ou adicionaria para melhorar a visualização?

Quais aspectos você avalia como mais positivos na ferramenta?

Tabelas para análise de dados inferidos

Gráficos com dados implícitos para identificar novos conhecimentos

Redes de indivíduos, mostrando as relações entre os atores do domínio

Identificação de indivíduos mais relevantes nos grupos e possibilidade de criar novos agrupamentos

() Outros

O que você gostaria de ressaltar em relação aos itens avaliados?

Ocorreram quaisquer problemas técnicos durante a utilização da ferramenta? Em caso positivo, citar quais foram.

Campo livre para sugestões e considerações

C MAPEAMENTO SISTEMÁTICO SOBRE VISUALIZAÇÃO DE INFORMAÇÃO

Foram conduzidos um mapeamento e uma revisão sistemática da literatura de acordo com as práticas sugeridas por Kitchenham (B.; S., 2007). O objetivo é identificar e avaliar estudos na literatura, considerando questões de pesquisa levantadas a fim de definir os avanços de uma linha de pesquisa específica.

O mapeamento e a revisão sistemática abordados neste artigo, tem como objetivo fazer o levantamento de estudos relacionados à área de visualização de dados.

C.1 QUESTÕES DE PESQUISA

As questões de pesquisa são importantes para a definição de um objetivo principal durante a revisão e o mapeamento, ou seja, alinhar um foco na qual se pretende responder ao fim da análise dos dados.

Para a execução do mapeamento sistemático, pretende-se responder algumas questões relacionadas à área, a fim de descobrir como está o estado da arte com relação à visualização de dados. Para isto, foram definidas as seguintes perguntas:

- **Questão 1:** Quais são os principais autores na área?
- **Questão 2:** Quantos estudos foram publicados ao longo dos anos?
- **Questão 3:** Quais veículos de publicação foram os mais utilizados para assuntos da área?

As perguntas relacionadas ao mapeamento sistemático tem como principal objetivo identificar um contexto mais geral sobre os estudos relacionados à visualização de dados.

O objetivo principal abordado durante a revisão sistemática é: Caracterizar abordagens/técnicas utilizadas para visualização de dados em Engenharia de Software e Banco de Dados. A definição da pesquisa é de acordo com a seguinte frase:

O propósito é caracterizar **o efeito de** técnicas e abordagens de visualização de dados **na** tomada de decisão **do ponto de vista** de gerentes de negócio e tomadores de decisão **no contexto de** Engenharia de Software e Banco de Dados.

A partir do objetivo definido, foram definidas questões de pesquisa que pretendemos conseguir responder após a realização desta revisão sistemática.

- **Questão 1:** Quais abordagens/técnicas são utilizadas para visualização de dados em Engenharia de Software e Banco de Dados?
- **Questão 2:** Como essas abordagens/técnicas influenciam a tomada de decisão de gerentes de negócio e tomadores de decisão?
- **Questão 3:** Qual a influência do domínio ou dos tipos de dados em relação à abordagem/técnica utilizada para visualização de dados?

Neste sentido, cada pergunta tem uma papel bem definido para o entendimento da área pesquisada. A questão 1, tem como principal função a obtenção das técnicas de visualização de dados mais utilizadas em diferentes contextos que envolvam Engenharia de Software e Banco de Dados. Com isso, é possível obter um conjunto de técnicas que são relevantes para a interpretação de dados.

Em seguida, a pergunta 2 inclui mais informações aos dados descobertos na questão 1. A questão 2 analisa, a partir das estratégias de visualização descobertas na primeira questão, como cada técnica influencia na tomada de decisão por parte de gerentes de negócio e tomadores de decisão. Ao obter essa resposta, espera-se que se possa descobrir como uma determinada visualização pode promover uma melhor interpretação dos dados, ou seja, qual o poder da perspectiva durante o auxílio no processo cognitivo de interpretação de textos, números, tabelas.

