JOSÉ GOMES NETO

**Modelo Navegacional Dinâmico, para implementação da
integração inter-estrutural de dados**

SÃO PAULO

2016

JOSÉ GOMES NETO

**Modelo Navegacional Dinâmico, para implementação da
integração inter-estrutural de dados**

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Ciências

Orientador: Prof. Dr. Jorge Rady de
Almeida Junior

SÃO PAULO

2016

JOSÉ GOMES NETO

**Modelo Navegacional Dinâmico, para implementação da
integração inter-estrutural de dados**

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Ciências

Área de Concentração: Engenharia de
Computação

Orientador: Prof. Dr. Jorge Rady de
Almeida Junior

SÃO PAULO

2016

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Gomes Neto, José

Modelo Navegacional Dinâmico, para implementação da integração inter estrutural de dados / J. Gomes Neto, J. R. de Almeida Junior -- versão corr. -- São Paulo, 2016.

146 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Banco de dados 2.Modelagem de dados 3.Modelos de dados pós relacionais 4.Integração inter-estrutural de dados I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t. III.de Almeida Junior, Jorge Rady

DEDICATÓRIA

À minha amada mãe, Leonor Eid,
pela determinação, caráter,
empenho, dedicação em minha
formação e com quem, embora
pudesse aprender muito mais,
sempre muito aprendo.

AGRADECIMENTOS

Ao Prof. Livre Docente Jorge Rady de Almeida Junior, orientador dessa Tese, pela paciência, paciência, paciência, paciência, competência e valiosa orientação prestadas ao desenvolvimento deste trabalho. Obrigado, Dr. !

À POLI/USP, por permitir a conclusão deste trabalho e ao PCS, por apoiar o desenvolvimento acadêmico de seus colaboradores.

Ao Centro Universitário da Fundação Educacional Inaciana, minha amada FEI, na pessoa do seu representante maior, Pe. Peters e seu reitor, Prof. Dr. Fábio do Prado, pelas coerentes cobranças por excelência em minha titulação acadêmica.

Ao sempre colega e líder, Prof. Dr. Flávio Tonidandel, ex-LAC, pela confiança.

Aos colegas do PCS e aos funcionários do xerox da POLI-PRODUÇÃO, pelas informações fornecidas e sobretudo, pelos sorrisos sempre oportunos.

Aos colegas do IPEN, em especial ao Prof. Livre Docente Delvonei Alves de Andrade, excepcional chefe de pesquisas, pela prontidão em me atender.

Aos colegas da FEI, em especial aos professores Taylor Mac Intyer Fonseca Jr, pela contagiante paixão à ciência dos materiais, Gilberto Ieno, pela humildade irreverente de sempre e a Sra. Maria das Graças de Paula, pela perene, sincera e pragmática simpatia (*Graça...*).

Ao Sr. Romildo Savassa, pelo incentivo em épocas pretéritas, onde minha fome quase venceu minha vontade de ser Engenheiro. Corinthiano, não deu tempo do Sr. assistir a defesa dessa Tese (15/Jun/2k16)...Descanse em paz, amigo.
(*Major...*)

À Professora Dra. Sônia Geraij Mokarzel, fantástica professora de Mecânica Quântica, Doutora incansável mesmo nos momentos mais difíceis e que sempre me incitou a estudar um pouco mais do que eu, preguiçosamente, queria. Sábado cedo não, certo Professora ? Salim...sábado cedo (*Está claro isso?*).

À meu ex-professor João Usberco, extraordinário professor de Química do Anglo Tamandaré, minha eterna referência na arte de lecionar (*Calma, não se empolga...*).

Ao professor Álvaro Puga Paz, notável matemático, com quem a cada contato muito aprendo, inclusive matemática (*Tudo ótemo...*).

À minha saudosa tia Aida Eid, pela convicção inconsciente de simplicidade, honradêz e honestidade. Já se vão 14 anos... (*Alemão c. de pão*).

À meu saudoso, notável, incomparável e inesquecível tio, Farid Eid, de quem com muito orgulho herdei o pseudônimo e que, mesmo após mais de três décadas da sua partida, se faz sempre presente em minhas atitudes. Saudade infinita, tio...um dia, entre colunas, tudo justo e perfeito(.:) Mavericks, tio? (...*tende...*).

À quase todos meus alunos e ex-alunos, pela inocência e incentivo, velados.

À aqueles que, direta ou indiretamente, contribuíram para que este trabalho não se realizasse. Continuem tentando, por favor !

A Deus e todos seus emissários, por tudo, em especial a Srta. Terezinha de Jesus Barra, amiguinha de um grande chefe indígena, pelos fortes ventos do noroeste Paulista, que sopram nas madrugadas da misteriosa represa do Broa !

*Não há arrastamento irresistível, quando
se tem a vontade de resistir*

Allan Kardec

RESUMO

Na última década, observaram-se substanciais mudanças nos tipos de dados processados, quando comparados à definição convencional de dados estruturados. Neste contexto, sistemas computacionais que em sua maioria acessam bases de dados convencionais, centralizadas, que armazenam dados estruturados, necessitam cada vez mais acessarem e processarem também dados não estruturados, distribuídos e em grandes quantidades. Fatores tais como versatilidade em abrigar dados não estruturados, coexistência, integração e difusão de dados complexos a velocidades superiores as velocidades até então observadas, restringem, em determinadas situações, o uso dos modelos de dados convencionais. Dessa forma, nesta Tese é proposto e formalizado um modelo de dados pós relacional, baseado nos conceitos de grafos complexos, também denominados, Redes Complexas. Por intermédio da utilização do modelo de grafos, define-se uma forma de se implementar uma integração inter-estrutural de dados, ou seja, os tradicionais dados estruturados, com os mais recentemente utilizados dados não estruturados, tais como os dados multimídia. Tal integração envolve todas as transações presentes em um banco de dados, ou seja, consulta, inserção, atualização e exclusão de dados. A denominação dada a tal forma de trabalho e implementação foi Modelo Navegacional Dinâmico – MND. Esse modelo representa diferentes estruturas de dados e sobretudo, permite que essas diferentes estruturas coexistam de forma integrada, agregando à informação resultante maior completeza e abrangência. Portanto, o MND associa os benefícios da junção da estrutura das Redes Complexas ao contexto de dados não estruturados, sobretudo no que tange à integração resultante de dados com estruturas distintas, conferindo assim às aplicações que necessitam desta integração, melhoria no aproveitamento dos recursos.

Palavras-chave: Dados não estruturados. Modelos de Dados. Redes Complexas. Integração de dados.

ABSTRACT

Over the last decade several changes in data processing have been observed when compared to the conventional structured data definition. In such context, computational systems accessing centralized databases need to process large, distributed, non-structured data as well. Factors like versatility in hosting data, coexistence, integration and diffusion of such complex data at high speeds can be, in some cases, troublesome when using conventional data models. In this work a post-relational, graph-based (also known as Complex Network) model, is presented. Such model enables the integration of both structured data and non-structured data, such as multimedia, allowing such structures to coexist. This integration involves all transactions found in a database, such as select, insert, delete and update data. The name given to this form of work and implementation was Navigational Model Dynamic – MND. This model represents different data structures and above all, allows these different structures to coexist in an integrated way, adding to the resulting information greater completeness and comprehensiveness. Hence, MND harnesses the benefits of Complex Network and non-structured data providing all relational data handling already available in other databases but also integration and better use of resources.

Key-words: Unstructured Data.Data Models.Complex Networks.Data Integration.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – MODELO PARA DESENVOLVIMENTO DE UM MÉTODO	20
FIGURA 2 – FUNDAMENTAÇÕES CONCEITUAIS PARA A ELABORAÇÃO DAS HIPÓTESES.....	22
FIGURA 3 – MODELO ADAPTADO PARA DESENVOLVIMENTO DE UM MÉTODO.....	23
FIGURA 4 – ESTRUTURA DO MODELO HIERÁRQUICO DE DADOS.....	26
FIGURA 5 – ESTRUTURA DO MODELO DE DADOS EM REDE.....	27
FIGURA 6 – ESTRUTURA BÁSICA DO MODELO DE DADOS RELACIONAL	29
FIGURA 7 – ESTRUTURA BÁSICA DO MODELO DE DADOS ORIENTADO A OBJETOS	30
FIGURA 8 – ESTRUTURA DO MODELO DE DADOS <i>NoSQL</i> , TIPO CHAVE-VALOR.....	31
FIGURA 9 – EXEMPLO DE APLICAÇÃO DO MODELO DE DADOS <i>NoSQL</i> , CHAVE-VALOR	32
FIGURA 10 – ESTRUTURA DO MODELO DE DADOS <i>NoSQL</i> , ORIENTADO A COLUNAS	33
FIGURA 11 – EXEMPLO DE APLICAÇÃO DO MODELO <i>NoSQL</i> , ORIENTADO A COLUNAS.....	33
FIGURA 12 – ESTRUTURA DO MODELO DE DADOS <i>NoSQL</i> , ORIENTADO A DOCUMENTOS.....	34
FIGURA 13 – EXEMPLO DE APLICAÇÃO DO MODELO DE DADOS <i>NoSQL</i> , ORIENTADO A DOCUMENTOS.....	35
FIGURA 14 – ESTRUTURA DO MODELO DE DADOS <i>NoSQL</i> , ORIENTADO A GRAFOS.....	36
FIGURA 15 – EXEMPLO DE APLICAÇÃO DO MODELO DE DADOS <i>NoSQL</i> , ORIENTADO A GRAFOS	36
FIGURA 16 – COMPARATIVO DOS MODELO DE DADOS <i>NoSQL</i>	39
FIGURA 17 – TEOREMA <i>CAP</i>	43
FIGURA 18 – ETAPAS DA PROPOSTA DE LIU.....	47
FIGURA 19 – TÉCNICA <i>SCORE</i>	48
FIGURA 20 – REPOSITÓRIO INCREMENTAL DE TARI	49
FIGURA 21 – MODELO DE DADOS UDMS	50
FIGURA 22 – MODELO DE DADOS AUDR	51
FIGURA 23 – REDE COMPLEXA DOS PRODUTOS COMUNITÁRIOS COMERCIALIZÁVEIS	54
FIGURA 24 – EXEMPLO DE GRAFOS SIMPLES	56
FIGURA 25 – EXEMPLO DE GRAFO COMPLEXO	56
FIGURA 26 – EXEMPLO DE MODELO DE REDE COMPLEXA ALEATÓRIA	60
FIGURA 27 – EXEMPLO DE REDE COMPLEXA DO TIPO <i>PEQUENO-MUNDO</i>	62
FIGURA 28 – EVOLUÇÃO DE REDES COMPLEXAS DO TIPO LIVRES DE ESCALA.....	64
FIGURA 29 – CONDENSADO DE <i>BOSE-EINSTEIN</i>	65
FIGURA 30 – REDE REPRESENTATIVA DO POEMA DE CARLOS DRUMOND DE ANDRADE	66
FIGURA 31 – GRAFO CONTENDO 8 NÓS DO TIPO <i>LABEL</i> E 14 ARESTAS	81
FIGURA 32 – MATRIZ DE ADJACÊNCIAS <i>AL</i>	82
FIGURA 33 – MATRIZ DE INCIDÊNCIAS <i>IL</i>	83
FIGURA 34 – MATRIZ DE PESOS <i>PL</i>	85
FIGURA 35 – MATRIZ DE ESTRUTURAS <i>EL</i>	86
FIGURA 36 – EXEMPLO DE GRAFO CONTENDO OS VALORES DE <i>SD</i> E <i>CC</i> 'S.....	89
FIGURA 37 – DIAGRAMA INTERATIVO DO MODELO NAVEGACIONAL DINÂMICO	90
FIGURA 38 – COMPOSIÇÃO DO MODELO NAVEGACIONAL DINÂMICO	91
FIGURA 39 – A INTEGRAÇÃO INTER- ESTRUTURAL	95
FIGURA 40 – MODELO APOIADO EM GRAFOS, ACESSADO PELA LINGUAGEM <i>CYPHER</i>	97
FIGURA 41 – MODELAGEM DA APLICAÇÃO <i>Hc</i>	99

FIGURA 42 – NÓ <i>LABEL PESQUISADOR</i> , SUAS RELAÇÕES E CONTEÚDOS	100
FIGURA 43 – NÓ <i>LABEL PUBLICAÇÕES</i> , SUAS RELAÇÕES E CONTEÚDOS	101
FIGURA 44 – NÓ <i>LABEL FUNCIONÁRIOS</i> , SUAS RELAÇÕES E CONTEÚDOS	101
FIGURA 45 – MODELO APOIADO EM GRAFOS, CONTENDO TAMBÉM DADOS NÃO ESTRUTURADOS	102
FIGURA 46 – NÓ <i>PUBLICAÇÕES</i> MODIFICADO, SUAS RELAÇÕES E CONTEÚDOS	103
FIGURA 47 – REPRESENTAÇÃO DO CONCEITO DA INTEGRAÇÃO INTER- ESTRUTURAL	104
FIGURA 48 – MATRIZ DE ADJACÊNCIAS <i>AL</i> , DO GRAFO APRESENTADO NA FIGURA 47	104
FIGURA 49 – MATRIZ DE INCIDÊNCIAS <i>IL</i> , DO GRAFO APRESENTADO NA FIGURA 47	105
FIGURA 50 – MATRIZ DE PESOS <i>PL</i> , DO GRAFO APRESENTADO NA FIGURA 47	105
FIGURA 51 – MATRIZ DE ESTRUTURAS <i>EL</i> , DO GRAFO APRESENTADO NA FIGURA 47	106
FIGURA 52 – TELA DE <i>LOGIN</i> DO SISTEMA <i>SACA</i>	108
FIGURA 53 – TELA <i>SACA-ALUNO</i>	108
FIGURA 54 – FOTOGRAFIA INVÁLIDA À INSCRIÇÃO EM 9/1/2016	110
FIGURA 55 – CONSULTA <i>Q4</i>	111
FIGURA 56 – TRECHO DO MODELO DE DADOS, ACESSADO PELA CONSULTA <i>Q4</i>	112
FIGURA 57 – EXEMPLO DE CONTEÚDO ARMAZENADO NO GRAFO DA FIGURA 56	112
FIGURA 58 – MATRIZ DE ADJACÊNCIAS	113
FIGURA 59 – MATRIZ DE INCIDÊNCIAS	113
FIGURA 60 – MATRIZ DE PESOS	114
FIGURA 61 – MATRIZ DE ESTRUTURAS	114
FIGURA 62 – ARQUIVO <i>FOTO1.JPEG</i>	116
FIGURA 63 – ARQUIVO <i>FOTO2.TIFF</i>	116
FIGURA 64 – ARQUITETURA HÍBRIDA DE INTEGRAÇÃO E-R – GRAFOS	121
FIGURA 65 – O <i>MND</i> NO CONTEXTO DOS MODELOS DE DADOS <i>NoSQL</i>	126

LISTA DE TABELAS

TABELA 1 – OS MODELOS DE DADOS E SUAS DEMANDAS EVOLUTIVAS.....	37
TABELA 2 – COMPARATIVO DO MODELO DE DADOS RELACIONAL COM OS MODELOS DE DADOS <i>NoSQL</i>	38

SUMÁRIO

	página
1 INTRODUÇÃO	16
1.1 MOTIVAÇÃO E JUSTIFICATIVA	17
1.2 OBJETIVO	18
1.3 METODOLOGIA DE PESQUISA	19
1.4 ETAPAS DESTA TESE	22
1.5 ORGANIZAÇÃO DO TRABALHO	24
2 OS MODELOS DE DADOS	25
2.1 O MODELO DE DADOS HIERÁRQUICO	26
2.2 O MODELO DE DADOS EM REDES	27
2.3 O MODELO DE DADOS RELACIONAL	28
2.4 O MODELO DE DADOS ORIENTADO A OBJETOS	29
2.5 OS MODELOS DE DADOS NOSQL	30
2.5.1 Modelo de dados do tipo chave/valor	31
2.5.2 Modelo de dados do tipo colunas	32
2.5.3 Modelo de dados do tipo orientado a documentos	34
2.5.4 Modelo de dados do tipo orientado a grafos	35
2.6 COMPARATIVO ENTRE OS MODELOS DE DADOS	37
2.7 A CONSISTÊNCIA EVENTUAL	42
3 DADOS NÃO ESTRUTURADOS	46
3.1 DADOS NÃO ESTRUTURADOS - LINHAS DE PESQUISA	46
3.1.1 Reestruturação das linguagens de consulta	47
3.1.2 Criação de novos sistemas gerenciadores de bancos de dados	49
3.1.3 Desenvolvimento de novos modelos de dados	50
3.1.4 Métodos propostos por fabricantes de gerenciadores relacionais	51
3.2 COMPARATIVO ENTRE AS LINHAS DE PESQUISA	52
4 REDES COMPLEXAS	53
4.1 REDES COMPLEXAS – HISTÓRICO	55
4.2 O PARADIGMA DAS REDES COMPLEXAS	57
4.3 GRANDEZAS ESTATÍSTICAS QUE CARACTERIZAM AS REDES COMPLEXAS	58
4.4 MODELOS DE REDES COMPLEXAS	59
4.4.1 Redes Aleatórias	59
4.4.2 Redes Pequeno-mundo	61
4.4.3 Redes Livres de Escala	63

4.5 APLICAÇÕES DAS REDES COMPLEXAS.....	64
4.5.1 Redes sociais	67
4.5.2 Redes de informação	67
4.5.3 Redes tecnológicas	68
4.5.4 Redes biológicas	68
5 APLICAÇÕES RELACIONADAS.....	69
5.1 WEB	69
5.2 BIG DATA	71
5.3 CROSS-MEDIA	73
5.4 INTERNET DAS COISAS	74
5.5 SUBSÍDIOS FORNECIDOS PELA REVISÃO	75
6 MODELO NAVEGACIONAL DINÂMICO, VISANDO A IMPLEMENTAÇÃO DA INTEGRAÇÃO INTER-ESTRUTURAL DE DADOS	78
6.1 VISÃO GERAL.....	78
6.2 PREMISSAS DO MODELO NAVEGACIONAL DINÂMICO.....	80
6.3 AS GRANDEZAS ESTATÍSTICAS DO MODELO NAVEGACIONAL DINÂMICO	80
6.3.1 Grau de entrada e saída.....	82
6.3.2 Distribuição de Graus	86
6.3.3 Coeficiente de Clusterização.....	88
6.4 O MODELO NAVEGACIONAL DINÂMICO	90
6.5 ACESSO AO MODELO NAVEGACIONAL DINÂMICO	96
6.6 PRIMEIRO EXEMPLO DE APLICAÇÃO DO MODELO NAVEGACIONAL DINÂMICO	99
6.7 SEGUNDO EXEMPLO DE APLICAÇÃO DO MODELO NAVEGACIONAL DINÂMICO	106
6.7.1 Supostas Novas demandas, atualmente não suportadas pelo sistema SACA	109
6.8 TRABALHOS RELACIONADOS AO TEMA PROPOSTO	117
6.8.1 A proposta de Mirzasoleiman.....	118
6.8.2 A proposta de Boguá.....	118
6.8.3 A proposta de Gómez.....	119
6.8.4 A proposta de Lombardo.....	119
6.8.5 A proposta de Jiang.....	120
6.8.6 A proposta de Yafooz.....	120
6.8.7 A proposta de Cavoto.....	121
6.8.8 Comparação com Produtos Comerciais.....	122
6.8.9 Análise das propostas	124

6.9 VANTAGENS / LIMITAÇÕES DO MODELO NAVEGACIONAL DINÂMICO	125
7 CONSIDERAÇÕES FINAIS.....	127
7.1 CONTRIBUIÇÕES E CONCLUSÕES	127
7.2 TRABALHOS FUTUROS	128
REFERÊNCIAS	130
APÊNDICE A – DIAGRAMA ENTIDADE - RELACIONAMENTO SACA.....	139
APÊNDICE B – MODELO NAVEGACIONAL SACA	140
APÊNDICE C – MODELO NAVEGACIONAL SACA MODIFICADO	141
APÊNDICE D – NÓS <i>LABELS</i> DO MODELO SACA MODIFICADO	142

1 INTRODUÇÃO

Analizando a evolução dos sistemas computacionais, verifica-se a dependência desses sistemas com os modelos de dados que os apoiam.

Esses modelos de dados, cada qual com suas respectivas peculiaridades, apresentam, através dos sistemas computacionais que os acessam, os dados, segundo as regras neles modeladas.

Observa-se então, a partir do surgimento de uma nova estrutura, ligada ao dado dito não estruturado, aumento significativo da interatividade entre usuários e sistemas computacionais que acessam dados complexos (dados estruturados e dados não estruturados).

Como algumas evidências do aumento da interatividade entre usuários e sistemas que acessam dados, estruturados ou não, tem-se a indústria dos jogos eletrônicos *on-line* superando, em faturamento, a indústria cinematográfica (TECMUNDO, 2011) e a legislação, considerando crimes digitais (SICA, L., 2013).

Dispositivos móveis e portáteis passam acessar e prover acesso à informação, conferindo às recentes redes sociais, adeptos em quantidades até então inimagináveis (ALVARENGA, P., RAMOS, A., 2004).

O uso do conteúdo *on-line*, disponível em grandes volumes de dados, somado à necessidade de difusão da nova estrutura de dados a altas velocidades, assim como sua integração com a estrutura anterior, demandam ajustes nos modelos de dados. Modelos de dados denominados *NoSQL*, concebidos majoritariamente para o processamento de dados não estruturados, surgem como complemento ao modelo de dados relacional e com estes modelos *NoSQL*, a quebra do paradigma escalabilidade horizontal, ou seja, descentralização do processamento em prol do armazenamento e disponibilidade, de grandes volumes de dados (BREWER, E.A., 2012).

O estudo aprofundado dos formatos de dados estruturados e não estruturados, é tarefa complexa e desafiadora. Neste sentido, um modelo de dados onde estes dois formatos interajam, fornece importante subsídio para o desenvolvimento de sistemas computacionais (BROWNE, J., 2016).

1.1 MOTIVAÇÃO E JUSTIFICATIVA

Ao se observar mais e mais avanços tecnológicos, é quase inevitável se estabelecer ligações entre as demandas e a necessária evolução das tecnologias existentes para suportá-las. Recentemente analisando, por exemplo, o crescimento das redes sociais, percebe-se a dependência do crescimento destas redes com a infra-estrutura que as apoia.

Redes sociais por si só, são um tipo de inovação tecnológica permitida por outra inovação tecnológica, a *Web*, por sua vez suportada pela evolução dos modelos de dados, e assim por diante.

Agregando a este contexto outros importantes conceitos tais como a Internet das Coisas, (*Internet of Things, IoT*), *Cross-Media*, Sistemas em Tempo Real (STR), sistemas para aparelhos móveis, geo-processamento, computação em nuvens, dispositivos sem fio, monitorações interativas, *Big Data* etc, surge a preocupação da coexistência e compatibilidade do dado contido nestes conceitos, com sua respectiva integridade (HUNG, S.Y., KU, C.Y., CHANG, C.M., 2003)

Particularmente no âmbito dos dados não estruturados, a integridade é, em alguns casos, associada às regras do modelo de dados relacional, modelo este utilizado em sistemas computacionais em praticamente todas as áreas do conhecimento humano.

Reconhecendo então a informação como um bem precioso, cujo valor é associado diretamente à precisão e disponibilidade dos dados a ela associados, garantir que novos modelos de dados contemplem novas estruturas de dados e que estes novos modelos coexistam com modelos de dados já existentes, motiva a princípio, a elaboração desta Tese.

A crescente demanda por sistemas computacionais confiáveis, previsíveis, de grande capacidade de armazenamento/processamento de dados e que demandam cada vez mais requisitos temporais rígidos, tem se tornado um desafio comum.

Hoje, embora os modelos de dados permitam a coexistência de dados estruturados e não estruturados, não contemplam a integração entre eles nem tampouco, os armazenam em Redes Complexas.

Estes problemas, não só impedem melhoria na qualidade da informação como um todo, como também interferem substancialmente na velocidade de difusão dos contextos que envolvem estruturas distintas.

Um modelo que considere a integração de estruturas distintas, armazenando-as em Redes Complexas, justifica a elaboração desta Tese.

1.2 OBJETIVO

Esta Tese objetiva propor um modelo de dados pós relacional, baseado nos conceitos de grafos complexos, também denominados, Redes Complexas (SOLOMONO, R.; RAPOPORT, A., 1951), (ERDÖS, R., RÉNYI, A, 1959).

O objetivo central deste trabalho é definir, por meio da utilização do modelo de grafos, uma forma de se implementar uma integração inter-estrutural de dados, ou seja, os tradicionais dados estruturados, com os mais recentemente utilizados dados não estruturados (por exemplo, os dados multimídia). Tal integração envolve todas as transações presentes em um banco de dados, ou seja, consulta, inserção, atualização e exclusão de dados. A denominação dada a tal forma de trabalho e implementação foi Modelo Navegacional Dinâmico - MND.

Esse modelo representa diferentes estruturas de dados e sobretudo, permite que essas diferentes estruturas coexistam de forma integrada, agregando à informação resultante maior completeza e abrangência.

Considerando que no modelo relacional, a integridade referencial é assegurada por restrições e a integridade transacional por propriedades *ACID* (Atomicidade, Consistência, Isolamento e Durabilidade) (CODD, E.F., 1970), no Modelo Navegacional Dinâmico, as integridades referencial e transacional não são consideradas, e é acrescentado, sob consistência eventual (regras *BASE*), o conceito da integração inter-estrutural de dados.

