

Andréa Sabedra Bordin

**FRAMEWORK BASEADO EM CONHECIMENTO PARA
ANÁLISE DE REDE DE COLABORAÇÃO CIENTÍFICA**

Tese submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina como requisito para a obtenção do Grau de Doutor em Engenharia e Gestão do Conhecimento.

Orientador: Prof. Dr. Alexandre Leopoldo Gonçalves.

Coorientador: Prof. Dr. João Artur de Souza.

Florianópolis
2015

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Bordin, Andréa Sabedra

Framework baseado em conhecimento para análise de rede de colaboração científica / Andréa Sabedra Bordin ; orientador, Alexandre Leopoldo Gonçalves ; coorientador, João Artur de Souza. - Florianópolis, SC, 2015.
333 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Inclui referências

1. Engenharia e Gestão do Conhecimento. 2. Rede de Colaboração Científica. 3. Análise de Redes Sociais. 4. Modelagem de Conhecimento. 5. Representação de Conhecimento. I. Gonçalves, Alexandre Leopoldo . II. Souza, João Artur de. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Andréa Sabedra Bordin

**FRAMEWORK BASEADO EM CONHECIMENTO PARA
ANÁLISE DE REDE DE COLABORAÇÃO CIENTÍFICA**

Esta Tese foi julgada adequada para obtenção do Título de Doutora em Engenharia e Gestão do Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Florianópolis, 06 de agosto de 2015.

Prof. Roberto Carlos dos Santos Pacheco, Dr.
Coordenador do PPGEGC/UFSC

Banca Examinadora:

Prof. Alexandre Leopoldo
Gonçalves, Dr.
Orientador - PPGEGC/UFSC

Prof. Frederico Luiz Gonçalves de
Freitas, Dr.
Examinador externo
CIn/UFPE

Prof. Jesus Pascual Mena Chalco,
Dr.
Examinador externo
CMCC/UFABC

Prof. Marisa Brascher Basílio
Medeiros, Dra.
Membro interno
PPGCIN / UFSC

Prof. José Leomar Todesco, Dr.
Membro interno
PPGEGC / UFSC

Prof. Aran Morales Tcholakian, Dr.
Membro interno
PPGEGC / UFSC

AGRADECIMENTOS

Gostaria de agradecer primeiramente ao professor Dr. Alexandre Leopoldo Gonçalves, que aceitou o meu pedido de orientação, esteve sempre disponível nos vários momentos onde solicitei a sua contribuição e é detentor da minha admiração em vários sentidos. Estendo esse agradecimento ao professor coorientador João Artur de Souza, cujas palavras, há alguns anos, foram decisivas para iniciar um novo ciclo de intensa aprendizagem.

Agradeço aos membros da banca Dr. Frederico Luiz Gonçalves de Freitas, Dra. Marisa Brascher Basílio Medeiros, Dr. José Leomar Todesco e Dr. Aran Bey Tcholakian Morales pelas delicadas e valiosas contribuições na qualificação e a todos novamente e conjuntamente com o Dr. Jesus Pascual Mena Chalco por aceitarem o convite para participar da banca de defesa desta tese.

Aos colegas do PPGEHC que, alguns no início, outros no fim, colaboraram comigo de alguma forma: Flávio Ceci, Cristiane Woszezenki, Rafael de Moura Speroni, Michele Andreia Borges, Danielly Inomata, Thales Nascimento, Kedma Duarte e especialmente à Silvia Bentancourt na reta final deste trabalho.

Aos professores do PPGEHC, Dr. Roberto Carlos do Santos Pacheco, Dra. Gertrudes Dandolini e Dr. Gregório Varvakis pelo tempo despendido e pelas valiosas contribuições, tanto na concepção do instrumento da avaliação, quanto na avaliação em si deste trabalho.

Aos colegas da Universidade Federal do Pampa (Unipampa), Dr. Fladimir Fernandes pela leitura e contribuições do documento de qualificação e Dr. Fábio Kepler pelo suporte na ferramenta ScriptLattes.

Em âmbito institucional gostaria de mencionar primeiramente o Instituto Stela e agradecer pelos anos de aprendizagem e pela oportunidade que viabilizou o período inicial de vivência deste ciclo de aprendizagem. À Unipampa, *campus* Alegrete, que através dos seus gestores viabilizaram o meu afastamento da Universidade, permitindo assim, que um novo momento de dedicação integral e aprendizagem acontecesse.

Por fim, à minha mãe, família e amigos, especialmente ao amigo Sérgio Melo de Almeida, que sempre torceram por mim.

RESUMO

Entender a configuração da colaboração científica em um ambiente de Pesquisa e Desenvolvimento (P&D) permite a tomada de decisão em vários níveis, desde o individual até o estratégico. Análise de Redes Sociais (ARS) é um dos métodos utilizados analisar redes de colaboração científica. Nele uma rede é usualmente representada como um grafo ou matriz, o que carece de semântica e não permite que sejam analisadas características adicionais dessa rede. Ao analisar exclusivamente a estrutura da rede através das métricas de ARS, não contemplando o conteúdo dos artefatos oriundos da colaboração, perde-se um insumo importante que pode contribuir nesta análise. Percebe-se também que tarefas de análise de rede de colaboração científica são intensivas em conhecimento, pois requerem o conhecimento de um especialista, nem sempre disponível, de como realizar as tarefas. Conduzido pela abordagem metodológica *Design Science Research*, este trabalho apresenta uma proposição de representação ontológica de conhecimentos necessários – conhecimento de domínio, conhecimento de tarefa e solução de tarefa – para a realização de um conjunto de dezoito tarefas de análise de rede de colaboração, juntamente com um conjunto de artefatos para armazenamento e recuperação de dados dessas representações combinados em um *framework* de conhecimento para análise de rede de colaboração científica. A abordagem de modelagem de conhecimento é inspirada na noção de *Problem-Solved Method (PSM)* e utiliza a técnica de decomposição de tarefa (*task-structure*). O *framework* oferece artefatos de representação de conhecimento e módulos de implementação que podem ser reutilizados em outros contextos ou aplicações. Ele foi avaliado junto a gestores de Programas de Pós-Graduação onde verificou-se que as tarefas são relevantes e que os resultados possuem um grau de confiança alto. As ações de gestão propostas para as tarefas indicam a aplicabilidade como instrumento de gestão pois seus resultados evidenciam a realidade de um contexto de colaboração científica, baseado em produções em coautoria, e podem ser utilizados como insumo para a tomada de decisão por gestores, pesquisadores e comunidade do ambiente analisado.

Palavras-chave: Rede de Colaboração Científica. Análise de Redes Sociais. Modelagem de Conhecimento. Representação de Conhecimento.

ABSTRACT

Understanding the configuration of scientific collaboration in a research and development (R&D) environment enables decision making at various levels, from the individual to the strategic. Social Network Analysis (SNA) is one of the methods used to analyze scientific collaboration networks. In it, a network is usually represented as a graph or matrix , which lacks semantics and do not allow additional features of this network to be analyzed. By analyzing exclusively the structure of the network while not contemplating the content of the artifacts resulting from the collaboration, one may lose an important input that can contribute to the analysis of such network. It may also be noted that the analysis tasks of scientific collaboration networks are knowledge intensive because they require the knowledge of an expert, not always available, as to how to perform the tasks. Based on the arguments presented and guided by methodological approach named Design Science Research, this work presents a proposition of ontological knowledge representation - domain knowledge, task knowledge and task solution - required to carry out a set of eighteen tasks for analysis of collaboration along with a set of artifacts for storing and retrieving data from these representations combined into a knowledge framework for scientific collaboration network analysis. The knowledge modeling approach is inspired by the notion of Problem Solved-Method (PSM) and uses the task decomposition technique (task-structure). The framework provides knowledge representation artifacts and implementation modules that can be reused in other contexts or applications. He was evaluated by Graduate Program coordinators where it was found that the tasks are relevant, and that the results have a high level of reliability. The actions of management proposals for the tasks indicate the applicability as a management tool because their results show the reality of a scientific collaboration context, based on co-authorship, and can be used as input for decision making by managers, researchers and community of the analyzed environment.

Keywords: Scientific collaboration network. Social Network Analysis. Knowledge Modeling. Knowledge Representation.

LISTA DE FIGURAS

Figura 1 - Rede de coautoria	26
Figura 2 - Tipologia da colaboração.....	48
Figura 3 - Interligação de diretorias	59
Figura 4 - Relacionamento não-direcional “mora perto”	63
Figura 5 - Relacionamento direcional e valorado.....	64
Figura 6 - Caminho e Caminhada.....	66
Figura 7 - Distância e Diâmetro	67
Figura 8 - Comparação de densidades.....	68
Figura 9 - Centralidade local e global	70
Figura 10 - Atores importantes da rede do clube de karatê de Zachary.72	
Figura 11 - Componentes de uma rede.....	74
Figura 12 - Rede de coautoria do PPGEGC no ano de 2012.....	77
Figura 13 - Método de resolução de problema Classificação Heurística	86
Figura 14 - Estrutura de tarefa para a tarefa de Diagnóstico.	88
Figura 15 - Arquitetura de um sistema de mineração de texto.....	97
Figura 16 - Visão geral das fases do algoritmo Lingo.....	101
Figura 17 - Dados de entrada na fase de indução de rótulo: (a) documentos, $d = 7$; (b) termos, $t = 5$; e (c) frases, $p = 2$	103
Figura 18 - Matriz termo-documento na fase de indução de rótulo	103
Figura 19 - Matrizes na fase de indução de rótulo: (a) matriz de conceitos abstratos U, (b) matriz dos candidatos a rótulo de cluster P, e (c) matriz dos candidatos a rótulo de cluster (conceito abstrato) M....	104
Figura 20 - Resultado final da alocação de conteúdo ao cluster	106
Figura 21 - Processo de <i>Design Science Research Methodology (DSRM)</i>	119
Figura 22 - Processo de condução da pesquisa	123
Figura 23 - Procedimentos da avaliação da pesquisa.	126
Figura 24 - Fluxo de conhecimento na análise de rede de colaboração científica.....	130
Figura 25 - Organização dos modelos de representação de conhecimento do <i>framework</i>	132
Figura 26 - Estrutura geral do <i>framework</i>	133
Figura 27 - Exemplo de árvore de decomposição de tarefa	142

Figura 28 - Modelagem da Tarefa 1 “Quem são os pesquisadores que mais colaboram? ”	144
Figura 29 - Modelagem da Tarefa 3 “Quem são os pesquisadores mais próximos aos demais pesquisadores? ”	146
Figura 30 - Modelagem da Tarefa 6 “Qual o nível geral de atividade de colaboração da rede? ”	148
Figura 31 - Modelagem da Tarefa 8 “Quem são os colaboradores do pesquisador “x”? ”	149
Figura 32 - Modelagem da Tarefa 9 “Quem são os pesquisadores que colaboram no tópico “x”? ”	152
Figura 33 - Ontologia de Tarefas	156
Figura 34 – Propriedade de objeto da Ontologia de Tarefa	157
Figura 35 - Ontologia de Métodos Complexos – Visão Geral	157
Figura 36 - Ontologia de Métodos Complexos – Subclasses de MetodoComplexo	158
Figura 37 - Ontologia de Métodos Complexos – Subclasses de Subtarefa	158
Figura 38 - Ontologia de Métodos Primitivos.....	160
Figura 39 - Propriedades de objeto das ontologias de Tarefas e Métodos	161
Figura 40 - <i>Snapshot</i> do diagrama de classes da ontologia VIVO – <i>Person</i>	166
Figura 41 - <i>Snapshot</i> do diagrama de classes da ontologia VIVO – <i>Article</i>	167
Figura 42 - <i>Snapshot</i> do diagrama de classes da ontologia Vivo – <i>Relationship</i>	168
Figura 43 – <i>SNAMetrics</i> : Ontologia de Métricas de Análise de Rede Social	169
Figura 44 - Diagrama de Propriedades de Objeto da <i>SNAMetrics</i>	170
Figura 45 - Transformação de XML em RDF	173
Figura 46 - Processo de extração de termos das produções bibliográficas	175
Figura 47 - Operacionalização do <i>framework</i> – Pré-processamento dos dados	179

Figura 48 – Instanciação das ontologias de domínio VIVO-ISF e <i>SNAMetrics</i>	180
Figura 49 – Operacionalização do <i>framework</i> – execução da aplicação	184
Figura 50 - Rede de coautoria dos professores pesquisadores do PPG do Cenário 1	188
Figura 51 - Rede de coautoria dos professores pesquisadores do PPG do Cenário 2	189
Figura 52 - Principais operações do scriptLattes	191
Figura 53 - Interface da aplicação web- <i>front-end</i> do <i>framework</i>	192
Figura 54 - Síntese das respostas da avaliação do <i>framework</i> na Dimensão 1 – Colaboração Científica	219
Figura 55 - Síntese das respostas da avaliação do <i>framework</i> na Dimensão 2 – Resultados	220
Figura 56 - Síntese das respostas da avaliação do <i>framework</i> na Dimensão 3 – Aplicabilidade em ações de gestão	221
Figura 57 - Mapa mental com as sugestões da avaliação	223
Figura 58 - Modelagem da Tarefa 2 “Quem são os pesquisadores que menos colaboram? ”	293
Figura 59 - Modelagem da Tarefa 4 “Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram? ”	296
Figura 60 - Modelagem da Tarefa 5 “Quais os grupos de pesquisadores que trabalham isoladamente? ”	297
Figura 61 - Modelagem da Tarefa 7 “Qual o nível médio de colaboração dos pesquisadores da rede? ”	299
Figura 62 - Modelagem da Tarefa 10 “Quem é o pesquisador que mais colabora no tópico “x”? ”	300
Figura 63 - Modelagem da Tarefa 11 “Quais grupos de pesquisadores colaboram no tópico “x”? ”	304
Figura 64 - Modelagem da Tarefa 12 “Qual o perfil das colaborações do pesquisador “x”? ”	308
Figura 65 - Modelagem da Tarefa 13 “Quem são os pesquisadores que possuem perfis similares a um perfil específico? ”	312

Figura 66 - Modelagem da Tarefa 14 “Qual o grau de similaridade dos perfis de colaboração de dois pesquisadores? ”	314
Figura 67 - Modelagem da Tarefa 15 “Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador “x”?..	317
Figura 68 - Modelagem da Tarefa 16 “. Quem são os pesquisadores que, com base na proximidade e na similaridade de seus perfis com o pesquisador “x”, podem colaborar com esse pesquisador? ”	320
Figura 69 - Modelagem da Tarefa 17 “Qual o perfil dos pesquisadores que mais colaboram? ”	323
Figura 70- Modelagem da Tarefa 18 “Qual o perfil dos grupos que trabalham isoladamente? ”	326