Por fim, a questão 3 tem como principal objetivo determinar se o tipo dos dados que precisam ser analisados ou domínio dos mesmo influencia na escolha da estratégia de visualização utilizada. A principal função desta pergunta é tentar definir quesitos que possam ser úteis para reconhecer a capacidade de recomendar estas estratégias de visualização para uma melhor descoberta de informações do usuário durante a análise dos dados.

Ao responder todas estas perguntas, pretende-se atingir o objetivo principal proposto. Após a definição das questões de pesquisa, é definido o PICOC (Population, Intervention, Comparison, Outcome, Context) referente à busca. A tabela [C.1](#) ilustra quais foram os termos utilizados.

Tabela C.1: Elementos PICOC

Population	decision maker, manager
Intervention	data visualization, data abstraction
Comparison	-
Outcome	model, technique, approach
Context	software engineering, database

C.2 ESTRATÉGIA DE BUSCA

A estratégia de busca utilizada nesta revisão foi a utilização de termos identificados no PICOC e, com a utilização das palavras-chave, construir uma String de busca e aplicá-la em várias bases de dados de artigos a fim de evitar algum viés durante o processo de busca.

Neste sentido, as palavras chave identificadas foram: "approach", "data abstraction", "data visualization", "decision maker", "manager", "model", "technique". Assim, a busca tentou abranger o maior número de termos relacionados com visualização de dados para apoio à tomada de decisão em análises de dados.

C.2.1 String de Busca

A string de busca é definida de acordo com uma combinação das palavras-chave encontradas. A estrutura básica da busca da String é feita com operadores Booleanos (AND/OR) a fim de construir uma combinação de frases que são fundamentais em estarem contidas nos trabalhos encontrados, com intuito de se obter uma maior aderência ao tema abordado.

Por fim, a String de busca ficou definida da forma à seguir. É importante ressaltar, que o formato da String pode variar um pouco dependendo da base pesquisada, visto algumas particularidades durante a pesquisa em cada uma delas.

("decision maker"OR "manager") AND ("data visualization"OR "data abstraction") AND ("model"OR "technique"OR "approach") AND ("software engineering"OR "database")

Essa String de busca pretende selecionar todos os trabalhos relevantes que envolvam modelos, técnicas e abordagens para visualização ou abstração de dados com intervenção em tomadores de decisão ou gerentes no contexto de banco de dados ou engenharia de software.

C.2.2 Seleção nas Bases

A String de busca foi aplicada em diferentes bases de dados de artigos no período de Agosto de 2016. Não foi imposta nenhuma restrição com relação ao ano das publicações, pois em algumas análises *ad hoc* foram detectados muitos estudos antigos, caracterizando uma área não muito recente, com estudos que avançam desde períodos como 1980. Estes estudos têm avançado muito devido à grande integração de meios tecnológicos que facilitam a captação dos mais diversos dados.

A tabela [C.2](#) ilustra os artigos obtidos após a aplicação da String de busca nas bases mais relevantes. Nesta tabela não se encontra a base IEEE, porém esta também foi utilizada mas obteve apenas 1 artigo como resultado e o mesmo não apresentava relevância, portanto, a base não foi incluída nas análises.

Tabela C.2: Artigos encontrados nas bases de dados

Base de Dados	Quantidade de Artigos	Quantidade de Artigos (sem Duplicados)
Scopus	47	39
Sciente Direct	198	196
Web of Science	49	35
ACM DL	105	105

Totalizando 399 artigos. Ao excluir os artigos duplicados, foi obtido um total de 375 artigos.

C.3 CRITÉRIO PARA SELEÇÃO DE ESTUDOS

O planejamento dos critérios de inclusão e exclusão levam em consideração os principais trabalhos que tem capacidade de responder os questionamentos levantados na parte inicial da revisão sistemática. Os critérios de exclusão servem para descartar, inicialmente, algum artigo que não contenha o mínimo de informação esperada. Os critérios de inclusão fundamentam o tipo de trabalho no qual se espera fazer a leitura, ou seja, para se incluir este trabalho no estudo são levantados quais os temas necessários para que o trabalho seja aceito.