Em síntese, o Modelo Navegacional Dinâmico não se utiliza das propriedades *ACID* como o modelo relacional e os demais modelos de dados apoiados em grafos, mas sim, de consistência eventual, para armazenar e integrar em tempo de execução e em um modelo pós relacional único, dados com diferentes estruturas, baseando-se para isso, em matrizes construídas pelo próprio Modelo Navegacional Dinâmico.

De forma a se compatibilizar a utilização dos modelos de grafos, no Modelo Navegacional Dinâmico, outro objetivo é o de adaptar as matrizes de representação dos modelos de grafos, para uso, por parte do Modelo Navegacional Dinâmico.

De forma a tornar clara a forma de trabalho do Modelo Navegacional Dinâmico, há o objetivo final de descrever sua utilização, por meio de dois estudos de caso.

1.3 METODOLOGIA DE PESQUISA

Metodologias são conjuntos de métodos que fornecem subsídios para que se atinja os objetivos esperados, através da abstração da realidade.

Segundo (YIN, R.K., 2010), cada diretriz direciona a pesquisa para determinados aspectos do trabalho, e estes aspectos devem ser examinados dentro do contexto pesquisado. A metodologia apresenta a definição dos métodos e a forma de pesquisa utilizada.

Diferentemente de (YIN, R.K., 2010), (OLIVEIRA, S.L., 2012) entende método como a forma de pensar para que se chegue a natureza de um determinado problema, quer para estudá-lo, quer para explicá-lo.

Neste trabalho, o método utilizado segue a linha de (YIN, R.K., 2010), apresentado na Figura 1. Segundo (YIN, R.K., 2010), um projeto de pesquisa está diretamente ligado à lógica que conecta diretamente os dados a serem coletados com as premissas iniciais do trabalho. Para (YIN, R.K., 2010), existe um dinamismo contínuo entre as partes a considerar na pesquisa.

Figura 1 – Modelo para desenvolvimento de um método

Fonte: Yin (2010)

A seguir, apresentam-se as definições dos elementos da Figura 1, segundo Yin:

- Plano: Identifica as questões ligadas a pesquisa, que justifiquem o estudo. Nesta fase, decide-se entre o estudo de caso ou outras formas de pesquisa.
- Projeto em avaliação: Esta etapa orienta o pesquisador na coleta, análise e interpretação dos dados.
- Preparação: Esta fase necessita que o pesquisador desenvolva habilidades para que o estudo de caso seja conduzido.
- Coleta: Evidências mencionadas anteriormente têm várias fontes, a saber: entrevistas, documentação, registro de arquivos, observação direta, artefatos físicos e observação participante. Deve-se contar com bons livros-texto de apoio, artigos etc, gerar um banco de dados do estudo de caso e encadear as evidências constantemente.
- Análise: Utilizam-se nesta etapa técnicas para dados qualitativos, quantitativos ou ambos. Elaboração de séries temporais, tabulação dos resultados, testes ou simulações são utilizadas para se formular as conclusões.
- Compartilhamento: Relatar determinado estudo de caso, é sinônimo de trazer seus resultados e constatações a uma condição final. O processo de compartilhamento de informações, corrobora com a revisão e colabora para que o nível da apresentação seja elevado, além de permitir que outros pesquisadores concluam acerca do assunto.

O conjunto de processos e/ou operações mentais, empregados na investigação científica, constituem o método científico adotado, ou ainda, o método científico é a linha de raciocínio adotada no processo de pesquisa.

Particularmente à esta Tese, construiu-se o método segundo uma metodologia científica dita dedutiva, a qual fornece bases lógicas para a investigação (GIL, A.C., 1999, LAKATOS, E.M., MARCONI, M.A., 1993): *“explicar o conteúdo das premissas por intermédio de uma cadeia de raciocínio descendente; analisando-se do geral para o particular, chega-se a uma conclusão (silogismo)”*.

Por conta da impossibilidade de se validar o método escolhido através de técnicas analíticas, baseadas, por exemplo, em coleta e análise de dados estatísticos, esta Tese utilizou a pesquisa explicativa, tipo *ex-post-facto* (SILVA, E.L., MENEZES, E.M., 2005). Como neste tipo de pesquisa o experimento somente ocorre após os fatos, as hipóteses são suposições com respostas plausíveis porém provisórias, para o problema da pesquisa.

A validação de um método é processo dinâmico e constante, começando nas fases de seleção, desenvolvimento, qualificação de instrumentos, materiais e pessoal, e continuando nas fases dos experimentos e transferência do método.

Um processo de validação bem definido, fornece indícios que o método escolhido e o sistema adotado, são adequados às intensões do trabalho (LAKATOS, E.M., MARCONI, M.A., 1993).

A demanda por um modelo que integre dados estruturados e não estruturados, justifica o desenvolvimento do método definido nesta Tese, embora atualmente não se possa validá-lo com dados reais por ainda não existir um modelo de dados disponível com tais características.

A Figura 2 apresenta as fundamentações conceituais nas quais se basearam as hipóteses desta Tese.

Figura 2 – Fundamentações conceituais para a elaboração das hipóteses

Fonte: Autor

1.4 ETAPAS DESTA TESE

Na fase do planejamento do estudo, foi definido o assunto a ser estudado: “Modelo Navegacional Dinâmico, para implementação da integração inter-estrutural de dados”.

O estudo do tema central da Tese, supracitado, foi iniciado em busca de possíveis métodos que já realizassem tal avaliação, por conta da necessidade de se definir um método que valide o modelo proposto e a integridade do dado resultante, na ocasião da comparação de dados com estruturas distintas.

Na fase de preparação, o estudo começa com minuciosa revisão da bibliografia e cuidadosas proposições que atendam aos objetivos da pesquisa. Também relevante, é a dedicação ao formalismo e aos procedimentos explícitos ao longo do período de realização da pesquisa (YIN, R.K., 2010).

Segundo (GIL, A.C., 1999), esta revisão pode ser realizada a partir de publicações, livros, *Internet* etc. Para tanto, uma revisão bibliográfica foi feita desde o início da pesquisa desta Tese, apresentada na seção seguinte e comentada na seção que antecede a proposta.

Segundo (YIN, R.K., 2010), compartilhar informações é direcioná-las para públicos específicos. Espera-se que as conclusões destes públicos acerca da pesquisa apresentada, contribuam para o crescimento da pesquisa e do próprio público.

Na Figura 1, (YIN, R.K., 2010) desconsidera que, após o compartilhamento, o fluxo contínuo retome as atividades.

A Figura 3, mostra o modelo de (YIN, R.K., 2010) adaptado a este trabalho.

Figura 3 – Modelo adaptado para desenvolvimento de um método

Fonte: Autor, adaptado de Yin (2010)

1.5 ORGANIZAÇÃO DO TRABALHO

Neste Capítulo constam a introdução ao trabalho, assim como sua motivação, justificativa, objetivo, metodologia de pesquisa e as etapas desta metodologia aplicadas nesta Tese.

No Capítulo 2 desta Tese, estão descritos alguns dos diversos tipos de modelos de dados já utilizados e em utilização, atualmente.

O Capítulo 3 contém o conceito de dados não estruturados e suas linhas básicas de pesquisa.

No Capítulo 4 constam os conceitos e exemplos de aplicações de Redes Complexas, insumo direto a este trabalho.

O Capítulo 5 contém as aplicações relacionadas, insumo indireto a este trabalho, assim como os subsídios fornecidos pelos capítulos até então mencionados.

O Capítulo 6 contém o modelo de dados proposto propriamente dito, suas premissas e dois exemplos de aplicação deste modelo, evidenciando vantagens e limitações dessa proposta .

No Capítulo 7, são apresentadas as conclusões, a contribuição deste trabalho e sugestões para trabalhos futuros.

2 OS MODELOS DE DADOS

Este capítulo apresenta os modelos de dados, desde o modelo hierárquico até os modelos *NoSQL*.

Antes da década de 1960, dados eram abrigados e acessados em estruturas denominadas arquivos indexados VSAM (*Virtual Storage Access Method*) (CÂNDIDO, F., 2007).

Na década de 1960, surge o modelo de dados hierárquico, estrutura única até então, onde os dados eram dispostos segundo uma hierarquia (KAMFONAS, J.M., 1952). Na mesma década, o comitê CODASYL (*Committee on Data Systems and Languages*), (TAYLOR, W.R., 1978), propõe ajustes no modelo de dados hierárquico, originando o modelo de dados em rede (FRANK, R.L., 1978).

Na década de 1970, surge o modelo de dados relacional (CODD, E.F., 1970), (DATE, C.J., 2004) cujo primeiro ajuste, denominado *MER-X* (Modelo Entidade Relacionamento Extendido), ocorreu na década de 1980 (SILBERSCHATZ, A. et al., 2006).

Na década de 1980, surge o modelo de dados orientado a objetos (KIM, W., 1990), inaugurando a família dos modelos de dados pós-relacionais.

Surge na década de 1990, o modelo de dados objeto-relacional, fruto da combinação dos conceitos encontrados no modelo de dados relacional e no modelo de dados orientado a objetos (STONEBREAKER, M, 2011). Na sequência, ainda na mesma década, 1990, modelos de dados em sua maioria relacionais, abrigavam dados dispostos em arquivos do tipo XML (*Extensiveble Markup Language*), e por conta das características estruturais desses arquivos, passaram a ser denominados modelos de dados semi-estruturados (HEUSER, C.A., 2014).

Na década seguinte, 2000, também na condição de pós relacionais, surgem os modelos de dados não estruturados, rotulados pela sigla *NoSQL* (*Not Only SQL*), e com eles, novos conceitos associados a consistência, disponibilidade e principalmente, escalabilidade de dados (BREWER, E.A., 2012).

2.1 O MODELO DE DADOS HIERÁRQUICO

Embora o modelo hierárquico de dados não esteja diretamente ligado à proposta desta Tese, os conceitos nele contidos têm ligação com ela.

Neste modelo, os dados são organizados de forma descendente, haja vista a representação no formato árvore, e são definidos como uma coleção de registros conectados entre si por meio de ligações (CÂNDIDO, F., 2007).

Cada registro desta coleção, contém uma coleção de campos onde cada campo contém um único valor. Uma ligação conecta exclusivamente dois registros, ou seja, o relacionamento entre um registro-pai ou segmento pai, hierarquicamente superior, e um registro-filho ou segmento filho, hierarquicamente inferior, tem cardinalidade 1:N (um para muitos), respectivamente.

A Figura 4 apresenta um exemplo da estrutura do modelo hierárquico de dados. Nesta Figura, observa-se as relações hierárquicas entre os diferentes níveis do modelo.

Figura 4 – Estrutura do modelo hierárquico de dados

Fonte: Autor, adaptado de ELMASRI (2005)

Porém, com o aumento crescente do volume dos dados e da complexidade dos sistemas, o modelo hierárquico de dados foi se mostrando inadequado, pois manter uma estrutura hierárquica de árvore se tornou tarefa trabalhosa, além da impossibilidade de se representar relacionamentos com cardinalidade N:N (muitos para muitos).

2.2 O MODELO DE DADOS EM REDES

Esse modelo, contemporâneo ao modelo de dados hierárquico, foi utilizado pela primeira vez em 1964, por Charles Bachman. Analogamente ao modelo hierárquico de dados, os conceitos deste modelo de dados têm ligação com a proposta desta Tese, sobretudo pelo diferencial que apresenta em relação ao modelo hierárquico, ou seja, a possibilidade de cardinalidade N:N entre diferentes hierarquias.

Neste modelo, a pesquisa torna-se mais flexível e o desempenho melhor, quando comparados ao modelo hierárquico, por não se depender de um único registro (nó) como inicializador da pesquisa (MANFRED, H. , 2004). Porém, o modelo de dados em redes, sob o ponto de vista estrutural, apresenta os mesmos problemas do modelo hierárquico, onde uma alteração efetuada em determinada classe hierárquica de dados, implica na criação de nova estrutura para suportar esta nova classe (FRANK, R.,L., 1978), além de, da mesma forma que no modelo de dados hierárquico, o acesso aos dados ocorrer através de linguagens denominadas hospedeiras.

A Figura 5 destaca o diferencial do modelo em redes, em relação ao modelo hierárquico.

Figura 5 – Estrutura do modelo de dados em rede

Fonte: Autor, adaptado de ELMASRI (2005)

2.3 O MODELO DE DADOS RELACIONAL

O modelo relacional, criado na década de 1970 pelo matemático britânico Edgar Frank Codd (CODD, E.F., 1970) e posteriormente aprimorado por Christopher J. Date (DATE, C.J., 2004) e Hugh Darwen (DARWEN, H., 2012), foi desenvolvido com a finalidade de suprir deficiências estruturais dos modelos hierárquicos e de redes, principalmente as deficiências associadas a dificuldades de manipulação dos dados e também por inconsistências causadas pela não versatilidade cardinal das associações hierárquicas, ou seja, dependência dos dados.

O modelo relacional, através de regras fundamentadas na teoria matemática de conjuntos com o acréscimo de funções apoiadas na álgebra relacional, permite processamento *ad hoc*¹ quando da captura, processamento e disponibilização da informação (SILBERSCHATZ, A. et al., 2006).

Devido a sua simplicidade, flexibilidade e desempenho satisfatórios, o modelo de dados relacional mostrou-se apropriado principalmente na solução de problemas de cunho comercial. Este modelo de dados representa os dados, como uma coleção de entidades interligadas.

As entidades (retângulos), interligadas por relacionamentos (losângos), são compostas por atributos que, necessariamente, contém o mesmo tipo de dado.

Regras oriundas da coexistência combinada de atributos e relacionamentos, ditas formas normais, garantem a integridade do dado através da eliminação de redundâncias, evitando tuplas (conjunto sequencial de colunas) repetidas (SILBERSCHATZ, A. et al., 2012).

A Figura 6 mostra a estrutura básica do modelo de dados relacional, por meio do diagrama Entidade-Relacionamento (E-R).

¹ Processamento dedicado, exclusivo

Figura 6 – Estrutura básica do modelo de dados relacional

Fonte: Autor

2.4 O MODELO DE DADOS ORIENTADO A OBJETOS

Quando criado, acreditava-se que o modelo de dados orientado a objetos, com o passar do tempo representaria grande parcela dos modelos de dados utilizados, por suprir na época, a então principal deficiência do modelo relacional, ou seja, representar de forma única dados de tipos distintos, minimizando por consequência, a perda por impedância, quer pela diferença da estrutura do dado oriundo do modelo relacional e sua representação em memória, quer pelas demandas oriundas de aplicações orientadas a objetos por dados mais elaborados.

Hoje, porém, verifica-se que os modelos de dados orientados a objetos são utilizados em aplicações específicas (KIM, W., 1990), enquanto que os modelos de dados relacionais continuam a sustentar os negócios tradicionais, onde estruturas de dados baseadas em relações são necessárias e suficientes.

O diagrama de classes da UML (*Unified Modeling Language*) é geralmente utilizado para representar o modelo de dados orientado a objetos (TAKAI, O.K., et. Al., 2005).

Análogo ao modelo de dados relacional, representado por entidades e relacionamentos, o modelo orientado a objetos é representado por classes e atributos.

A Figura 7 apresenta a estrutura básica do modelo de dados orientado a objetos, por meio do diagrama de classes.

Figura 7 – Estrutura básica do modelo de dados orientado a objetos

Fonte: Autor, adaptado de PLÁCIDO, D.G., CUNHA, D.P., (2008)

A conjunção dos tipos de dados distintos, segregados e representados por atributos no modelo relacional, tem, no modelo orientado a objetos, concepção integrada, diferenciando-se da representação semântica, observada no modelo relacional.

No modelo de dados orientado a objetos, a manipulação ocorre por intermédio dos métodos definidos pelas classes, nas quais os objetos estão.

Hierarquicamente falando, os objetos são organizados em supertipos que por sua vez possuem tipos e subtipos. Assim, o relacionamento entre os objetos é possível desde que haja referências estabelecidas entre eles.

Porém, diferentemente do modelo relacional, fortemente fundamentado em teorias matemáticas consistentes, este modelo carece de fundamentação matemática formal (SILBERSCHATZ, A. et al., 2006).

2.5 OS MODELOS DE DADOS NOSQL

Posterior aos modelos de dados relacional e orientados a objetos, segundo o Professor Pesquisador Valdemar Setzer, (SETZER, V.W., 2010), modelos de dados *NoSQL* podem ser classificados como modelos relacionais fracamente normalizados, denominados por ele, MRNN (Modelos Relacionais Não Normalizados).

Segundo (BRITO, R., 2011), os modelos de dados *NoSQL* são divididos em quatro tipos: Tipo Chave-valor, Tipo coluna, Tipo orientado a documentos e Tipo orientado a grafos.

2.5.1 Modelo de dados do tipo chave/valor

Considerado o mais escalável dos modelos de dados pós-relacionais, os conceitos deste modelo de dados são intrínsecos aos demais modelos *NoSQL* e suporta maiores volumes de dados quando comparado aos demais modelos, por ser baseado em coleções de chaves únicas, com valores associados a estas chaves (FERREIRA, E, 2011). Neste modelo, existe uma tabela *hash* baseada em eventos *get* e *put* que permite, por conta da estrutura simples, melhor desempenho quando comparado aos demais modelos *NoSQL*.

A Figura 8 apresenta a estrutura do modelo de dados tipo chave/valor.

Figura 8 – Estrutura do modelo de dados *NoSQL*, tipo chave-valor

Chave1	Valor1a
	Valor1b
	Valor1c

Fonte: Autor

A Figura 9 apresenta um exemplo de aplicação para o modelo de dados *NoSQL* do tipo chave/valor.

Figura 9 – Exemplo de aplicação do modelo de dados *NoSQL*, chave-valor

Fonte: Autor

2.5.2 Modelo de dados do tipo colunas

Neste modelo, existe um mapa indexado por três chaves (uma linha, uma coluna e um *timestamp*) que combinados, retornam uma *string* valorada. (NASCIMENTO, J, 2011).

A Figura 10 apresenta a estrutura do modelo de dados *NoSQL*, orientado a colunas.

Figura 10 – Estrutura do modelo de dados *NoSQL*, orientado a colunas

Fonte: Autor, adaptado de Fowler (2012).

A Figura 11 apresenta um exemplo de aplicação para o modelo de dados *NoSQL*, orientado a colunas.

Figura 11 – Exemplo de aplicação do modelo *NoSQL*, orientado a colunas

Fonte: Autor, adaptado de Fowler (2012).

2.5.3 Modelo de dados do tipo orientado a documentos

Este modelo de dados é o que mais se aproxima do modelo relacional, quando se compara documentos deste modelo, com tuplas do modelo relacional.

Neste modelo de dados, os documentos, unidades básicas de armazenamento, não precisam possuir estrutura pré-definida nem comum entre si, contendo apenas o conteúdo mais relevante do documento, além da possibilidade de um documento conter outros documentos aninhados (STEPPAT, N., 2011).

A Figura 12 apresenta a estrutura do modelo de dados orientado a documentos.

Interessante observar a semelhança deste modelo de dados com modelos de dados hierárquicos e de redes, apresentados nas seções 2.1 e 2.2, respectivamente.

Figura 12 – Estrutura do modelo de dados *NoSQL*, orientado a documentos

Fonte: Autor

A Figura 13 apresenta um exemplo de aplicação para o modelo de dados *NoSQL*, orientado a documentos.

Figura 13 – Exemplo de aplicação do modelo de dados *NoSQL*, orientado a documentos

Fonte: Autor

2.5.4 Modelo de dados do tipo orientado a grafos

Neste modelo, os dados são armazenados em nós de um grafo e ligados pelas arestas a outros nós, conforme suas relações (SOUZA, T., ROCHA, A., 2011).

Diferentemente dos outros modelos *NoSQL*, neste modelo a consistência transacional é assegurada por propriedades *ACID*, logo, apresenta menor escalabilidade e desempenho (FERREIRA, E., 2011).

Algumas bibliografias rotulam este modelo de dados como *NewSQL*, associando a ele características dos modelos *NoSQL* quanto ao desempenho e escalabilidade, e características do modelo relacional, quanto a consistência transacional (HOFF, T., 2012).

A Figura 14 apresenta a estrutura básica do modelo de dados *NoSQL*, do tipo grafos.

Figura 14 – Estrutura do modelo de dados *NoSQL* , orientado a grafos

Fonte: Autor

A Figura 15 apresenta um exemplo de aplicação para o modelo de dados *NoSQL*, orientado a grafos.

Figura 15 – Exemplo de aplicação do modelo de dados *NoSQL*, orientado a grafos

Fonte: Autor

2.6 COMPARATIVO ENTRE OS MODELOS DE DADOS

A Tabela 1 apresenta os modelos de dados, classificando-os quanto às principais demandas de suas evoluções, ou seja, complexidade de acesso, escalabilidade horizontal por aumento do volume de dados e facilidade em representar dados de diferentes estruturas.

Tabela 1 – Os modelos de dados e suas demandas evolutivas

Modelos	Complexidade	Escalabilidade	Representatividade
<i>NoSQL</i>	Variável / Baixa	Alta	Alta
Orientado a objetos	Variável	Variável/Baixa	Variável/Alta
Relacional	Variável	Variável	Variável/Baixa
Em redes	Alta	Mínima	Nula
Hierárquico	Alta	Mínima	Nula

**Evolução
no tempo**

Fonte: Autor, adaptado de SOUZA, T., ROCHA, A., FERREIRA, E.(2011)

Analisando a Tabela 1, observa-se que, à medida que o tempo passa, os modelos de dados apresentam acessos menos complexos (mais amigáveis), maior escalabilidade horizontal e tendem a representar melhor dados com estruturas distintas.

A Tabela 2 apresenta o detalhamento dos modelos de dados *NoSQL*, comparando-os ao modelo de dados relacional. Neste comparativo, são consideradas funcionalidades tais como: desempenho em consultas envolvendo dados, estruturados ou não; escalabilidade por aumento significativo do volume de dados não estruturados; flexibilidade e complexidade nas formas de acesso; consistência do modelo; e fundamentação teórica estrutural.

Tabela 2 – Comparativo do modelo de dados relacional com os modelos de dados *NoSQL*

Modelos	Desempenho	Escalabilidade	Flexibilidade	Complexidade	Consistência	Teoria
Orientado a grafos	Variável	Variável/Alta	Alta	Alta	Variável	Dos grafos
Orientado a documentos	Alto	Variável/Alta	Alta	Baixa	Baixa	Mínima
Tipo chave / valor	Alto	Alta	Alta	Mínima	Baixa	Mínima
Tipo colunas	Alto	Alta	Média	Baixa	Baixa	Mínima
Relacional	Variável	Variável	Baixa	Média	Alta	De conjuntos

Fonte: Autor, adaptado de BRITO, R.(2011)

Analizando a Tabela 2, observa-se que os modelos *NoSQL* apresentam boa escalabilidade, e que esta característica é, na mesma proporção, observada no respectivo desempenho. Da mesma forma, existe uma relação direta entre a teoria que apoia o modelo de dados e a consistência dos dados abrigados neste modelo, fatores estes inversamente proporcionais ao desempenho.

Vale notar que no caso particular de modelos *NoSQL* orientados a grafo, algumas características tais como desempenho e escalabilidade se assemelham ao modelo de dados relacional, embora no modelo orientado a grafos a representatividade de dados com diferentes estruturas (Tabela 1), seja maior.

A Figura 16 complementa a Tabela 2, comparando a complexidade de acesso (*front-end*) e o desempenho dos modelos de dados relacional e *NoSQL*, com o respectivo volume armazenado.

Figura 16 – Comparativo dos modelo de dados *NoSQL*

Fonte: Autor, adaptado de STEPPAT, N. (2011).

Analizando a Figura 16, observa-se que o modelo de dados relacional, embora atualmente suporte a maioria dos sistemas computacionais que acessam dados, mesmo tendo *interface* de acesso menos complexa, não representa o modelo de dados mais utilizado quando há a necessidade de armazenamento de grandes volumes de dados, dados estes em sua maioria não estruturados.

Ainda considerando a Figura 16, como, por exemplo, a partir do arquivo de áudio de um filme (dado não estruturado), abrigado em um gerenciador de banco de dados que opera sob regras do modelo de dados relacional, identificar o registro das pessoas (dado estruturado) mencionadas neste filme?

No questionamento do parágrafo anterior, a incompatibilidade entre as diferentes estruturas, não nos permite essa comparação. As características da linguagem SQL, não permitem a recursividade demandada por funcionalidades de interfaces complexas e impõem substancial queda de desempenho por conta das perdas por impedâncias (FOWLER, J.H., CHRISTAKIS, N.A., 2008), perdas estas caracterizadas pelo tempo gasto, necessário para compatibilizar as requisições majoritariamente oriundas de aplicações orientadas a objetos para o modelo de dados relacional e vice-versa.

Entendendo melhor as perdas por impedância, parte-se do fato que o modelo relacional organiza os dados bidirecionalmente, em atributos e tuplas e as aplicações orientadas a objetos os organiza em classes de objetos. Ao acessar-se bancos de dados relacionais a partir de linguagens orientadas a objetos, as classes de objetos da linguagem devem ser traduzidas para as tabelas dispares do banco de dados relacional (FINN,M., 2007).

Segundo (SILBERCHATZ, A. et. al., 2006), as perdas por impedância existem por conta das diferenças estruturais entre o modelo de dados relacional e as linguagens de programação orientadas a objetos, e se mostram críticas, quando do armazenamento e recuperação de dados complexos.