LISTA DE QUADROS

Quadro 1 - Tipos de Artefatos.....	117
Quadro 2 - Tipos de Grupos Focais em Design Science Research.....	121
Quadro 3 - Classes de Problemas e Tipos de Artefatos.....	124
Quadro 4 - Conjunto de perguntas/tarefas de análise de colaboração científica.....	138
Quadro 5 - Quadro explicativo da modelagem da Tarefa 1	144
Quadro 6 - Quadro explicativo da modelagem da Tarefa 3	146
Quadro 7 - Quadro explicativo da modelagem da Tarefa 6	148
Quadro 8 - Quadro explicativo da modelagem da Tarefa 8	150
Quadro 9 - Quadro explicativo da modelagem da Tarefa 9	152
Quadro 10 - Composição de ontologias da ontologia VIVO-ISF	165
Quadro 11 - <i>Snapshot</i> do XSLT	174
Quadro 12 - Avaliação da tarefa 1.....	195
Quadro 13 - Avaliação da tarefa 2.....	198
Quadro 14 - Avaliação da tarefa 3.....	199
Quadro 15 - Avaliação da tarefa 4.....	200
Quadro 16 - Avaliação da tarefa 5.....	201
Quadro 17 - Avaliação da tarefa 6.....	203
Quadro 18 - Avaliação da tarefa 7.....	204
Quadro 19 - Avaliação da tarefa 8.....	205
Quadro 20 - Avaliação da tarefa 9.....	206
Quadro 21 - Avaliação da tarefa 10.....	208
Quadro 22 - Avaliação da tarefa 11.....	209
Quadro 23 - Avaliação da tarefa 12.....	210
Quadro 24 - Avaliação da tarefa 13.....	212
Quadro 25 - Avaliação da tarefa 14.....	213
Quadro 26 - Avaliação da tarefa 15.....	214
Quadro 27 - Avaliação da tarefa 16.....	215
Quadro 28 - Avaliação da tarefa 17.....	216
Quadro 29 - Avaliação da tarefa 18.....	217
Quadro 30 - Quadro explicativo da modelagem da Tarefa 2	293
Quadro 31 - Quadro explicativo da modelagem da Tarefa 4	296
Quadro 32 - Quadro explicativo da modelagem da Tarefa 5	298
Quadro 33 - Quadro explicativo da modelagem da Tarefa 7	299
Quadro 34 - Quadro explicativo da modelagem da Tarefa 10	301
Quadro 35 - Quadro explicativo da modelagem da Tarefa 11	305
Quadro 36 - Quadro explicativo da modelagem da Tarefa 12	310
Quadro 37 - Quadro explicativo da modelagem da Tarefa 13	312
Quadro 38 - Quadro explicativo da modelagem da Tarefa 14	315
Quadro 39 - Quadro explicativo da modelagem da Tarefa 15	317

Quadro 40 - Quadro explicativo da modelagem da Tarefa 16	321
Quadro 41 - Quadro explicativo da modelagem da Tarefa 17	324
Quadro 42 - Quadro explicativo da modelagem da Tarefa 18	327

LISTA DE TABELAS

Tabela 1 - Revisão do tópico “Colaboração científica”	253
Tabela 2 - Revisão do tópico “Análise de redes sociais”	254
Tabela 3 - Revisão do tópico “Rede de coautoria”	255
Tabela 4 - Revisão dos tópicos “Análise de redes sociais” e “Tecnologias semânticas”	258
Tabela 5 - Revisão do tópico “Sistemas baseados em conhecimento”	260

SUMÁRIO

1 INTRODUÇÃO.....	23
1.1 CONSIDERAÇÕES INICIAIS	23
1.2 APRESENTAÇÃO DO PROBLEMA	27
1.3 PERGUNTA DE PESQUISA	31
1.4 OBJETIVOS.....	31
1.4.1 Objetivo Geral	31
1.4.2 Objetivos Específicos	32
1.5 JUSTIFICATIVA	32
1.6 ADERÊNCIA DA TESE AO PROGRAMA DE PÓS-GRADUAÇÃO E ÁREA DE PESQUISA.....	39
1.7 ORGANIZAÇÃO DO TRABALHO	40
2 FUNDAMENTAÇÃO TEÓRICA.....	41
2.1 COLABORAÇÃO CIENTÍFICA.....	41
2.1.1 Definição, Categorização e Natureza das Atividades	42
2.1.2 Estágios da Colaboração.....	43
2.1.3 Formas de Colaboração	48
2.1.4 Benefícios da Colaboração.....	49
2.1.5 Custos da Colaboração	49
2.1.6 Medição da Colaboração	50
2.1.7 Métodos de Estudo da Colaboração	51
2.2 ANÁLISE DE REDE SOCIAL	53
2.2.1 História do Desenvolvimento da Área de ARS	54
2.2.2 Dados Sociais: Tipologia e Forma de Tratamento.....	57
2.2.3 Unidades e Níveis de Análise.....	60
2.2.4 Teoria de Grafos e Matrizes	61
2.2.5 Centralidade.....	68
2.2.6 Subgrupos	72
2.2.7 ARS em Redes de Colaboração Científica	74
2.3 SISTEMAS BASEADOS EM CONHECIMENTO	78
2.3.1 Representação do Conhecimento.....	79
2.3.2 SBC's e a Engenharia do Conhecimento.....	80
2.3.3 Método de Resolução de Problema (Problem Solving Method – PSM)	83
2.3.4 Abordagens de Modelagem de Conhecimento	85
2.3.5 Ontologias	89
2.3.6 Ontologias e Métodos de Resolução de Problemas (PSM)	94
2.4 MINERAÇÃO DE TEXTO.....	96
2.4.1 Pré-Processamento	97
2.4.2 Abordagem de clusterização Lingo	99
2.5 TRABALHOS RELACIONADOS	108

3	METODOLOGIA	113
3.1	EMBASAMENTO TEÓRICO.....	114
3.2	ENQUADRAMENTO METODOLÓGICO	122
3.2.1	Procedimentos de Avaliação da Pesquisa	125
4	FRAMEWORK DE CONHECIMENTO.....	129
4.1	MODELO DE TAREFA E MODELO DE SOLUÇÃO DE TAREFA.....	135
4.1.1	Identificação das Tarefas de Análise de Colaboração Científica.....	136
4.1.2	Modelagem das Tarefas de Análise de Colaboração Científica.....	141
4.1.3	Representação do Modelo de Tarefa e Solução de Tarefa	155
4.2	MODELO DE DOMÍNIO.....	163
4.2.1	Ontologia de Colaboração Científica	164
4.2.2	Ontologia de Métricas de Análise de Redes Sociais	168
4.3	MÓDULOS DE PRÉ-PROCESSAMENTO DOS DADOS E INSTANCIAÇÃO DAS ONTOLOGIAS	171
4.3.1	Requisitos de Dados.....	171
4.3.2	Módulo de Instanciação da Ontologia de Colaboração Científica VIVO.....	172
4.3.3	Módulo de Instanciação da Ontologia de Métricas de ARS.....	176
4.4	MÓDULOS DE RECUPERAÇÃO DOS DADOS.....	177
4.5	PASSOS PARA OPERACIONALIZAÇÃO DO FRAMEWORK.....	177
4.5.1	Instanciação das ontologias de domínio de colaboração científica e métricas de análise de rede social.....	177
4.5.2	Execução do framework.....	181
4.6	CONSIDERAÇÕES FINAIS.....	185
5	AVALIAÇÃO DO FRAMEWORK	187
5.1	CENÁRIO DE DEMONSTRAÇÃO E AVALIAÇÃO	187
5.1.1	Cenário de Demonstração e Avaliação 1.....	187
5.1.2	Cenário de Demonstração e Avaliação 2.....	189
5.2	OPERACIONALIZAÇÃO.....	190
5.2.1	Coleta e Extração dos Dados.....	190
5.2.2	Pré-Processamento dos Dados	191
5.2.3	Instanciação das Ontologias	192
5.2.4	Execução do Framework.....	192
5.3	DEMONSTRAÇÃO E AVALIAÇÃO	193
5.3.1	Perguntas e Respostas da Dimensão 1	194
5.3.2	Perguntas e Respostas da Dimensão 2 e Dimensão 3.....	194
5.4	ANÁLISE DOS RESULTADOS	218
6	CONSIDERAÇÕES FINAIS.....	225

REFERÊNCIAS	231
GLOSSÁRIO	251
APÊNDICE A – Revisão Sistemática	253
APÊNDICE B - Roteiro da Entrevista do Grupo Focal	265
APÊNDICE C - Termo de Consentimento Livre e Esclarecido.....	279
APÊNDICE D - Instrumento de Verificação das Perguntas de Análise de Colaboração Científica	281
APÊNDICE E - Sumário das Respostas do Instrumento de Verificação das Perguntas de Análise de Colaboração Científica	287
APÊNDICE F - Modelagem da Estrutura das Tarefas de Análise de Colaboração Científica.....	293
ANEXO A - Documentação da Ontologia VIVO-ISF	329
ANEXO B – Ontologia SemSNA	333

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

A colaboração científica é um processo social, vivenciado por atores do contexto científico, tais como pesquisadores, professores, alunos, etc. É uma característica da ciência moderna e inúmeras são as suas vantagens.

A colaboração em pesquisa pode se manifestar de forma explícita pela divulgação dos seus resultados, com todos os colaboradores assinando a coautoria de um artigo, ou de forma latente ou implícita, sendo fruto de uma conversa informal entre pesquisadores.

Os primeiros estudos sobre o tema datam do final da década de 1950 e, desde então, diversos autores têm se dedicado ao estudo da colaboração científica em todos os seus níveis – entre indivíduos, instituições, países, setores da sociedade, áreas do conhecimento (SONNEWALD, 2007).

Como tema de pesquisa a colaboração científica está consolidada e é abordada por várias áreas de conhecimento, tais como, Ciência da Informação, Psicologia, Administração, Ciência da Computação, Sociologia, Políticas de Pesquisa, Estudos Sociais da Ciência, Filosofia e, também, em muitas outras disciplinas científicas onde ocorre colaboração.

Em ciência e tecnologia a colaboração é um problema de estudo que desperta interesse mundial. Isso se demonstra pela instituição, em janeiro de 2000, em Berlin, de uma rede mundial de pesquisa sobre o assunto, denominada “*Collaboration in Science and in Technology – COLLNET*”¹. Entre os objetivos da COLLNET está o estudo de diversos aspectos da colaboração e como ela acontece nas diferentes áreas do conhecimento. A rede incentiva a participação de membros provenientes de várias nacionalidades, para possibilitar que os estudos sejam realizados por meio de grande diversidade de metodologias e pontos de vista intelectual e cultural (KRETSCHMER; LIMING; KUNDRA, 2001).

A motivação para pesquisar o tema colaboração científica advém do contexto atual do mundo científico, cada vez mais interdisciplinar, e dos inúmeros benefícios que ela provê.

Para Hara et al. (2003), a colaboração é um componente de pesquisa crítico no mundo da “*big science*”², o qual envolve projetos de

¹ <http://www.collnet.de/>

² *Big Science* é um termo usado por cientistas e historiadores da ciência para descrever uma série de mudanças na ciência que ocorreram em

grande porte dominados por problemas complexos, rápidas mudanças da tecnologia, crescimento dinâmico do conhecimento e conhecimento altamente especializado.

Sonneward (2007) relata que a colaboração tem o potencial para resolver problemas científicos complexos e promover várias agendas políticas, econômicas e sociais, tais como democracia, desenvolvimento sustentável e compreensão cultural e integração.

Os métodos comumente utilizados para estudar colaboração tem sido: bibliométricos (GLÄNZEL, 2002; GLÄNZEL; SCHUBERT, 2004); análise de redes sociais (BARABÁSI et al., 2002; NEWMAN, 2001, 2004; WAGNER; LEYDESDORFF, 2005; MILOJEVIC, 2010); métodos qualitativos de observação e entrevistas (HARA et al., 2003); e pesquisas (*surveys*) (BIRNHOLTZ, 2006).

Segundo Abbasi, Altmann e Hwang (2010), a coautoria é o indicador de colaboração científica mais visível e acessível e tem sido usada frequentemente para medir a atividade colaborativa, especialmente em estudos bibliométricos e estudos de análise de rede social.

Entretanto, alguns autores defendem que a colaboração científica se manifesta de outras formas. De Haan (1997) recomenda seis indicadores para medir a colaboração entre os cientistas: a coautoria, a editoria compartilhada de publicações, a coordenação compartilhada em projetos de doutorado, a escrita em conjunto de propostas de pesquisa, a participação em programas de pesquisa formais e a organização compartilhada de conferências científicas. Katz e Martin (1997) e Laudel (2002) ainda argumentam que muitas colaborações não resultam em publicações coautoradas.

Estudos utilizando métodos bibliométricos de coautoria têm se concentrado principalmente sobre os efeitos da colaboração na atividade científica, bem como sobre os aspectos organizacionais e institucionais de colaboração olhando para os autores, centros de pesquisa, instituições e países como unidades de análise. Estudos utilizando métodos de rede, por outro lado, têm focado principalmente na construção de redes de colaboração e na compreensão das estruturas, mecanismos e processos que conduzem às estruturas de rede observadas.

países industrializados antes e depois da II Guerra Mundial, quando o progresso científico se caracterizou pela existência de projetos de grande porte geralmente financiados por governos nacionais ou grupos de governos.

Nesta tese a inspiração para a investigação da colaboração científica se concentra inicialmente no método de Análise de Redes Sociais (ARS) ou *Social Network Analysis (SNA)*.

O método de Análise de Redes Sociais é fruto da abordagem estrutural das Ciências Sociais, que se preocupa em examinar as ligações ou relacionamentos entre os objetos de estudo. A abordagem estrutural não se limita ao estudo de relações sociais humanas, ou seja, nessa abordagem importantes relações sociais podem vincular indivíduos sociais que não são humanos, como os animais, os grupos, as organizações, os estados, entre outras (FREEMAN, 2004).

Mais do que um método, explorado em áreas de conhecimento diversas, como na Química Molecular, onde os analistas de rede analisam a forma como diferentes tipos de átomos interagem entre si para formar diferentes tipos de moléculas, a ARS é uma área de pesquisa consolidada conforme denotado por Wasserman e Faust (1994), Scott (2000) e Freeman (2004).

A ARS está fundamentada na teoria de grafos, onde uma rede é representada por um grafo $G = (V, E)$, formado por Vértices (V) e Arestas (E). Cada vértice ou nodo representa um ator e cada aresta representa a relação existente entre dois atores integrantes da rede.

Através das medidas de ARS é possível identificar aspectos tais como, a) padrões de relacionamento entre os atores de uma rede; b) a conectividade entre os mesmos; c) a formação de agrupamentos; d) a evolução da rede ao longo do tempo e, e) o fluxo de comunicação, informação e conhecimento dentro da rede (WASSERMAN; FAUST, 1994).

A colaboração científica quando estudada sob o prisma de uma rede social, coloca pesquisadores como nodos ou vértices da rede e os relacionamentos de colaboração entre eles como arestas ou *links* dessa rede, caracterizando assim uma rede de colaboração científica. Sendo a coautoria de artigos, o indicador de colaboração mais utilizado nos estudos de colaboração, esse tipo de rede também é chamada de rede de coautoria (*co-authorship network*).

A análise de uma rede da colaboração científica fornece respostas a importantes questionamentos que fazem do contexto da pesquisa científica, tais como: Qual o nível de colaboração da comunidade de pesquisa que está sendo analisada? Quem são os grupos que trabalham desconectados (isoladamente)? Quais os autores que mais colaboram? As respostas a essas questões fornecem insumos para a tomada de decisão em diversos níveis organizacionais num contexto científico, os

quais podem variar desde o pesquisador até o gestor da maior instância a qual ele está vinculado.

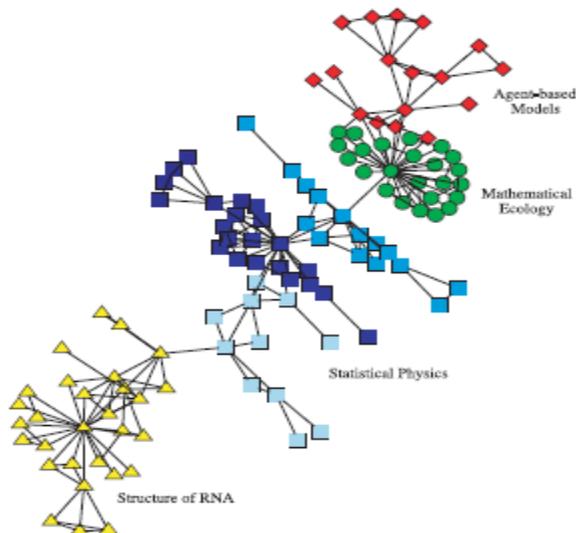
Existem diversas métricas de ARS que podem fornecer essas respostas. Métricas que analisam a rede como um todo, métricas que analisam subgrupos de atores e métricas que analisam especificamente a posição do ator dentro da rede.