No que concerne aos critérios de inclusão e exclusão, foram levantados tópicos específicos em cada um, para que estes guiassem o próximo passo de seleção dos estudos levantados.

C.3.1 Critérios de Exclusão

Estes critérios tem o foco de remover os estudos que não são pertinentes ou que não possuem conteúdo que aparente ser relevante.

- Artigo Duplicado
- Artigo Secundário
- Artigo Não Disponível
- Artigo que não esteja em português/inglês
- Somente Artigos
- Artigo em que a abstração/visualização de dados não seja o foco

Artigos duplicados não interessam devido ao fato de que já vão ser abordados em alguma outra base. Artigos secundários não interessam nesta fase da revisão pelo motivo de buscar apenas pesquisas primárias realizadas com foco nas estratégias de visualização. Artigos que não estejam disponíveis também são descartados devido à necessidade de se obter o texto completo para leitura completa posteriormente. Neste processo da análise, é fundamental que os artigos estejam na língua nativa (português) ou em inglês. Nesta revisão serão lidos apenas artigos, excluindo qualquer outro tipo de documento, visto que não há tempo hábil para leituras de livros, por exemplo, sendo que serão aceitos apenas artigos em que a abstração/visualização dos dados sejam o foco do artigo.

C.3.2 Critério de Inclusão

Os critérios de inclusão são os principais motivos para a escolha dos artigos durante o processo de seleção. Estes critérios visam listar quais são os focos no qual os estudos se encaixam, sempre buscando responder as questões de pesquisa propostas. Além disso, os critérios de inclusão podem ser visto como motivos para a escolha do estudo, ou seja, quais as características mais importantes que se espera encontrar nos estudos selecionados inicialmente.

- Aplicações de técnicas de visualização de dados para apoio à tomada de decisão
- Comparação entre técnicas de visualização de dados

- Recomendação de técnicas de visualização para domínios particulares
- Utilização de abstrações de dados para representação da informação

Para cada item houve um fator que incentivou a sua escolha como um critério de inclusão, estes são: para o primeiro item, o foco é abranger artigos que apliquem estratégias de visualização dentro de seu contexto, visando uma fundamentação da sua escolha (que é detalhado no processo de avaliação da qualidade mais adiante). Para o item 2, artigos que comparem técnicas de visualização, são relevantes devido ao fato de comparar proposições que podem se destacar mais em relação às outras, visando uma fundamentação por parte da avaliação no momento de se escolher qual abordagem de visualização é mais eficiente para a tomada de decisão e análise de dados. O item 3 corresponde ao critério de analisar estudos onde o autor utiliza alguma espécie de técnica de recomendação para checar a viabilidade de se utilizar técnicas de visualização em domínios específicos. Além disso, buscar respostas sobre qual a influência na escolha para as recomendações de estratégias diferentes, assim como domínio, tipo de dados, recurso tecnológico. Esse item também visa uma fundamentação mais concreta no quesito da avaliação da viabilidade da estratégia de recomendação. Por fim, o último item tem como objetivo obter trabalhos que contenham utilizações de abstração de dados para representar informação, o objetivo principal desta pergunta é a identificação de quais as principais abstrações podem ser utilizadas no momento de conseguir representar estes dados.

C.4 PROCESSO DE SELEÇÃO DOS ESTUDOS

A seleção dos estudos foi dividida em duas etapas: leitura de títulos e abstracts; leitura completa do artigo. Na etapa de leitura de títulos e abstracts, a ideia principal é filtrar artigos que não tenham relação com o que se busca nos estudos. Essa filtragem auxilia em obter apenas os trabalhos que sejam realmente relevantes e estejam alinhados com o que foi definido nos critérios de inclusão. No contexto desta revisão, foram selecionados muitos artigos que apenas apresentavam uso de ferramentas para auxiliar alguma proposta, o que não é o foco desta revisão. Após esta primeira análise foram selecionados 50 artigos, e destes artigos, alguns foram selecionados como duvidosos.