Dados não estruturados tais como imagens, sons, mensagens contínuas, reportagens, páginas de navegadores ou a composição destes dados, são, no modelo de dados relacional, armazenados da mesma forma que, por exemplo, um valor numérico.

O fornecedor Oracle, por exemplo, criou em versões mais recentes de seus gerenciadores de bancos de dados (10G, 11G e 12C), apoiados sob regras do modelo de dados relacional, *data-types* próprios para este fim, denominados BLOB (*Binary Large Object*) e CLOB (*Character Large Object*), coexistindo assim dados não estruturados, abrigados nestes *data-types* com dados estruturados, no mesmo modelo relacional, mas alerta sobre eventuais dificuldades associadas a estes tipos de dados no caso de procedimentos associados a salva-guarda/recuperação dos dados, manipulação de conteúdos compartilhados e sobretudo, desempenho em sistemas distribuídos, quando do aumento do volume de dados não estruturados armazenados nestes *data-types* (KEVIN,L., BOB, B., 2004).

Essa conjunção de fatores, demanda ao modelo relacional tempo adicional de processamento, tempo este proporcionalmente aumentado em acessos de grandes volumes de dados.

Modelos de dados mais recentes, originalmente projetados para dados distribuídos e de natureza complexa, como os modelos de dados *NoSQL*, apropriados para grandes volumes de dados, altamente escaláveis horizontalmente, disponibilizando os dados em tempos computacionais menos dilatados e principalmente, capazes de coexistir estruturas distintas, representam a tendência natural de evolução do modelo de dados relacional, cujo propósito na ocasião da sua concepção era abrigar e prover acesso a dados estruturados de forma bidimensional, haja vista nesta ocasião, a inexistência de dados não estruturados.

Referencialmente falando, observa-se nos modelos de dados *NoSQL*, formas particulares no tratamento dos dados, adequando seu uso à aplicação e não necessariamente utilizando-os sobre regras menos flexíveis, como as regras impostas pelo modelo relacional, adaptando neste último, a aplicação ao modelo de dados.

Transacionalmente falando, nos modelos de dados *NoSQL*, o conceito *ACID* é, na maioria das vezes, substituído pelos conceitos do teorema *CAP* (*Consistency, Availability e Partition Tolerance*), (BREWER, E.A., 2012) e pelo conceito *BASE* (*Basically Available, Soft State and Eventual Consistency*) (VOGELS, W., 2009), onde, no primeiro conceito, (BREWER, E.A., 2012) demonstra ser impossível para um modelo de dados distribuído ser simultaneamente consistente, estar sempre disponível e intolerante a falhas, como sugere o modelo de dados relacional.

Assim, o teorema *CAP* sugere somente ser possível a combinação das três propriedades (Consistência, Disponibilidade e Tolerância a falhas no particionamento), duas a duas, ou seja, quando combinadas consistência e disponibilidade, combinação mais próxima dos modelos de dados relacionais, a tolerância a falhas em sistemas com dados distribuídos não é garantida. Quando combinadas disponibilidade com tolerância a falhas, a consistência não é assegurada, assim como a combinação de consistência e tolerância a falhas não assegura disponibilidade dos dados.

Por sua vez, regras do conceito *BASE* sugerem modelos de dados disponíveis sob demanda, sem a necessidade de estarem consistentes o tempo todo, daí o termo consistência eventual.

2.7 A CONSISTÊNCIA EVENTUAL

A consistência eventual é uma das características do modelo de dados proposto. De acordo com as 3 (três) propriedades do Teorema *CAP* (*Consistency*, *Availability* e *Partition-Tolerancy*), um sistema computacional distribuído garante, simultaneamente, apenas duas destas três propriedades.

Segundo a hipótese de (BREWER, E.A., 2000), posteriormente comprovada por (GILBERT, S. et. al., 2002), em um sistema computacional distribuído é desejável obter-se consistência, disponibilidade e tolerância ao particionamento dos dados, porém, essas propriedades, são improváveis de serem obtidas ao mesmo tempo.

Segundo (WEI, Z., 2009), para que um sistema seja considerado consistente, deve-se garantir que, após a realização de uma transação, o sistema mantenha a integridade dos dados, ou seja, a transação não pode alterar as regras do modelo de dados.

A disponibilidade, segundo (BROWNE, J., 2009), deve assegurar que um sistema computacional ofereça, sempre que solicitado por seus usuários, acesso às informações.

A tolerância ao particionamento dos dados, segundo (GILBERT, S.; LYNCH, N., 2002), está ligada a capacidade de um sistema computacional permanecer em operação, permitindo operações de leitura e escrita na base de dados, após os dados serem segregados em redes diferentes. No entanto, segundo (OREND, K., 2010), um sistema computacional tolerante ao particionamento, somente pode assegurar consistência, caso ocorram interrupções na sua disponibilidade. Esta propriedade é observada, por exemplo, na ocasião de uma operação de escrita. Neste caso, o sistema deve garantir que cada operação de escrita termine, somente se o dado for replicado para todos os nós segregados, o que, em situações tais como falhas do *hardware* e/ou quedas de conexão entre os nós, não ocorre.

A tentativa de se garantir simultaneamente as três propriedades mencionadas, ou seja, Consistência, Disponibilidade e Tolerância ao particionamento, segundo (MONSON-HAEFEL, R., 2009) seria possível, mas acarretaria aumento significativo nos custos computacionais, quando comparado aos custos necessários para manter-se a simultaneidade de duas destas propriedades.

Então, se os dados distribuídos de um determinado sistema computacional devem estar sempre disponíveis, a consistência é improvável. Da mesma forma, sistemas computacionais consistentes e tolerantes ao particionamento, apresentam indisponibilidade.

Como exemplo de sistemas computacionais que priorizam as propriedades consistência e disponibilidade a despeito da tolerância ao particionamento, pode-se citar sistemas apoiados no modelo relacional de dados (NETO, J.G., 2008).

Quanto a sistemas computacionais que disponibilizam dados sob consistência eventual, como o modelo proposto por este trabalho, pode-se citar os modelos de dados do tipo *NoSQL*, apresentados na seção 2.5.

A Figura 17 apresenta os componentes do Teorema *CAP* e suas intersecções.

Figura 17 – Teorema *CAP*

Fonte: Autor, adaptado de BREWER (2012)

A região 1 (um) da Figura 17, denominada CA (*Consistency – Availability*), representa a coexistência da componente consistência e da componente disponibilidade, encontrados em modelos de dados relacionais. Nesses modelos de dados, a consistência do modelo e a disponibilidade dos dados são fatores fundamentais. Operam desta forma os gerenciadores relacionais *DB-2*, *Oracle*, *MSSQL-Server*, *Informix*, *Postgree*, *My-sql*, dentre outros relacionais.

A região 2 (dois) da Figura 17, denominada CP (*Consistency – Partition-Tolerance*), representa a coexistência da componente consistência e da componente tolerância ao particionamento. Em termos práticos, significa continuar sendo necessário se ter consistência com a base de dados em operação, ainda que haja uma falha de rede que particione (separe) os dados desta base em duas ou mais redes independentes.

Analogamente, pode-se imaginar um modelo escalável horizontalmente, sendo acessado por diversos nós de um mesmo *cluster*, com dados armazenados em *storages* distintos, mas com a disponibilidade prejudicada. Operam desta forma os gerenciadores *BigTable*, *MongoDB*, dentre outros *NoSQL*.

A região 3 (três) da Figura 17, denominada AP (*Availability – Partition-Tolerance*), representa a coexistência da componente disponibilidade com a componente tolerância ao particionamento. Em termos práticos, são os modelos que não admitem ficar fora de operação e para isso, abrem temporariamente mão da consistência, denominados assim, modelos eventualmente consistentes.

Aliado às componentes do Teorema CAP, exatamente nesta condição, o conceito *BASE* (*Basically Available, Soft State and Eventual Consistency*), se aplicado, destaca seu principal quesito sob a óptica da proposta deste trabalho, ou seja, consistência eventual. Porém, para isso, as escritas precisam ser sempre aceitas para serem sincronizadas posteriormente, gerando a inevitável e as vezes inadmissível janela de inconsistências. Operam desta forma os gerenciadores do tipo *NoSQL Amazon-Dynamo DB* e *Cassandra*.

Interpretando o significado das intersecções da Figura 17, observa-se que a forte demanda por escalabilidade horizontal, torna a consistência e a disponibilidade fatores disjuntos, e no caso de ausência total de consistência, mutuamente excludentes.

Então, os modelos de dados que supostamente representam a simultaneidade dos conceitos existentes na região *I* da Figura 17, ou seja, são intolerantes a falhas de particionamento, incondicionalmente consistes e sempre disponíveis, contrariam as premissas do Teorema *CAP*.

Sobre a consistência eventual do modelo proposto, caso, por alguma anomalia, ocorram escritas descartadas, o sistema não as efetivará na base de dados por não saber se serão ou não processadas. Analogamente, no caso de sistemas com dados distribuídos, haverá descarte da informação quando do restabelecimento das partições, pois o modelo proposto não permitirá a efetivação de requisições, por não ter certeza em poder processá-las.

Assim, uma forma simplificada de se entender o modelo de dados proposto sob a óptica da consistência eventual, é tê-lo menos consistente no caso de demanda extrema por alta disponibilidade, pois, no modelo proposto, estima-se que a janela de inconsistência seja diretamente proporcional ao volume dos dados armazenados, ou seja, corresponde ao modelo de dados proposto ficar inconsistente por um período maior ou um período menor, dependendo diretamente do volume de dados armazenado.

3 DADOS NÃO ESTRUTURADOS

É crescente o interesse da comunidade científica por pesquisas ligadas a dados não estruturados, dados estes em sua maioria oriundos da plataforma *Web*, hoje majoritariamente abrigados em modelos de dados relacionais e em grandes quantidades.

Por definição (COLUMBYA UNIVERSITY, 2015), dado não estruturado é o dado que não reside na tradicional estrutura bidimensional linha-coluna, do modelo de dados relacional.

O termo dado semi-estruturado também é frequentemente associado ao contexto de dados não estruturados, mas carecem de definição formal. Segundo (FICO, 2015), dados semi-estruturados são dados incompletos, irregulares e não necessariamente associados a um esquema de banco de dados, ou seja, com estruturas não previamente conhecidas.

Arquivos *XML* (*Extensible Markup Language*) e arquivos *JSON* (*Java Script Object Notation*), ambos também utilizados na programação de *API's* (*Application Programming Interface*), são exemplos de fontes de dados semi-estruturados.

3.1 DADOS NÃO ESTRUTURADOS - LINHAS DE PESQUISA

Dados não estruturados são sub-divididos em dois grupos: Textuais e não textuais. Representam dados não estruturados do grupo textuais as mensagens instantâneas, páginas *Web*, arquivos de textos etc, e os não textuais, fotografias, filmes, arquivos audíveis etc.

Diferentes abordagens foram feitas no sentido de representar dados não estruturados, sobretudo os não textuais. Essas abordagens, possuem três linhas básicas de pesquisa:

- Reestruturação das linguagens de consulta;
- Criação de novos sistemas gerenciadores de bancos de dados;e
- Desenvolvimento de novos modelos de dados.

3.1.1 Reestruturação das linguagens de consulta

Esta linha de pesquisa é baseada em soluções que otimizam consultas a dados não estruturados contidos em bases de dados que operam sob o modelo relacional de dados.

Assim, são executadas consultas a estas bases por intermédio de aplicativos *front-end*, obedecendo a premissa de que toda consulta deve ser eficiente no que tange ao processamento e apresentar qualidade nos resultados, ou seja, integridade e desempenho, respectivamente (LIU, J. et al., 2006), (ROY, P., MOHANIA, M., 2007). O método de Liu, (LIU, J. et al., 2006), denominado *SQoUT* utiliza linguagem *SQL* sobre dados não estruturados, em modelos relacionais. Consiste em extrair os dados para, na sequência, remodelá-los, consistindo-os sob as regras do modelo de dados relacional para então, disponibilizá-los via consultas em linguagem *SQL*.

A Figura 18 apresenta as etapas da técnica *SQoUT*, proposta por Liu.

Figura 18 – Etapas da proposta de Liu

Fonte: Autor, adaptado de Liu et. al. (2006)

Ainda na linha de reestruturação das linguagens de consultas, (ROY, P., MOHANIA, M., 2007) apresentam a técnica por eles denominada *SCORE* (*Symbiotic Context Oriented Information Retrieval*).

Esta técnica, mostrada na Figura 19, executa consultas SQL em dados estruturados, porém, ao decompor os resultados das consultas SQL em palavras-chave, utiliza essas palavras como termos de busca em documentos não estruturados, estabelecendo assim, ligações entre os termos encontrados pelo mecanismo de busca e os resultados das consultas SQL. Em seguida, estes resultados são apresentados ao usuário.

Figura 19 – Técnica SCORE

Fonte: Autor, adaptado de Roy (2007)

3.1.2 Criação de novos sistemas gerenciadores de bancos de dados

Esta linha de pesquisa consiste na aplicação de algoritmos, estatísticos e heurísticos, em bases de dados que operam sob o modelo de dados relacional, onde dados não estruturados do tipo textuais, mais especificamente arquivos, são comparados a dados estruturados existentes no modelo relacional, gerando ou não um repositório relacional provisório, cujo objetivo é estruturar dados outrora não estruturados e destes repositórios, extrair informações até então desconhecidas (CHU, E. et al., 2007), (MANSURI, I.R. et al, 2006).

De forma semelhante, (TARI, L. et al, 2010) utilizam o modelo de dados relacional como repositório de informações extraídas de arquivos não estruturados do tipo textuais, para posterior utilização, subdividindo o procedimento entre o processamento do texto (*TPDB*) e extração de dados (*PTQL*), este último com capacidade de criar relações semânticas entre as palavras do texto e as consultas SQL. Os resultados finais, caso existam, são disponibilizados ao usuário através do denominado por eles, repositório incremental de Tari.

A Figura 20, apresenta o repositório incremental, proposto por (TARI, L. et al, 2010).

Figura 20 – Repositório incremental de TARI

Fonte: Autor, adaptado de TARI et. al. (2010)

3.1.3 Desenvolvimento de novos modelos de dados

Esta linha de pesquisa consiste em desenvolver novos modelos de dados para gerenciamento de dados não estruturados. Esse novo modelo é utilizado para organizar os dados não estruturados, como uma nova estrutura de dados independente do atual modelo de dados, porém também relacional.

(DOAN, A. et al., 2009), da Universidade de Wisconsin , propuseram o *UDMS*, (*Unstructured Data Management Systems*), por sua vez baseado em um modelo de dados denominado *DGE* (*Data Generation Exploration*), cujos elementos principais são o sistema, os dados e o usuário, análogo ao modelo relacional de dados. Segundo (DOAN, A. et al., 2009), este modelo, baseado em buscas por palavras-chave, revela informações ocultas em dados não estruturados, muitas vezes importantes ao usuários.

A Figura 21 apresenta as etapas do modelo *UDMS*, proposto por DOAN et. al.

Figura 21 – Modelo de dados UDMS

Fonte – Autor, adaptado de DOAN, (2009)

Análogo ao modelo *UDMS*, (LIU, J. et al, 2006 ; LI, Y. et al, 2008) desenvolveram um sistema de gerenciamento de dados não estruturados, denominado *AUDR* (*Advanced Unstructured Data Repository*).

A Figura 22 apresenta o modelo de dados *AUDR*, proposto por Liu et.al.

Figura 22 – Modelo de dados *AUDR*

Fonte – Autor, adaptado de LIU, (2008)

Com base tetraédrica, o modelo *AUDR* é direcionado ao gerenciamento de arquivos de mídia, especialmente imagens e áudio, difíceis de serem geridos no modelo relacional de dados. Neste modelo, *BAF* (*Basic Attribute Facet*) representa o conteúdo dos atributos básicos do arquivo de mídia tais com identificação do autor, identificação do arquivo, tamanho do arquivo, formato do arquivo, extensão do arquivo etc. *RDF* (*Raw Data Facet*) representa a imagem, propriamente dita. *LFF* (*Low-level Feature Facet*) contém os parâmetros físicos do arquivo de mídia tais como *descriptor-color-layout*, *feature-FCTM*, *descriptor-edge-histogram* etc (PARK, S.J. et. al, 2000) e *SFF* (*Semantic Feature Facet*), o respectivo conteúdo semântico.

3.1.4 Métodos propostos por fabricantes de gerenciadores relacionais

Desenvolvedores de sistemas gerenciadores de banco de dados relacionais tais como *Oracle*, *Sybase*, *Microsoft* etc, propõe dois métodos para manipulação de dados não estruturados: Métodos tradicionais e métodos não tradicionais.

Métodos tradicionais armazenam dados textuais no banco relacional em colunas cujos tamanhos são fixos (NETO, J.G., 2008) ou em arquivos externos ao banco de dados, em disco rígido. Métodos não tradicionais, introduzem tipos de dados binários, no próprio modelo relacional de dados, para armazenamento destes dados.

No modelo de dados utilizado no *Microsoft SQL-Server*, estes tipos de dados denominam-se *filestream* (AZEMOVIC, J. et al., 2010). No modelo de dados *Oracle*, estes tipos de dados denominam-se *BLOB* (*Binary Large Object*) e *CLOB* (*Character Large Object*), além da técnica conhecida por *Oracle-Text*, que consiste da indexação / classificação de documentos, mas que para grandes volumes se mostra desvantajosa por exigir que parâmetros sejam inseridos pelos usuários (BEALL, J., 2008), (FUNG, B.C.M. et al., 2003), (SHARMA, A. et al., 2009).

3.2 COMPARATIVO ENTRE AS LINHAS DE PESQUISA

Analizando as propostas das três diferentes linhas de pesquisa citadas na seção anterior, observa-se os aspectos enfatizados por cada uma delas, suas semelhanças e diferenças, quando comparadas com a proposta deste trabalho.

A semelhança da primeira linha de pesquisa com a proposta deste trabalho, está apenas no acesso ao dado não estruturado, pois, diferentemente desta linha de pesquisa, este trabalho propõe um modelo de dados pós relacional, eventualmente consistente, com estruturas integradas e não apenas coexistentes.

A semelhança da segunda linha de pesquisa com a proposta deste trabalho, está também no acesso ao dado não estruturado, particularizando este acesso a dados não estruturados do tipo textuais. Este trabalho, diferentemente da segunda linha de pesquisa, propõe não só a coexistência de dados de estruturas distintas, como dito anteriormente, mas também a integração destes dados em um modelo de dados pós relacional único, sem a existência de repositórios provisórios.

A linha de pesquisa “Desenvolvimento de novos modelos de dados” é a mais próxima da proposta deste trabalho. Embora também com ênfase a dados não estruturados, essa linha de pesquisa propõe um modelo apoiado nas regras do modelo relacional de dados, se distanciando assim, da proposta desta Tese.

4 REDES COMPLEXAS

Atualmente, Redes Complexas são utilizadas na descrição e estudo dos mais variados tipos e gêneros de sistemas (NEWMAN, M., et. al., 2006).

Porém, algumas destas Redes Complexas só foram possíveis de serem modeladas recentemente, tais como redes sociais (*site Facebook*, *site Orkut* etc) e redes biológicas (cadeias de ácidos *DNA*, genomas, etc), principalmente por conta das limitações computacionais impostas pelo *hardware* dos equipamentos, na ocasião do processamento das informações.

O que principalmente caracteriza as Redes Complexas é seu crescimento imprevisível, mesmo que na ocasião da sua concepção, as regras que ligam seus componentes tenham sido pré estabelecidas, regras estas definidas pelo problema a ser modelado (NEWMAN, M. et. al., 2006).

Como exemplo, pode-se citar redes sociais. Nestas Redes Complexas, indivíduos são representados por nós e a ligação entre estes nós, pode em um problema representar o grau da amizade entre estes indivíduos e em outro problema, o grau de indvidamento, destes mesmos indivíduos.

Desta forma, regras pré estabelecidas originarão interações que por sua vez, caracterizarão a Rede Complexa quanto às suas propriedades, através das suas respectivas grandezas estatísticas. A análise destas grandezas, permitirá concluir-se acerca do comportamento do sistema como um todo representado por uma Rede Complexa.

Na Figura 23, apresenta-se um exemplo de Rede Complexa do tipo aleatória, representativa de produtos comercializáveis por determinada comunidade (WANG, P., et. al., 2009). O objetivo desta modelagem, é fornecer subsídios para apoiar a tomada de decisões estratégicas ligadas a economia dessa comunidade.

Nesta modelagem, regras pré-estabelecidas (a partir de qual valor o tamanho de um nó passa ser significativo) associam o tamanho de cada nó (neste exemplo, junção ponderada das grandezas estatísticas Distribuição de Graus, Coeficiente de Clusterização e Distribuição de Pesos, com maior ênfase à grandeza Coeficiente de Clusterização) com a relevância dos produtos comercializáveis na comunidade.

As ligações entre estes nós representam a utilização de cada produto como insumo para a produção de outro produto. Assim, nós maiores, quando associados, tem maior representatividade que nós maiores dissociados, que por sua vez tem maior importância que nós menores associados, e assim por diante.

Desta análise, conclui-se, por exemplo, prioridades na aquisição de determinados produtos a despeito de outros, quais as consequências financeiras diretas da falta de determinados produtos para a comunidade como um todo, qual produto adquirir em maior quantidade etc.

Figura 23 – Rede complexa dos produtos comunitários comercializáveis

Fonte: Autor, adaptado de HIDALGO, C.A., et. al.(2007)

4.1 REDES COMPLEXAS – HISTÓRICO

Estudar Redes Complexas é sinônimo de interdisciplinaridade, pois a abrangência do tema não se restringe apenas às ciências exatas como física, química, engenharia, matemática, computação, etc, mas também, à biologia e ciências sociais, estas últimas precursoras da utilização do modelo de Redes Complexas para modelar problemas cotidianos.

Euler, em meados do século 18 (1735), propôs resolver o problema das pontes de *Königsberg* utilizando um modelo de redes e com sua proposta, originou a teoria dos grafos. Por conta da aceitação do modelo de Euler, passou-se a se estudar redes pela teoria dos grafos.

Mais tarde, século 20 (1998),² Jerrold Grossman preparou uma lista completa sobre a matemática de Paul Erdős, matemático Húngaro, contemporâneo de Euler, cuja colaboração à matemática em problemas ligados à análise combinatória, teoria dos números, teoria dos conjuntos, análise matemática, teoria das probabilidades e teoria dos grafos, foi notável. Na teoria dos grafos especificamente, colaborou com trabalhos ligados a matemática discreta e algoritmos. Segundo Erdős, grafos são estruturas compostas por nós (ou vértices) e arestas (ou arcos).

As arestas ligam os nós, estabelecendo entre eles alguma relação. Quando o grafo é direcionado, dito dígrafo, cada aresta tem uma orientação, ou seja, a aresta que conecta dois nós tem um sentido, estabelecendo assim o nó origem e o nó destino. Além disso, o nó que tem caminho estabelecido para si mesmo é dito nó autônomo ou cíclico, caso contrário, acíclico. Mensagens de texto, *e-mails* e chamadas telefônicas, são exemplos clássicos de dígrafos.

Quando no início do século 20 (1930), os cenários analisados através da teoria dos grafos passaram a apresentar aumento significativo no número de nós, mas principalmente na ligação entre eles, estes grafos foram denominadas Redes Complexas. Então, Redes Complexas também são grafos cuja estrutura é formada por nós interligados (BARABASI, A.L., 2003).

A Figura 24 apresenta um exemplo de grafo simples. Nesta representação, os nós apresentam baixo grau de clusterização e baixa distribuição de graus, e são ligados pelas arestas de forma unidirecional.

⁴ Disponível em: <http://www.oakland.edu/enp> (último acesso em 04/07/16)

Figura 24 – Exemplo de grafos simples

Fonte: Autor

A Figura 25 apresenta um exemplo de grafo complexo ou Rede Complexa. Nesta representação, os nós apresentam maior grau de clusterização, maior distribuição de graus, maior correlação de graus e maior coeficiente de aglomeração, quando comparados ao grafo da Figura 24, e são ligados pelas arestas, de forma bidirecional.

Figura 25 – Exemplo de grafo complexo

Fonte: Autor

As Redes Complexas nasceram da necessidade de se analisar o comportamento dos indivíduos em relação à sociedade que pertenciam ou a influência da sociedade no comportamento individual. Nessas pesquisas, os indivíduos eram representados pelos nós e a interação deles com a sociedade a que pertenciam, representado pelas arestas. Tais pesquisas, ainda pouco desenvolvidas, concentravam-se no nó mais frequentemente acessado e seu respectivo número de conexões.

Objetivavam identificar, por exemplo, indivíduos que melhor se relacionavam com outros indivíduos da mesma sociedade, os mais influentes e a relação entre o bom relacionamento e a influência.

Porém, grandezas estatísticas mais elaboradas, passaram a melhor caracterizar as Redes Complexas.

Com a evolução significativa da capacidade de processamento dos computadores interligados em rede, aliada ao avanço da tecnologia da informação como um todo, pesquisas outrora focadas em pequenas sociedades (Redes Complexas com alguns milhares de nós, algumas centenas de arestas), passam a considerar milhões ou ainda bilhões de nós e arestas, diferentes das anteriores pela abrangência, robustez e consequente complexidade de análise.

Redes robustas (grandes quantidades de nós e arestas), como a rede internacional de computadores (*Internet*), podem ter seu modelo representado por Redes Complexas, onde, por exemplo, para determinada análise, os computadores pertencentes a esta rede são representados pelos nós e os meios que os interligam, representados pelas arestas.