Métricas de ARS, além de analisarem a situação atual da rede, também podem ser utilizadas para fazer a predição de novos relacionamentos.

Newman (2004a) destaca que o estudo da colaboração científica sob o enfoque de rede revela muitas características interessantes das comunidades acadêmicas. A Figura 1, por exemplo, exibe uma rede coautoria onde os nodos representam os autores e as arestas representam a coautoria entre os autores.

Somente pela estrutura visual da rede é possível identificar os autores com maior número de coautorias, a distância entre dois autores, comunidades de autores, entre outros elementos.

Figura 1 - Rede de coautoria



Fonte: Newman (2004a)

Além das descobertas apoiadas na estrutura visual, com a aplicação das diversas métricas que a ARS fornece, suportado pelo

embasamento matemático das mesmas, é possível descobrir diversas outras características.

No entanto, a revisão inicial de literatura na área de ARS demonstrou a existência de muitos trabalhos de análise de rede de colaboração científica. Essa constatação conduziu a uma revisão nas áreas de ARS e Web Semântica, que apontou para o problema da falta de semântica na representação de uma rede social, o qual se tornou o ponto de partida da construção deste trabalho de pesquisa.

Semântica neste caso refere-se a dar significado a algo. Essa definição de semântica apoia-se na própria definição de Web Semântica dada por Berners-Lee (2001), afirmando que a “web semântica é uma extensão da web atual, na qual é atribuído à informação um significado bem definido, habilitando pessoas e computadores a trabalhar em melhor cooperação”.

No contexto da ARS, a falta de semântica está relacionada aos elementos principais de uma rede: nodos e arestas. Consequentemente, isso conduziu ao entendimento de que as descobertas possíveis através da ARS podem ser enriquecidas adicionando semântica aos elementos envolvidos na representação da rede.

Metodologias e técnicas de modelagem e representação do conhecimento, advindas da Engenharia do Conhecimento, mas popularizadas pela Web Semântica³ se propõem a fornecer semântica ou significado aos elementos representados.

A seguir são apresentados todos os argumentos que conduziram à formulação do problema de pesquisa desta tese.

1.2 APRESENTAÇÃO DO PROBLEMA

Os argumentos ou pilares da problemática foram construídos a partir do aprofundamento da revisão de literatura e estão divididos em três partes:

1) **Representação da rede de colaboração:**

Na Análise de Redes Sociais (ARS) uma rede é entendida simplesmente como uma representação em grafo onde os nodos estão interligados por arestas, não existindo nenhuma tipificação desses elementos (nodos e arestas). A falta de elementos que expressem a semântica dos atores (nodos) e relacionamentos (arestas) da rede é considerada pelos autores apresentados a seguir um fator limitante.

Para Chelmis e Prassana (2011), as representações em grafo e as análises realizadas em cima dessas representações exploram de uma

³ W3C : <http://www.w3.org/standards/semanticweb/>

maneira superficial os complexos tipos de relacionamento e carecem de semântica.

Chen, Wei e Qingpu. (2010) e Li et al. (2010) destacam que a ARS tradicional considera somente as características estruturais da rede, tratando nodos e relações de forma homogênea, não permitindo que mais informações sobre os mesmos sejam analisadas. Além disso, a maioria das métricas de ARS são aplicadas a redes com somente um tipo de relação. Entretanto, nodos e relações presentes num contexto social possuem muitas características passíveis de serem exploradas.

Erétéo et al. (2008) afirmam que as representações baseadas em grafo estão preocupadas somente com a sintaxe, ou seja, elas não oferecem semântica e possibilitam uma exploração superficial dos tipos de nodos e das relações. Em outros trabalhos, Erétéo et al. (2009a) e Erétéo, Gandon e Buffa (2009b) argumentam que os dados de redes sociais são ricos e difusos e não podem ser representados usando grafos crus e algoritmos de análise de redes sociais sem perda de conhecimento.

Chen, Wei e Qingpu. (2010) e Erétéo et al. (2011) concordam no argumento de que *frameworks* ou tecnologias que se propõem a fornecer semântica aos dados, como os relacionados a Web Semântica⁴, podem ser utilizados para solucionar o problema da representação e da troca de conhecimento em redes sociais.

Neste trabalho considera-se que a colaboração científica, quando modelada e analisada sob a ótica de uma rede social (ARS), herda as limitações semânticas expostas pelos autores citados acima. Isso significa que a representação dos atores e os relacionamentos de uma rede de colaboração científica nas suas formas originais (grafo e matriz) e as consequentes análises que podem ser feitas a partir dessa representação são carentes de elementos semânticos que possam, por exemplo, identificar os tipos e as características de atores e os diferentes tipos de relacionamento de colaboração que possam existir numa rede de colaboração científica.

Entende-se também, assim como sugerido por Chen, Wei e Qingpu (2010) e Erétéo et al. (2011), que a utilização de algumas

⁴ São exemplos de *frameworks* ou tecnologias relacionados a Web Semântica: modelo de grafo tipado (*Resource Description Framework - RDF*), linguagem de consulta (*Simple Protocol and RDF Query Language - SPARQL*), *frameworks* de definição de esquema (*Resource Description Framework Schema - RDFS* e *Ontology Web Language - OWL*), etc.

tecnologias relacionadas à Web Semântica, sejam uma forma de resolver a falta de semântica na representação de uma rede de colaboração científica.

2) Análise do conteúdo produzido pela rede

As métricas de ARS analisam a rede sob o aspecto estrutural, indicando posições de centralidade de atores na rede, densidade da rede, caminhos entre atores, etc. A análise do conteúdo dos artefatos resultante dos relacionamentos, tais como os artigos produzidos em coautoria em uma rede de colaboração científica, pode complementar a análise da rede feita pelas métricas de ARS.

Para Chelmiss e Prassana (2011) existe uma sinergia entre a estrutura social da rede e o conteúdo publicado por ela e a ARS, por si só, não é suficiente para entender essa sinergia, ou seja, a análise do conteúdo da rede social também é crucial.

No que refere-se a colaboração científica, o conteúdo das colaborações pode ser explorado de forma a revelar, por exemplo, o perfil de pesquisa do ator que mais colabora (medido com a Centralidade de Grau). A análise do conteúdo dos artefatos resultantes das colaborações - artigos, projetos, teses, dissertações, etc. - permite qualificar ou fornecer um significado adicional à análise estrutural fornecida pela ARS.

Os trabalhos de Yang et al. (2012) e Xu et al. (2012) exploram a dimensão estrutural de uma rede social conjuntamente com a dimensão do conteúdo relacionado ou produzido por essa rede, com o objetivo de recomendar especialistas (acadêmicos). Os autores argumentam que abordagens de recomendação baseadas em rede social usam somente as informações de conectividade de rede e desconsideram a semântica relacionada aos conhecimentos das pessoas. Eles acreditam que as relações sociais e a informação semântica relacionada ao conteúdo das áreas de pesquisa dos especialistas são dois fatores que devem ser combinados.

Técnicas de mineração de dado textual, tais como extração de informação, indexação e análise de agrupamentos podem ser utilizadas para revelar informações que complementem a análise estrutural.

3) Entendimento das métricas da ARS

A ARS fornece uma série de medidas que podem ser difíceis de serem compreendidas e aplicadas na análise de um domínio de rede, principalmente por não especialistas em análise de redes.

Todas as medidas de ARS resultam valores que por si só, são apenas números. O entendimento do número exige o entendimento da medida, de como ela é calculada, mas também da explicação sociológica

da medida na rede em questão. Por exemplo, um determinado nodo da rede tem Centralidade de Grau igual a 10. Como essa medida é calculada? E mais importante, o que ela indica? Ela é calculada pelo número de nodos adjacentes. Ela indica que esse nodo é conectado, ou colabora cientificamente no caso de uma rede de colaboração científica, com 10 pesquisadores.

Segundo Scott (2000), apesar de haver um interesse crescente na utilização dos métodos de análise de redes sociais, os mesmos têm sido vistos como inatingíveis para muitos investigadores, que encontraram dificuldades para entender a linguagem técnica e matemática em que muitos destes métodos são discutidos.

Esse argumento é abordado no trabalho de Balancieri (2010), onde o autor afirma que os métodos de ARS, para serem efetivos, necessitam de um especialista em ARS que possa traduzir os índices produzidos para a linguagem do domínio do problema analisado. Portanto, a dependência do especialista em ARS limita a aplicabilidade do método a situações onde este especialista está disponível.

No domínio de uma rede de colaboração científica, por exemplo, um grau de centralidade alto indica que o ator tem um alto nível de colaboração e um grau de proximidade alto indica que o ator é o mais próximo dos demais atores da rede. Esses exemplos denotam que a aplicação das métricas de ARS só é útil se contextualizada ao domínio que está sendo analisado e isso demanda o entendimento das métricas de ARS, normalmente detido por especialistas em ARS.

No entanto, entende-se que encontrar especialistas no domínio com conhecimento suficiente de ARS não é trivial, o que sugere a existência de mecanismos que representem o conhecimento do especialista de ARS na realização de tarefas de análise de rede de colaboração.

Assim como Balancieri (2010), entende-se que a Engenharia do Conhecimento oferece apoio à solução desse problema através de metodologias e técnicas de modelagem e representação do conhecimento. Através de metodologias e técnicas, o Engenheiro do Conhecimento pode extrair, modelar e representar o conhecimento do especialista em ARS de como resolver essas tarefas. Indo mais além, esse processo de extração, modelagem e representação pode ser estendido para qualquer tarefa intensiva em conhecimento de análise de colaboração científica.

Na Engenharia do Conhecimento, conhecimento de domínio é o conhecimento sobre uma área específica de interesse e conhecimento de

tarefa está relacionado com a maneira de resolver problemas usando esse conhecimento de domínio (MIZOGUCHI; VAN WELKENHUYSEN; IKEDA, 1995, STRUSS, 2008).

Partindo dessa definição operacional e dos pilares da problemática que evidenciaram algumas lacunas ou oportunidades de pesquisa, categoriza-se os problemas que esta tese pretende investigar nos seguintes pontos:

- a) Representação do conhecimento de domínio de uma rede de colaboração científica;
- b) Processamento e análise dos artefatos resultantes da colaboração como complemento à análise estrutural da Análise de Rede Social.
- c) Representação do conhecimento de tarefa necessário para a realização de tarefas de análise de colaboração;

A combinação desses pontos remete à pergunta de pesquisa dessa tese que é descrita na seção a seguir.

1.3 PERGUNTA DE PESQUISA

Como representar o conhecimento de domínio e o conhecimento das tarefas de análise de uma rede colaboração científica de modo que esta representação possa auxiliar o processo de análise da referida rede?

1.4 OBJETIVOS

A seguir são apresentados os objetivos gerais que respondem a pergunta de pesquisa e os objetivos específicos necessários para o alcance dos objetivos gerais desta tese.

1.4.1 Objetivo Geral

O1: Propor e desenvolver um conjunto de representações dos conhecimentos envolvidos na análise de uma rede de colaboração científica, bem como um conjunto de artefatos para armazenar e recuperar dados dessas representações combinados em um *framework*⁵.

O2: Avaliar a contribuição do *framework* no processo de análise de uma rede de colaboração científica, no que tange a relevância das tarefas de análise propostas, aos resultados fornecidos e a sua aplicabilidade como instrumento de gestão.

⁵ A definição de *framework* encontra-se no Glossário.

1.4.2 Objetivos Específicos

- a) Representar o conhecimento de domínio de uma rede de colaboração científica e das métricas de Análise de Rede Social (ARS) através de ontologias;
- b) Identificar necessidades ou tarefas de análise de rede de colaboração científica relevantes;
- c) Representar o conhecimento de tarefas e de solução de tarefas de análise de rede de colaboração científica através de ontologias;
- d) Implementar os métodos que realizam as tarefas propostas no *framework*;
- e) Implementar mecanismos para armazenar e recuperar os dados das representações ontológicas;
- f) Propor um *framework* que agregue as representações ontológicas e os mecanismos de armazenamento e recuperação de dados;
- g) Demonstrar e avaliar o *framework* em cenários de colaboração científica distintos;

1.5 JUSTIFICATIVA

A apresentação da justificativa, ou relevância dessa pesquisa, se dá em duas dimensões: social e científica. A relevância social caracteriza a contribuição que a pesquisa traz para sociedade. A relevância científica caracteriza a contribuição ou avanço na área de conhecimento onde está inserida.

Na dimensão social entende-se que essa pesquisa é relevante porque auxilia a gestão da pesquisa nas organizações de P&D.

Genericamente a função de gestão da pesquisa (research management) em contextos de P&D consiste em coletar os resultados (output) de pesquisa; armazenar em um sistema de informação; extrair resultados ou indicadores diversos tais como, produtividade, colaboração, entre outros; disponibilizar e utilizar esses indicadores para a tomada de decisões estratégicas, as quais normalmente incluem a definição de políticas e metas de P&D.

A gestão da pesquisa é fundamental para o desenvolvimento de uma estratégia de pesquisa eficaz, para a construção de programas de pesquisa sólidos, para o crescimento das atividades de pesquisa e para o alinhamento das prioridades institucionais com os critérios de financiamento de agência de fomento (KATERNDAHL, 2012).

De acordo com o relatório intitulado *The Future of Research Management* produzido em 2010 pelo *National Council of University Research Administrators (NCURA)* dos Estados Unidos e pela *European Association of Research Managers and Administrators (EARMA)*, as atividades de gestão de pesquisa têm enfrentado grandes desafios, relacionados principalmente a critérios como a existência de colaboração científica, exigidos pelas agências de financiamento desses países. Conforme tal relatório, a competição por financiamentos de pesquisas está cada vez mais acirrada e os critérios de avaliação das propostas medem até que ponto os projetos são multidisciplinares e colaborativos.

O relatório observa que, mesmo quando não existe um critério pré-estabelecido relacionado à colaboração, as agências de fomento procuram identificar a existência de colaboração de algum pesquisador de renome no campo em questão. Além disso, destaca que existe uma tendência crescente de surgimento de fundos de financiamento que privilegiam a colaboração de pesquisa internacional.

Compreender as redes de colaboração científica e conseguir extrair informações relevantes dessas redes é, portanto, uma atividade de gestão de pesquisa vital para a promoção da produtividade acadêmica de sistemas e indivíduos. Tal compreensão pode permitir o direcionamento a editais de fundos de financiamento, impulsionar pesquisas existentes, propiciar o surgimento de novas pesquisas, desenvolver programas de pós-graduação, ajudar a desenvolver estratégias para maximizar a produtividade e diagnosticar causas de baixa produtividade, entre outros fatores.

A colaboração é capaz de promover a atividade de pesquisa, a produtividade, o impacto da pesquisa e, assim, deve ser incentivada e apoiada pelos meios de gestão de pesquisa e políticas de ciência. Os benefícios, no entanto, não aparecem automaticamente, o que evidencia a necessidade de mecanismos de monitoramento quantitativo de insumos e dos resultados das pesquisas (GLÄNZEL; SCHUBERT, 2004).