Para uma seleção mais minuciosa, foram lidas conclusões e quais eram os focos dos artigos, com isso, foram removidos e restaram 20 artigos para a segunda etapa.

Na etapa de leitura completa, é preciso avaliar a qualidade dos artigos apresentados, ou seja, se eles possuem todas as características e contribuições esperadas para que o trabalho possa auxiliar em responder as questões de pesquisa propostas. Neste sentido, foram desenvolvidas perguntas de avaliação de qualidade. Estas são essenciais para o controle de qualidade dos artigos, buscando se realmente atendem ao propósito de auxiliar nas respostas das questões de pesquisa.

C.5 AVALIAÇÃO DA QUALIDADE

O estudo da avaliação da qualidade foi modelado de acordo com uma avaliação baseada em 3 escalas: resposta "sim" com peso 1, "parcialmente" com peso 0.5 e "não" com peso 0. Espera-se artigos que contenham, pelo menos, metade dos pontos da avaliação ($6 / 2 = 3$). A escolha se dá ao fato de que as perguntas realizadas são fundamentais para assegurar que o trabalho esteja realmente focado no que se pretende observar.

Tabela C.3: Perguntas para avaliação da Qualidade

Pergunta	Opções		
Realizou algum experimento empírico para validação?	sim	parcialmente	não
Utilizou técnicas de visualização de dados/abstração da informação	sim	parcialmente	não
Fundamenta a escolha das técnicas utilizadas?	sim	parcialmente	não
Explica a motivação para utilização de visualização de dados?	sim	parcialmente	não
Define o problema de maneira clara?	sim	parcialmente	não
Demonstra qual o recurso tecnológico utilizado para visualização?	sim	parcialmente	não

O resultado esperado é de que, pelo menos, as perguntas 2, 4 e 5 sejam respondidas com afirmações (alguma podendo ser parcial), portanto o resultado esperado para a fundamentação do problema é de, aproximadamente, 2 pontos. As questões 1, 3 e 6 são destinadas ao processo de fundamentação de escolhas envolvendo a estratégia de visualização abordada no artigo, preocupação com o contexto do usuário além de realizar avaliações para os resultados encontrados. Espera-se que o autor consiga responder, pelo menos, alguma destas questões de maneira eficiente ou abordar mais de uma, esperando-se 1 ponto como resultado da resposta destas questões.

Ao fim desta análise, foi escolhido que a metade dos pontos é um valor que pode assegurar a qualidade do trabalho para que se consiga obter as respostas desejadas para a pesquisa.

C.6 EXTRAÇÃO DE DADOS E ESTRATÉGIA DE RESUMO

A extração de dados contém quais os atributos mais importantes que se pretendem extrair dos artigos selecionados. Estes atributos de interesse podem ser de vários formatos como por exemplo: datas, números inteiros ou reais, Strings, entre outros.

Nesta revisão, os principais dados dos artigos selecionados serão: nome da estratégia de visualização utilizada, o domínio e o tipo dos dados. Todos os campos são do tipo String, ou seja, apenas serão armazenados os respectivos nomes.

C.7 CONDUÇÃO DA REVISÃO SISTEMÁTICA

Nesta seção é descrito como foi conduzido o processo de revisão e mapeamento, levando em conta a quantidade de artigos que foram selecionados e os filtros aplicados nos mesmos durante o processo de seleção.

Durante o mapeamento, foram selecionados alguns gráficos relevantes para a análise dos artigos. Estes gráficos são exibidos abaixo.