Dentro da própria rede *Internet*, pode-se representar indivíduos de determinada rede social e seu grau de interatividade com seus pares e/ou com a rede toda.

Redes como cadeias alimentares, cadeias metabólicas, redes neuronais, redes sociais, redes tecnológicas etc, também são exemplos de Redes Complexas.

4.2 O PARADIGMA DAS REDES COMPLEXAS

As redes aleatórias foram, por muito tempo, consideradas como o único modelo de Rede Complexa (BARABASI, A.L., 2003; NEWMAN, M.E.J., 2003).

A quebra do paradigma, que passou a classificar a rede como Rede Complexa ou não (Rede Regular), ocorreu primeiramente, da comparação das redes aleatórias com as então recentes, redes do mundo real (quaisquer Redes Complexas que não possam ser consideradas como Rede Complexa aleatória).

As Redes Complexas que representam situações cotidianas (mundo real), embora distintas das redes aleatórias, foram consideradas Redes Complexas, por também se enquadrarem na mesma definição de Rede Complexa, ou seja, apresentarem grandezas estatísticas com valores que caracterizam-na como Rede Complexa. Embora semelhantes pela definição, o principal fator de diferenciação entre estes dois tipos de Redes Complexas, aleatórias e do mundo real, é a robustez (quantidade de nós, arestas e conexões), porém não o único.

Redes Complexas não seguem um padrão estrutural regular, embora não exista um consenso no que tange essa regularidade e consequentemente, uma definição formal sobre este padrão.

De forma geral, Redes Complexas são estruturas peculiares, não observadas em redes regulares, dinâmicas, que possuem leis próprias de formação e propriedades específicas que as caracterizam (NEWMAN, M.E.J., 2003), (COSTA, L., et al., 2007).

4.3 GRANDEZAS ESTATÍSTICAS QUE CARACTERIZAM AS REDES COMPLEXAS

A análise das grandezas estatísticas de uma Rede Complexa permite concluir-se acerca de suas propriedades e consequentemente, do comportamento do sistema por ela modelado.

A seguir, apresenta-se um resumo das principais grandezas estatísticas diretamente ligadas às Redes Complexas do modelo proposto nesta Tese. Na sessão 6.3, apresentam-se detalhes dessas grandezas, seus parâmetros, assim como a justificativa de uso.

Grau de entrada e saída: Quantifica as arestas que chegam e saem de um nó, ou seja, grau de entrada e grau de saída, respectivamente. Por definição, o somatório do grau de entrada e do grau de saída, define o grau de conexão k de um nó (NEWMAN, M., et al., 2006).

Distribuição de Graus: O grau de determinado nó em uma Rede Complexa é a medida do número de arestas que se conectam a ele. Trata-se de uma Função Densidade de Probabilidade, que indica a probabilidade de determinado nó ter grau fixo (NEWMAN, M., et al., 2006).

Por exemplo, nas Redes Complexas do tipo aleatórias, a distribuição de graus obedece à distribuição de *Poisson*. Nas demais Redes Complexas, a distribuição de graus segue a lei de potência, ou seja, a função matemática associada, apresenta decaimento exponencial (KLEIN, J.S., 2011).

Coeficiente de Clusterização: O cálculo do Coeficiente de Clusterização indica quanto os nós de uma vizinhança de determinado nó, estão conectados entre si. Este valor varia entre 0 (zero) e 1 (um), onde 0 corresponde à vizinhança desconectada (totalmente dependente do nó analisado para se manter conectada) e 1 corresponde à vizinhança totalmente conectada (totalmente independente do nó analisado para se manter conectada) (NEWMAN, M., et al., 2006).

4.4 MODELOS DE REDES COMPLEXAS

Nesta seção, são abordados os três modelos teóricos de Redes Complexas, ou seja, Redes Complexas aleatórias e a subdivisão das Redes Complexas classificadas até então como Redes Complexas do mundo cotidiano.

4.4.1 Redes Aleatórias

Inicialmente desenvolvida por Solomon e Rapoport (SOLOMONO, R.; RAPOPORT, A., 1951) e posteriormente estudada por Erdős e Rényi (ERDÖS, R., RÉNYI, A, 1959), corresponde ao modelo mais simples de Rede Complexa, onde arestas não direcionadas são adicionadas de forma aleatória entre os nós existentes, ou seja, os nós são conectados aleatoriamente a uma proporção de arestas conhecidas.

Segundo Rényi (ERDÖS, R., RÉNYI, A, 1959), quão mais robusta a rede, ou seja, quão maior a quantidade de nós e arestas, maior a probabilidade de ser aleatória.

A Equação 1 mede, segundo Rényi (ERDÖS, R., RÉNYI, A, 1959), quão uma rede complexa tem característica aleatória.

$$u = \frac{C/Crg}{l/lrg} \quad (\text{Equação 1})$$

Na Equação 1, C é o coeficiente de clusterização médio, l a distância média, considerando todos os nós existentes na rede estudada, Crg e lrg são, respectivamente, os valores estimados para C e l , função do volume de dados armazenado na Rede Complexa. Se o resultado desta equação se aproximar de 1 (um), a rede analisada é dita aleatória, caso o resultado seja muito maior que 1 (um), a rede será caracterizada como rede de pequeno mundo (WALSH, T., 1999).

A Figura 26 apresenta um exemplo de um trecho de Rede Complexa do tipo aleatória. Neste exemplo, como apresentado na Equação 2, u aproxima-se de 1 (um) por conta das distâncias curtas ($lrg \sim l$) e do alto grau de clusterização ($Crg \sim C$).

Figura 26 – Exemplo de modelo de Rede Complexa aleatória

Fonte: Autor

O cálculo de u , apresentado a seguir na Equação 2, supõe que, no grafo apresentado pela Figura 26, C tenha valor igual a 1, Crg tenha valor igual a 0,98, l tenha valor igual a 1 e lrg tenha valor igual a 0,99.

$$u = \frac{C/Crg}{l/lrg} = \frac{1/0,98}{1/0,99} \sim 1,01 \quad (\text{Equação 2})$$

4.4.2 Redes Pequeno-mundo

Redes Complexas ditas pequeno-mundo, assumiram importância fundamental no estudo das Redes Complexas, quer por permitir a análise do comportamento de alguns sistemas naturais, quer pela possibilidade de observar-se de forma acentuada, o valor da grandeza Distribuição de Graus, importante indício de surgimento de redes ditas de escala livre (NEWMAN, M., et al., 2006).

Este modelo de Rede Complexa, a princípio nasceu da necessidade de se modelar redes sociais para o estudo da proliferação de doenças (WATTS, D.J.; STROGATZ, S.H., 1998).

(WATTS, D.J.; STROGATZ, S.H., 1998), contemporâneos de Erdős e Rényi, propuseram um padrão semelhante das redes aleatórias, onde grande parte das conexões ocorrem entre nós mais próximos. A este subconjunto de nós mais próximos conectados, denominou-se pequeno-mundo.

Observando-se as redes representativas das situações cotidianas (mundo real), duas características principais levaram à criação de Redes Complexas ditas pequeno-mundo:

A primeira característica é o fato da distância média entre os nós crescer logarithmicamente com o número total de nós, ou seja, à medida que nós são inseridos, suas distâncias crescem lentamente.

A segunda característica é o alto coeficiente de clusterização médio dos nós, ou seja, a vizinhança de um nó é altamente conectada entre si.

Nesse modelo teórico, a distância média entre dois nós quaisquer da rede não ultrapassa um pequeno número de nós desta mesma rede (BUCHANAN, M., 2002).

Assim, redes de pequeno-mundo se caracterizam pela conexão entre nós através de um caminho mínimo, também conhecido por caminho geodésico. A este caminho mínimo, estarão associadas o menor número possível de arestas.

Segundo (NEWMAN, M., et al., 2006), nestas Redes Complexas o coeficiente de clusterização médio de um nó é grandeza importante na sua caracterização, por ter ligação direta com o caminho geodésico médio, e é calculado pela relação da quantidade de conexões existentes entre os vizinhos do nó e a máxima quantidade de conexões possíveis na rede.

A Figura 27 apresenta um exemplo de rede complexa do tipo pequeno-mundo, contendo 8 (oito) nós.

Figura 27 – Exemplo de Rede Complexa do tipo *pequeno-mundo*

Fonte: Autor

Considerando-se o grafo da Figura 27, observa-se a existência de 12 (doze) ligações unidirecionais em um total possível de 28 (vinte e oito), e uma ligação bidirecional, em um total possível de 56 (cinquenta e seis). Observa-se também, em relação ao nó 1, por exemplo, arestas com pesos distintos, (arestas 1-2, 1-7 e 1-8=0,12 e 1-3=0,96), denotando assim, maior ou menor relação funcional (relevância da ligação) entre o nó 1 e seus nós adjacentes.

O cálculo do Coeficiente de Clusterização (importante grandeza estatística para Redes Complexas do tipo pequeno-mundo), para um determinado nó de um grafo, pode ser obtido pela relação do número de conexões dos vizinhos deste nó e o número possível de conexões entre os vizinhos deste nó.

A principal limitação deste modelo de Rede Complexa, está no fato de nós conectados entre si não estarem necessariamente conectados a outros nós integrantes de outro subconjunto (pequeno-mundo) da mesma rede, influenciando assim, o cálculo do caminho geodésico médio. Desta forma, apenas nós com arestas estabelecidas são considerados.

4.4.3 Redes Livres de Escala

Redes Complexas do tipo livres de escala, da mesma forma que as Redes Complexas do tipo pequeno-mundo, apresentam crescimento logarítmico da distância entre os nós, à medida que o número de nós da rede aumenta.

Diferentemente das Redes Complexas aleatórias, onde o número de nós é fixo e as arestas são inseridas aleatoriamente, as redes livres de escala crescem dinamicamente não de forma aleatória mas sim, sob o conceito de ligação preferencial, conceito este que rege a forma como novas arestas são inseridas na rede.

Matematicamente, quando um novo nó é adicionado à rede complexa do tipo escala livre, a probabilidade deste novo nó se ligar a um nó de grau mais elevado (grandeza estatística Distribuição de Graus), é proporcionalmente maior do que se ligar a um nó de grau menos elevado (BARABÁSI, A.L., 2003).

Outra forma de se entender a ligação preferencial, é a tendência de um nó recente se conectar a um nó menos recente com alto grau de conectividade, ditos *label* ou *hub*. Por consequência, a rede na qual estes nós, recente e menos recente, estão inseridos, acaba por conter poucos *labels* e muitos nós com poucas conexões.

Então, redes que apresentam esta característica são denominadas redes livre de escala ou ainda, redes de escala livre. Redes livres de escala são redes cuja distribuição de graus segue a lei de distribuição de potência (KLEIN, J.S., 2011).

Segundo (NEWMAN, M.E.J., 2003), se observa a presença destas redes em vários segmentos como em redes metabólicas, redes *Web* e em redes de artigos científicos.

A Figura 28 apresenta a evolução de uma rede livre de escala, a medida que um novo nó é inserido (círculo com diâmetro maior).

Figura 28 – Evolução de Redes Complexas do tipo Livres de Escala

Fonte: Autor, adaptado de BARABASI, (2003)

Por exemplo, este novo nó inserido pode representar novo artigo científico, em um contexto de artigos científicos pré-existent, e a evolução da Rede Complexa na qual este novo artigo foi inserido, pode representar a aderência deste novo artigo a artigos já existentes da mesma linha de pesquisa e/ou do mesmo instituto de pesquisas e/ou do mesmo grupo de pesquisadores etc.

4.5 APLICAÇÕES DAS REDES COMPLEXAS

No campo da física teórica, por exemplo, mais especificamente da mecânica quântica, redes ditas de *Bose-Einstein* utilizam-se dos conceitos de Redes Complexas para estudarem uma fase da matéria formada por elementos denominados bósons (partículas de *spin* inteiro), fase esta observada apenas em temperaturas muito baixas, próximas do zero absoluto (- 273 °C).

Segundo esta teoria, próximo destas temperaturas, uma grande massa de átomos atinge seu menor estado quântico, permitindo assim que efeitos quânticos sejam macroscopicamente visualizados (MOKARZEL, S.G., 2000).

A Figura 29 apresenta o crescimento da massa de átomos, anteriormente mencionada, a medida que a temperatura dessa massa tende ao zero absoluto.

Figura 29 – Condensado de *Bose-Einstein*

Fonte: Autor, adaptado de MOKARZEL, S.G.(2000)

Nesta aplicação, nós representam massas de átomos que atingem o menor estado quântico e as arestas, a interferência entre as massas destes átomos. Assim, a medida que novos nós vão surgindo ou desaparecendo, a Rede Complexa que representa o fenômeno se reconfigura, se recharacteriza por conta da variação dos valores das grandezas estatísticas do fenômeno e o fenômeno como um todo é analisado, objetivo primeiro da modelagem.

Em um outro exemplo, os autores (ANTIQUERA, L., et al. 2005a) aplicam conceitos de Redes Complexas em outra área do conhecimento humano: modelar textos com o intuito de avaliar sua qualidade literária.

Analogamente, um texto corresponde a Rede Complexa, cada palavra deste texto corresponde a um nó desta Rede Complexa e a relação entre pares de palavras, corresponde as arestas direcionadas.

Neste exemplo, a grandeza distribuição de pesos é a grandeza estatística mais representativa do fenômeno analisado pois, cada aresta contém seu peso e este peso pondera a quantidade de vezes que determinada palavra ocorre no texto.

Para melhor se entender como uma Rede Complexa representa um texto, segue um texto do poeta Carlos Drumond de Andrade, denominado “no meio do caminho” (ANTIQUERA, L., et al. 2005b).

No meio do caminho tinha uma pedra
 tinha uma pedra no meio do caminho
 tinha uma pedra
 Nunca me esquecerei desse acontecimento
 na vida das minhas retinas tão fatigadas
 Nunca me esquecerei que no meio do caminho
 tinha uma pedra
 tinha uma pedra no meio do caminho
 no meio do caminho tinha uma pedra

A Figura 30 mostra a Rede Complexa correspondente ao poema “no meio do caminho”.

Figura 30 – Rede representativa do poema de Carlos Drumond de Andrade

Fonte: Autor, adaptado de Antiqueira (2005b)

A Rede Complexa correspondente ao texto “no meio do caminho”, é representada matematicamente por uma matriz quadrada de ordem N , denominada Matriz de Adjacências W , onde N , neste exemplo, corresponde ao número de palavras distintas.

Os pesos das arestas vão sendo incrementados na matriz W a medida que as palavras correlatas (pares de palavras) são lidas da Rede Complexa.

Para melhor avaliação da capacidade de se analisar textos por meio de Redes Complexas, grandezas estatísticas tais como graus de saídas dos dígrafos³, média do coeficiente de aglomeração de cada nó e o caminho mínimo médio de todos os pares de nós, também são extraídas.

Resumidamente, mais alguns exemplos de aplicação das Redes Complexas são mostrados a seguir.

4.5.1 Redes sociais

Redes sociais são construídas para se estudar o relacionamentos entre pessoas. Nestas Redes Complexas, os nós representam as pessoas e as arestas suas relações. Essas relações podem revelar probabilisticamente futuras amizades, doenças, casamentos, negócios, fóruns de pesquisa, contatos etc.

(LILJEROS, F., et al., 2001).

4.5.2 Redes de informação

Redes de informação constroem-se a partir de bases de conhecimento formal. Nestas Redes Complexas, os nós representam informações e as arestas a relação entre as informações.

Em (REDNER, S., 1998), foi estudado um modelo de Rede Complexa para representar as citações entre artigos acadêmicos.

³ Como em um dígrafo, a média dos graus de entrada coincide com a média dos graus de saída, somente uma das médias é considerada.

As características dessas Redes Complexas permitiram entender a dependência entre a grandeza Distribuição de Graus dos artigos, descrita por uma lei de distribuição de potência (KLEIN, J.S., 2011) e o *ranking* de classificação de acordo com o número de citações (PRICE, D.D.S., 1965).

4.5.3 Redes tecnológicas

Redes tecnológicas são Redes Complexas utilizadas para modelar a geração e/ou distribuição de recursos tais como: água, eletricidade, transportes, telefonia (fixa), *Internet* (cabramento e roteadores), entre outros.

Em (GOVINDAN,R., TANGMUNARUNKIT, H., 2000), foi proposto um modelo de Rede Complexa para aumentar a fidelidade dos roteadores. O mesmo estudo também faz uma análise sobre o mapa físico da *Internet*.

4.5.4 Redes biológicas

Alguns sistemas biológicos tais como sistema vascular, sistema nervoso, sistema circulatório entre outros, são modelados como Redes Complexas.

Como exemplo, em (SPORNS, O., 2002), foi analisado o córtex cerebral de primatas, através do estudo das propriedades de sua respectiva Rede Complexa.

5 APLICAÇÕES RELACIONADAS

Aplicações ligadas a dados não estruturados, majoritariamente os dados oriundos da plataforma *Web*, são em potencial, fontes de demanda ao modelo proposto por este trabalho.

O tráfego de dados oriundo de mídias distintas (*Cross-Media*), assim como o conceito da interação entre objetos distintos via *Internet* (Internet das Coisas), se relacionam com este trabalho, porém não de forma tão direta como as aplicações *Web*.

5.1 WEB

A *Web*, *World Wide Web*, surge em 1989 e representa importante marco na era da informação, se não, o principal.

Nas últimas duas décadas, período aproximado do surgimento comercial em grande escala de sistemas operacionais providos de *interface* gráfica, a *Web* vem evoluindo consideravelmente. Com essa evolução, a interação entre os usuários tomou formas até então desconhecidas, mudando assim a dinâmica dos relacionamentos pessoais e profissionais.

As redes sociais, por exemplo, criaram fatores propícios ao compartilhamento de dados, impulsionando a propagação da informação de forma rápida e ampla através de um modelo acessível, em qualquer lugar do mundo.

Embora se observe maturidade quando do uso de aplicações *Web*, ainda existe grande dificuldade por parte das instituições, em entender aplicações práticas das mídias sociais, sobretudo as que contém dados não estruturados, no âmbito interno.

Aumento da produtividade e maior eficácia operacional por conta da melhoria no fluxo interno de informações, são exemplos da utilização adequada desta plataforma, hoje sub-dividida em 3 (três) partes: *Web 1.0*, *Web 2.0* e *Web 3.0*.

Segundo (O'REILLY, T., 2005), disseminador do termo *Web 2.0*, a principal diferença quando comparada ao modelo anterior, *Web 1.0*, está no conceito de *Web* como plataforma, envolvendo nesta plataforma *wikis*, aplicativos baseados em folksonomia (forma particular de marcar a informação), tecnologia da informação e principalmente, redes sociais. Ainda segundo o autor, esta nova versão não se caracterizava somente por conter inovações tecnológicas mas sim, pela forma interativa de utilização por parte de usuários e desenvolvedores.

Em 2006, Tim Beners-Lee, cientista britânico criador da plataforma *Web* (BERNERS-LEE, T., 1997), introduziu o conceito de *Web* semântica⁴, onde mostrou através de um *framework* comum, ser possível o compartilhamento e reutilização de dados por diferentes aplicações, usuários, comunidades e instituições. Desta forma, segundo Lee, os dados também poderiam ser acessados e entendidos pelos próprios computadores.

Amit Agarwal (AGARWAL, A., 2009) compartilha as idéias de Lee e entende a *Web 3.0* como *Web* semântica, a introdução da inteligência artificial nas máquinas para estabelecer maior eficiência na interação com humanos, e a interatividade móvel gerada por *notebooks*, *tablets* e *smartphones*.

As plataformas *Web*, sobretudo as presentes em aparelhos móveis, através de navegadores versáteis e de alto desempenho, evidenciam a importância de interfaces navegacionais interativas e/ou colaborativas, principalmente quando da interação com dados distribuídos não estruturados (CASANOVA, M.A., MOURA, A.V., 1999).

A junção da plataforma *Web* aos conceitos de Redes Complexas, sobretudo as Redes Complexas que armazenam e disponibilizam dados não estruturados, cada qual com suas peculiaridades, corroboram para que a base tecnológica, na qual se apoia a proposta deste trabalho, seja construída de forma fundamentada e seja tecnologicamente viável.

⁴ Do francês *sémantique*, semântica é o ramo da linguística que estuda o significado das palavras

5.2 BIG DATA

Opondo-se ao que a própria denominação sugere, a tecnologia *Big Data* não é apenas caracterizada pelo grande volume de dados (SCHÖNBERGER, V.M., CUKIER, K, 2013).

Big Data representa, sob condições extremas de tempo, espaço e estrutura, uma nova maneira de armazenar, interligar e disponibilizar dados com volumes gigantescos, estruturas diferentes oriundas de fontes distintas e tempos reduzidos.

Big Data tende a descobrir, através por exemplo, da mineração heurística de dados não estruturados, relacionamentos escondidos e seus significados quando da análise de riscos financeiros, produtividade industrial, pureza de processos siderúrgicos etc.

Big Data é apoiado em três conceitos simultâneos e dependentes:

- Velocidade
- Volume
- Variedade

Velocidade – Para algumas aplicações, como detecção de fraudes, autorizações de crédito, análise financeira de ações em bolsas de valores, análise de diagnósticos médicos etc, a latência transacional de sistemas históricos analíticos (OLAP- *On Line Analytical Processing*) como os observados em estruturas *D.W.* (*Data Warehouse*) e *L.D.W.* (*Logical Data Warehouse*), somada a cargas de dados não atualizadas, podem, por exemplo, fazer a diferença entre maior ou menor incremento financeiro ao negócio ou ainda, a viabilidade ou não deste (SCHÖNBERGER, V.M., CUKIER, K, 2013).

Sistemas apoiados em tecnologia *Big Data* tendem a fornecer informações de forma mais completa e rápida, por se apoiarem em modelos de dados pós relacionais do tipo *NoSQL*, cuja escalabilidade e desempenho são fatores nativos e prioritários.

Volume – O volume médio de informação é da ordem de Terabytes (TB) e em alguns casos, Petabytes (PB) (GARCIA, M., 2014). O *site Google*, apoiado no modelo de dados *NoSQL* do tipo documento, contabilizou em Setembro de 2013, 12 Terabytes de *tweets* gerados diariamente. Em 2012, contabilizou o tráfego de cerca de 2,9 Exabytes (EB) e para 2020, estima um volume em torno de 40 EB (GARCIA, M., 2014). O *site Google* processa diariamente, cerca de três bilhões de pesquisas em todo mundo, sendo desse total, cerca de 15% inéditas. Rastreia vinte bilhões de *sites* diariamente e armazena desses, aproximadamente, 100 PB de informação. Importante destacar que, segundo o *Google*, mais de 90% de toda a informação gerada até o momento, está concentrada nos últimos dois anos e estão diretamente ligadas a dados não estruturados (GARCIA, M., 2014).

Os *sites Amazon* e *Facebook*, apoiados no modelo de dados *NoSQL* do tipo coluna, movimentam mensalmente cerca de 50 PB de informação (GARCIA, M., 2014).

O *site Youtube*, apoiado no modelo de dados *NoSQL* chave-valor registra, a cada minuto, *uploads* de vídeos cujo somatório aproxima-se de 48h (GARCIA, M., 2014).

O *site Tweeter*, apoiado no modelo de dados *NoSQL* do tipo coluna, contabiliza diariamente mais de 500 milhões de mensagens com média de 5.700 *TPS* (*Tweets Per Second*), cujo maior volume registrado até então aproxima-se de 143.200 *TPS* (GARCIA, M., 2014).

Variedade – A variedade neste contexto, classifica a informação em duas categorias: estruturada e não estruturada.

Informação estruturada – Informação que possui algum padrão ou formato conhecido, de modo a permitir através dos meios existentes, sua conversão para o padrão relacional de dados (NETO, J.G., 2008).

São exemplos de informações estruturadas, arquivos texto, estruturas de bancos de dados, planilhas eletrônicas etc. (SCHÖNBERGER, V.M., CUKIER, K, 2013).

Informação não-estruturada – Analogamente, não possuem padrão ou formato pré-estabelecido. São exemplos de informações não-estruturadas, vídeos, mensagens instantâneas, páginas *Web*, áudios, fotos etc.

Partindo do pré-suposto que a proposta primeira desta Tese enfatiza um modelo de dados pós relacional onde dados estruturados e não estruturados coexistam, integrando-se, e que destes dados, a parcela não estruturada é a representante majoritária do todo, os conceitos definidos como bases da tecnologia *Big Data*, especialmente a parcela ligada a variedade de estruturas, mais que agregar valor, se confundem com os objetivos deste trabalho.

5.3 CROSS-MEDIA

Conceitualmente falando, *Cross-Media* ou mídia cruzada define-se como a distribuição de conteúdos sob mídias distintas, de forma que haja de cada mídia, a contribuição necessária e suficiente para que a interação e o direcionamento do usuário para as próximas mídias, ocorram.

(BOUMANS, J.,2004) elencou cinco peculiaridades da mídia cruzada:

- objetiva uma produção integrada;
- envolve mais de uma mídia;
- o conteúdo é entregue em múltiplos meios/plataformas tais como computadores pessoais, telefones móveis, televisores, televisores interativos etc;
- mais de uma mídia é necessária para apoiar uma única mensagem / narrativa / meta;
- a mensagem / narrativa / meta é distribuída em múltiplas plataformas e a interação apoiada pode acontecer nestas diferentes plataformas.