Na Europa e nos Estados Unidos a função do administrador ou gerente de pesquisa (*research manager*) é mais consolidada e existem associações nesses países, tais como *NCURA* na Europa e *EARMA* nos Estados Unidos, já mencionadas anteriormente, que fomentam discussões e especialização desses profissionais. Logo, é mais comum a existência de escritórios ou órgãos de gestão de pesquisa dentro das organizações de P&D, que se preocupam em monitorar editais de financiamento, assessorar o desenvolvimento de propostas, ajudar na

execução das atividades burocráticas do projeto, monitorar indicadores de produtividade e colaboração, entre outros, deixando o pesquisador com mais tempo disponível para executar suas atividades de pesquisa.

No Brasil, essa é uma realidade mais emergente, com algumas iniciativas em universidades, na forma de escritório de projetos que se propõem a divulgar editais nacionais e internacionais e auxiliam o pesquisador no desenvolvimento de propostas e posteriormente na prestação de contas (atividades administrativas) e na divulgação da pesquisa. A Universidade Estadual Paulista (UNESP), através do Escritório Central de Apoio à Pesquisa (Ecap), e a Universidade Estadual de Campinas (UNICAMP), através da Unidade de Apoio ao Pesquisador, são exemplos de iniciativas brasileiras⁶.

Ainda, em nível nacional, o cenário de monitoramento sistemático de dados referentes a produção intelectual, para futuras análises e tomada de decisão, também parece ser incipiente e carecedor de mecanismos mais inteligentes. O sistema Coleta de Dados Capes, agora um módulo de coleta de dados da Plataforma Sucupira⁷, tem o objetivo de coletar dados de programas de pós-graduação para avaliação, mas também serve para prover à Capes informações necessárias ao planejamento dos seus programas de fomento e delineamento de suas políticas institucionais. Esse sistema, alimentado pelos programas de pós-graduação, por exigência da Capes, está dentro de um contexto de gestão de pesquisa, mas não oferece, principalmente para os programas pós-graduação, recursos de análise de produção e colaboração mais robustos e inteligentes.

A questão da exigência mencionada anteriormente remete a outra percepção, a de que em muitas universidades as atividades de gestão são feitas para cumprimento de exigências de agências de fomento ou outros órgãos. Não existe, portanto, uma consciência ampla dos potenciais benefícios que atividades de gestão de pesquisa, apoiadas por dados atualizados e processados de forma inteligente, possam trazer a comunidade acadêmica. Por outro lado, é possível que essa consciência não seja tão ampla pela inexistência, em muitos contextos, de mecanismos que utilizem os *outputs* de pesquisa e que gerem resultados (dados) interessantes para a tomada de decisão.

O *framework* proposto nessa tese tem o objetivo de auxiliar a gestão de pesquisa nos contextos de P&D, oferecendo uma proposta

⁶ <http://agencia.fapesp.br/13704>

⁷ <https://sucupira.capes.gov.br/sucupira/>

baseada na Engenharia do Conhecimento para representação e recuperação do conhecimento de redes de colaboração científica. Com essa proposta pretende-se oferecer um mecanismo para fornecer respostas a diversas questões de análise da colaboração científica, contribuindo para o entendimento de tal contexto.

O desenvolvimento de uma proposta fundamentada na Engenharia do Conhecimento para representação e análise de redes de colaboração científica traz alguns desafios e se justifica na dimensão científica pelas contribuições de pesquisa que pretende entregar.

As contribuições de pesquisa estão diretamente relacionadas às limitações expostas na seção de apresentação do problema, pois buscam propor alternativas de solução para tais limitações.

Em relação ao problema da falta de semântica envolvida nas representações de rede baseadas em matriz ou grafo, a qual possibilita uma exploração superficial dos elementos (nodos e relações) dessa rede, entende-se que é necessário um novo modelo de representação e armazenamento que descreva ou tipifique esses elementos com mais atributos, bem como suporte os diversos tipos de relacionamentos que podem existir em uma rede. Por exemplo, um nodo em uma rede de colaboração científica representado na forma tradicional, corresponde a um único tipo, é um “pesquisador”, e pode ter como atributo o nome ou abreviação do pesquisador. Em um modelo que permite a descrição ou tipificação dos dados e relacionamentos, esse nodo pode ter uma hierarquia de tipos e ter muitos atributos, além de tipos de relacionamentos diversos com outros nodos.

Modelos desse tipo se justificam porque conseguem capturar a realidade de sistemas modelados em rede com todas as suas características e relacionamentos. Além disso, eles propiciam que esse conjunto de dados possa ser integrado, através de uma com outros conjuntos de dados, criando uma grande rede semântica. Dessa forma também é possível recuperar os dados de forma mais inteligente.

A Engenharia do Conhecimento, ou dentro a partir de uma abordagem mais atual, as tecnologias de Web Semântica ou Web dos Dados, através do modelo *RDF (Resource Definition Framework)*, *RDFS (Resource Definition Framework Schema)* e Ontologias *OWL*, permitem a construção de um vocabulário ou uma representação de dados padrão que expressa de maneira mais rica o significado de conceitos e relacionamento, propicia ligações entre diferentes conjuntos de dados e promove a integração e comunicação desses dados. Segundo Krafft (2010), representações desse tipo podem servir como uma solução eficaz para o aumento da demanda de comunicação mais ampla

e aprofundada entre os cientistas da comunidade acadêmica em todo o mundo.

Em relação ao problema da análise do conteúdo originado da colaboração, que pode ser uma fonte adicional no processo de análise de rede de colaboração, entende-se que este conteúdo requer primeiramente ser representado, mas também necessita ser processado, inspecionado ou minerado para a descoberta de informações que complementem a análise de rede.

A utilização de técnicas que descubram os termos mais representativos de uma produção bibliográfica (fruto da colaboração entre pesquisadores), quando aplicadas no conjunto das produções de um pesquisador, propiciam, por exemplo, a criação do perfil de pesquisa (vetor de termos) desse pesquisador, a caracterização do perfil dos grupos de pesquisa dentro da rede, etc.

No *framework* proposto nesta tese a maior parte dos conceitos relacionados a uma rede de colaboração científica, bem como os artefatos diversos originados pela colaboração, com todas as suas características e relacionamentos são representados conjuntamente por meio de uma ontologia de domínio. Além disso, os resultados de um processo de mineração da descrição dos artefatos (título da produção, palavras-chave) produzidos por um pesquisador para formar o perfil de pesquisa do mesmo também são representados nessa ontologia de domínio.

Em relação ao problema do entendimento das métricas de ARS no processo de análise de rede e consequente necessidade de um especialista em ARS, é entendido que a busca por uma solução que represente o conhecimento desse especialista, de como executar ou realizar as tarefas de análise de colaboração científica, é uma possibilidade factível diante das metodologias de modelagem de conhecimento que a Engenharia do Conhecimento oferece.

Ao utilizar uma metodologia de modelagem de conhecimento, o engenheiro do conhecimento elícita junto ao especialista de ARS todos os passos de inferência ou raciocínio e requisitos do domínio de rede necessários para executar as tarefas de análise de colaboração, de forma a criar blocos de construção (conjuntos de passos) que são comuns a várias tarefas, permitindo com isso a reutilização desses blocos.

Este trabalho visualizou nessas lacunas uma oportunidade para a proposição de uma série de artefatos que representam com mais expressividade o conhecimento de domínio de uma rede de colaboração científica e o conhecimento das tarefas intensivas em conhecimento de

análise de colaboração dessa rede, compondo, assim, o *framework* de conhecimento para análise da colaboração científica.

Na Engenharia do Conhecimento um sistema que modela e representa o conhecimento de domínio separado do conhecimento de como resolver problemas utilizando o conhecimento de domínio é denominado de Sistema Baseado em Conhecimento (SBC).

Para Studer, Benjamins e Fensel (1998) a construção de um SBC implica na construção de um modelo de computador com o objetivo de implementar as capacidades de resolução de problemas comparável à de um especialista de domínio. A ideia não é criar um modelo cognitivo para simular os processos cognitivos de um especialista em geral, mas criar um modelo que ofereça resultados semelhantes na resolução de problemas na área em questão.

A construção de sistemas inteligentes passou de uma abordagem baseada na transferência de conhecimento para uma abordagem baseada na modelagem do conhecimento, a partir do trabalho de Newell (1982), onde o autor propôs a noção de “nível de conhecimento”, que busca descrever o raciocínio em termos de objetivos a serem alcançados, as ações necessárias para atingir esses objetivos e os conhecimentos necessários para executar essas ações.

A motivação principal do paradigma de modelagem de conhecimento é modelar sistemas em termos do conhecimento que eles têm, mais do que os seus detalhes de implementação. Uma descrição em nível de conhecimento de um processo de resolução de problemas abstrai os detalhes relacionados com a implementação do processo de raciocínio e resulta na noção de um Método de Resolução de Problemas (*Problem-Solving Method* ou *PSM*).

Portanto, inspirado na abordagem de modelagem de conhecimento e na noção de *PSM*, o *framework* proposto neste trabalho representa o comportamento ou conhecimento de um especialista na solução de tarefas de análise de colaboração, tarefas que demandam o conhecimento de análise de redes e também de análise do conteúdo produzido pela rede. Para isso, é proposta uma modelagem e representação do conhecimento de domínio da rede de colaboração, dos artefatos resultantes da colaboração e das medidas de ARS, conjuntamente com uma modelagem e representação do conhecimento das tarefas de análise de colaboração e dos passos (métodos) necessários para resolver essas tarefas.

A contribuição para a área de Engenharia do Conhecimento se dá essencialmente na modelagem e representação do conhecimento em nível de domínio e de resolução de problemas (tarefas e métodos) nesse

domínio, os quais podem ser reutilizados em outros contextos ou aplicações e que, quando combinados, se propõem a fornecer um mecanismo que permitirá análises mais ricas e inteligentes acerca do domínio modelado, trazendo com isso vantagens no processo de tomada de decisão no contexto de ARS.

O conhecimento de domínio nessa pesquisa, diz respeito à rede de colaboração científica, aos artefatos produzidos pela mesma e as métricas de ARS, que da maneira como foram modelados e representados, podem ser reutilizadas em outros contextos ou aplicações.

O conhecimento de tarefas e métodos, modelados através da técnica de modelagem de conhecimento denominada de Análise de Estrutura de Tarefa (*Task-Structure Analysis*) proposta por Chandrasekaran, Johnson e Smitt (1992), onde uma tarefa define "o que tem que ser feito" e métodos definem "como executar uma tarefa", permite que uma tarefa seja especificada de forma que tanto a tarefa quanto os seus métodos possam ser reutilizados em outros contextos ou aplicações.

A ratificação da contribuição dessa proposta para a Engenharia do Conhecimento, advém de O'Connor et al. (2009), no qual os autores destacam que a reutilização do chamado conhecimento de controle, que é o conhecimento de como executar uma tarefa, é um grande desafio da engenharia de sistemas baseados em conhecimento. Para os autores a engenharia de sistemas baseados em conhecimento ainda é extremamente difícil e, na prática, reutilizar conhecimento não é mais fácil agora, do que era quando as pesquisas na área começaram.

Além disso, para complementar o desafio, O'Connor et al. (2009) destacam que não existe um conjunto disponível e amplo de *PSMs* implementados e não existem ambientes de desenvolvimento robustos de *PSM*.

A convergência de todas as lacunas ou oportunidades de pesquisa encontradas no contexto da Análise de Redes Sociais (ARS), aplicada a Colaboração Científica, quando traduzidas para uma proposta fundamentada na modelagem e representação de conhecimentos da Engenharia do Conhecimento, tornam esse trabalho inédito e relevante para as áreas abordadas nesta pesquisa.

1.6 ADERÊNCIA DA TESE AO PROGRAMA DE PÓS-GRADUAÇÃO E ÁREA DE PESQUISA

Essa tese foi desenvolvida no âmbito do Programa de Pós-graduação em Engenharia e Gestão do Conhecimento (PPGEGC,) com foco na área de concentração Engenharia de Conhecimento (EC).

O PPGEGC é um programa de natureza interdisciplinar criado em meados de 2004 e desde sua concepção tem caracterizado seu foco de pesquisa e formação no conhecimento e nos processos que lhe tornam fator gerador de valor na sociedade contemporânea. Além do conhecimento enquanto produto, o PPGEGC também trata dos processos que o tornam gerador de valor.

Sua estruturação em áreas de concentração é justamente decorrente da atribuição de missões inter-relacionadas aos processos de codificação/formalização (área de Engenharia do Conhecimento); planejamento e gerência (área de Gestão do Conhecimento); difusão, comunicação e compartilhamento (área de Mídia e Conhecimento) do conhecimento. O PPGEGC, portanto, considera que seu objeto de pesquisa e formação é essencialmente interdisciplinar (PACHECO; TOSTA; FREIRE, 2010).

A disciplina Engenharia do Conhecimento surgiu do trabalho inicial com sistemas especialistas na década de setenta. Com a crescente popularidade dos sistemas baseados em conhecimento (como passaram a ser chamados), também surgiu a necessidade de uma abordagem sistemática para a construção de tais sistemas, semelhantes às metodologias de Engenharia de Software (STUDER et al., 1998).

Ao longo dos anos, a disciplina Engenharia de Conhecimento evoluiu com o desenvolvimento de teorias, métodos e ferramentas para o desenvolvimento de aplicações intensivas em conhecimento. Segundo Schreiber (2007), ela fornece uma orientação sobre quando e como aplicar técnicas de representação de conhecimento para resolver problemas particulares.

A intensificação do papel do conhecimento como elemento estratégico nas organizações, a partir dos anos noventa, fez com que a geração, a codificação e a gestão do conhecimento organizacional tornassem-se tarefas essenciais às organizações. Nesse contexto, a Engenharia do Conhecimento se consolidou.

No PPGEGC os objetivos da área de Engenharia do Conhecimento incluem a pesquisa e o desenvolvimento de técnicas e ferramentas para a formalização, codificação e gestão do conhecimento; de métodos de análise da estrutura e processos conduzidos por

profissionais em atividades de conhecimento intensivo; e a pesquisa e desenvolvimento de sistemas de conhecimento.

Sendo assim, esta tese é primeiramente aderente a este programa interdisciplinar porque propicia a interação entre diferentes disciplinas, tais como Colaboração Científica, Análise de Redes Sociais e Engenharia do Conhecimento em prol de uma proposição que fornece insumos para auxiliar a gestão em organizações de P&D.

Além disso, ela se caracteriza como uma pesquisa da área de Engenharia do Conhecimento, quando aborda a gênese dessa área, objetivando modelar e representar os diferentes tipos de conhecimentos envolvidos em uma rede de colaboração científica, combinando-os em um *framework* ou sistema de conhecimento útil a gestão da colaboração científica.

1.7 ORGANIZAÇÃO DO TRABALHO

O trabalho está organizado da seguinte forma: o capítulo 2 aborda o referencial teórico que deu suporte a proposição do *framework*, o capítulo 3 apresenta a metodologia - baseada na abordagem *Design Science Research* - de construção deste trabalho, o capítulo 4 apresenta o *framework* proposto, o capítulo 5 o processo de avaliação através de entrevista com grupos focais e, por fim, o capítulo 6 apresenta as considerações finais do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Os tópicos apresentados nesse capítulo e que fundamentam essa pesquisa são: colaboração científica, análise de redes sociais e sistemas baseados em conhecimento. A seguir será apresentado o referencial teórico de cada um dos tópicos.