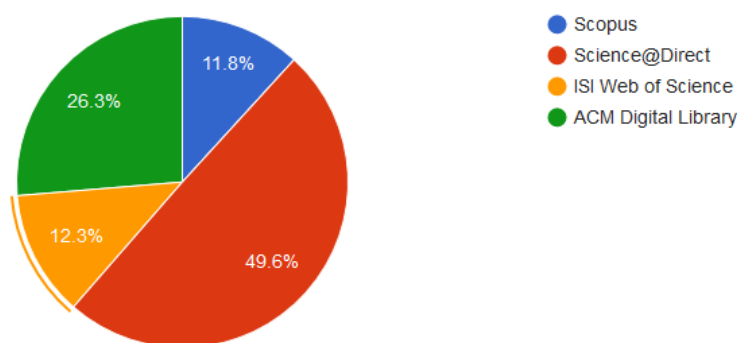


Figura C.1: Quantidade de Artigos por Base de Dados

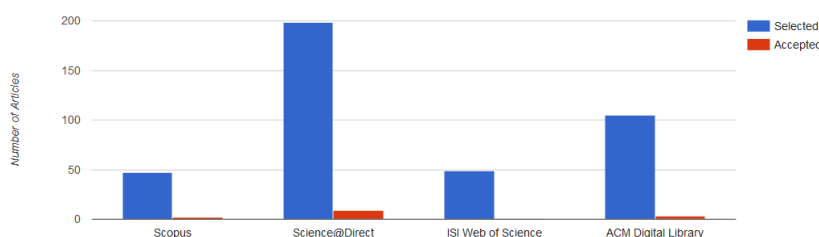


Figura C.2: Relação entre Selecionados e Escolhidos por Base de Dados

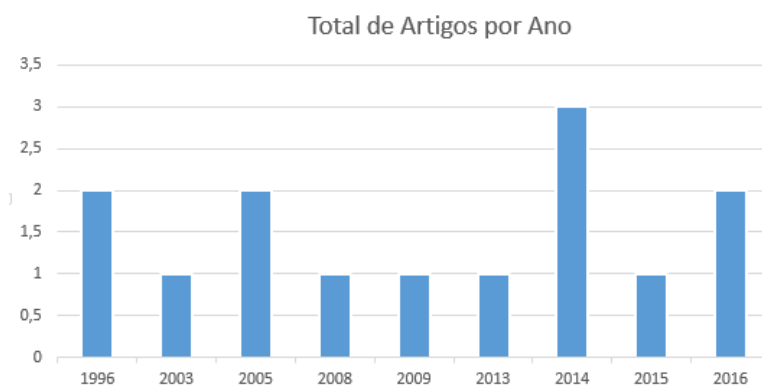


Figura C.3: Relação de artigos por ano (antes do Quality Assessment)