Dispositivos de fácil acesso, existentes atualmente, como telefones celulares e suas múltiplas variações, televisores, *tablets*, vídeo-games etc, ganham a cada dia novas funcionalidades de modo a obterem cada vez mais, maior capacidade de processamento e inter-comunicação (multimodalidade).

Assim, o desenvolvimento de aplicações que permeiem diversas mídias apresenta uma tendência natural, sobretudo quando da necessidade de processar dados não estruturados, haja vista o grande volume destes dados no contexto *cross-media*, quando comparados com o volume de dados estruturados.

Porém, esta tendência exige maior controle no que tange à produção e sobretudo, a distribuição de conteúdos por conta da sua complexidade.

Partindo do princípio que existem tipos distintos de conteúdos, a escolha da mídia mais adequadas para cada conteúdo é fundamental.

Por exemplo, (FINCHER, S.,2003), propõe parâmetros de um algoritmo para equacionar a escolha de mídias mais adequadas para serviços do governo Brasileiro. O algoritmo é baseado nas propriedades de cada mídia e nas características do serviço, tais como sincronismo, restrição de acesso, cobertura, segmentação etc.

Peculiaridades que também destacam a importância da escolha mais adequada da mídia e de sua aplicação de forma integrada, podem ser observadas nos aspectos ligados a complementação, corroboração e concomitância (MIYAMARU, F., et al., 2008).

5.4 INTERNET DAS COISAS

A sigla *IoT* (*Internet of Things*) não tem aceitação universal. Terminologias como *IcO* (*Internet connected Objects*), *WoT* (*Web of Things*), Computação Pervasiva, Computação Ubíqua e *M2M* (*Machine to Machine*), são também encontrados na literatura especializada com o mesmo significado básico, ou seja, comunicação entre objetos por rede e comunicação de objetos com a *Internet*.

Com a constante evolução da capacidade computacional em equipamentos cada vez mais compactos, hoje o conjunto *transponder / tag*, unidades teóricas básicas da *IoT*, deixou de ser apenas uma etiqueta eletrônica para se transformar em verdadeiras centrais de processamento, recebendo, analisando, armazenando, processando e transmitindo dados, sobretudo os dados não estruturados.

Desta forma, aplicações mais sofisticadas surgiram, como por exemplo a monitoração de variáveis de estado (temperatura, umidade, pressão, velocidade, vazão, radiação etc) em equipamentos de difícil acesso, quando do acoplamento de sensores a estes equipamentos, integrando o funcionamento destes equipamentos a sistemas que acessam dados e/ou tomam decisões.

Com esta integração, aplicações baseadas em tecnologia RFID, *Cross-Media* e *Mobile* passam a compor também o conceito de *IoT*, enfatizando por conta do grande volume presente no contexto *IoT*, os dados não estruturados.

Neste universo diferente e recente, a interação e iteração com as coisas da vida cotidiana, permitem que processos sejam influenciados pelo contexto das informações (AMAZONAS, J.R., 2011).

5.5 SUBSÍDIOS FORNECIDOS PELA REVISÃO

Os estudos outrora citados, sobretudo os referentes aos modelos de dados pós relacionais, suas ligações com dados de diferentes estruturas, com a teoria matemática dos grafos complexos e com as tendências da plataforma *Web*, hoje representada pelas demandas agressivas dos conceitos *Big Data*, em muito colaboraram para o direcionamento desta Tese, servindo inclusive, no caso das Redes Complexas, de subsídio à ela.

Plataformas funcionais tais como a *Web*, ressurgem e são rapidamente adaptadas a novas tendências de acesso, ainda mais rápidas, versáteis, portáteis e completas. A natureza do dado, hoje caracterizado pela conjunção de estruturas simples e estruturas complexas, estas últimas mais recentes, demandam modelos de dados mais rápidos, versáteis e sobretudo escaláveis, onde volumes até então inimagináveis são realidade em Redes Complexas, como por exemplo, as redes sociais.

Modelos de dados matematicamente consagrados, tais como o modelo de dados relacional, representado por seus respectivos gerenciadores de bancos de dados relacionais, por mais de três décadas hegemônicos tanto em sistemas centralizados como em sistemas distribuídos, ganham nova roupagem por conta de demandas agressivas, inimagináveis na ocasião de suas criações originais.

Estas demandas, impostas pela nova natureza do dado, onde desempenho, coexistência / processamento de diferentes estruturas e escalabilidade horizontal, são pré-requisitos à continuidade dos sistemas apoiados por esses modelos.

O uso de modelos de dados no apoio a tomadas de decisão, a preocupação com a integridade do dado de falha, a difusão de dados a volumes e velocidades expressivas e sobretudo, as tentativas de suprir as demandas apresentadas por novas estruturas de dados díspares da estrutura na qual foi baseado o projeto do modelo de dados relacional, hoje adaptado à natureza destes novos dados, enfatizam a real necessidade de um modelo de dados híbrido, consistente mesmo que eventualmente e de bom desempenho, onde estruturas diferentes coexistam, integrando-se, se necessário.

Os modelos de dados até então propostos e apresentados, embora evidenciem a necessidade de se representar também dados não estruturados, concentram-se direta ou indiretamente na estrutura pré-existente no modelo de dados relacional.

Estes modelos, geram ou não uma estrutura provisória e/ou intermediária, também relacional, e embora confirmem alta escalabilidade horizontal por em sua maioria abrirem mão da consistência transacional *ACID*, propriedade encontrada no modelo de dados relacional, exigem invariavelmente a presença de especialistas de dados, interação de usuários ou ainda não raramente, ajustes nas aplicações, para remodelarem e então disponibilizarem os resultados obtidos.

Neste contexto, vale destacar alguma exceção aos modelos de dados apoiados em grafos, que, embora também sejam escaláveis, gozem de consistência e sejam altamente disponíveis, aos moldes dos modelos anteriormente citados, não integram estruturas distintas.

A demanda atual, representada majoritariamente por Redes Complexas do tipo redes sociais e acentuada pelo uso crescente de aparelhos móveis, condiciona o desenvolvimento deste trabalho, cuja proposta está em um modelo de dados pós relacional, apoiado na teoria matemática dos grafos, eventualmente consistente, condicionalmente disponível e híbrido, a medida que abriga e integra, sob demanda, dados estruturados e não estruturados.

6 MODELO NAVEGACIONAL DINÂMICO, VISANDO A IMPLEMENTAÇÃO DA INTEGRAÇÃO INTER-ESTRUTURAL DE DADOS

Este capítulo apresenta o Modelo Navegacional Dinâmico, MND, propriamente dito, a Rede Complexa que caracteriza o modelo de dados proposto e a ligação deste modelo com dados estruturados e não estruturados.

Apresentam-se também os trabalhos relacionados ao tema proposto e duas aplicações do Modelo Navegacional Dinâmico.

O propósito primeiro do MND, é a coexistência com a integração inter-estrutural, de dados com diferentes estruturas, sob os conceitos e regras das Redes Complexas.

Nos modelos de dados apresentados no *Capítulo 2*, a integridade referencial, quando existente, é assegurada por restrições e a integridade transacional, pelas propriedades *ACID*. No Modelo Navegacional Dinâmico, a integridade referencial não é considerada e a integridade transacional, eventualmente consistente, complementada pela integração inter-estrutural de dados.

6.1 VISÃO GERAL

A demanda por armazenamento de volumes crescentes de dados e a disponibilização desses dados em tempos tão reduzidos quanto possíveis, é realidade considerada nesta Tese.

À medida que o crescimento significativo do volume de dados é contínuo e majoritariamente atribuído aos dados não estruturados, a navegação em modelos escaláveis que os armazena e a integração de dados não estruturados com dados estruturados, em tempos computacionais compatíveis com as regras dos sistemas que os acessa, torna-se iminente, quer pela necessidade de se identificar em Redes Complexas detalhes menos evidentes delas, mas intrínsecos a elas e confirmados pelas grandezas estatísticas que as qualificam, quer pela demanda por tempos menores no acesso aos resultados de eventuais demandas por integração de dados.

Modelos escaláveis horizontalmente, como o *MND*, capazes de absorverem aumento expressivo do volume armazenado, sem que essa absorção cause algum prejuízo a seus dados por interferir nas regras sistêmicas que os controlam, são cada vez mais necessários.

O *MND* reúne características únicas, ora por integrar, sob demanda, dados com diferentes estruturas, ora por condicionar a Rede Complexa que representa o modelo de dados à consistência eventual e não, à propriedades *ACID*, como os demais modelos apoiados em grafos.

Neste ponto vale destacar as principais contribuições apresentadas pelo Modelo Navegacional Dinâmico, ou seja:

- A integração inter-estrutural de dados proposta, a qual possibilita a busca e comparação de dados oriundos de estruturas tradicionais (dados estruturados) com dados multimídia (dados não estruturados);
- A integração inter-estrutural utiliza o Modelo de Grafos, sendo que tal forma de uso não foi identificada em outros trabalhos do gênero;
- No Modelo de Grafos utilizado no Modelo Navegacional Dinâmico, não se adotam as propriedades *ACID*, mas sim as propriedades *BASE*, como nos modelos Chave-Valor, de Colunas e de Documentos;
- Propõe-se, além das matrizes já utilizadas no Modelo de Grafos, uma nova matriz (a Matriz de Estruturas), além de se propor uma nova maneira de se atribuir valores na Matriz de Pesos;
- Comparando-se o Modelo Navegacional Dinâmico com diversos trabalhos acadêmicos, pode-se dizer que nenhum deles apresenta características que possam estabelecer uma comparação direta com o modelo aqui descrito (descrição na seção 6.7);
- Finalmente, comparando-se o Modelo Navegacional Dinâmico com alguns produtos comerciais atualmente disponíveis, pode-se destacar que tais produtos, ou não implementam uma integração inter-estrutural, ou não possibilitam uma escalabilidade horizontal significativa, além de fazerem uso do modelo relacional de dados (descrição na seção 6.7).

6.2 PREMISSAS DO MODELO NAVEGACIONAL DINÂMICO

As premissas do *MND*, são a conjunção da navegabilidade com a integração inter-estrutural de dados, ambas em tempo de execução da consulta, sobre a Rede Complexa que representa o modelo de dados.

A navegabilidade é apoiada na leitura dos valores das grandezas estatísticas da Rede Complexa que armazena os dados representados pelo modelo de dados, sendo essa Rede Complexa representada pelas Matrizes de Adjacências, de Incidências, de Pesos e de Estruturas. Essas matrizes, armazenadas no dicionário de dados, são construídas pelo *MND* à medida que nós e/ou arestas são inseridos e/ou removidos da Rede Complexa que representa o modelo de dados.

A integração inter-estrutural, deve ser entendida como o reconhecimento de informações, em dados não estruturados, que, após identificadas e tratadas, possam ser consideradas estruturadas, diferentemente de outras propostas que sugerem funcionalidades semelhantes, e que são apresentadas na seção 6.7.

6.3 AS GRANDEZAS ESTATÍSTICAS DO MODELO NAVEGACIONAL DINÂMICO

Independentemente de ser caracterizado como Rede Regular ou Rede Complexa, a medida que um grafo se forma, os nós e as ligações entre eles (arestas) precisam ser registradas em matrizes, tais como Matriz de Adjacências (matriz que registra a ligação ou não entre os nós do grafo), Matriz de Vizinhanças (matriz complementar à Matriz de Adjacências, que permite verificar a distância entre nós não adjacentes), Matriz de Distâncias (matriz que permite verificar a distância entre todos os nós do grafo), Matriz de Incidências (matriz que registra as arestas que ligam os nós) e Matriz de Pesos (matriz que permite verificar a relação funcional entre dois nós).

Neste trabalho, as matrizes que representam a Rede Complexa do modelo de dados proposto, são as matrizes de Adjacências, de Incidências, de pesos e de Estruturas. Da leitura dos valores registrados nessas matrizes, obter-se-á informações acerca das grandezas estatísticas *Grau de entrada e saída* e *Distribuição de graus*, caracterizando assim, a Rede Complexa do modelo proposto.

Caracterizar o grafo do modelo de dados proposto como Rede Complexa, é associar, por conta da interpretação dos valores de suas respectivas grandezas estatísticas, registradas nas matrizes de Adjacências, de Incidências e de Estruturas, a qual modelo teórico de Rede Complexa esse grafo mais se aproxima.

A seguir, justifica-se, utilizando-se como exemplo, o grafo apresentado na Figura 31, grafo este contendo os nós *Labes* L1 a L8 ($n=8$) e as arestas a1 a a14 ($m=14$), porque as grandezas estatísticas *Grau de entrada e saída*, *Distribuição de graus* e *Coeficiente de Clusterização* são necessárias à caracterização da Rede Complexa do Modelo Navegacional Dinâmico.

Figura 31 – Grafo contendo 8 nós do tipo *Label* e 14 arestas

Fonte: Autor

6.3.1 Grau de entrada e saída

Considerando o grau de entrada e saída (*OD*), este tem importância para o modelo proposto, à medida que verifica, através da leitura dos valores da Matriz de Incidências ($n \times m$), nós $\times$ arestas, e da Matriz de Adjacências ($n \times n$), nós $\times$ nós, respectivamente, a existência ou não, no modelo, dos nós mencionados na consulta e, em caso positivo, se existe ou não conexão entre estes nós, e o tipo de conexão.

a-) A Matriz de Adjacências

A Figura 32 apresenta a Matriz de Adjacências *AL*, com seus respectivos índices, correspondentes aos nós *Labels* do grafo apresentado na Figura 31.

Figura 32 – Matriz de Adjacências *AL*

Nós Labels		L1	L2	L3	L4	L5	L6	L7	L8
<i>AL</i> =	L1	0	0	1	0	0	0	1	1
	L2	1	0	0	0	0	0	0	0
	L3	1	0	0	1	1	0	0	0
	L4	0	1	0	0	1	1	0	0
	L5	0	0	0	0	0	1	1	0
	L6	0	0	0	0	0	0	1	0
	L7	0	0	0	0	0	0	0	0
	L8	0	0	0	0	0	1	0	0

Fonte: Autor

Na Matriz de Adjacências, ao se definir 2 (dois) nós quaisquer, o valor 1 (um) representa conexão entre estes dois nós, independentemente do peso de suas arestas, e o valor 0 (zero), representa desconexão entre estes 2 (dois) nós.

A menos que exista(m) laços, ou seja, nós conectados a si próprios (possibilidade não considerada no modelo proposto), a diagonal principal da Matriz de Adjacências, sempre apresentará valores nulos (zeros).

Observa-se na matriz AL , por exemplo, completa desconexão entre os nós *Labels* $L7$ e $L8$, diferentemente dos nós *Labels* $L1$ e $L3$, cuja conexão é bidirecional.

Assim, o nó cujo valor da grandeza estatística *Grau de entrada e saída* seja nulo, ou seja, nó existente mas desconectado, inviabiliza a consulta que o menciona.

b-) A Matriz de Incidências

A Figura 33 apresenta a Matriz de Incidências IL , com seus respectivos índices, correspondentes aos nós *Labels* e arestas, do grafo apresentado na Figura 31.

Figura 33 – Matriz de Incidências IL

		Arestas													
		$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$	$a8$	$a9$	$a10$	$a11$	$a12$	$a13$	$a14$
$IL =$	$L1$	1	1	0	0	0	0	1	0	0	1	1	0	0	0
	$L2$	1	0	0	0	0	0	0	0	0	0	0	1	0	0
	$L3$	0	0	0	0	0	1	0	0	1	1	1	0	0	0
	$L4$	0	0	0	0	1	1	0	0	0	0	0	1	1	0
	$L5$	0	0	0	1	1	0	0	1	1	0	0	0	0	0
	$L6$	0	0	1	1	0	0	0	0	0	0	0	0	1	1
	$L7$	0	0	1	0	0	0	1	1	0	0	0	0	0	0
	$L8$	0	1	0	0	0	0	0	0	0	0	0	0	0	1
		Nós Labels													

Fonte: Autor

Como em um grafo, cada aresta é incidente a exatamente 2 (dois) nós, ao observar-se a Matriz de Incidências IL , observa-se dois valores 1 (um) a cada coluna dessa matriz.

Analogamente, linhas da Matriz de Incidências que apresentam apenas valores 0 (zero), representam *Labels* isolados que, para o modelo proposto, correspondem a *Labels* desconectados ou com distância infinita a qualquer que seja o outro nó considerado.

c-) A Matriz de Pesos

Na Matriz de Pesos, os valores da grandeza estatística peso, associados às arestas que ligam nós *Labels* ligados aos nós que armazenam dados não estruturados, são constantes, e calculados pela expressão ($Wp = L \cdot m^3$), onde L é o número de nós *Labels* da Rede Complexa considerada e m , o número de arestas que os interligam. Justifica-se essa expressão, por conta do seu valor ser único e diretamente associado à estrutura do grafo em questão. Por essa expressão, o valor calculado para Wp no exemplo em desenvolvimento é :

$$Wp = (L \cdot m^3) = (8 \cdot 14^3) = 21.952 \quad (\text{Equação 3})$$

É necessário mencionar que o custo computacional cresce, a medida que a ordem da matriz de Pesos e o parâmetro Wp , crescem.

Figura 34 – Matriz de Pesos PL

Arestas

	$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$	$a8$	$a9$	$a10$	$a11$	$a12$	$a13$	$a14$
$L1$	21952	21952	0	0	0	0	21952	0	0	21952	21952	0	0	0
$L2$	2	0	0	0	0	0	0	0	0	0	0	3	0	0
$L3$	0	0	0	0	0	3	0	0	2	2	2	0	0	0
$L4$	0	0	0	0	5	4	0	0	0	0	0	6	1	0
$L5$	0	0	0	6	6	0	0	2	3	0	0	0	0	0
$L6$	0	0	2	3	0	0	0	0	0	0	0	0	7	2
$L7$	0	0	1	0	0	0	7	4	0	0	0	0	0	0
$L8$	0	8	0	0	0	0	0	0	0	0	0	0	0	3

Nós Labels

Fonte: Autor

A Matriz de Pesos é importante ao *MND*, a medida que contém os valores dos pesos das arestas incidentes aos nós *Labels* que armazenam dados não estruturados e também o valor dos pesos das arestas que incidem sobre outros nós, *Labels* ou não, que contém dados estruturados, apresentando assim o panorama completo da Rede Complexa quanto a essa grandeza.

d-) A Matriz de Estruturas

Criada por esta proposta, a Figura 35 apresenta a Matriz de Estruturas *EL*, correspondente às arestas incidentes apenas aos nós *Labels* com dados não estruturados, do grafo apresentando na Figura 31. Observa-se, pela leitura dos valores das linhas da matriz *EL*, que apenas o nó *Label L1* armazena dados não estruturados.

Figura 35 – Matriz de Estruturas *EL*

The diagram illustrates the structure of the *EL* matrix. A box labeled "Arestas" (Edges) points to the column headers, and a box labeled "Nós Labels" (Node Labels) points to the row headers.

	<i>a1</i>	<i>a2</i>	<i>a3</i>	<i>a4</i>	<i>a5</i>	<i>a6</i>	<i>a7</i>	<i>a8</i>	<i>a9</i>	<i>a10</i>	<i>a11</i>	<i>a12</i>	<i>a13</i>	<i>a14</i>
<i>L1</i>	1	1	0	0	0	0	1	0	0	1	1	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: Autor

6.3.2 Distribuição de Graus

Esta grandeza estatística (*SD*) tem importância para o *MND* por condicionar a relevância do nó (*n*) existente na Rede Complexa, com a quantidade de arestas (*m*), conectadas a este nó.

No modelo proposto, por definição, arestas que conectam nós que armazenam dados não estruturados, são identificadas por apresentarem valores da grandeza estatística peso acentuadamente superiores ($Wp = L \cdot m^3$), quando comparados a arestas que conectam nós que armazenam dados estruturados.

Considerando que as arestas, no modelo proposto, representam as regras que relacionam nós de grau mais alto (nós com grande quantidade de arestas conectadas a ele, ditos nós *Labels*), com nós de grau mais baixo (nós com pequena quantidade de arestas conectadas a ele), comparar o valor da grandeza estatística *Distribuição de graus* de um nó (ou simplesmente grau do nó) com o valor do grau médio do grafo a que este nó pertence, estabelece no modelo proposto, em conjunto com a grandeza estatística peso, prioridades nas relações funcionais entre os dados.

A leitura das linhas da Matriz de Incidências, atribui ao grau do nó em questão, valor diretamente ligado ao somatório de valores 1 (um), existentes nessa linha.

Por exemplo, observando-se a matriz *IL*, apresentada na Figura 33, observa-se que o nó *Label L1* possui grau 5 (cinco), que os nós *Labels L3, L4, L5 e L6* possuem grau 4 (quatro), que o nó *Label L7* possui grau 3 (três) e que os nós *Labels L2 e L8* possuem grau 2 (dois).

A Equação 4 apresenta uma forma simplificada de se calcular o valor médio da grandeza estatística *Distribuição de graus*, atribuindo-se à *Distribuição de graus* de cada *Label* do grafo, o valor do grau médio (***k***) do grafo todo.

$$k = 2 \cdot \frac{m}{n} \quad (\text{Equação 4})$$

No grafo apresentado na Figura 31, $m=14$, $n=8$. Logo, $k = 3,5$.

$$k = 2 \cdot \frac{14}{8} = 3,5$$

Embora o valor de ***k*** não seja considerado pelo *MND*, vale notar a diferença, por exemplo, do valor atribuído ao grau do nó *Label L1* (5,0), somatório dos valores 1 (um) da linha do nó *Label L1* da Matriz de Incidências, apresentada na Figura 33, e o valor que a este nó *Label L1* seria atribuído (3,5), se considerado fosse o grau médio do grafo no qual está inserido.

Caso o *MND* considerasse o valor de ***k*** a despeito do valor lido na Matriz de Incidências, a diferença entre esses valores seria suficiente para que, equivocadamente, um nó de grau superior fosse considerado como sendo um nó de grau inferior e vice-versa, comprometendo assim a relação funcional entre os nós *Labels* e seus subordinados.

6.3.3 Coeficiente de Clusterização

Assim como a grandeza estatística *Distribuição de graus* representa determinado nó quanto à sua conectividade, a grandeza estatística *Coeficiente de Clusterização* (*CC*) representa a conectividade da vizinhança de determinado nó.

Embora também calculada pelo Modelo Navegacional Dinâmico, essa grandeza estatística tem importância secundária para o modelo proposto, por mensurar quão conectada está a vizinhança de um determinado nó e conseqüentemente, quão o caminho para se chegar a este nó, via sua vizinhança, é mais viável, ou não.

Saber acerca do *CC*, é entender qual a tendência de arestas vizinhas do nó analisado se conectarem a nós distantes do nó analisado e sobretudo, a relevância do nó mais distante, no mesmo grafo, para o nó analisado.

A Equação 5 apresenta o cálculo da grandeza estatística *Coeficiente de Clusterização*, considerando o *Label L1*, do grafo da Figura 31.

$$CC1 = \frac{\text{número de arestas vizinhas ao Label L1}}{\text{número possível de arestas vizinhas ao Label L1}} \quad (\text{Equação 5})$$

$$CC1 = \frac{0}{5} \sim 0$$

Caso o valor associado ao ***CC1*** seja igual a 1 (um), conclui-se que todos os nós da vizinhança do nó *L1* estão conectados entre si. Se o valor do ***CC1*** for 0 (zero), significa que somente o nó *L1* mantém sua vizinhança conectada.

Analogamente, os *Coeficientes de Clusterização* dos outros nós do grafo, são calculados e apresentados no grafo apresentado pela Figura 36.

O cálculo do **CCg**, *Coeficiente de Clusterização* global, ou seja, para o grafo todo, é o resultado da média aritmética dos *Coeficientes de Clusterização* de cada nó deste grafo, também apresentado no grafo da Figura 36.

A Equação 6 apresenta o cálculo do *Coeficiente de Clusterização* global para o grafo da Figura 36.

$$CCg = \frac{(0 + 0 + 0,33 + 0,20 + 0,50 + 0,20 + 0 + 0)}{8}$$

$$CCg \sim 0,15$$

(Equação 6)

Figura 36 – Exemplo de grafo contendo os valores de *Sd* e *CC*'s

Fonte: Autor

Resumidamente, caracterizar uma Rede Complexa, é associar a relevância de suas grandezas estatísticas e seus respectivos valores, ao volume de dados armazenados nesta Rede Complexa para só então, concluir-se acerca de suas peculiaridades, comportamentos e prováveis tendências.

Assim, pelas definições dos modelos teóricos de Redes Complexas, apresentados na seção 4.4.2 (Redes Pequeno-mundo), somadas a relevância das grandezas estatísticas *Grau de entrada e saída*, *Distribuição de graus* e *Coeficiente de clusterização*, classifica-se a Rede Complexa do *MND* como Rede Complexa do tipo pequeno-mundo, quer por permitir analisar a tendência de evolução do *MND* para outras configurações de Redes Complexas, por exemplo, do tipo livres de escala, sem que nesta situação haja perda das propriedades funcionais do modelo por conta do acréscimo de novos *Clusters*, quer por também permitir analisar-se o comportamento de sistemas naturais, tais como redes sociais.

6.4 O MODELO NAVEGACIONAL DINÂMICO

No *MND*, Modelo Navegacional Dinâmico, dados, estruturados e não estruturados, são armazenados em estruturas de grafos ditas Redes Complexas. Para que o Modelo Navegacional Dinâmico opere eficientemente, modelo esse apresentado em blocos pela Figura 37 e em detalhes, pela Figura 38, faz-se necessário que critérios de iterações e interações sejam observados, ora com a verificação de parâmetros básicos, ora com o cálculo e posterior leitura dos valores das grandezas estatísticas associadas ao modelo de dados, registrados nas matrizes que representam a estrutura da Rede Complexa na qual o *MND* está inserido.