2.1 COLABORAÇÃO CIENTÍFICA

A afirmação de que a ciência é uma atividade eminentemente social, visto que os resultados produzidos por ela são utilizados pela e em prol da sociedade foi feita inicialmente feita por Bernal (1939). Além dele, muitos autores destacam que a atividade científica em si é um processo social, fruto da interação explícita e implícita entre pesquisadores (KATZ; MARTIN, 1997; ZIMAN, 1979). Essa interação social em prol de resultados científicos é conhecida como colaboração científica.

Segundo Stevens e Campion (1994) uma tendência histórica em direção à especialização da ciência trouxe a necessidade de uma colaboração multidisciplinar para reunir os conhecimentos, competências e habilidades necessárias para o avanço da pesquisa. Isso pode ser comprovado em diversos estudos bibliométricos feitos nas últimas duas décadas, os quais mostram um aumento contínuo no número de trabalhos com coautoria em diversas disciplinas científicas, bem como dentro e entre países e áreas geográficas. (GROSSMAN, 2002; WAGNER; LEYESDORFF, 2005)

O interesse pelos processos que permeiam a colaboração científica, tais como os motivos que levam pesquisadores a colaborar, os benefícios, os custos, dificuldades, assim como indicadores de colaboração, evidencia-se a partir da década de 60 (SMITH, 1958; PRICE; BEAVER, 1966; BEAVER; ROSEN, 1978, 1979).

Esse interesse se mantém uma vez que existe um consenso de que colaboração no meio acadêmico é uma necessidade e traz consigo muitas vantagens. Katz e Martin (1997) corroboram com esse argumento afirmando que existe um crescente interesse pela colaboração e que há uma noção generalizada de que ela traz “coisas boas”. Simonin (1997) relata que tem havido um aumento na colaboração como ferramenta da ciência e destaca a necessidade de desenvolvimento do *know-how* em colaboração.

Para Milojevic (2010), Persson, Glänzel e Danell (2004) e Wagner e Leyesdorff (2008), uma das características que definem ciência moderna nas últimas décadas é a colaboração. A seguir será apresentada uma série de conceitos que permeiam esse tópico de estudo.

2.1.1 Definição, Categorização e Natureza das Atividades

Hara (2003) destaca que os pesquisadores têm dado pouca atenção para a definição de colaboração. Em muitos casos, o termo "colaboração", é usado de forma intuitiva e intercambiável com outros termos como "cooperação" e "coordenação".

O autor destaca uma série de definições no contexto organizacional e individual, e conclui, no qual é acompanhado por Katz e Martin (1997), que o conceito pressupõe ao menos dois elementos: 1) trabalhar conjuntamente por um objetivo em comum e 2) compartilhar conhecimento.

Partindo dos elementos que a colaboração pressupõe, pode-se dizer que a colaboração científica acontece quando existe trabalho e compartilhamento em prol de um conhecimento científico.

Como a colaboração científica é abordada em diversos campos de investigação, uma variedade de terminologias, abordagens e métodos de investigação podem ser encontrados na literatura. O termo colaboração científica também é referido como colaboração de pesquisa (*research collaboration*) (KATZ e MARTIN, 1997; GLÄNZEL; SCHUBERT, 2004; LIAO et al., 2012), colaboração de P&D (*R&D collaboration*) (CHOMPALOV; GENUTH; SHRUM, 2002) e ciência de equipe (*team science*).

Diversos tipos ou categorias de colaboração emergem das características do contexto científico. Colaboração com foco disciplinar é chamada de inter, multi, trans, cross e intra disciplinar; Colaboração com um foco geográfico é denominada colaboração internacional ou colaboração remota; Colaboração interinstitucional, colaboração participativa ou universidade-comunidade, colaboração universidade-indústria, possuem um com foco organizacional. Termos como colaboração ciência-sociedade, colaboração em grande escala (*big science*, equipes de equipes) determinam outros escopos de estudos (ZIMAN, 2000; SONNENWALD, 2007).

Dentro de um contexto de colaboração científica, as tarefas muitas vezes têm um alto grau de incerteza, mais do que em outros tipos de trabalho. Tarefas de pesquisa podem ser compartilhadas entre os cientistas de várias maneiras. Algumas tarefas são divisíveis e podem ser realizadas sequencialmente ou simultaneamente (WHITLEY, 2000). Outras tarefas podem ser feitas em conjunto, onde todos devem completar a tarefa. Por exemplo, em uma colaboração em ciências naturais, um cientista pode sugerir uma questão de pesquisa e desenvolver amostras de dados. Um segundo cientista pode analisar as

amostras usando instrumentação científica especializada. A pesquisa sugere que as demandas da tarefa, os recursos disponíveis, interação do grupo, o grau de dependência funcional entre os cientistas, e o grau de dependência estratégica determine como as tarefas são alocadas e compartilhadas entre os cientistas em uma colaboração (WHITLEY, 2000).

2.1.2 Estágios da Colaboração

Para auxiliar a compreensão do processo de colaboração científica, Sonnenwald (2007) apresenta um *framework* no qual a colaboração científica possui quatro estágios: criação, formulação, sustentação e conclusão.

a) Estágio de Criação

O estágio de criação se concentra em fatores que proporcionam ou motivam e impactam o início da colaboração. Os fatores são classificados em cinco categorias: científico, político, socioeconômico, acessibilidade de recursos, redes sociais e fatores pessoais. Katz e Martin (1997) em um conhecido e bem referenciado estudo sobre a colaboração em pesquisa apresentam uma síntese de fatores que contribuem ou motivam a colaboração e que se enquadram nessas categorias. Em Beaver (2001) também pode ser encontrado uma lista de dezoito propósitos pelos quais os pesquisadores colaboram.

Na categoria de fatores científicos, a oportunidade de descobrir novos conhecimentos e resolver problemas complexos em tempo hábil, motiva muitos cientistas a considerar a colaboração. Katz e Martin (1997) argumentam que é crescente a necessidade de especialização dentro de certos campos científicos, especialmente aqueles onde a instrumentação necessária é muito complexa. Isso quer dizer que um pesquisador isolado já não consegue conduzir pesquisas em algumas áreas, motivando portanto a colaboração. Isso pode ser mais bem observado em áreas como Física de Alta Energia, onde para se realizar um experimento, é preciso reunir especialistas de diversas áreas.

Outro fator motivador está relacionado com a crescente importância da pesquisa interdisciplinar, visto que alguns dos avanços científicos mais significativos surgem como resultado da integração ou fusão de campos previamente isolados.

O desejo de aumentar a popularidade científica, a visibilidade e o reconhecimento; a necessidade de ganhar experiência ou formar pesquisadores aprendizes da forma mais eficaz possível; a necessidade de trabalhar em estreita proximidade física com os outros, a fim de

beneficiar de suas habilidades e conhecimento tácito são outros fatores motivadores apontados por Katz e Martin (1997).

Sonnenwald (2007) destaca que também há aspectos potencialmente negativos na colaboração científica. Há preocupação de que colaborações são por vezes usadas para esconder uma conduta antiética. Por exemplo, as colaborações entre países avançados e países em desenvolvimento podem ocorrer de forma a conduzir ensaios clínicos antiéticos, pesquisas que envolvam recursos naturais que são proibidos em países avançados, etc.

Aspectos negativos relacionados a responsabilidades podem aparecer. Quando muitos cientistas colaboram, ninguém pode se sentir responsável pelo trabalho. Além disso, colaborações podem tornar-se poderosos grupos de pressão, influenciando a políticas de pesquisa e decisões de financiamento em benefício próprio.

Na categoria dos motivadores políticos, o entendimento entre os países pode ser aumentado através da colaboração científica informal e formal. Sonnenwald (2007) exemplifica que durante a Guerra Fria, cientistas nos EUA e URSS estabeleceram e mantiveram relações que foram valiosas na promoção do fim da Guerra Fria.

A categoria dos motivos socioeconômicos está relacionada com a percepção de que a curto prazo, negócios ou empresas podem receber benefícios econômicos através de colaboração de pesquisa, obtendo créditos fiscais e acesso a financiamento público que de outra forma não estaria disponível para eles. Este último é conseguido através da colaboração com universidades em programas de pesquisa universidade-indústria patrocinados por governos. Os países também buscam colaboração para apoiar o desenvolvimento econômico nacional e regional. Muitos países têm programas de pesquisa que requerem colaboração entre universidades e indústrias, incluindo as pequenas e médias empresas.

Os motivos relacionados a acessibilidade de recursos se caracterizam pela necessidade de muitas pesquisas de ter acesso a instrumentos caros, dados científicos exclusivos, recursos naturais e sociais escassos e grandes quantidades de financiamento científico. Nessa linha Katz e Martin (1997) abordam o aumento de custos de pesquisa como um fator motivador de colaboração. Os autores argumentam que em muitos campos, os custos de instrumentação científica aumentaram consideravelmente com a introdução de novas tecnologias, tornando impossível para as agências de financiamento proporcionar as facilidades de pesquisa necessárias para todos os grupos

de pesquisa que atuam na área, forçando assim os pesquisadores a procurarem trabalhar em colaboração.

Por fim, Sonnenwald (2007) destaca que as redes sociais e fatores pessoais fornecem a base para a colaboração. As redes sociais podem alcançar as fronteiras disciplinares, organizacionais e nacionais. A colaboração frequentemente surge e se perpetua através das redes sociais. São nas suas redes que os cientistas olham para obter ideias sobre novos projetos de pesquisa e para identificar e selecionar colaboradores. Os fatores pessoais desempenham um papel na criação e manutenção de redes sociais e posteriormente colaborações. Compatibilidade pessoal, incluindo abordagens semelhantes para a ciência, estilos de trabalho semelhantes, respeito mútuo, confiança e capacidade de conviver também são utilizados para identificar e selecionar colaboradores.

Katz e Martin (1997) também apresentam como fator motivador as facilidades de comunicação cada vez mais atraentes, tanto em termos físicos de deslocamentos entre laboratórios de estados/países, como em termos virtuais, através das novas tecnologias de informação e comunicação. Os autores destacam que a colaboração é um processo intrinsecamente social e por isso pode haver muitos outros muitos fatores que contribuem para o início da mesma.

b) Estágio de formulação

Durante o estágio de formulação, os cientistas iniciam e planejam projetos de pesquisa de forma colaborativa. Sonnenwald (2007) diz que a literatura sugere que visão de pesquisa, objetivos e tarefas, liderança e estrutura organizacional, uso de tecnologia de informação e comunicação (TIC), propriedade intelectual e outras questões legais devem ser considerados em maior detalhe do que em uma pesquisa feita individualmente.

Uma visão de pesquisa e problemas complexos podem motivar os cientistas a colaborar. No entanto, na pesquisa colaborativa as visões e objetivos científicos são muitas vezes mais complexos do que aqueles enfrentados pela pesquisa feita individualmente. Por causa disso, as tarefas de pesquisa devem ser claramente definidas e assumidas pelos pesquisadores individualmente.

Os estudos de colaborações bem sucedidas mostram que a liderança é importante para o sucesso da colaboração. Colaborações são mais bem sucedidas quando a liderança tem experiência de gestão de projetos e é respeitada pelos participantes.

Colaborações podem ser organizadas de diferentes formas. Quatro tipos de organizações surgem a partir da análise de Chompalov,

Genuth and Shrum (2002): burocrática, sem liderança, não especializada e participativa. Colaborações burocráticas têm uma "hierarquia de autoridade, regras e regulamentos escritos, responsabilidades formalizadas, e uma divisão especializada do trabalho". Colaborações sem líderes têm administração, mas não tem liderança científica. Em colaborações não especializadas há gerenciamento hierárquico, mas menos formalização e diferenciação de papéis e responsabilidades. Colaborações participativas são igualitárias na medida em que não há um líder científico ou administrativo.

As TICs podem facilitar a colaboração científica e dar origem a novos tipos de colaboração, especialmente quando os cientistas não podem estar juntos presencialmente. Estudos identificaram algumas formas de colaborações científicas remotas com base em sua utilização das TICs: sistemas de instrumentos compartilhados, sistemas de dados comunitários, comunidade virtual de prática, comunidade virtual de aprendizagem, centros de pesquisa distribuídos.

É importante reconhecer que os direitos de propriedade intelectual e outras questões legais impactam a colaboração. Direitos de propriedade intelectual e sua globalização estão em fluxo e cada vez mais impulsionados por interesses privados. Questões relativas à propriedade intelectual podem surgir a partir de colaborações e as relações jurídicas entre os participantes de uma colaboração devem ser negociadas na fase de formulação para evitar incompreensões e conflito ou antes da propriedade intelectual com potencial de mercado ter sido criada.

c) Estágio de sustentação

Depois que uma colaboração é formulada e o trabalho começa, a colaboração precisa ser sustentada durante algum período de tempo, para que possa atingir os seus objetivos. Os desafios desse estágio podem ser identificados e tratados através de um processo contínuo de avaliação em que a estrutura e as tarefas de organização, comunicação e aprendizagem são examinadas e evoluem.

Dependendo do tamanho da colaboração e da diversidade dos participantes pode-se levar um tempo para chegar a um entendimento do funcionamento do trabalho compartilhado, da estrutura organizacional e das práticas de gestão.

Os cientistas também podem descobrir diferenças inesperadas. Por exemplo, perceberem que não têm normas compartilhadas com relação à participação dos alunos ou ao compartilhamento de informações sobre a pesquisa com pessoas de fora do grupo.

A confiança entre os cientistas é um componente integral de colaboração e visões conflitantes de cooperação e de competição podem surgir durante esta fase.

O tamanho de uma colaboração, as distâncias geográficas entre os cientistas participantes, a interdependência das tarefas e a competitividade podem exacerbar estes desafios. Desafios podem fazer uma colaboração mais forte e mais eficaz quando eles são tratados construtivamente.

A aprendizagem é um componente integral de colaboração científica, principalmente nesse estágio. Cientistas precisam aprender uns com os outros para desenvolver uma compreensão comum de trabalho em relação ao projeto de pesquisa, e como eles podem integrar seus conhecimentos especializados para criar novos conhecimentos.

A comunicação é outro componente fundamental de colaboração nesta etapa. Sem comunicação, tarefas não são coordenadas e os cientistas não aprendem uns com os outros.

d) Estágio de conclusão

No estágio de conclusão, os resultados de sucesso da colaboração idealmente devem emergir. Pode haver diferentes tipos de resultados de sucesso e a difusão e publicação dos resultados podem ajudar outros a aprenderem com a colaboração.

Um resultado de sucesso importante é a criação de novos conhecimentos científicos, incluindo novas questões de pesquisa e propostas, bem como novas teorias e modelos, os quais são tradicionalmente medidos pela contagem de publicações e citações. Outros resultados de sucesso podem ser menos visíveis do que as contagens de publicações e citações, mas são importantes. Estes incluem desenvolvimentos em vários segmentos: carreira, educacional, administrativo, negócios e desenvolvimento sociopolítico.

A divulgação dos resultados da investigação é um componente importante de toda pesquisa científica. O método tradicional de divulgação dos resultados de uma colaboração é através da coautoria de publicações. Sonnenwald (2007) aponta que, no caso de pesquisas interdisciplinares, pode ser um desafio encontrar um fórum adequado para publicar os resultados que não pertencem claramente a uma disciplina ou outra.