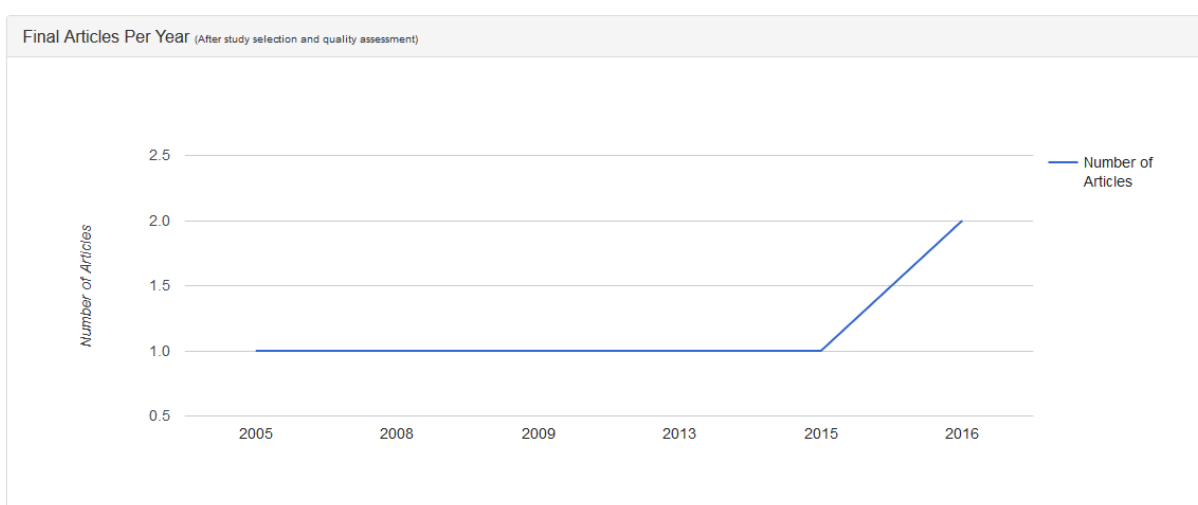


Figura C.4: Relação de artigos por ano (após o Quality Assessment)

Os principais journals encontrados durante o mapeamento são: *Automation in Construction* e *Decision Support Systems*, ambos com 2 artigos publicados.

Com relação aos autores, não foi encontrado nenhum que tenha se destacado com relação à possuir mais publicações, apenas Hyunwoo Park que teve 2 artigos selecionados após a leitura completa dos artigos.

Por fim, o total de artigo que tiveram uma qualidade adequada foram 7. Estes são: (THULKE et al., 2005), (RUSSELL et al., 2009), (PARK et al., 2016), (GRIERSON et al., 2015), (CHENG; TEIZER, 2013), (SAVOSKA et al., 2008), (PARK; BASOLE, 2016).

C.7.1 Resposta Q1: Técnicas de visualização utilizadas

Através da leitura dos artigos, foi descoberto que existem vários tipos de visualização, além da grande quantidade de estratégias novas implementadas. Muitas técnicas novas

Tabela C.4: Condução da Revisão Sistemática

Filtro	Total de Artigos
Artigos selecionados	399
Leitura de títulos e abstracts	50
Artigos Indisponíveis e leitura de conclusão	14
Excluídos nos critérios de avaliação	7

são desenvolvidas a fim de atender as necessidades crescentes no contexto da tomada de decisão.

Muitos autores citam as referências de técnicas mais básicas, como gráficos 2D. Dentre estes gráficos, podemos observar os gráficos de barras, gráficos de pizza, plotagem, entre outros. Os estudos mostraram que as técnicas de visualização podem variar, desde representações mais básicas até representações mais avançadas, construídas com a composição de 1 ou mais técnicas, ampliação das dimensões.

Grande parte dos artigos cita a utilização de técnicas 3D como uma melhoria na exibição da informação, além disso, as técnicas de visualização não são utilizadas sozinhas. Estão associadas às estratégias outros princípios que complementam suas funcionalidades.

Duas questões debatidas de maior interesse foram: a utilização de *Dashboards* como meio de representação de múltiplas visualizações na tela e a utilização de técnicas de interação do usuário com a visualização.

No que concerne a utilização de *Dashboards*, há indícios de que sua utilização pode ser importante para a tomada de decisões, pois sua interface pode ser facilmente customizável, auxiliando o processo de interação com o usuário além de conseguir pode exibir diversas técnicas ao mesmo tempo.

No sentido de técnicas de interação, muitas são debatidas durante os artigos. Algumas das técnicas citadas são: zoom, mudança de visualizações, filtrar dados, entre outras. Estas técnicas facilitam no quesito de amplificar a perspectiva do usuário, possibilitando-o ter visões diferentes de um mesmo conjunto de dados a fim de maximizar sua interpretação.

C.7.2 Resposta Q2: Como técnicas de visualização influenciam a tomada de decisão

É possível notar no artigo (THULKE et al., 2005) que uma visualização, quando utilizada de forma errada pode induzir a formulação de uma hipótese falsa. Neste artigo, em particular, o autor chega à conclusão que a utilização de visualizações em 2D omitem informações que são exibidas nas estratégias que contém mais de duas dimensões.