Figura 37 – Diagrama interativo do Modelo Navegacional Dinâmico

Fonte: Autor

Figura 38 – Composição do Modelo Navegacional Dinâmico

Fonte: Autor

Concomitantemente à construção da Rede Complexa que representa o modelo *MND*, o *MND* cria, a princípio, em seu dicionário de dados, a Matriz de Adjacências *AL* (Figura 32) e a Matriz de Incidências *IL* (Figura 33), cujos conteúdos são, respectivamente, os nós *Labels* e as arestas incidentes nesses nós.

Após a Matriz de Adjacências *AL* (Figura 32) e a Matriz de Incidências *IL* (Figura 33) criadas, o *MND* cria, também no dicionário de dados e baseado na Matriz de Incidências *IL* (Figura 33), a Matriz de Pesos *PL* (Figura 34) e a Matriz de Estruturas *EL* (Figura 35).

Figura 35A criação dessas quatro matrizes é necessário e suficiente, a medida que as verificações ligadas a existência ou não de nós *Labels* e arestas, conexão entre nós *Labels*, peso das arestas incidentes aos nós *Labels* e a estrutura do dado armazenado em nós de graus distintos, inerentes ao funcionamento do *MND*, ocorrem através da leitura dos índices dessas matrizes.

Assim, são apresentadas a seguir, de forma segregada e na sequência que ocorrem, ou seja, da identificação dos nós e arestas (etapa 1), à apresentação dos resultados obtidos (etapa C), as etapas correspondentes ao funcionamento do *MND*.

Após a consulta ser submetida, o *MND* verifica a existência de todos os nós *Labels* e todas as arestas mencionados na consulta. Essa verificação ocorre com a leitura dos nós *Labels* e das arestas, encontrados na Matriz de Incidências *IL* (Figura 33), independentemente da estrutura do dado armazenado nesses nós *Labels* (etapa 1).

Caso essa verificação apresente divergências entre os nós *Labels* e arestas, mencionados na consulta, e os nós *Labels* e arestas, existentes no *MND*, a consulta é interrompida e finalizada, apresentando ao usuário executante as ressalvas associadas ao impeditivo encontrado, no sentido de, na consulta, ter-se mencionado indevidamente nós *Labels* e arestas inexistentes no *MND*.

Em caso positivo, o *MND*, direciona a consulta à etapa (2).

Na etapa (2), o *MND* verifica, primeiramente, se a consulta menciona ou não, estruturas distintas de dados, condição necessária mas não suficiente para que haja a integração inter-estrutural de dados. Essa verificação ocorre, observando-se as linhas da Matriz de Estruturas *EL* (Figura 35), correspondentes ao nós *Labels* mencionados na consulta, já identificados na etapa (1).

Como exemplo, destacado na Figura 35, a presença do valor 1 (um) na linha correspondente ao nó *Label L1*, mostra a existência de dados não estruturados armazenados em nós subordinados a esse *Label*. Analogamente, a presença do valor 0 (zero) nas linhas correspondentes aos nós *Labels L2* a *L8*, mostra a existência de dados estruturados, armazenados em nós subordinados a esses nós *Labels*.

Caso a consulta mencione somente uma estrutura de dados, ou seja, dados estruturados ou dados não estruturados, o *MND* não aciona o Algoritmo Navegacional Dinâmico, *AND*, responsável pela integração inter-estrutural de dados.

Se somente dados estruturados forem mencionados na consulta, o resultado é apresentado (etapa C) e a consulta, finalizada.

Se somente dados não estruturados forem mencionados na consulta, o *MND* não aciona o *AND* e a consulta é interrompida e finalizada, apresentando ao usuário executante as ressalvas associadas ao impeditivo encontrado, haja vista que o *MND* não considera a integração de dados não estruturados.

Caso a consulta mencione estruturas distintas de dados, confirmando assim a provável demanda por integração inter-estrutural de dados, se faz necessário verificar a viabilidade desta provável integração, ou seja, se os dados estão dispostos no *MND*, da forma como são mencionados na consulta (etapa 3). Essa verificação ocorre comparando-se os nós *Labels* e arestas, da consulta, com os valores obtidos na leitura dos campos (intersecção de linhas com colunas) da Matriz de Pesos *PL* (Figura 34), identificando ou não, no *MND*, a ligação das arestas com os nós *Labels*, assim como a estrutura do dado armazenado nesses nós *Labels*.

Confirmada a viabilidade da integração inter-estrutural de dados, o *MND* direciona a consulta à etapa (4). Caso haja divergências entre a consulta e o *MND*, o *MND* interrompe e finaliza a consulta, apresentando ao usuário executante as ressalvas associadas ao impeditivo encontrado.

Na etapa (4), o *MND* verifica se os nós *Labels*, mencionados na consulta, são ou não distintos. Essa verificação ocorre, observando-se, na Matriz de Pesos *PL* (Figura 34), se as arestas em questão (colunas da Matriz de Pesos *PL*, Figura 34), incidem ou não sobre o mesmo nó *Label* (linhas da Matriz de Pesos *PL*, Figura 34).

Em caso positivo, o *AND* é acionado, a integração inter-estrutural de dados é realizada e o resultado da integração inter-estrutural de dados, armazenado.

Se na mesma consulta não houver mais demanda por integração inter-estrutural de dados (etapa 5), o resultado da consulta é apresentado e a consulta, finalizada. Caso contrário, o *MND* direciona a consulta à etapa (3), refazendo parcialmente o fluxo até que todas as integrações sejam executadas para somente assim, finalizar-se a consulta.

Caso os nós *Labels* sejam distintos, o *MND* direciona a consulta à etapa (4a).

A etapa (4a) verifica, se os nós *Labels* distintos estão ou não, conectados entre si.

Essa verificação ocorre observando-se ou não, o valor 1 (um), em pelo menos um, dos dois campos da Matriz de Adjacências *AL* (Figura 32), que relacionam os nós *Labels* envolvidos.

Caso os nós *Labels* distintos estejam conectados entre si, o *AND* é acionado, a integração inter-estrutural de dados é realizada e o resultado da integração inter-estrutural de dados, armazenado. Se na mesma consulta não houver mais demanda por integração inter-estrutural de dados (etapa 5), o resultado da consulta é apresentado e a consulta, finalizada. Caso contrário, o *MND* direciona a consulta à etapa (3), refazendo parcialmente o fluxo até que todas as integrações sejam executadas para somente assim, finalizar-se a consulta.

Caso os dados a serem integrados pertençam a nós *Labels* distintos desconectados entre si, o *MND* não aciona o *AND* e a consulta é finalizada, apresentando ao usuário executante as ressalvas associadas ao impeditivo encontrado.

É apresentado na Figura 39, o detalhamento da etapa *B* do *MND*, ou seja, o *AND*, responsável pela integração inter-estrutural de dados. Nesta etapa, o dado estruturado é passado como parâmetro ao mecanismo de reconhecimento de padrões (BHATTACHARYYA, A., 1943) (MAHALANOBIS, P. C. , 1936), quer diretamente, como valor a ser identificado no dado não estruturado, quer indiretamente, através do conteúdo de um nó que o armazene. De posse do dado estruturado, o mecanismo de reconhecimento de padrões procura, no dado não estruturado, a informação a ser considerada estruturada para na sequência, fornecer o resultado dessa comparação, ou seja, a integração inter-estrutural dos dados.

Figura 39 – A integração inter- estrutural

Fonte: Autor

6.5 ACESSO AO MODELO NAVEGACIONAL DINÂMICO

O acesso ao *MND*, pode ser realizado por intermédio da linguagem *Cypher*.

Embora outras linguagens como por exemplo o *SQL* também possam ser utilizadas para acesso ao modelo proposto, a linguagem *Cypher*, com suas cláusulas de definição e manipulação de dados apoiadas por métodos *CRUD* (*Create, Retrieve, Update e Delete*) (DAMAS, L., 2007) apresenta vantagens se utilizada como linguagem de acesso a dados armazenados em grafos, como no modelo de dados proposto.

Baseada nas linguagens *Haskell, Python* e *SPARQL* (W3C, 2008), a linguagem *Cypher* é adequada para acesso a dados armazenados em estruturas apoiadas em grafos (NEO4J, 2014), como a estrutura proposta neste trabalho, por ser declarativa, possuir *síntaxe* simples e enfatizar em suas cláusulas, a coleção de dados a acessar e não, como acessar, diferente por exemplo, do *SQL*, utilizado nos gerenciadores de bancos de dados relacionais.

Para a proposta deste trabalho, a linguagem *Cypher* apresenta vantagens também quando comparada a linguagens mais frequentemente utilizadas para acessar dados em grafos, tais como as linguagens *Java, Gremlin, JRuby* etc, pois, nestas linguagens, diferentemente da linguagem *Cypher*, as consultas necessariamente, precisam ser modificadas quando a estrutura do modelo for alterada (mudança no número de nós, mudança no valor das grandezas estatísticas, mudança nas arestas etc).

O modelo apresentado na Figura 40, é utilizado como exemplo para demonstrar a estrutura básica da linguagem *Cypher*. Neste exemplo, a modelagem objetiva associar, direta ou indiretamente, os laços de amizade (aresta *Amigo*) entre os componentes (nós) do grafo que representa o modelo de dados.

Figura 40 – Modelo apoiado em grafos, acessado pela linguagem *Cypher*

Fonte: Autor

A consulta Q1, escrita em linguagem *Cypher* e executada sobre o modelo da Figura 40, no gerenciador de banco de dados *Neo4J* (NEO4J, 2014), gerenciador este baseado em um modelo de dados apoiado em grafos como o modelo proposto por este trabalho, apresenta o componente *Salim* e todos seus amigos indiretos.

```
MATCH (salim {name: 'Salim'})-[amizade:amigo]->()-[amizade:amigo]->(amgind)
RETURN Salim.name, Amgind.name
```

Consulta Q1

A seguir, o resultado da consulta Q1.

Salim.name

Amgind.name

"Salim"

"Yda"

"Salim"

"Farid"

2 rows

A consulta Q2, análoga a consulta Q1, porém um pouco mais elaborada, apresenta um filtro (ordenação). Este filtro restringe a consulta Q2 a todos amigos diretos de quaisquer componentes existentes no modelo da Figura 40, cujo nome comece com a letra Y.

```
MATCH (amg)-[amizade:amigo]->(amgdir)
WHERE amg.name IN ['Youssef', 'Salim', 'Sálua', 'Yda', 'Farid'] AND
amgdir.name =~ 'Y.*'
RETURN amg.name, amgdir.name
```

Consulta Q2

A seguir, o resultado da consulta Q2.

amg.name	amgdir.name
"Salim"	"Youssef"
"Sálua"	"Yda"

2 rows

A consulta Q3, não diretamente ligada ao grafo apresentado na Figura 40, destaca a linha que contém a integração inter-estrutural de dados.

Nesse exemplo, a integração ocorre entre o nó *nome*, do *Label Pesquisador* e o nó *filme*, do *Label Publicações*, respectivamente.

```
MATCH (pesq,pess)-[:publicar],[:pesquisar]->(pesqdir)
WHERE pesq.name IN ['Pesquisador', 'Publicações'] AND
Titu.value >= '20.*'
MERGE (name.Pesquisador,Publicações {name: 'nome', 'filme'})
RETURN pesq.name, titu.name, titu.count
```

Consulta Q3

6.6 PRIMEIRO EXEMPLO DE APLICAÇÃO DO MODELO NAVEGACIONAL DINÂMICO

Uma aplicação provável é utilizada para se demonstrar conceitualmente, uma aplicação do modelo proposto. Esta aplicação, doravante denominada *Hc*, consiste em modelar um sistema que apresente o índice *H*, do centro de pesquisas *c*.

Esse índice, tem seu valor diretamente associado ao número de citações de publicações de determinado pesquisador, do centro de pesquisas *c*, por outros pesquisadores do mesmo centro de pesquisas *c*.

Em termos práticos, o centro de pesquisas *c* objetiva classificar seus pesquisadores quanto a suas respectivas produções científicas, além de verificar a influência destas produções científicas em produções científicas de outros pesquisadores do mesmo centro de pesquisas. Para isso, entende-se necessário organizar as produções científicas dos últimos 20 anos quanto ao autor principal da publicação, a qual departamento do centro de pesquisas *c* esse autor pertence, qual o título da produção científica, o ano da publicação, a instituição patrocinadora do evento científico onde ocorreu a publicação da produção científica e sobretudo, quantos acessos, oriundos do centro de pesquisas *c*, foram feitos a estas produções científicas.

A Figura 41 representa o grafo dos nós *Labels* da aplicação *Hc*.

Figura 41 – Modelagem da aplicação *Hc*

Fonte: Autor

No modelo da Figura 41, são armazenados a princípio, apenas dados estruturados. Nesse modelo, a regra de modelagem assegura que, caso o pesquisador não tenha artigos publicados, será ainda assim considerado pesquisador, porém, outros funcionários, integrantes do centro de pesquisas c, que não são pesquisadores, não podem ter publicações científicas registradas.

A Figura 42, a Figura 43 e a Figura 44 apresentam, cada qual, um exemplo detalhado de dados armazenados nos nós subordinados aos nós *Labels* do modelo apresentado na Figura 41.

Figura 42 – Nó *Label Pesquisador*, suas relações e conteúdos

Fonte: Autor

Figura 43 – Nó *Label Publicações*, suas relações e conteúdos

Fonte: Autor

Figura 44 – Nó *Label Funcionários*, suas relações e conteúdos

Fonte: Autor

O nó *Label Pesquisador* é vizinho do nó *Label Publicações* e do nó *Label Funcionários*, que por sua vez, é vizinho do nó *Label Pesquisador* mas não é vizinho do nó *Label Publicações*.

Para efeitos de demonstração conceitual do *MND*, o modelo da Figura 45 apresenta uma modificação, quando comparado ao modelo da Figura 41. Essa modificação, fruto da alteração da demanda *Hc* anterior, e necessária para que o modelo anterior passe a operar a integração inter-estrutural de dados, consiste em remover, do nó *Label Publicações*, a aresta *EVENTO* (Figura 43) com seus respectivos nós de grau inferior e em seu lugar, inserir a aresta *FILME* (Figura 46), com seus respectivos nós de grau inferior. Com isso, o grafo que a princípio armazenava apenas dados estruturados (Figura 41), passa a armazenar também, dados não estruturados (Figura 45).

O destaque na Figura 45, evidencia esta modificação, ou seja, a coexistência de dados não estruturados e estruturados no mesmo modelo de dados, ligados, neste exemplo, pelo dígrafo *Publicar*, arestas *da1* e *da2*, e consequentemente, a possibilidade de integração inter-estrutural destes dados, quando do acesso a informações abrigadas nos nós *Labels Pesquisador* e *Publicações*.

Figura 45 – Modelo apoiado em grafos, contendo também dados não estruturados

Fonte: Autor

A Figura 46, apresenta um exemplo detalhado de dados armazenados no nó *Label Publicações*, pertencente ao modelo da Figura 45.

Figura 46 – Nó *Publicações* modificado, suas relações e conteúdos

Fonte: Autor

Na Figura 47, o destaque indica que consultas envolvendo os nós *Labels Pesquisador* e *Publicações*, podem demandar integração inter-estrutural de dados, e que esta integração ocorreria, por exemplo, no caso da consulta necessitar identificar em determinado filme, os registros internos dos pesquisadores citados nele, navegando assim entre *Pesquisador.RI* (parcela estruturada da integração) e *Publicações.FILME* (parcela não estruturada da integração), respectivamente.

A consulta apresentada sobre o modelo da Figura 47 poderia ser ainda mais específica necessitando, por exemplo, identificar no filme *Filme1.avi*, o pesquisador com R.I. 2000. Neste caso, a navegação ocorreria entre *Pesquisador.RI.2000* e *Publicações.FILME.filme1.avi*.

Figura 47 – Representação do conceito da integração inter- estrutural

Fonte: Autor

As matrizes que representam o trecho do modelo da Figura 45, apresentado na Figura 47, são apresentadas a seguir.

Figura 48 – Matriz de Adjacências AL , do grafo apresentado na Figura 47

	Le	Lu	$ne1$	$ne2$	$ne3$	$nu4$	$nu5$	$nu6$
$AL =$	Le	0	1	1	1	0	0	0
	Lu	1	0	0	0	1	1	1
	$ne1$	1	0	0	0	0	0	0
	$ne2$	1	0	0	0	0	0	0
	$ne3$	1	0	0	0	0	0	0
	$nu4$	0	1	0	0	0	0	0
	$nu5$	0	1	0	0	0	0	0
	$nu6$	0	1	0	0	0	0	0

Fonte: Autor

Figura 49 – Matriz de Incidências IL , do grafo apresentado na Figura 47

	<i>da1</i>	<i>da2</i>	<i>a1</i>	<i>a2</i>
<i>Le</i>	1	1	1	0
<i>Lu</i>	1	1	0	1
<i>ne1</i>	0	0	1	0
<i>IL = ne2</i>	0	0	1	0
<i>ne3</i>	0	0	1	0
<i>nu4</i>	0	0	0	1
<i>nu5</i>	0	0	0	1
<i>nu6</i>	0	0	0	1

Fonte: Autor

Figura 50 – Matriz de Pesos PL , do grafo apresentado na Figura 47

	<i>da1</i>	<i>da2</i>	<i>a1</i>	<i>a2</i>
<i>Le</i>	20	20	20	0
<i>Lu</i>	128	128	0	128
<i>ne1</i>	0	0	15	0
<i>PL = ne2</i>	0	0	15	0
<i>ne3</i>	0	0	15	0
<i>nu4</i>	0	0	0	128
<i>nu5</i>	0	0	0	128
<i>nu6</i>	0	0	0	128

Fonte: Autor

Figura 51 – Matriz de Estruturas *EL*, do grafo apresentado na Figura 47

	<i>da1</i>	<i>da2</i>	<i>a1</i>	<i>a2</i>
<i>Le</i>	0	0	0	0
<i>Lu</i>	1	1	0	1
<i>ne1</i>	0	0	0	0
<i>EL = ne2</i>	0	0	0	0
<i>ne3</i>	0	0	0	0
<i>nu4</i>	0	0	0	1
<i>nu5</i>	0	0	0	1
<i>nu6</i>	0	0	0	1

Fonte: Autor

6.7 SEGUNDO EXEMPLO DE APLICAÇÃO DO MODELO NAVEGACIONAL DINÂMICO

Esta seção apresenta o sistema computacional *SACA*, Sistema de Avaliação Contínua do Aluno de Administração, utilizado, após supostas modificações, para elucidar uma forma de se utilizar em termos práticos, o *MND*.

O sistema *SACA*, desenvolvido em meados do ano de 2014 pelo autor desta Tese, é escrito em linguagem *Java* e acessa exclusivamente dados estruturados, em uma base de dados relacional *Oracle 11G*.

Os dados não estruturados utilizados no sistema *SACA* (arquivos com imagens), são armazenados externamente ao banco de dados, e são gerenciados pelo sistema operacional.

Na avaliação dos usuários do sistema *SACA*, ou seja, os gestores acadêmicos do Centro Universitário da FEI, os resultados obtidos com a utilização do sistema são positivos.

A princípio, restrito a um total aproximado de 150 (cento e cinquenta) alunos, o sistema *SACA* objetiva identificar variações no rendimento acadêmico dos alunos do curso de Bacharelado em Administração de empresas, do Centro Universitário da Fundação Educacional Inaciana, *FEI*, *Campus* São Bernardo do Campo.

Uma vez constatadas, através do uso do sistema SACA, variações no rendimento acadêmico, quer por interpretações gráficas, quer por associações analíticas de outras ocorrências cujas variações de rendimento apresentam valores próximos, identifica-se, também através do sistema, a origem mais provável das variações no rendimento acadêmico e com isso, as correções devidas.

No sentido de analisar a origem destas variações de rendimento acadêmico, parâmetros sistêmicos tais como alunos/turmas que cursaram as mesmas disciplinas em períodos distintos ou não, no mesmo horário ou não, mais/menos frequentes, com o mesmo docente ou não, submetidos às mesmas avaliações ou não, são comparados a possíveis causas tais como troca recente e/ou frequente de docentes, docentes cujas aulas apresentam ausências mais frequentes

dos alunos, preferência dos alunos por departamentos distintos que ofereçam a mesma disciplina, frequência com que alunos questionam determinado docente etc.

As conclusões oriundas dessas comparações, externas ao sistema SACA, colaboram para que o diagnóstico acerca da variação de rendimento dos alunos, fornecido pelo sistema, seja mais abrangente e assertivo.

A Figura 52 apresenta a atual tela inicial do sistema SACA.

Observa-se na Figura 52 que, por conta de demandas posteriores à versão inicial, o sistema, além do *Campus* São Bernardo do Campo, passou a contemplar também o *Campus* São Paulo, pertencente ao mesmo Centro Universitário.

Figura 52 – Tela de *login* do sistema SACA

Fonte: Autor

Como exemplo de utilização do sistema SACA, a Figura 53 compara o aproveitamento de uma aluna, formada no ano de 2015, com o aproveitamento da turma que essa aluna pertence.

Figura 53 – Tela SACA-Aluno

Fonte: Autor

6.7.1 Demandas atualmente não suportadas pelo sistema SACA

Embora com pequenos ajustes posteriores à sua concepção, estruturais e/ou estéticos, tais como a inclusão do *Campus* São Paulo e a atualização do logotipo do Centro Universitário, às funcionalidades do sistema SACA continuam atendendo as demandas iniciais.

Supostamente, a situação seguinte, distinta da realidade na qual o sistema SACA foi concebido, apresenta um cenário novo a ser contemplado.

Imagina-se então, o sistema SACA como solução acadêmica difundida em âmbito nacional, onde a adesão de novas Faculdades, Centros Universitários, Universidades e Institutos de Pesquisas deva ocorrer em um curto período de tempo e de forma contínua. Por consequência das supostas novas demandas, o sistema SACA não terá mais seu uso restrito aos gestores acadêmicos, como na versão atual. Nesta suposta nova versão, o sistema SACA estará disponível também aos alunos, nas plataformas *Web* e *mobile*.

Por supostas exigências legais dos órgãos que fiscalizam os critérios para inserção de novos alunos no sistema SACA, a inscrição destes novos alunos deverá ocorrer ao longo de um curto e pré-determinado período e exclusivamente via *Internet*, haja vista a inviabilidade de inscrições manuais por conta do alto volume de inscrições previstas. Na ocasião da inscrição, o sistema SACA deverá, além de outras verificações, ser capaz de validar a data da inscrição e rejeitar fotografias, obrigatórias à inscrição, cujas datas no formato *dd/mm/aaaa*, sejam superiores a 2 (dois) anos.

Considerando que uma suposta solicitação de inscrição ocorreu em 9/1/2016 e que o período de inscrições está compreendido no intervalo de 00:00h de 04/01/2016 até 23:59h de 29/02/2016, o destaque da Figura 54 apresenta um exemplo de fotografia inválida, por ter data superior a 2 (dois) anos (o limite, neste caso, é 29/02/2014).

Figura 54 – Fotografia inválida à inscrição em 9/1/2016

Fonte: Autor

As supostas novas demandas atribuídas ao sistema SACA, evidenciam a necessidade desse sistema contemplar um número maior de alunos, e têm por consequências diretas significativo aumento no volume de dados, estruturados e não estruturados, além de proporcional aumento no número de acessos a esses dados, quer pela ocasião das novas inscrições, onde se faz necessário integração de dados com estruturas distintas (reconhecer na fotografia, o parâmetro estruturado data), quer pelo uso das supostas novas funcionalidades do sistema.

Observando-se o novo cenário imposto pelas supostas novas demandas, identificam-se e justificam-se os seguintes tópicos:

- A necessidade do aumento da capacidade computacional (*“...o sistema SACA deverá contemplar um número expressivamente maior de alunos e seus respectivos cursos...”*).
- A necessidade iminente de escalabilidade horizontal (*“...onde a adesão de novas Faculdades, Centros Universitários, Universidades e Institutos de Pesquisas deva ocorrer em um curto período de tempo e de forma contínua.”*).
- A necessidade de integração de dados estruturados com dados não estruturados (*“...e rejeitar fotografias, obrigatórias à inscrição, cujas datas no formato dd/mm/aaaa, sejam superiores a 2 (dois) anos.”*).
- A necessidade de versatilidade no acesso e de alta disponibilidade dos dados (*“...o sistema SACA estará disponível também aos alunos, nas plataformas Web e mobile.”*).

- A valorização da consistência eventual em prol da disponibilidade (“...o sistema SACA como solução acadêmica difundida em âmbito nacional...”)

A demanda por escalabilidade horizontal é suportada por esta proposta, a medida que os modelos orientados a grafos, apresentam maior escalabilidade horizontal, quando comparados ao modelo relacional de dados, sobretudo para grandes volumes de dados, como na suposta demanda apresentada.

A demanda por integração de dados estruturados (data no formato *dd/mm/aaaa*) com dados não estruturados (identificação do dado no formato *dd/mm/aaaa*, na imagem associada) é suportada por esta proposta, à medida que em tempo de execução, o conceito da integração inter- estrutural é aplicado, dispensando assim procedimentos apartados do modelo de dados.

Nesta proposta, uma vez implementada, a integração inter-estrutural validaria ou não a fotografia e conseqüentemente, a suposta inscrição *on-line*.