Chegar a um consenso em relação à inclusão e a ordem de autoria também pode ser um desafio, e estas dificuldades aumentam à medida que aumenta a concorrência. Perguntas como as que seguem emergem nesse estágio: Quem, entre os alunos, técnicos de laboratório e pesquisadores, deve ser incluído como um autor? O que constitui uma

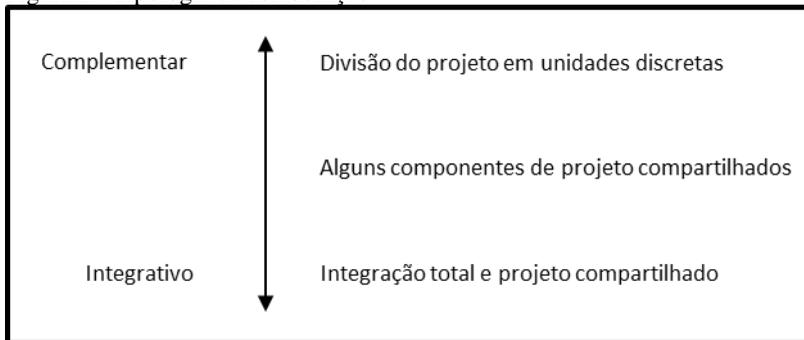
contribuição significativa merecendo coautoria em relação a inclusão na seção de agradecimentos de um jornal?

2.1.3 Formas de Colaboração

A colaboração pode assumir várias formas, desde a oferta de conselhos gerais a conhecimentos para a participação ativa em uma parte específica de pesquisa. Estas contribuições colaborativas também podem variar em nível desde o substancial para o quase insignificante (KATZ; MARTIN, 1997). Muitas vezes alguém que fornece um material para uma pesquisa é visto como colaborador, outras vezes a colaboração acontece quando colaboradores de diferentes instituições compartilham dados e ideias em conferências e listas de discussão.

Segundo Hara et al. (2003) a colaboração se manifesta através de diferentes formas de interação entre cientistas. Essas formas ou níveis de trabalho em equipe vão desde o nível complementar até o nível integrativo, como pode ser verificado na Figura 2.

Figura 2 - Tipologia da colaboração



Fonte: Adaptado de Sonnenwald (1999).

O nível complementar de colaboração exige consciência dos conhecimentos e habilidades de todos os participantes e de como esses elementos podem complementar a pesquisa. Ou seja, esse nível requer consciência e complementaridade, em vez de compatibilidade de personalidades, sendo assim mais fácil de estabelecer. Os cientistas estão à procura de um elemento que complemente e se encaixe com suas pesquisas. Eles podem trabalhar no mesmo projeto, mas não necessariamente trabalham em estreita colaboração. Cada um é responsável por seu próprio pedaço da pesquisa, contribuindo para o projeto através do fornecimento de seus resultados particulares. O

resultado total ou final é maior do que qualquer membro poderia realizar individualmente

O nível de colaboração integrativo exige que as pessoas trabalhem em conjunto durante todo o processo de pesquisa de forma a desenvolver ideias e desafiar os pressupostos de cada um, respeitando e confiando uns nos outros, tanto em nível pessoal como profissional.

2.1.4 Benefícios da Colaboração

Katz e Martin (1997) apontam os benefícios da colaboração. Para os autores o primeiro tipo de benefício está relacionado ao compartilhamento de conhecimentos, habilidades e técnicas inerentes a essa prática. O segundo tipo de benefício relaciona-se a transferência de conhecimentos e habilidades já que a colaboração é uma maneira de transferir conhecimento, principalmente, o conhecimento tácito. Para exemplificar, os autores argumentam que habilidades sociais e de gestão também são necessárias para se trabalhar em equipe e elas não podem ser ensinadas em sala de aula, ou seja, a melhor maneira de aprendê-las é engajando colaboradores em atividades colaborativas.

Em terceiro lugar, a colaboração pode trazer um choque de pontos de vista, uma fertilização de ideias, que por sua vez pode gerar novas ideias ou perspectivas em relação a pesquisadores que trabalham individualmente. Além disso, os benefícios de trabalhar com outras pessoas não se limitam às relações com os colaboradores imediatos. A colaboração também tem o efeito de 'ligar' o pesquisador em uma rede mais ampla de contatos na comunidade científica, possibilitando uma maior visibilidade.

2.1.5 Custos da Colaboração

Manter um contexto colaborativo pode implicar em alguns custos adicionais e eles podem ter várias formas: financeiro, tempo e gerencial (KATZ; MARTIN, 1997).

Embora a colaboração resulte em economia para as agências de fomento à pesquisa, ela pode implicar alguns custos, como aqueles relacionados a deslocamento de pessoas e equipamentos e estadia, no caso de colaboração em nível interinstitucional e internacional.

A variável tempo também é uma geradora de custos. Ela está presente já na preparação de uma proposta conjunta, ou seja, na definição dos problemas de pesquisa e no planejamento da abordagem. Diferentes partes da pesquisa podem ser realizadas em locais diferentes e isso contribui para os custos de tempo porque deve-se manter todos os colaboradores plenamente informados da progressão da pesquisa bem

como das decisões de quem deve fazer o quê. As diferenças de opinião são quase inevitáveis, logo é necessário tempo para resolver essas diferenças. Escrever resultados de pesquisa em conjunto, também pode levar mais tempo, na medida em que há divergências sobre os resultados e seu significado, ou sobre quem deve ser incluído entre os coautores e em que ordem eles devem ser listados.

Um terceiro aspecto a abordar são os custos em termos de aumento de administração, já que com mais pessoas e talvez instituições envolvidas, mais procedimentos gerenciais são demandados para conduzir a pesquisa. Além disso, quando duas ou mais instituições estão colaborando, muitas vezes há o problema de conciliar as diferentes culturas de gestão, sistemas financeiros, regras sobre direito de propriedade intelectual, etc. Todas essas variáveis precisam ser conciliadas, de forma a não perturbar a colaboração.

2.1.6 Medição da Colaboração

A publicação com vários autores, chamada de publicação de coautoria tem sido, ao longo das décadas, largamente utilizada como uma unidade de medida da atividade colaborativa (MOED et al., 1991; BRAUN et al., 1992; DE HAAN, 1997; GLÄNZEL, 2002; GLÄNZEL; SCHUBERT, 2004; MAIA; CAREGNATO, 2008). Smith (1958) foi um dos primeiros a verificar um aumento de artigos com vários autores e sugeriu que esses artigos fossem usados como uma medida aproximada para avaliar a colaboração entre os grupos de pesquisadores. Muitos outros pesquisadores o seguiram e advogaram na linha da coautoria como medida de colaboração (PRICE; BEAVER, 1966; BEAVER; ROSEN, 1978, 1979).

Katz e Martin (1997), entretanto, observaram que com o uso da coautoria apenas alguns dos aspectos mais tangíveis de trabalho colaborativo podem ser quantificados, enquanto outros certamente não podem. Mesmo uma avaliação qualitativa da colaboração é extremamente difícil por causa da relação indeterminada entre atividades quantificáveis e contribuições intangíveis. Laudel (2002) corroborou com esse argumento relatando que cerca de metade das colaborações são invisíveis porque elas não resultam em publicações em coautoria nem em reconhecimentos formais em textos científicos.

Apesar das observações em relação ao aspecto intangível das colaborações, estudos que utilizam indicadores tangíveis e com mais frequência a coautoria, são muitos mais frequentes. Nessa linha, De Haan (1997) recomendou outros tipos de indicadores em seu estudo. Ele

destacou seis indicadores para medir colaboração entre cientistas sociais: coautoria, editoração compartilhada de publicações, supervisão compartilhada em projetos de PhD, escrita conjunta de propostas de pesquisa, participação em programas de pesquisa formais e organização compartilhada de conferências científicas.

Apesar de existirem propostas de outros indicadores, parece haver um consenso de que coautoria é o indicador de colaboração mais visível e com isso ele tem sido mais utilizado nos estudos sobre o tema (GLÄNZEL; SCHUBERT, 2004; MILOJEVIĆ, 2010). Embora ele não seja perfeito, Subramanyam (1993) relata algumas vantagens. Primeiro, é invariante e verificável; dado o acesso ao mesmo conjunto de dados, outros investigadores devem ser capazes de reproduzir os resultados. Em segundo lugar, é um método relativamente barato e prático para a quantificação da colaboração. Além disso, o tamanho da amostra que é possível analisar usando esta técnica pode ser muito grande e os resultados devem, portanto, ser estatisticamente mais significativos do que os obtidos a partir de estudos de caso.

2.1.7 Métodos de Estudo da Colaboração

Os métodos mais comuns para estudar colaboração tem sido bibliométrico, análise de redes social ou análise de rede e métodos qualitativos de observação, entrevista e questionários.

Ao utilizar métodos bibliométricos, os estudos sobre colaboração (SUBRAMANYAM, 1983; GLÄNZEL, 2002; GLÄNZEL; SCHUBERT, 2004) normalmente usam medidas estatísticas como a média para descobrir o número médio de autores que contribuem em um artigo, indicando assim o nível de colaboração. Para visualizar mudanças nos padrões de coautoria, a distribuição da frequência de coautores sobre os artigos também pode ser empregada, além de outras correlações como a comparação entre o nível de colaboração encontrado em publicações em periódicos com o fator de impacto do periódico e a taxa média de citações pelo número de coautores.

Métodos de análise de rede estudam a estrutura das redes de colaboração com base nas relações de coautoria dos indivíduos (NEWMAN, 2001a, 2001b, 2001c, 2004a, 2004b; BARABÁSI et al., 2002; WAGNER; LEYDESDORFF, 2005; KATERND AHL, 2012; GONZÁLEZ-ALCAIDE, 2012; TAMANINI et al., 2013; LEE et al., 2013). Nesse tipo de método, os autores são os nodos ou vértices da rede e as relações de coautoria são as arestas da rede. A partir desse modelo em rede ou grafo várias métricas podem ser aplicadas e contextualizadas. Medidas que focam na posição dos indivíduos ou

colaboradores dentro da rede permitem inferir, por exemplo, que o autor com mais laços ou relações é o autor que mais colabora. Outras métricas ou algoritmos permitem detectar *clusters* ou comunidades que são grupos de nodos conectados mais densamente, além da detecção de componentes isolados na rede. O nível de conexão ou densidade da rede é outra medida comumente apresentada nos estudos que usam esse método.

2.2 ANÁLISE DE REDE SOCIAL

Muitos sistemas podem ser modelados e estudados como uma rede, que é basicamente um conjunto de nodos ou vértices ligados por *links* ou arestas. Exemplos incluem sistemas de redes sociais, como rede de colaboração (NEWMAN 2001a, 2001b, 2001c), redes tecnológicas como a Internet (FALOUTSOS, 1999), a World Wide Web (BRODER et al., 2000) e redes biológicas como redes neurais (WATTS; STROGATZ, 1998) e redes metabólicas (FELL; WAGNER, 2000).

Segundo definição da *International Network for Social Network Analysis (INSNA)*, a Análise de Redes Sociais (ARS) está focada em descobrir os padrões de interações das pessoas ou unidades de análise. A análise de redes é baseada na noção intuitiva que estes padrões são características importantes da vida dos indivíduos que os detêm. Analistas de redes acreditam que a maneira como um indivíduo vive depende em grande parte de como o indivíduo está amarrado dentro de uma larga rede de conexões sociais. Além disso, muitos acreditam que o sucesso ou fracasso de sociedades e organizações frequentemente depende do padrão de suas estruturas internas. Logo, eles procuram descobrir os vários tipos de padrões e tentam determinar as condições em que esses padrões surgem para então descobrir suas consequências.

A análise de redes tem importantes aplicações no estudo de comportamento organizacional (KRACKHARDT, 2010), relações interorganizacionais (WASSERMAN; FAUST, 1994), na difusão de doenças contagiosas como HIV (KELLY et al., 2006), em questões de segurança nacional, mapeando redes de terroristas (WILL; GNIADK; MEMON, 2010), na difusão de informação (XU; LIU, 2010), etc.

No contexto organizacional Krackhardt (2010) destaca que estudos tradicionais sobre carreira de sucesso focam em treinamento e educação, enquanto que a perspectiva de rede social enfatiza a conectividade do empregado com outros dentro da organização. Enquanto liderança é frequentemente pensada como um conjunto de habilidades e competências, a análise de rede social foca nas relações do líder com seus seguidores ou no papel de ponte que o líder fornece para outros grupos, o que poderia aumentar ou inibir a efetividade desse líder.

Ainda no contexto das organizações Wasserman e Faust (1994) dizem que a análise de redes sociais tem sido utilizada para examinar como as organizações interagem com outras, caracterizando as muitas conexões informais que ligam executivos, bem como associações e conexões individuais entre empregados em diferentes organizações. Os autores argumentam, por exemplo, que o poder nas organizações muitas

vezes está mais relacionado com o grau de um indivíduo que é central dentro de uma rede do que com o título de um cargo.

Essa seção está organizada em várias subseções que abordam a história do nascimento e desenvolvimento da área de análise de redes sociais, as características dos dados utilizados em ARS, tais como a tipologia e a forma de tratamento dos dados, conceitos fundamentais de teoria de grafos necessários para o entendimento dos demais capítulos, o conceito de centralidade de um ator na rede, métodos de detecção e análise de subgrupos dentro de um grafo e no final a aplicação da ARS no contexto das redes de colaboração científica.

2.2.1 História do Desenvolvimento da Área de ARS

A área de análise de redes sociais tem seus princípios nas ciências sociais e no interesse em estudar as relações sociais dos indivíduos. Segundo Freeman (2004), esse interesse não é recente, ele data do início do século passado, embora evidências de interesse nesse tipo de estudo são anteriores a essa data possam ser encontradas na literatura.

Nas ciências sociais, a abordagem estrutural, que se baseia no estudo da interação entre os atores sociais, é chamada de análise de redes sociais. As relações que os analistas de rede sociais geralmente estudam são aquelas que ligam seres humanos individuais. Entretanto, importantes relações sociais podem vincular os indivíduos sociais que não são humanos, como os animais e podem também vincular os atores que não são indivíduos, tais como grupos, organizações, estados, etc. A abordagem de rede social se baseia na intuitiva noção de que o padrão de relações sociais no qual os atores estão inseridos tem consequências importantes para esses atores.

Nos primórdios da pesquisa em redes sociais, alguns precursores da área foram motivados por uma intuição estrutural baseada nos laços que ligavam os atores sociais. Eles deixaram essa intuição mais explícita ao documentar suas pesquisas. Algumas pesquisas também se basearam na coleta de dados e na análise sistemática dos padrões sociais dos mesmos. Outras pesquisas basearam-se na construção de imagens gráficas que representavam os padrões de interação e desenvolveram procedimentos para construção de tais imagens. A quarta abordagem utilizada se baseou no uso de matemática e/ou modelos computacionais, logo alguns trabalharam na definição e escrita das propriedades matemáticas dos padrões sociais.

Freeman (2004) relata que antes do surgimento da análise de rede social moderna, os pesquisadores usavam uma ou mais combinações

dessas quatro abordagens para a condução de pesquisa estrutural em fenômenos sociais e que posteriormente, essas abordagens foram integradas em um paradigma organizado para pesquisa e juntas elas caracterizaram a análise de rede social moderna.

Segundo o mesmo autor a área de análise de redes sociais é definida não apenas pelas quatro abordagens discutidas acima, mas também por uma variedade extremamente ampla e sempre crescente de aplicações. Devido a essa generalidade, a análise de redes cruza as fronteiras das disciplinas tradicionais. Ela reúne sociólogos, antropólogos, matemáticos, economistas, cientistas políticos, psicólogos, cientistas de comunicação, estatísticos, epidemiologistas, cientistas da computação, especialistas em comportamento organizacional, físicos, etc. Essas pessoas vêm com diferentes conhecimentos, mas todos eles compartilham um compromisso comum com a perspectiva estrutural presente na abordagem de rede.