Além disso, todos os autores afirmam que as técnicas de visualização realmente auxiliam no processo cognitivo da descoberta de novas informações. Visto que abstrações visuais são interpretadas de maneiras mais simples em relação à conjuntos de textos, números ou tabelas. Além disso, é uma área pesquisada há tempos, porém com os avanços tecnológicos atuais e o crescimento no volume de dados gerados nos mais diversos domínios, a visualização de dados têm crescido o seu papel no sentido de prover ferramentas para o auxílio na descoberta de informações em dados complexos.

As visualizações podem influenciar positivamente ou negativamente, neste sentido, é importante a conclusão da resposta para a questão 3, que define uma abordagem mais voltada para o contexto em que a visualização é aplicada.

C.7.3 Resposta Q3: Qual a influência do domínio na escolha da estratégia de visualização

A resposta desta questão é muito abrangente, mas de acordo com a leitura nos estudos selecionados, foi possível perceber que muitos domínios podem influenciar na escolha de uma melhor visualização. os artigos (PARK; BASOLE, 2016), (PARK et al., 2016), (RUSSELL et al., 2009), (CHENG; TEIZER, 2013) e (THULKE et al., 2005) reforçam muito isso. Pois nestes artigos, a escolha da estratégia de visualização pode melhorar ou piorar a perspectiva por parte do usuário.

Por exemplo, nos dados de construção, são relacionadas estratégias que utilizem uma maior capacidade de representação e reconhecimento do ambiente, (CHENG; TEIZER, 2013) mostra em seu trabalho uma abordagem que utiliza realidade virtual na representação dos dados, a grande vantagem destas representações é a capacidade de amplificar a percepção do usuário em perspectivas diferentes.

Alguns dos trabalhos observados motivaram à chegar a conclusão em domínios espe-

cíficos, é válido caracterizar um próximo estudo buscando observar este comportamento em um domínio mais específico, ou seja, fechar mais a abrangência da String de busca.

C.8 AMEAÇAS À VALIDADE

Ameaças à validade são fatores que podem influenciar a revisão sistemática de forma a causar algum viés durante a escolha dos estudos para análise. Não conhecer as bases mais influentes no domínio de visualização de dados pode influenciar com relação aos artigos escolhidos, a tentativa de mitigar esta ameaça foi a pesquisa nas principais bases, visando selecionar o maior número de artigos relevantes possível.

A String de busca pode ter problemas por faltarem alguns termos relacionados à visualização de dados que foram identificados após o fim da revisão sistemática. Estes novos termos devem ser inseridos em uma nova revisão sistemática a fim de complementar a atual.

C.9 CONSIDERAÇÕES FINAIS

Foram realizados um mapeamento e uma revisão sistemática no sentido de caracterizar técnicas de visualização de dados, assim como sua influência na tomada de decisão de gerentes e tomadores de decisão. A revisão promoveu conhecimentos para gerar indícios de respostas para algumas perguntas de pesquisa gerada a partir deste objetivo principal.

As principais descobertas referem-se à capacidade de técnicas de visualização auxiliarem no desempenho do processo cognitivo em análise de grandes volumes de dados. Além disso, muitas técnicas são desenvolvidas ao longo dos anos, algumas novas e outras mais clássicas. Algumas abordagens novas se mostram mais efetivas do que as abordagens tradicionais observadas.

Outra observação relevante é a falta de preocupação com o recurso tecnológico do usuário. Maioria dos trabalhos não se preocupam com questões de resolução de tela no momento de apresentar a estratégia de visualização pro usuário, o que pode ser um problema no momento de realizar a análise de determinados tipos de gráficos. Muitos artigos usam a abordagem de utilização de *Data Warehouses* combinada com *Dashboards* para a plotagem de dados na tela.

A visualização de dados é uma área de estudo crescente. Muitos autores têm dado

atenção nesse sentido buscando melhorias na qualidade da representação das informações de uma maneira que exija o mínimo de esforço possível por parte do gerente ou tomador de decisão.