A seguir, é apresentada a aplicação do Modelo Navegacional Dinâmico, proposto nesta Tese, como uma solução para as supostas novas demandas ao sistema SACA. Para isso, se fez necessário converter o modelo Entidade-Relacionamento, no qual o sistema SACA está apoiado, em um modelo apoiado em grafos complexos, no qual se apoia esta proposta.

Detalhes do modelo Entidade-Relacionamento do sistema SACA , assim como seu correspondente modelo de grafos, podem ser encontrados nos Apêndices A B e C deste trabalho, respectivamente.

A Figura 55, apresenta a consulta Q4, feita por meio da linguagem *Cypher*, utilizada para validar ou não, a suposta inscrição de novos alunos.

Figura 55 – Consulta Q4

```
MATCH (can,canins-[:inscrição],[:inscrição]->(inscrição)
WHERE can.name IN ['Candidato', 'Candidato-inscrito'] AND
MERGE (name.Candidato, Candidato inscrito {name: 'foto','limitefoto'})
RETURN can.name, canins.name
```

Fonte: Autor

A Figura 56, apresenta apenas os *Labels* acessados pela consulta Q4, necessários a validação ou não, da suposta inscrição de novos alunos, e suficientes para demonstrar-se a utilização do modelo *MND*.

Figura 56 – Trecho do modelo de dados, acessado pela consulta Q4

Fonte: Autor

A Figura 57, apresenta um exemplo de conteúdo armazenado nos nós *Labels* *Candidato* e *Candidato-inscrito*, apresentados na Figura 56.

Figura 57 – Exemplo de conteúdo armazenado no grafo da Figura 56

Fonte: Autor

A Figura 58 apresenta a Matriz de Adjacências, do grafo apresentado na Figura 57.

Figura 58 – Matriz de Adjacências

	<i>LA</i>	<i>LB</i>	<i>n1</i>	<i>n2</i>	<i>n3</i>
<i>LA</i>	0	1	1	1	0
<i>LB</i>	1	0	0	0	1
<i>n1</i>	1	0	0	0	0
<i>n2</i>	1	0	0	0	0
<i>n3</i>	0	1	0	0	0

Fonte: Autor

A Figura 59 apresenta a Matriz de Incidências, do grafo apresentado na Figura 57.

Figura 59 – Matriz de Incidências

	<i>a1</i>	<i>a2</i>	<i>a11</i>	<i>a12</i>
<i>LA</i>	1	1	1	0
<i>LB</i>	1	1	0	1
<i>n1</i>	0	0	1	0
<i>n2</i>	0	0	1	0
<i>n3</i>	0	0	0	1

Fonte: Autor

A Figura 60 apresenta a Matriz de Pesos, do grafo apresentado na Figura 57 (no grafo completo, apresentado no *Apêndice B*, que contém o trecho apresentado na Figura 57, $L = 11$ e $m = 20$, logo $L \cdot m^3 = 88.000$).

Figura 60 – Matriz de Pesos

	<i>a1</i>	<i>a2</i>	<i>all</i>	<i>al2</i>
<i>LA</i>	88000	88000	88000	0
<i>LB</i>	88000	88000	0	19
<i>n1</i>	0	0	88000	0
<i>n2</i>	0	0	88000	0
<i>n3</i>	0	0	0	19

Fonte: Autor

A Figura 61 apresenta a Matriz de Estruturas, do grafo apresentado na Figura 57.

Figura 61 – Matriz de Estruturas

	<i>a1</i>	<i>a2</i>	<i>all</i>	<i>al2</i>
<i>LA</i>	1	1	1	0
<i>LB</i>	0	0	0	0
<i>n1</i>	0	0	1	0
<i>n2</i>	0	0	1	0
<i>n3</i>	0	0	0	0

Fonte: Autor

Como mencionado anteriormente, a consulta *Q4*, nesse exemplo, objetiva validar ou não, inscrições de novos alunos. Para isso, é necessário e suficiente que a fotografia do candidato, com data no formato *dd/mm/aaaa*, tenha no máximo dois anos, a contar de 29/02/2016, ou seja, fotografias com datas anteriores a 29/02/2014 são consideradas inválidas e consequentemente, a inscrição do candidato é rejeitada.

A seguir, descrevem-se os detalhes da execução da consulta *Q4* sobre o grafo da Figura 57, sob as regras do *MND*.

A consulta *Q4* não apresenta erros (sintáticos e léxicos) e o usuário executor da consulta *Q4*, tem as permissões de acesso solicitadas pela consulta *Q4*.

Os nós *Labels* mencionados na consulta Q4, nó *Label Candidato (LA)* e nó *Label Candidato-inscrito (LB)*, são confirmados com a leitura dos nós *Labels* da Matriz de Adjacências (Figura 58) (*etapa 1*).

Identificados no *MND* os nós *Labels LA* e *LB*, confirma-se a existência de dados não estruturados no nó *Label LA* e dados estruturados no nó *Label LB*, por observar-se as linhas da Matriz de Estruturas (Figura 61), correspondentes aos nós *Labels NA* e *NB*.

Confirmados os nós *Labels LA* e *LB* e as distintas estruturas (dados não estruturados e dados estruturados), indiretamente armazenadas por esses nós *Labels*, a leitura da Matriz de Adjacências (Figura 58), confirma que os nós subordinados aos nós *Labels LA / LB* são os nós *n1 / n2* e *n3*, respectivamente.

A leitura da Matriz de Incidências (Figura 59), identifica as arestas incidentes aos nós *n1* e *n2* (aresta *a11*, *FOTO*) e *n3* (aresta *a12*, *LIMITEFOTO*).

Na sequência, identifica-se na Matriz de Pesos (Figura 60), através dos pesos das arestas *a11* (*FOTO*) e *a12* (*LIMITEFOTO*), que os nós *n1* e *n2*, ligados ao nó *Label LA* pela aresta *a11*, armazenam dados não estruturados e que o nó *n3*, ligado ao nó *Label LB* pela aresta *a12*, armazena dados estruturados.

Identificados no *MND LA, LB, n1, n2, n3, a1, a2, a11* e *a12*, observa-se na consulta Q4, demanda por integração inter-estrutural de dados, com a aresta *a11* (*FOTO*), incidindo sobre os nós *n1* e *n2*, que armazenam os dados não estruturados, e a aresta *a12* (*LIMITEFOTO*), incidindo sobre o nó *n3*, que armazena os dados estruturados (*etapa 2* e *etapa 3*).

A leitura dos valores da Matriz de Pesos (Figura 60), mostra que os nós *n1/n2* e *n3* pertencem a nós *Labels* distintos, pois, os nós *n1* e *n2*, que armazenam dados não estruturados, se conectam ao nó *Label LA (Candidato)* via aresta *a11* (*FOTO*) e o nó *n3*, que armazena dados estruturados, se conecta ao nó *Label LB (Candidato-inscrito)* via aresta *a12* (*LIMITEFOTO*) (*etapa 4*).

A leitura dos valores da Matriz de Adjacências, (Figura 58), confirma que os *Labels* distintos *LA* e *LB*, estão conectados bidirecionalmente (dígrafo) (*etapa 4a*).

Então, a integração inter-estrutural é acionada, ou seja, a data 29/02/2014, conteúdo estruturado do nó *n3*, é passado como parâmetro ao mecanismo de reconhecimento de padrões, que a procura no conteúdo não estruturado dos nós *n1* e *n2*, ou seja, procura nos arquivos *Foto1.jpeg* e *Foto2.tiff*, datas superiores a 29/02/2014 (*etapa B*).

As

Figura 62 e 63 apresentam, respectivamente, os conteúdos armazenados nos nós *n1* e *n2*

Figura 62 – Arquivo *Foto1.jpeg*

Fonte: Autor

Figura 63 – Arquivo *Foto2.tiff*

Fonte: Autor

Como a consulta *Q4* não demanda outra integração inter-estrutural de dados (*etapa 5*), os resultados da consulta *Q4* são apresentados, ou seja, a inscrição, cuja fotografia corresponde ao arquivo *Foto1.jpeg* é rejeitada, a inscrição cuja fotografia corresponde ao arquivo *Foto2.tiff* é aceita, e a consulta *Q4*, finalizada (*etapa C*).

6.8 TRABALHOS RELACIONADOS AO TEMA PROPOSTO

Inicialmente, os estudos que fazem uso de Redes Complexas, priorizaram as ciências biológicas e o comportamento humano. Soluções desse tipo, dependentes do apoio computacional, não ultrapassam uma década de uso.

Exemplos de pesquisas ligando Redes Complexas à biologia e ao cérebro humano, podem ser encontrados em (RAJAPAKSE, I. et al., 2012), (LUCA, D.M. et al., 2006), (BELLEC, P. et al., 2006), (ACHARD, S. et al., 2006), (ACHARD, S. ; BULLMORE, E., 2007), (BULLMORE, E. ; SPORNS, O., 2009), (WANG, L. et al., 2009) e (LYNALL, M.E.; BASSETT, D.S., 2010).

Observa-se porém, ao longo dos últimos cinco anos, significativo aumento das pesquisas que utilizam Redes Complexas para solucionarem e/ou entenderem problemas de cunho geral, desde estudos ligados à geografia (TAKAHASHI, T. et al., 2011) e indústria (LOZANO, S. et al., 2012), até estudos ligados a vírus em aparelhos telefônicos móveis (WANG, P. et al., 2009), algoritmos que identificam vulnerabilidades em *links* das próprias Redes Complexas (IGOR, M. et al., 2010), análises em sistemas de *software* (SUBELI, L., BAJEC, M., 2012), modelagem de sistemas computacionais complexos (EMMERICH, T. et. al., 2013) etc.

Recentemente, (WANG, L. et al., 2012) apoiaram em Redes Complexas, um estudo sobre o núcleo (*Kernel*) do sistema operacional Linux.

São apresentadas, a seguir, pesquisas que se aproximam da proposta deste trabalho, ora por estudarem dados estruturados e não estruturados em Redes Complexas, ora por estudarem o comportamento de dados estruturados e não estruturados quanto ao acesso e armazenamento em diferentes arquiteturas e em diferentes modelos de dados.

6.8.1 A proposta de Mirzasoleiman

Apoiado na plataforma *Apache Hadoop* (DEAN, J. et. al., 2004), (MIRZASOLEIMAN, B. et. al., 2012) entendem que o algoritmo denominado *GreeDi*, proposto por eles, confere maior segurança (integridade com desempenho) quando do acesso a dados não estruturados do tipo não textuais, armazenados em bases relacionais. Para isso, os autores associam a qualidade das imagens ao desempenho das consultas que acessam essas imagens, ora armazenadas em sistemas relacionais distribuídos, ora armazenadas em sistemas relacionais centralizados. Na sequência, os autores comparam os diferentes desempenhos obtidos pelas consultas. Desta forma, pode-se dizer que (MIRZASOLEIMAN, B. et. al., 2012) realizaram pesquisas estudando a ligação de dados não estruturados com seus respectivos modelos de dados.

A proposta de (MIRZASOLEIMAN, B. et. al., 2012), se liga à proposta deste trabalho a medida que estuda a ligação de dados não estruturados com seus respectivos modelos de dados.

6.8.2 A proposta de Boguá

(BOGUÁ, M. et. al., 2013), estudam a dinâmica de propagação de processos relacionados entre si e associam a vulnerabilidade dos dados destes processos, ou seja, a distorção da informação contida nesses processos, à forma como estes dados são propagados.

Evidenciam o conceito denominado por eles difusão assimétrica, comparando, por exemplo, a dinâmica da proliferação de determinada epidemia com a dinâmica da difusão das informações preventivas acerca desta epidemia. Segundo os autores, são processos independentes, porém dinamicamente relacionados.

A proposta de (BOGUÁ, M. et. al., 2013), se liga à proposta deste trabalho, a medida que liga Redes Complexas a dados não estruturados.

6.8.3 A proposta de Gómez

Em (GOMÉZ, S. et. al., 2013), podem ser encontradas pesquisas que evidenciam a preocupação dos pesquisadores com a velocidade de difusão dos distintos tipos de dados, em Redes Complexas. Embora considerem que parte da variação na velocidade de difusão possa ser associada à estrutura do dado transmitido, priorizam ajustes na estrutura do modelo de dados envolvendo as grandezas estatísticas Coeficiente de Clusterização, Distribuição de Graus e Grau de entrada e saída, em prol de melhor desempenho nas consultas.

A proposta de (GOMÉZ, S. et. al., 2013), próxima da proposta de BOGUÁ, M. et al., se liga à proposta deste trabalho a medida que estuda a ligação de Redes Complexas a dados estruturados e não estruturados.

6.8.4 A proposta de Lombardo

(LOMBARDO, S. et al., 2012), alertam sobre problemas de desempenho quando do acesso a dados estruturados em bases de dados que operam sob as regras dos modelos de dados *NoSQL*. Segundo os autores, mostra-se ineficiente sob a óptica do desempenho, acessar dados estruturados em bases de dados *NoSQL*, mesmo que nestas bases o desempenho esteja assegurado pela inexistência de integridade referencial. Ainda segundo os mesmos autores, parte desta ineficiência ocorre, ora devido à natureza complexa dos acessos, ora devido à incompatibilidade estrutural dos dados envolvidos. Então, (LOMBARDO, S. et al., 2012), recomendam utilizar bases de dados *NoSQL*, como linguagem de acesso a dados não estruturados e não como forma de armazenamento e disponibilização destes dados.

A proposta de (LOMBARDO, S. et al., 2012), se liga à proposta deste trabalho a medida que estuda dados estruturados abrigados em modelos de dados pós-relacionais do tipo *NoSQL*.

6.8.5 A proposta de Jiang

(JIANG, W. et. al., 2012) discorrem sobre a importância da disponibilidade de dados não estruturados armazenados em bases de dados que operam sob regras do modelo relacional de dados, quando do armazenamento destes dados, em *storages*. Os autores apoiam sua pesquisa nos fundamentos do teorema *CAP*, principalmente na premissa tolerância da tolerância a falhas em sistemas distribuídos (*Partition Tolerance*), e estendem a análise utilizando o paradigma *BASE* onde, segundo este paradigma, os dados podem estar disponíveis o tempo todo, mas não necessariamente consistentes o tempo todo (eventualmente consistente). Concluem a análise comentando a integridade transacional do modelo relacional, assegurada por propriedades *ACID*.

A proposta de (JIANG, W. et. al., 2012), se liga à proposta deste trabalho a medida que estuda o desempenho quando do acesso a dados não estruturados abrigados em modelos de dados relacionais.

6.8.6 A proposta de Yafooz

Semelhante à parte final da proposta de (LOMBARDO, S. et al., 2012), (YAFOOZ, W.M.S. et. al., 2013), defendem o modelo *NoSQL* como linguagem de acesso a dados não estruturados em bases relacionais, observando o desempenho de consultas *SQL* quando do acesso a dados estruturados e dados não estruturados, para posteriormente, comparar este desempenho com o desempenho de consultas onde inexistam dados não estruturados. Embora de forma superficial, os autores também mencionam a importância da integração destas diferentes estruturas.

A proposta de (YAFOOZ, W.M.S. et. al., 2013), se liga à proposta deste trabalho a medida que estuda modelos de dados não relacionais como linguagem de acesso a dados não estruturados.

6.8.7 A proposta de Cavoto

Em um estudo apresentado à Universidade Estadual de Campinas (*UNICAMP*), (CAVOTO, P.R.N., 2015) propõe uma arquitetura híbrida que integra gerenciadores de banco de dados relacionais com gerenciadores apoiados no modelo de grafos.

Esse estudo, cuja abordagem é aplicada à biodiversidade com ênfase na *Web 3.0*, integra, em um gerenciador de banco de dados dito híbrido, as maiores potencialidades dos dois modelos de dados envolvidos.

A

Figura 64 apresenta o diagrama explicativo da arquitetura integrada, proposta por (CAVOTO, P.R.N., 2015).

Figura 64 – Arquitetura híbrida de integração E-R – Grafos

Fonte: Autor , adaptado de CAVOTO, P.R.N.(2015)

A proposta de (CAVOTO, P.R.N., 2015) se liga a proposta deste trabalho a medida que estuda a integração dos modelos de dados relacionais com os modelos de dados apoiados em grafos.

6.8.8 Comparação com Produtos Comerciais

Diferentemente de alguns produtos comerciais que, assim como o *MND*, tratam majoritariamente dados não estruturados, tais como *Oracle-Spatial*, *DB2-Spatial Extender*, *Microsoft-Spatial-Denali* e *PostGIS*, com finalidades voltadas ao geo-processamento e cujas características são resumidamente abordadas a seguir, o *MND* é apoiado em matrizes geradas pelo próprio *MND*, em seu dicionário de dados, matrizes essas necessárias e suficientes ao seu funcionamento.

Neste contexto, onde grandes volumes de dados, predominantemente não estruturados, são processados em pequenos intervalos de tempo, o *MND* prioriza, a despeito da consistência, a disponibilidade da informação, independentemente da informação estar armazenada de forma distribuída ou centralizada.

O fornecedor *Oracle*, por exemplo, através da técnica *Oracle-Spatial*, implementa em seu recente produto *Oracle 12C*, apoiado nas regras do modelo relacional de dados, o tratamento preferencial a dados geo-espaciais distribuídos, condicionando a substancial melhoria de desempenho no acesso a esses dados não estruturados, ao uso do *hardware* utilizado no processamento, assim como o sistema operacional (*ORACLE 12C DATABASE*, 2014).

Em termos práticos, a técnica *Oracle-Spatial*, através do produto *Oracle 12C*, apresenta melhor desempenho no acesso a dados geo-espaciais armazenados em nuvem, sobretudo aos dados modelados sob técnicas voltadas à *B.I.* (*Business Intelligence*), se aplicações processarem esses dados, armazenados no equipamento denominado *Exadata*, sob sistema operacional *Oracle-Linux*, ambos (*hardware* e *software*), desenvolvidos pelo fornecedor *Oracle*.

Do fornecedor *IBM*, a técnica *DB2-Spatial-Extender* também ligada ao armazenamento e disponibilização de dados geo-espaciais, armazena e processa dados estruturados e não estruturados, sem integrá-los (*DB2 Extender*, 2012).

Apoiada nas normas *ISO SQL/MM* e *Open Geospatial Consortium* (OGC), utiliza os recursos da linguagem *SQL* para análise de dados espaciais, estruturados e não estruturados, armazenados sob as regras do modelo relacional de dados.

Em termos práticos, a técnica *DB2-Spatial-Extender*, é uma extensão do gerenciador relacional *DB-2* que, através de um navegador customizado, dito *Geobrowser*, explora informações geográficas, tais como as localizações de edifícios de escritórios, de pessoas em prováveis zonas de inundação, de incêndios, queimadas etc, correlacionando essas localizações e pessoas, à atributos geométricos armazenados em entidades do modelo relacional, tais como pontos, linhas, polígonos etc.

Segundo o fornecedor *IBM*, estas peculiaridades, majoritariamente direcionadas à grandes centros urbanos, adicionam inteligência ao banco de dados, por integrarem informações espaciais aos dados de negócios.

Análogas a essa técnica, porém disponíveis em plataformas distintas do sistema operacional *Z/Os* (sistema operacional *IBM Main-Frame*), o fornecedor *IBM* disponibiliza outros produtos tais como *Informix Spatial Datablade* e *Netezza Spatial* (IBM DEVELOPER WORKS, 2012).

O fornecedor *Microsoft* apresenta a técnica *Microsoft-Spatial-Denadi*, complementar ao gerenciador de banco de dados relacional *SQL-Server*, versão 2012. Essa técnica também se utiliza da linguagem *SQL*, é compatível com os critérios do *OGC* e reconhece padrões geométricos através das técnicas denominadas *circular-string* e *linestring*, ambas apoiadas em índices espaciais.

Segundo o fornecedor *Microsoft*, os recursos espaciais disponibilizados pelo *Microsoft-Spatial-Denadi*, além de serem os primeiros do setor para sistemas de banco de dados relacionais, representam um marco significativo na evolução do suporte a dados espaciais, por conta de, através das técnicas *circular-string* e *linestring*, conseguirem apoiar plenamente e com desempenho satisfatório, geometrias circulares e elípticas (MICROSOFT SQL-SERVER, 2011).

A técnica *PostGIS* (POSTGIS, 2005), análoga as demais técnicas apresentadas, também é uma extensão do modelo relacional *PostgreSQL* e se destaca das demais técnicas existentes, segundo o fornecedor *Postgre*, por suportar feições geométricas complexas em suas tabelas. Assim, o banco de dados geográfico proposto, é a junção do modelo relacional de dados convencional *PostgreSQL* com o módulo geográfico *PostGIS*.

Compatível com os padrões OGC, é também compatível com a maioria dos sistemas de informações geográficas (*SIG*) e *Webmapping* comerciais tais como *uDIG*, *QGIS*, *OpenJump*, *ArcGIS*, *KosmoSIG*, *gvSIG*, *GRASS*, *TerraView*, *Spring*, *MapInfo* e *AutoCAD* (MEDEIROS, A.M.L., 2012).

6.8.9 Análise das propostas

As propostas apresentadas se aproximam da proposta deste trabalho, à medida que relacionam dados com estruturas distintas a modelos de dados, embora não mencionem e/ou recomendem diretamente, um modelo de dados pós relacional que considere a escalabilidade.

Estima-se que o modelo de dados pós relacional proposto neste trabalho, por ser apoiado em Redes Complexas, se mostre indiferente à forma de armazenamento dos dados, embora estudiosos da tecnologia *Big Data* recomendem, em prol da escalabilidade, aproximar-se o quão possível, a parte do processamento computacional (processadores e memória) da parte do armazenamento (discos magnéticos), contra recomendando assim, *storages* (GARCIA, M., 2014).

A proposta deste trabalho não recomenda uma nova linguagem de acesso a dados complexos nem tampouco a integração de dados oriundos de gerenciadores distintos, por técnicas visuais ou analíticas, com propósitos ligados a áreas específicas tais como nuclear, acadêmica, geo-processamento, financeira etc, como observado em algumas das propostas anteriores apresentadas na seção 6.7, mas sim, através de uma linguagem pré-existente, propõe em tempo de execução, eventual integração de dados com diferentes estruturas.

Assim como nas linhas de pesquisa apresentadas na seção 6.7, em nenhum estudo apresentado até então, observou-se a proposta de um modelo de dados pós relacional, apoiado nos conceitos de Redes Complexas, onde dados de estruturas distintas coexistam e se integrem em tempo de execução, embora os estudos apresentados, cada qual com suas respectivas ênfases, evidenciem de forma direta ou indireta, esta demanda.

6.9 VANTAGENS / LIMITAÇÕES DO MODELO NAVEGACIONAL DINÂMICO

A evolução dos modelos de dados, apresentada na *Tabela 1*, evidencia a tendência de modelos de dados posteriores surgirem como opção à modelos de dados anteriores, porém não necessariamente, substituindo-os.

Peculiaridades do *MND*, mostram-se vantajosas quando comparadas, por exemplo, a outros modelos de dados que não possuem suas funcionalidades.

Como consequência das vantagens destas peculiaridades, pode-se citar a maior autonomia do modelo de dados, por exemplo, por dispensar o uso de *software's* apartados ao modelo de dados, específicos para a leitura de arquivos de imagens e/ou áudios.

Outro fator positivo associado ao modelo de dados proposto, é a escalabilidade horizontal proporcionada pela estrutura apoiada em Redes Complexas.

A premissa da consistência eventual, ao mesmo tempo que confere desempenho satisfatório à manipulação de dados com diferentes estruturas, limita o uso do modelo proposto, condicionando-o à sistemas computacionais onde a janela de inconsistência, consequência direta da consistência eventual, possa ser tolerada.

Um fator limitante do modelo proposto, é observado quando o acesso requerer integração entre dados não estruturados. Neste caso, a premissa da integração inter-estrutural não é utilizada, mesmo observando-se no modelo, os dados não estruturados mencionados na consulta.

Outro fator limitante associado ao modelo proposto, pode ser observado pela inexistência de registros históricos de consultas, impedindo o modelo de dados de operar, se necessário, apoiado em estatísticas geradas por acessos anteriores, cujos trajetos podem ser mais curtos (tempos menores).

Desta forma, estima-se que o modelo proposto, com suas vantagens e limitações, apresente melhor desempenho quando comparado a modelos de dados *NoSQL* apoiados em grafos que operam sob propriedades *ACID*, e que também, permita armazenar maiores volumes de dados, sobretudo não estruturados.

A Figura 65 apresenta a inserção do modelo de dados *MND*, no contexto dos modelos de dados *NoSQL*.

Figura 65 – O *MND* no contexto dos modelos de dados *NoSQL*

Fonte: Autor

O gráfico apresentado na Figura 65, mostra que o *MND*, quando comparado aos modelos de dados *NoSQL* orientados a grafos, tende a apresentar melhor desempenho na solução das consultas e também, maior capacidade de armazenamento de dados. Essas vantagens ocorrem, majoritariamente, por conta do *MND* operar sob consistência eventual.

Embora também apoiado em grafos, logo, também altamente escalável horizontalmente como os modelos apoiados em grafos, o *MND* não opera sob propriedades *ACID*, como os demais modelos apoiados em grafos, tendendo assim, armazenar maiores volumes em menores intervalos de tempo e navegar mais rapidamente pelo grafo, resolvendo assim mais rapidamente as consultas.

Analogamente, quando comparado aos modelos de dados *NoSQL* apoiados em documentos, colunas e chave-valor, sua estrutura de grafos apresenta menor escalabilidade horizontal e conseqüentemente, menor capacidade de armazenamento, além de menor desempenho computacional quando da solução de consultas, embora esses modelos também sejam, como o *MND*, eventualmente consistentes.