Alguns pesquisadores ao examinarem as origens da análise de redes sociais concordam que ela começou no início de 1930 com o trabalho de Jacob Moreno intitulado "Who Shall Survive?" (HUMMON; CARLEY, 1993; WASSERMAN; FAUST, 1994). Esse trabalho é considerado uma introdução a Sociometria e destacado como um momento decisivo na história da análise de redes sociais.

Entretanto outros escritores argumentam que a análise de rede social não começou até o início dos anos 70, quando Harrison White iniciou os trabalhos com essa área em Harvard (BERKOWITZ, 1982; SCOTT, 2000). Durante essa época, White, junto com seus orientandos, produziu um número surpreendente de importantes contribuições para a teoria das redes sociais (WHITE, 1970; WHITE; BOORMAN; BREIGER, 1976).

Em seu livro, Freeman (2004) destaca em detalhes importantes trabalhos que aconteceram antes de 1930, durante a década de 30 e desenvolvimentos que ocorreram entre 1940 e 1960. Cada um desses períodos será abordado a seguir:

a) Antes de 1930

Auguste Comte (século XIX) foi o primeiro estudioso a propor uma forma de olhar para a sociedade em termos de interconexões entre os atores sociais. Comte mostrou como as partes do sistema social estão interligadas: "As famílias se tornam tribos e tribos se tornar nações." Ele definiu a sociedade usando os tipos de termos estruturais que são encontrados hoje em análise de redes sociais. Percebe-se, portanto, que ele teve uma influência indireta sobre o desenvolvimento do campo. A

maioria dos outros proeminentes sociólogos do século XIX e início do XX abraçou a perspectiva estrutural de Comte.

Todas as contribuições seguintes vieram de psicólogos desenvolvimentistas e educacionais durante a década de 1920. Durante esses anos, grandes doações foram feitas para institutos de bem-estar infantil. O resultado deste financiamento foi um enorme aumento na quantidade de pesquisas voltadas para estudo das relações interpessoais das crianças. Estas obras (ALMACK, 1922; WELLMAN, 1926; BOTT, 1928) não são, na maior parte, reconhecidas no campo da pesquisa de rede social, mas elas conseguiram inovar em uma série de ideias e práticas de rede importantes.

Freeman (2000) relata que as imagens gráficas tiveram um lugar importante em estudos estruturais desde os primeiros tempos. Imagens que mostravam a proximidade de parentesco e construídas no formato de árvore já eram desenhadas no século IX.

2) 1930 - O nascimento da análise de redes sociais: Sociometria

No início da década de 1930 um amplo esforço de pesquisa chamado Sociometria foi iniciado. Foi o primeiro trabalho que incluiu todas as quatro características abordadas anteriormente e que definem a análise de rede social.

Jacob Levy Moreno foi o precursor no desenvolvimento da Sociometria. Naquela época, Moreno estava preocupado com as propriedades psicológicas dos indivíduos. Ele usava perguntas sociométricas para determinar os sentimentos das pessoas em relação às outras. Por exemplo, ele pedia aos membros de uma população definida, para nomear quem eles gostariam de viver, trabalhar ou passar o tempo de lazer. A abordagem de Moreno usava estas escolhas para descobrir grupos e posições dos indivíduos dentro desses grupos. Segundo Freeman (2004) neste ponto, a Sociometria obteve a atenção da elite da comunidade científica americana em ciência social.

Nessa época um outro esforço de pesquisa sobre estrutura social começou a ser desenvolvido em Harvard. Centrado na Escola de Administração de Negócios, envolveu um número relativamente grande de professores. Segundo Freeman (2004) e Scott (2000) os esforços de Harvard quase nunca são reconhecidos em revisões históricas de análise de redes sociais. Após o trabalho de Moreno e da Universidade de Harvard, o campo de análise de redes sociais pareceu passar por um período que pode ser chamado de "idade das trevas". Durante um período de 30 anos, de cerca de 1940 até cerca de 1970, nenhum centro de pesquisa nessa área surgiu.

3) O período de 1970 até o presente: O renascimento em Harvard

Harrison White foi o pesquisador responsável pelo renascimento da área em Harvard, aplicando a perspectiva estrutural em uma gama de problemas. Seu trabalho em cadeia de vagas em organizações resultou em um livro (WHITE, 1970).

Freeman (2004) considera que a maior contribuição de White foi como professor, doutrinando uma geração inteira de estudantes dentro da perspectiva estrutural. Juntos, eles tornaram Harvard um centro de pesquisa estrutural e muitos estudantes de doutorado publicaram na área de redes sociais. Essa situação fez com que ao final dos anos 70, a análise de redes sociais se tornasse amplamente reconhecida entre os cientistas sociais.

2.2.2 Dados Sociais: Tipologia e Forma de Tratamento

Segundo Scott (2000), apesar de haver um interesse crescente pela utilização dos métodos de análise de redes sociais, eles têm sido vistos como inatingíveis para muitos pesquisadores, que encontraram dificuldades para entender a linguagem técnica e matemática em que muitos destes métodos são discutidos. Para o autor somente se o pesquisador tem um claro entendimento da lógica de um método é que ele pode fazer um julgamento sociológico sobre sua relevância para uma determinada pesquisa.

A primeira tarefa do processo de entendimento deve ser a definição do tipo de dado para o qual a análise de rede social é a mais adequada (SCOTT, 2000). Os principais tipos de dados são "dados de atributos" e "dados relacionais". Dados de atributos referem-se às atitudes, opiniões e comportamentos e são considerados propriedades, qualidades ou características dos indivíduos ou grupos que podem ser quantificados e analisados através de procedimentos estatísticos. Os dados relacionais são os contatos, os laços, as ligações, as reuniões e conexões de grupo, que relacionam um ator a outro e por isso não podem ser reduzidos a propriedades dos atores. Os métodos apropriados para dados relacionais são os de análise de rede, em que as relações são tratadas para expressar as ligações que existem entre os agentes.

O uso dos métodos de análise de redes sociais depende da disponibilidade dos dados relacionais. Logo é importante que os mesmos sejam coletados, organizados e armazenados para futuras análises.

a) Coleta de dados: Existem problemas importantes na seleção dos dados relacionais. A identificação das fronteiras/limites da seleção é um deles. Por exemplo, qual a definição de amizade forte em um grupo? Não é mesma para todos os elementos do grupo. Essa situação deve ser contornada com a identificação dos limites da situação sob investigação, porém, não os limites naturais ou óbvios, mas os limites oriundos de um estudo teórico mais profundo da situação pesquisada. Ou seja, pesquisadores devem ser envolvidos em um processo de elaboração conceitual e construção de um modelo e não simplesmente na coleta de dados pré-formatados.

Uma vez que as fronteiras do estudo são definidas, a pesquisa deve definir a população alvo de estudo. Duas abordagens são identificadas: posicional e reputacional. Na abordagem posicional o pesquisador define posições ou grupos de interesse (sala de aula, grupo de trabalho, etc.) e posteriormente colhe amostras entre os membros dessas posições ou grupos. Um problema conhecido nessa abordagem é a determinação das posições a serem escolhidas. Na abordagem reputacional, os elementos são escolhidos pela sua reputação no contexto da pesquisa, são nomeados por outros agentes com conhecimento para executar essa tarefa. Uma variação da estratégia reputacional é feita através da técnica de *snowballing*, onde um número de informantes indica um grupo de pessoas e esse grupo sua vez indicada outro grupo e assim segue, culminando com a formação do grupo a ser estudado. Nesse tipo de técnica a relação social em si é usada como uma cadeia para construção do grupo de estudo.

b) Organização dos dados relacionais

Os dados coletados geralmente são organizados na forma de uma matriz. No caso dos dados relacionais e quando os atores participam em várias afiliações eles devem ser dispostos em uma matriz do tipo “caso por afiliação”, onde as linhas são os casos, agentes ou atores analisados e as colunas são afiliações caracterizadas por eventos, organizações ou atividades que os agentes estão envolvidos. Na matriz I da **Figura 3**, as linhas são as companhias e as colunas são os diretores que trabalham nas companhias. Nas células a presença ou ausência do agente em determinada afiliação ou relacionamento é marcado com 1 ou 0.

A partir dessa matriz sociogramas ou grafos podem ser desenhados, contudo, quando o conjunto de dados a ser representado não é pequeno, o desenho e visualização de sociogramas se torna uma

tarefa complicada. Nesse caso, é recomendada a decomposição dos dados da matriz original em duas matrizes.

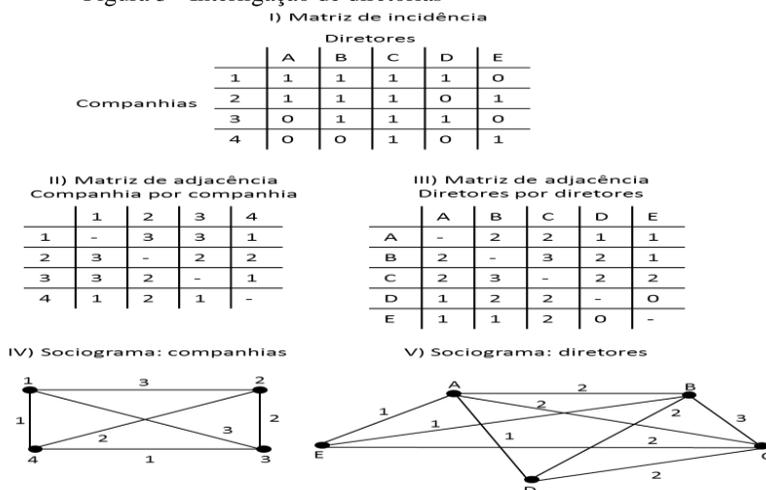
A primeira matriz, chamada de caso por caso, é uma matriz quadrada onde os casos ou agentes são listados duas vezes, primeiramente nas linhas e depois nas colunas e o conteúdo das células pode ser 1 ou 0 indicando a existência ou não de relacionamento entre os dois agentes em determinada afiliação. A matriz II da Figura 3 representa o relacionamento de companhia por companhia.

A segunda matriz, também quadrada, é chamada de afiliação por afiliação, tem as afiliações também listadas nas linhas e colunas e o conteúdo das células indica o número de agentes que as duas afiliações possuem em comum.

A matriz original (matriz I na Figura 3 é chamada de matriz de incidência e a matrizes resultantes são chamadas de matrizes de adjacência (matriz II e III na Figura 3). A maioria das técnicas de análise de redes envolvem a manipulação de matrizes de adjacência.

A Figura 3 ilustra os dados referentes a interligação de diretores em várias companhias. A organização dos dados relacionais é representada através de matrizes de incidência e adjacência. Nesse caso, o conteúdo da matriz de adjacência companhia x companhia contém o número de diretores em comum nas mesmas companhias. Os sociogramas das matrizes de adjacência também são exibidos.

Figura 3 - Interligação de diretorias



Fonte: Adaptado de Scott (2000, p.45)

c) Armazenamento dos dados relacionais

Muitas características fundamentais de redes podem ser analisadas através da manipulação direta de matrizes, em operações de transposição, adição e multiplicação. A álgebra matricial, no entanto, é bastante complexa para a maioria dos pesquisadores. Apesar de matrizes serem úteis para a organização e armazenamento de dados relacionais, softwares especializados permitem uma abordagem mais fácil e direta para análise de rede (SCOTT, 2004).

Softwares ou ferramentas de análise de rede são usados para representar graficamente os vértices (nodos) e arestas (relações) em uma rede e analisar os dados da rede. A maioria dos softwares de análise tem módulos para visualização da rede, onde a exploração dos dados é feita através de exibição de nós e relações em vários layouts, atribuindo cores, tamanho e outras propriedades avançadas para nós e relações.

A representação visual da rede é importante para compreender os dados da rede e transmitir o resultado da análise. A visualização teve e sempre terá um papel importante na área de análise de redes, pois desde o início imagens de redes têm sido usadas tanto para desenvolver insights estruturais e de comunicar esses insights para outros (FREEMAN, 2000).

Existem muitos softwares especializados em análise de redes sociais, tais como UCINET⁸ (BORGATTI; EVERETT; FREEMAN, 2002), Pajek⁹ (BATAGELJ; MRVAR, 2003) e Gephi¹⁰ (BASTIAN; HEYMANN; JACOMY, 2009), considerados os mais utilizados. Pajek e Gephi são softwares livre.

Em *International Network for Social Network Analysis (INSNA)* pode ser encontrada uma relação de softwares recomendados por membros dessa associação. Huisman e Van Duijn (2004) fizeram uma análise detalhada das diversas ferramentas disponíveis naquele período. Passmore (2011) também produziu uma extensa lista de ferramentas utilizadas para análise de redes sociais.

2.2.3 Unidades e Níveis de Análise

Na análise de rede dois conceitos devem ser inicialmente esclarecidos: unidade de análise e nível de análise. A unidade de análise refere-se à estrutura que está sendo analisada, podendo ser pessoas,

⁸ <https://sites.google.com/site/ucinetsoftware/home>

⁹ <http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm>

¹⁰ <https://gephi.github.io/>

organizações, países, etc. O nível de análise refere-se a agregações de características relacionais de interesse que podem ser feitas (WASSERMAN; FAUST, 1994).

Existem basicamente quatro níveis de análise de uma rede social e para cada um dos níveis existem métodos específicos de análise:

a) Nível de análise que considera a rede como um todo: Para analisar a rede como um todo, medidas de conectividade, tais como, Densidade, Distância, Diâmetro e Componente são normalmente utilizadas;

b) Nível de análise que considera as propriedades de um conjunto de atores na rede: Existem várias medidas que permitem detectar subgrupos de atores que são homogêneos em relação a algumas características da rede, tais como Clique, N-clique, K-plex, etc.

c) Nível de análise que considera as propriedades dos pares de atores e as relações entre eles: Para analisar pares de atores e as relações entre eles, podem ser utilizadas medidas de distância e alcançabilidade de atores;

d) Nível de análise que considera o ator individualmente: Para analisar o ator individualmente dentro de um grupo, medidas como Centralidade e Prestígio podem ser aplicadas.

Algumas dessas medidas serão apresentadas nas seções seguintes.

2.2.4 Teoria de Grafos e Matrizes

No contexto de análise de rede sociais a compreensão dos conceitos básicos da teoria de grafos contribui para melhorar a sofisticação de análise de um pesquisador, assegurando que ele escolha os métodos adequados (SCOTT, 2000).

Ainda segundo Scott (2000) a teoria de grafos é uma abordagem matemática que fornece uma linguagem formal para descrever redes e suas características. Ela oferece uma tradução de dados matriciais em conceitos formais e teoremas que podem ser diretamente relacionados com as características essenciais das redes sociais.

A teoria de grafo fornece a representação de uma rede social como um modelo de um sistema social composto de atores e relações entre esses atores. Um grafo é um modelo de uma rede social onde pontos, também chamados de nodos, são usados para representar os atores, e linhas conectando os pontos são usadas para representar os laços entre os atores (WASSERMAN; FAUST, 1994).

A teoria de grafo tem sido útil em análise de redes sociais por muitas razões, dentre as quais se pode destacar: primeiro, a teoria de grafo fornece um vocabulário que pode ser usado para nomear e denotar

precisamente muitas propriedades sociais estruturais; segundo, a teoria de grafo fornece operações matemáticas e ideias através das quais muitas dessas propriedades podem ser quantificadas e medidas; e terceiro, dado esse vocabulário e matemática, a teoria de grafo fornece a habilidade para provar teoremas sobre grafos, logo, sobre representações de estruturas sociais (HARARY; NORMAN; CARTWRIGHT, 1965).