7 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as contribuições e conclusões acerca do MND, assim como sugestões de trabalhos futuros, na área de modelagem de dados.

7.1 CONTRIBUIÇÕES E CONCLUSÕES

A principal contribuição do MND, está na coexistência e integração, em tempo de execução, de dados com estruturas distintas, em um modelo de dados pós relacional único, sob consistência eventual.

A estrutura das Redes Complexas, presente nessa proposta, viabiliza a coexistência de dados estruturalmente distintos. Essa coexistência, por sua vez, permite, através do Algoritmo Navegacional Dinâmico, a integração inter-estrutural de dados, ou seja, a interação dos tradicionais dados estruturados com os dados não estruturados, multimídia, mais recentemente utilizados. Essa integração, por sua vez, confere às aplicações que se utilizam do MND, maior completeza, abrangência e versatilidade, não observadas em modelos similares.

A operação do MND contém as transações observadas em outros modelos que apoiam gerenciadores de bancos de dados, tais como inserção, leitura, atualização e exclusão de dados.

Baseado em três matrizes adaptadas, que ligam os conceitos do modelo de grafos ao MND, além da forma peculiar de atribuir valores a essas três matrizes, ou seja, matriz de Adjacências, de Incidências e de Pesos, o MND apresenta uma quarta matriz, denominada Matriz de Estruturas. Assim como as demais matrizes, a Matriz de Estruturas também é construída pelo próprio MND, e é diretamente ligada à integração de dados estruturalmente distintos.

Vale ressaltar que, tal qual outros modelos de dados pós relacionais apoiados em grafos, o MND apresenta maior escalabilidade horizontal quando comparado a outros modelos *NoSQL* e tende a apresentar melhor desempenho, por não operar sob propriedades *ACID* mas sim, sob regras *BASE*, diferenciando-se dos demais modelos de dados apoiados em grafos.

É importante destacar que novas demandas não suportadas pelo modelo relacional de dados surgem, conforme exemplificado pelos estudos de caso de uso, das seções 6.6 e 6.7. Nesses estudos, de fato há a demanda por integração de estruturas tradicionais com estruturas menos tradicionais, como proposto neste trabalho.

Assim, o esforço empregado na realização deste trabalho, pode ser utilizado como opção a sistemas computacionais que processam sobretudo, dados não estruturados, e que possam operar, em prol da disponibilidade, com dados eventualmente consistentes.

Conclui-se portanto, a viabilidade deste trabalho, quer pelo cunho científico deste, quer pelas iminentes e crescentes demandas à arquitetura por ele proposta.

7.2 TRABALHOS FUTUROS

Há alguns pontos a serem melhor explorados nesta proposta.

O dimensionamento da janela de inconsistência, por exemplo, é um desses pontos. Por conta da consistência eventual, intrínseca ao modelo proposto, a janela de inconsistência, diretamente ligada à consistência eventual, não é considerada, quer pelo desconhecimento do comportamento da função que correlaciona o tempo de acesso com o volume de dados armazenado, quer pela variação de desempenho no acesso a dados com estruturas distintas, haja vista que determinadas aplicações, sobretudo as aplicações transacionais de alta latência, não podem, em sua maioria, considerar dados eventualmente consistentes.

Outro aspecto que pode constituir objeto de pesquisa, está no tratamento de dados integralmente não estruturados, ou seja, fazer com que o Algoritmo Navegacional Dinâmico passe a tratar consultas contendo apenas dados não estruturados.

Outro campo de pesquisa, diretamente ligado a implementação do Algoritmo Navegacional Dinâmico e por consequência, indiretamente ligado a janela de inconsistência e a integração de dados não estruturados, está na existência de registros históricos de consultas, hoje não considerado nesta proposta.

Essa funcionalidade, poderia, por exemplo, associar às consultas submetidas ao MND, um plano de acesso otimizado e previamente conhecido, conferindo assim às consultas, melhor desempenho e/ou maior previsibilidade no que tange a seus tempos de execuções.

REFERÊNCIAS

- ACHARD, S., SALVADOR, R., WHITCHER, B., SUCKLING, J., BULLMORE, E. **A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs**. The Journal of Neuroscience, 2006. v. 26 . 63 p.
- ACHARD, S., BULLMORE, E. **Efficiency and cost of economical brain function networks**. PloS computacional biology, 2007. v. 3.
- AGARWAL, A. **Web 3.0 concepts explained in plain English**. Digital Inspiration, 2009.
- ALVARENGA, P., RAMOS, A. **Evolução, aplicações, barreiras e tendências do m-business**. Encontro Nacional da Engenharia de Produção, Florianópolis, 2004.
- AMAZONAS, J.R. **The role of the Internet of Things(IoT) for business process management in the social and economic spheres**. Universidade de São Paulo, 2011.
- ANTIQUEIRA, L., NUNES, M. G. V., JR., O. N. O., and COSTA, L. F. **Complex networks in the assessment of quality text**, (2005a).
- ANTIQUEIRA, L., Nunes, M. G. V., JR., O. N. O., and COSTA, L. F. **Modelando textos como redes complexas**. In XXV Congresso da Sociedade Brasileira de Computação (III Workshop em Tecnologia da Informação e da Linguagem Humana - TIL 2005), (2005b).
- AZEMOVIC, J. and MUSIC, D., **Comparative analysis of efficient methods for storing unstructured data into database with accent on performance**. Education Technology and Computer (ICETC), 2th Internacional Conference, 2010.
- BARABÁSI, A. L. **Linked: How everything is connected to everything else and what it means for business, science and everyday life**, 2003.
- BEALL, J., **The Weaknesses of Full-Text Searching**, The Journal of Academic Librarianship, 2008. v. 34, n. 5, 438-444 p.
- BELLEÇ, P., PERLBARG, V., JBABDI, S. et al. **Identification of large-scale networks in the brain using fMRI**. Neurolmage, 2006. v. 29, 1231-1243 p.
- BERNERS-LEE, T. **Word-Wide Computer**. Communications of the ACM, 1997, v. 40.
- BHATTACHARYYA, A. **"On a measure of divergence between two statistical populations defined by their probability distributions"**. Bulletin of the Calcutta Mathematical Society, 1943, 35-99-109 p.

BOGUÁ, M., SATORRAS, R.P., VESPIGNANI, A., **Epdemic Spreding in Complex Networks with Degree Correlations**. Statistical Mechanics of Complex Networks, 2013. v. 625 127-147 p.

BOUMANS, J., **Crossmedia - E-Content Report 8**, ACTeN Anticipating Content Technology Needs, 2004.

BREWER, E. A. **Towards robust distributed systems. (Invited Talk). Principles of Distributed Computing (PODC)**, Portland, Oregon, 2000.

BREWER, E. A. **CAP twelve years later: How the "rules" have changed**, IEEE Explore, 2012. v. 45, p. 23-29.

BROWNE, J. **Brewer's CAP Theorem**, 2009. Disponível em: <<http://www.julianbrowne.com/article/viewer/brewers-cap-theorem>>. Último acesso em: 22 Fev. 2016.

BRITO, R., **Bancos de Dados NoSQL x SGBDs Relacionais: Análise Comparativa**, 2011. Disponível em <<http://www.infobrasil.inf.br/userfiles/27-05-S4-1-68840-Bancos%20de%20Dados%20NoSQL.pdf>> Último acesso em: 10 Nov 2011.

BUCHANAN, M. **Nexus - small world and the groundbreaking science of network**. W. Norton Company, 2002.

BULLMORE, E. and SPORNS, O., **Complex brain networks: graph theoretical analysis of structural and functional systems**. Nature Reviews Neuroscience, 2009.

CÂNDIDO, F., MVS/VSAM. São Paulo, S.P.: Manual técnico – TSystems, 2007.

CASANOVA, M.A., MOURA, A.V., **Princípios de Sistemas de Gerência de Bancos de Dados Distribuídos**. São Paulo: Érica, 1999.

CAVOTO, P.R.N., **Arquitetura Híbrida de Integração entre Banco de dados Relacional e de Grafos: Uma Abordagem Aplicada à Biodiversidade**. Tese de Mestrado, UNICAMP, 2015.

CHU, E., BAID, A., CHEN, T., DOAN, A. and NAUGHTON, J., **A relational approach to incrementally extracting and querying structured in nostruturedo data**. Proceeding of the 33rd Internacional conference on Very large databases, 2007. v. VLDB Endowment.

CODD, E.F., **Relational Model of Data for Large Shared Data Banks**. Association for Computing Machinery, 1970. v. 13, p. 377-387.

COLUMBYA UNIVERSITY, **INFORMATION AND KNOWLEDGE STARTEGY**, 2015, Disponível em: < http://www.webopedia.com/TERM/U/unstructured_data.html> Último acesso em: 28 Out. 2015.

COSTA, L., RODRIGUES, F. A., TRAVIESO, G., and BOAS, P. R. V., **Characterization of complex networks: A survey of measurements**, 2007, p. 167-242.

DAMAS, L., **SQL, Structured Query Language** ; Tradução Elvira Maria Antunes Uchôa. Ed. LTC, Rio de Janeiro, 2007.

DARWEN, H., **An Introduction to Relational Database Theory**, Ventus Publishing ApS, 2012. v. 1.

DATE, C.J., **Introdução a sistemas de banco de dados – Tradução da oitava edição**, Campus, 2004.

DB2 EXTENDER, 2012, Disponível em: <<http://www-03.ibm.com/software/products/en/db2spaextl>> Último acesso em: 6 Jun. 2016.

DEAN, J. , GHEMAWAT, S. **MapReduce: Simplified Data Processing on Large Clusters. Commun**, 2004 107–113 p.

DOAN, A., NAUGHTON, J.F., BAID, A., CHAI, X., CHEN F., CHEN, T., CHU, E., DEROSE P., GAO, B., GOKHALE, C., HUANG, J., SHEN, W. and VUONG, B.Q., **The case for a structured approach to managing unstructured data**, 2009.

ELMASRI, R., NAVATHE, S.B., **SISTEMAS DE BANCO DE DADOS**, Pearson, 4th ed. São Paulo, 2005.

EMMERICH, T., BUNDE, A., HAVLIN, S., LI, G., LI, D., **Complex networks embedded in space: Dimension and scaling relations between mass, topological distance, and Euclidian distance**, Physical Review, 2013. v. 87.

ERDÖS, P. , RÉNYI, A., **On random graphs**, 1959.

FERREIRA, E., **Introdução ao NoSQL – PARTE I**, 2011 Disponível em < <http://escalabilidade.com/2010/03/08/Introducao-ao-nosql-parte-i/>> Último acesso em: 8 Nov. de 2011

FICO. **C Big Data: Overhyped or Underexploited?**, 2015. v.1. Disponível em:<http://subscribe.fico.com/big-data-overhyped-or-underexploited?CID=7018000000etZk&gclid=CL_grYTpq8kCFU0XHwodPFYEcA> Último acesso em: 25 Nov. 2015.

FINCHER, S., **Perspectives on HCI patterns: concepts and tools (introducing PLML)**. Interfaces, 2003. p. 26-28.

FINN, M. (2007). **Como resolver a impedância em banco de dados**. Disponível em: <<http://www.intersystems.com.br>> Último acesso em: 17 Nov. 2015.

FOWLER, J. H.& CHRISTAKIS, N. A. **The dynamic spread of happiness in a large social network** 2008. p. 338.

FRANK, R.L., **CODASYL Data-Base Management Systems**, Computer Science Departament , Univesity of Utah, Salt Lake City, Utah, 1978.

FUNG, B.C.M., WANGY, K., ESTER, M., **Hierarchical Document Clustering Using Frequent Itemsets**, Proceedings of the SIAM Internacional conference on data mining, 2003. v. 30.

GARCIA, M., **E esse tal Big Data**, 2014. Disponível em: <<http://cetax.com.br>>. Último acesso em: 27 Jun. 2016.

GIL, A.C., **Métodos e técnicas de pesquisa social**. São Paulo:Atlas,1999.

GILBERT, S., LYNCH, N., **Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services**, 2002. v. 33, p. 51-59.

GOMÉZ, S., GUILERA, A.D., GARDE, N.G., VICENTE, C.J.P., MORENO, Y., ARENAS, A., **Diffusion Dynamics on Multiplex Networks**, Physical Review Letters, 2013. v. 1.

GOVINDAN, R., TANGMUNARUNKIT, H., **Heuristics for internet map discovery**. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. 2000. v. 3, p. 1371–1380. Disponível em: <http://dx.doi.org/10.1109/INFCOM.2000.832534>, Último acesso em: 17 Jun. 2016.

HEUSER, C.A., **Projeto de banco de dados** Disponível em: <<http://www.buscapede.com.br/projeto-de-banco-de-dados-carlos-alberto-heuser-8577803821.html>>, 2014. Último acesso em: 19 Nov. 2015.

HIDALGO, C. A., KLINGER, B., BARABASI, A.L., and HAUSMANN, R. **The product space conditions the development of nations**: Science, 2007. p.317-482.

HOFF, T., **Google Spanner's Most Surprising Revelation**: NoSQL is Out and NewSQL is In, 2012.

HUNG, S.Y., KU, C.Y., CHANG, C.M.,

Criticals factors of WAP services adoption: an empirical study: Eletronic Commerce Research and Applications, 2003. p.42-60.

IBM DEVELOPER WORKS, 2012. Disponível em: <https://www.ibm.com/developerworks/br/data/tutorials/dm-1202db2spatialdata1/>.
Último acesso em: 6 Jun. 2016.

IGOR, M., FILIPOSKA, S., GRAMATIKOV, S., TRAJANOV, D., KOCAREV, L., **Novell Algorithms and Techniques in Telecommunications and Networking**, 2010. p. 211-216.

JIANG, W., ZHANG, L., QIANG, W., JIN, H., PENG, Y., **MyStore: A High Available Distributed Storage System for Unstructured Data**, International Conference on Embedded Software, IEEE, 2012.

KANFOMAS, J.M., **Recursive Hierarchies: The Relational Taboo**, 1952. The Relational Journal.

KEVIN, L., BOB, B., **Oracle Database 10g – Manual do DBA**, 2004. p. 531.

KIM, W., **Introduction to Object-Oriented Databases**. The MIT Press, 1990.

KLEIN, J.S., **Comunicação e Redes Leis de Potência**, UFABC, 2011.

LAKATOS, E.M., MARCONI, M.A., **Fundamentos de Metodologia Científica**. São Paulo: Atlas, 1993.

LI, Y., THAI, M.T., WU, W. **Wireless sensor networks and applications**. New York, NY, USA. Springer Science+Business Media, LLC, 2008.

LILJEROS, F.; EDLING, C. R.; AMARAL, L. A.; STANLEY, E. H.; A BERG, Y. **The web of human sexual contacts**. Nature, Nature Publishing Group, 2001. v. 411, p. 907–908.

LIU, J., DONG, X. and HALEVY, A., **Answering structured queries on unstructured data**, Proc. Of WebDB, 2006. v.6.

LOMBARDO, S., DI NITTO, E., ARDAGNA, D., **Symbolic and Numeric Algorithms for Scientific Computing.(SYNASC)**, 2012. p. 443-448.

LOZANO, S., BUZNA, L., GUILERA, A.D., **Role of network topology in the synchronization of power systems**. The European Physical Journal B- Condensed Matter and Complex Systems, 2012. v. 85, p. 1-8.

LYNALL, M.E., BASSETT, D.S., **Functional connectivity and brain networks in Schizophrenia**. The journal of Neuroscience, 2010.

LUCA,D.M., BECKMANN, C.F., STEFANO, D.N., MATTEWUS, P.M., SMITH, S.M., **Resting state networks define distinct models of long-distance interactions in the human brain**. Neurolmage, 2006. v. 29, p. 1359-1367.

MANFRED, H., **CA-IDMS Task Management**. Computer Associates, Ukiua Roadshow, Langley, BERKS, 2004.

MANSURI, I.R., and SARAWAGI, **Integrating and unstructured data into relational databases**. Data engineering ICDE06. Proceeding of the 22nd Internacional Conference on IEEE, 2006.

MAHALANOBIS, P. C. **"On the generalised distance in statistics"**. Proceedings of the National Institute of Sciences of India, 1936 p. 49–55, Retrieved 2012-05-03.

MEDEIROS, A. M. L., **APLICAÇÕES GEOGRÁFICAS DO POSTGRESQL E SEU MÓDULO POSTGIS**. Revista FOSSGIS Brasil, Coluna Banco de Dados Geográfico, 2012. p. 25 - 27.

MICROSOFT SQL-SERVER – **New Spatial Features in SQL Server Code-Named "Denali"**, SQL Server Technical Article, 2011.

MIRZASOLEIMAN, B., BABAEI, M., JALILI, M., **Immunizing complex networks with limited budget**, EPL, 2012.

MIYAMARU, F., LEITE, L., BERTUZZI, A., FILGUEIRAS, L. V. L., **Mídia cruzada em serviços de governo: conceito e aplicação**, 2008, p. 140-145.

NETO, J.G., **Modelo híbrido de banco de dados relacional, de alto desempenho e capacidade de armazenamento, para aplicações voltadas a engenharia nuclear**. Dissertação de Mestrado, IPEN, Universidade de São Paulo, 2008.

.

MOKARZEL, S.G., **Decoerência e dissipação de sistemas quânticos: Técnicas e Aplicações**. Tese de doutoramento RE21868, Universidade de São Paulo, 2000.

MONSON-HAEFEL, R. **Things Every Software Architect Should Know. Collective Wisdom from the Experts**, 2009. p. 116-117.

NASCIMENTO, J. **Introdução ao NoSQL**, 2011. Disponível em <<http://www.nosql.com.br/introducao-ao-nosql.html>>, Último acesso em: 8 Nov. de 2011.

NEO4J, **The Neo4J Manual v.2.3.1 – 8.1 What is Cypher?** 2014. Disponível em <<http://neo4j.com/docs/stable/cypher-introduction.html>>, Último acesso em: 4 Jan. de 2016.

NEWMAN, M.E.J. **The structure and function of complex networks**, 2003.v. 45, p. 167–256.

NEWMAN, M.; BARABASI, A.-L.; WATTS, D. J. **The structure and dynamics of networks**. [S.I.]: Princeton University Press, 2006.

OLIVEIRA, S.L. DE. **Tratado de metodologia científica: projeto de pesquisa, TGI, TCC, monografias, dissertações e teses**. São Paulo:Pioneira Thomson Learning, 2012.

ORACLE 12C DATABASE, 2014 **Oracle Spatial and Graph: Advanced Data Management** Disponível em <<http://www.oracle.com/technetwork/database/options/spatialandgraph/spatial-and-graph-wp-12c-1896143.pdf>> , Último acesso em: 31 Mai. de 2016.

O'REILLY, TIM. **What Is Web 2.0- design Patterns and Business Models for Next Generation of Software**, 2005.

OREND, K., **Analysis and Classification of NoSQL Databases and Evaluation of their Ability to Replace an Object-relational Persistence Layer**, 2010. Dissertação (Mestrado em Engenharia de Software para Sistemas de Informação Empresariais). Universidade técnica de Munique, Baviera

PARK, S.J., PARK, D.K., WON, C.S., **Core experiments on MPEG-7 edge histogram descriptor** MPEG document M5984, Geneva, 2000.

POSTGIS 2.2.1 **dev Manual SVN Revision (14234)**, 2005 Disponível em <http://postgis.net/stuff/postgis-2.2.pdf>, Último acesso em: 31 Mai. de 2016.

PLÁCIDO, D.G., CUNHA, D.P., **Impedância em Bancos de Dados**, UNESC, 2008.

PRICE, D.D.S., **Network of scientific papers**, 1965, p. 510–515.

RAJAPAKSE, I., GROUDINE, M., MESBAHI, M., **What Can Systems Theory of Networks Offer to Biology**. PLoS Compute Biol, 2012. v. 8.

ROY, P. and MOHANIA, M., **SCORE: symbiotic context oriented information retrieval**. Advanced in Data and Web Management Springer Berlin Heidelberg, 2007. p. 30-38.

REDNER, S., **How popular is your paper ? an empirical study of the citation distribution**. The European Physical Journal B - Condensed Matter and Complex Systems, 1998.

SCHÖNBERGER, V.M., CUKIER, K. – **BIG DATA – A revolution that will transform how we live,work, and think**. AN EAMON DOLAN BOOK, Boston, N.Y., 2013.

SETZER, V.W., 2010. Disponível em < http://pt.wikipedia.org/wiki/Valdemar_Setzer>
Último acesso em: 26 Mar. 2014.

SHARMA, A. and DHIR, R., **A wordset based document clustering algorithm for large dataset**. Methods and Models in Computer Science, Proceeding of International Conference on IEEE, 2009. v. ICM2CS.

SICA, L., 2013. Disponível em < <http://www.cartaforense.com.br/conteudo/artigos/lei-carolina-dieckmann--aspectos-penais/11242>>, Último acesso em: 18 Jan. 2014.

SILVA, E.L.; MENEZES, E. M.. **Metodologia da Pesquisa e Elaboração de Dissertação**. Florianópolis. 4ª ed., UFSC, 2005.

SILBERSCHATZ, Abraham, KORTH, Henry F., SUDARSHAN, S. **Sistema de Banco de Dados**. 5ª ed., São Paulo, Elsevier, 2006.

SILBERSCHATZ, Abraham, KORTH, Henry F., SUDARSHAN, S. **Sistema de Banco de Dados**. 6ª ed., São Paulo, Elsevier, 2012.

SOUZA, T., ROCHA, A., **NoSQL Princípios e Características**, 2011. Disponível em: <<http://http://www.slide.net/anderochaip/artigo-nosql>>, Último acesso em: 11 Nov. 2011.

SPORNS, O. **Graph theory methods for the analysis of neural connectivity patterns**. *Complex*, 2002, v. 8, p. 56–60.

SOLOMONO, R. and RAPOPORT, A., **Connectivity of random nets**. *Bulletin of Mathematical Biophysics*, 1951.

STEPPAT, N., **Bancos de dados não relacionais e o movimento NoSQL**, 2011. Disponível em <<http://blog.caelum.com.br/bancos-de-dados-nao-relacionais-e-o-movimento-nosql/>>, Último acesso em: 9 Nov. de 2011.

STONEBREAKER, M., **NewSQL: An Alternative to NoSQL and Old SQL for New OLTP Apps**. Communications of the ACM Blog, 2011.

SUBELI, L , BAJEC, M., **Software systems through complex networks science: review, analysis and applications**. First International WorkShop on Software Mining, 2012, p. 9-16.

TAKAI, O.K., ITALIANO, I.C., FERREIRA, J.E., **INTRODUÇÃO A BANCO DE DADOS**, São Paulo, DCC-IME-USP, 2005.

TAYLOR, W.R., **CODASYL Data-Base Management Systems**, IBM Research Laboratory, San Jose, California, 1978.

TARI, L., TU, P.H., HAKENBERG, J., CHEN, Y., SON, T.C., GONZALEZ, G. and BARAL, C., **Parde Tree Databases for Information Extraction: DATA ENGINEERING**, 2010.

TAKAHASHI, T., TOMIOKA, R., YAMANISHI, K., **Discovering Emerging Topics in Social Streams via Link Anomaly Detection**, 2011.

TECMUNDO, **O tamanho da indústria dos video games**, 2011 Disponível em < <http://www.tecmundo.com.br/infografico/9708-o-tamanho-da-industria-dos-video-games-infografico-.htm>> Último acesso em: 09 Jul. 2016.

VOGELS, W. **Eventually consistent**: Communications of the ACM 52, 2009.

W3C Website. SPARQL Query Language for RDF, **W3C Recommendation**, 2008. Disponível em: < <http://www.w3.org/TR/rdf-sparql-query/>>. Último acesso em: 4 Jan. 2016

WALSH, T., **Search in a small world**, Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. 1999, p. 1172-1177.

WANG, L., ZHU, C., HE, Y., ZANG, Y., CAO, Q.J., **Altered small-world brain functional networks in children whit attention-deficit/hyperactivity disorder**. Human Brain , Wiley Online Library , 2009.

WANG, L., ZHENGW, W., YANG, C., ZHANG, L., YE, Q., **Linux Kernels as complex networks**: A novel method to study evolution. In Software Maintenance, IEEE, Internacional Conference, 2012. p. 41-50.

WANG, P., GONZALEZ, M., HIDALGO, C.A., BARABASI, A.L., **Understanding the spread patterns of mobile phone viruses**, 2009.

WATTS, D. J. and STROGATZ, S. H., **Colletive dynamics of small-world networks**, 1998. p.440–442.

WEI, Z., **Scalable Transactions for Web Applications in the Cloud**. In proc. of the Euro-Par Conference, 2009.

YAFOOZ, W.M.S., ABIDIN, S.Z.Z., OMAR, N., IDRUS, Z., **Systems, Process & Control**. ICSPC, IEEE Conference, 2013. p. 198-203.

YIN,R.K. **Estudo de caso – Planejamento e Métodos**. 4th ed. , São Paulo, 2010.

APÊNDICE A – DIAGRAMA ENTIDADE - RELACIONAMENTO SACA

Fonte: Autor

APÊNDICE B – MODELO NAVEGACIONAL SACA

Fonte: Autor

APÊNDICE C – MODELO NAVEGACIONAL SACA MODIFICADO

Fonte: Autor

APÊNDICE D – NÓS *LABELS* DO MODELO SACA MODIFICADO

Fonte: Autor