Para Scott (2000) a representação visual de grafos é de importância secundária na teoria dos grafos já que muitas vezes é difícil traçar um diagrama claro e compreensível para grandes conjuntos de pontos com padrões complexos de conexão. Ao expressar as propriedades do grafo de uma forma matemática mais abstrata, é possível dispensar a necessidade de desenhar um sociograma.

Matrizes são uma forma alternativa de representar dados de rede. Uma matriz contém as mesmas informações que um grafo e é mais útil para sumarizar e analisar computacionalmente os dados. As operações matriciais são amplamente utilizadas em análise de redes sociais e são o principal tipo de representação para a maioria dos softwares de análise de redes (WASSERMAN; FAUST, 1994).

A seguir serão apresentados vários conceitos e medidas relacionados a teoria de grafos: tipos de grafo, adjacência, grau, caminhada, caminho, distância, diâmetro e densidade, extraídos de (WASSERMAN; FAUST, 1994; SCOTT, 2000).

a) Tipos de grafos

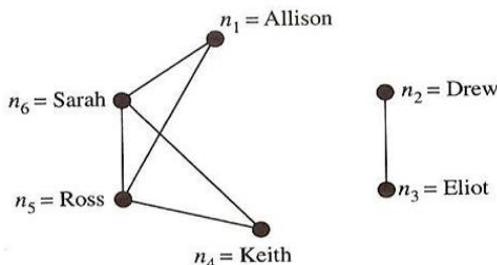
O tipo de um grafo é determinado pelas relações que ele representa. Um grafo pode representar relações não-direcionais, direcionais e elas podem ser sinalizadas ou valoradas.

Uma relação não-direcional é aquela onde existe uma reciprocidade na relação, por exemplo, a coparticipação em organizações ou grupos informais, algumas relações de parentesco como “é casado com”, “é parente de”, etc. A Figura 4 ilustra o relacionamento de 6 crianças através da relação não-direcional “mora perto”.

Figura 4 - Relacionamento não-direcional “mora perto”.

	Ator	Mora perto:
n1	Allison	Ross, Sarah
n2	Drew	Eliot
n3	Eliot	Drew
n4	Keith	Ross, Sarah
n5	Ross	Allison, Keith, Sarah
n6	Sarah	Allison, Keith, Ross

- 11 = (n1, n5)
- 12 = (n1, n6)
- 13 = (n2, n3)
- 14 = (n4, n5)
- 15 = (n4, n6)
- 16 = (n5, n6)



Fonte: Adaptado de Wasserman e Faust (1994, p. 96)

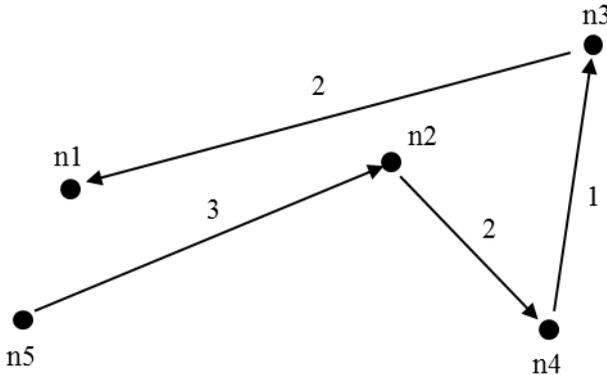
Uma relação é direcional se os laços são direcionados de um ator para outro ator. A exportação/importação de bens de uns país para outro é um exemplo de relação direcional. Amizade é outro exemplo, pois a escolha de uma pessoa como amiga não implica na reciprocidade por parte da pessoa que foi escolhida. Nesse tipo de grafo, chamado de dígrafo, a direção das relações é especificada através de uma seta.

Uma relação é sinalizada quando pode ser interpretada como positiva ou negativa no quesito afeição, avaliação ou significado. Relações do tipo “amor” e “ódio” entre pessoas de um grupo podem ser representadas como um grafo sinalizado, seja ele direcional ou não-direcional. Nesse tipo de grafo a relação recebe um sinal positivo (+) ou negativo (-).

Uma relação é valorada quando a força ou intensidade da relação necessita ser representada. Como exemplo de relações valoradas pode-se citar: a frequência de interações entre as pessoas de um grupo, a quantidade de dinheiro trocado numa transação financeira entre países, a taxa ou nível de amizade entre as pessoas de um grupo, etc. Relações valoradas podem ser representadas tanto de forma não-direcional como

direcional. A Figura 5 mostra o exemplo de um grafo direcionado valorado.

Figura 5 - Relacionamento direcional e valorado.



$$\begin{array}{ll}
 l_1 = (n_3, n_1) & v_1 = 2 \\
 l_2 = (n_2, n_4) & v_2 = 2 \\
 l_3 = (n_4, n_3) & v_3 = 1 \\
 l_4 = (n_5, n_2) & v_4 = 3
 \end{array}$$

Fonte: Adaptado de Wasserman e Faust (1994, p. 142)

b) Adjacência

A adjacência é a expressão utilizada quando dois atores representados por pontos estão diretamente relacionados ou conectados um ou outro. Em outras palavras, dois pontos que são ligados diretamente por uma linha são considerados adjacentes. Na Figura 4 os pontos adjacentes a n1 são n5 e n6.

c) Grau

Os pontos para o qual um determinado ponto é adjacente são chamados de sua vizinhança. O número total de pontos na vizinhança de um ponto é denominado grau. Assim, o "grau" de um ponto é uma medida numérica da dimensão da vizinhança de um ponto.

Em um grafo não-direcionado o grau é denotado por $d(n_i)$, onde n_i é o ponto ou nodo para o qual está se calculando o grau. O grau de um nodo pode variar de 0, no caso do nodo ser isolado, até $g - 1$, onde g é o

número total de nodos. Na Figura 4, por exemplo, o grau de n_1 é denotado por $d(n_1) = 2$.

Em grafos direcionados cada linha deve ser considerada juntamente com a sua direção. O fato de, por exemplo, A escolher B como amigo não quer dizer que haverá uma correspondência de B para A. Por esta razão, o "grau" de um nodo de um grafo direcionado compreende dois distintos elementos, chamados de "*indegree*" e "*outdegree*". O *indegree* de um nodo é o número total de nodos que têm linhas voltadas para ele. O *outdegree* de um nodo é o número total de nodos para os quais ele direciona linhas. Na Figura 5, por exemplo, o *indegree* de n_1 é denotado por $d_i(n_1) = 1$ e o *outdegree* é denotado por $d_o(n_1) = 0$.

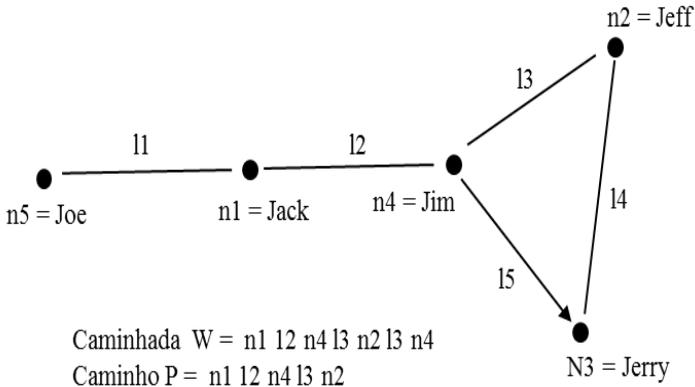
A medida fornecida pelo grau pode ser muito informativa para muitas aplicações. O grau de um nodo é a medida da "atividade" do ator que ele representa.

c) Caminhada e Caminho (*Walk e Path*)

Os pontos de um grafo podem ser diretamente ligados por uma linha, ou podem ser indiretamente ligados através de uma sequência de linhas. Uma sequência de linhas em um grafo é uma "caminhada" (*walk*) e é denotada por W . Uma caminhada no qual cada ponto e cada linha são diferentes é chamada de "caminho" (*path*). O comprimento tanto de uma caminhada como de um caminho é medido pelo número de linhas que o compõem.

Em estudos de redes sociais é importante saber se é possível acessar algum nodo a partir de outro nodo e qual a melhor maneira com que isso pode ser feito. Por exemplo, em uma organização pode ser interessante saber como uma determinada informação sai de um funcionário e alcança os outros funcionários e quantas linhas são necessárias para tal. Além disso, poderia se considerar todas as possíveis rotas para a informação seguir de um funcionário para outro. A Figura 6 ilustra o conceito de caminho e caminhada.

Figura 6 - Caminho e Caminhada



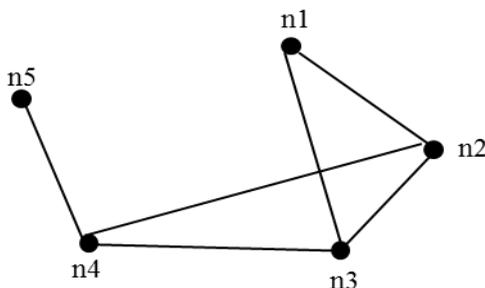
Fonte: Adaptado de Wasserman e Faust (1994)

d) Geodésico, Distância e Diâmetro

Podem existir diversos caminhos entre um par de nós. O caminho mais curto entre dois nós é chamado de 'geodésico'. A distância geodésica ou simplesmente distância entre dois nós é o comprimento do geodésico entre eles. A distância é denotada por $d(i,j)$. A distância indica o número de intermediários que um recurso, tais como a informação, deve passar até chegar ao seu destino final.

O diâmetro de um grafo é a maior distância geodésica entre todos os pares de nós presentes no grafo. É denotado por $\max d(i,j)$. A Figura 7 exibe todas as distâncias entre os pares de nós e o diâmetro do grafo.

Figura 7 - Distância e Diâmetro



$$\begin{aligned}
 d(1,2) &= 1 \\
 d(1,3) &= 1 \\
 d(1,4) &= 2 \\
 d(1,5) &= 3 \\
 \max d(1,5) &= 3
 \end{aligned}$$

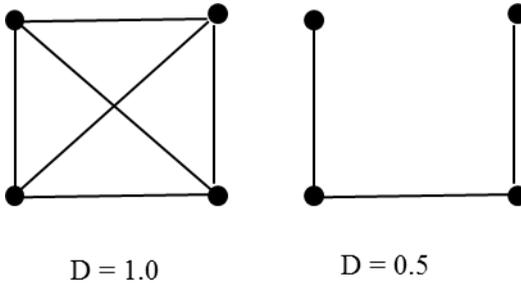
Fonte: Adaptado de Wasserman e Faust (1994).

e) Densidade

Um grafo "completo" é aquele em que todos os pontos são adjacentes um ao outro, ou seja, cada ponto é ligado diretamente a todos os outros pontos. O conceito de densidade mede o quão longe deste estado de completude o grafo está. Quanto mais pontos estão ligados uns aos outros, mais denso será o grafo. Segundo Scott (2000) um dos conceitos mais utilizados em teoria dos grafos é o de densidade, que descreve o nível geral de ligação entre os pontos de um grafo.

A densidade é denotada por $l/n(n-1)/2$, onde l é o número de linhas existentes no grafo e $n(n-1)/2$ é o número máximo de linhas que podem estar presentes no grafo, sendo n o número de pontos no grafo. O resultado dessa medida pode variar entre 0 e 1. A Figura 8 exhibe o valor da densidade para dois grafos distintos.

Figura 8 - Comparação de densidades



Fonte: Adaptado de Scott (2000).

2.2.5 Centralidade

A centralidade de indivíduos ou organizações em suas redes sociais foi uma das primeiras ideias apresentadas por analistas de redes sociais. As origens imediatas dessa ideia encontram-se no conceito sociométrico de "estrela", ou a pessoa que é mais popular em seu grupo ou que está no centro das atenções.

Medidas de centralidade tentam descrever as propriedades de localização de um ator na rede. Estas medidas levam em consideração as diferentes maneiras com que um ator interage e se comunica com o restante da rede. Os atores mais importantes ou mais proeminentes estão normalmente localizados em posições estratégicas dentro da rede. (WASSERMAN; FAUST, 1994).

Um ator é proeminente se seus laços o fazem visível para os outros atores na rede. Knoke e Burt (1983) definem dois tipos de proeminência: centralidade e prestígio.

A centralidade não se preocupa se a proeminência de um ator existe em função dos laços que recebe ou dos laços que transmite, o que importa é que o ator está envolvido em laços ou relações. Apesar dessa característica tornar essa medida mais apropriada para relações não-direcionais, ela também pode ser calculada para relações direcionadas. Contudo, nesse caso, deve-se levar em conta também os laços transmitidos pelo ator. O prestígio, ao contrário, leva em conta os laços recebidos pelo ator, o que torna essa medida aplicada somente para relações direcionadas.

As propriedades formais de centralidade foram inicialmente investigadas por Bavelas (1950) e a partir de seu trabalho pioneiro, uma

série de conceitos concorrentes de centralidade foram propostos. Como resultado desta quantidade de medidas formais de centralidade, existe uma confusão considerável na área (SCOTT, 2000).

Quando se aborda o conceito de centralidade é importante destacar que existe uma distinção entre centralidade de ponto "local" e "global". Um ponto é localmente central, se tem um grande número de ligações com os outros pontos no seu ambiente imediato, por exemplo, se tem uma grande "vizinhança" de contatos diretos. Um ponto é globalmente central, quando possui uma posição de importância estratégica na estrutura global da rede. A centralidade local está preocupada com a importância de um ponto focal na sua vizinhança, enquanto centralidade global diz respeito à proeminência do ponto dentro de toda a rede.

a) Centralidade local (Centralidade de grau)

A maneira mais simples e direta de medir a centralidade local de ponto, também conhecida como centralidade de grau (*degree centrality*) é pelos graus dos pontos no grafo. Um ponto é localmente central se ele tem um alto grau.

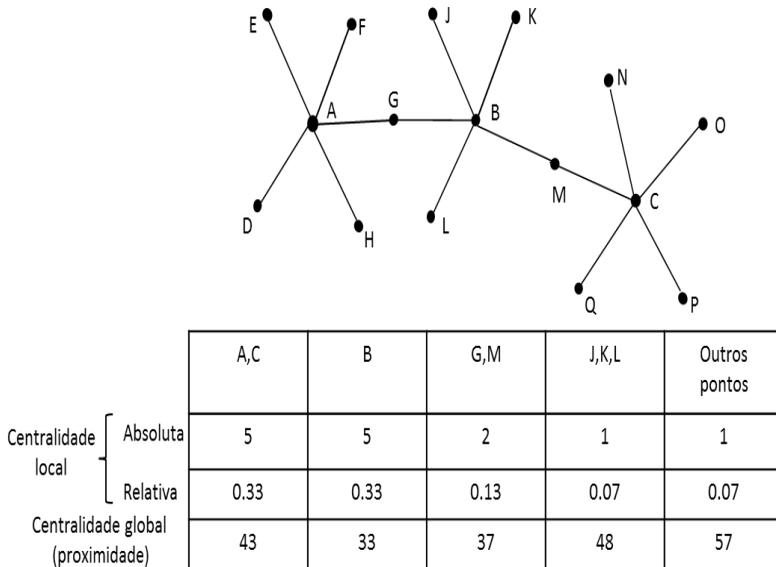
Ser um ponto central significa estar bem conectado ou "no meio das coisas". No processo de comunicação que acontece em uma rede social, uma pessoa que está em uma posição que permite o contato direto com muitos outros deve começar a ver-se e ser visto pelos outros como um grande canal de informação.

Segundo Freeman (1979), em certo sentido pode-se dizer que a pessoa é um ponto focal de comunicação, pelo menos no que diz respeito aos outros com quem está em contato e com isso é provável que venha a desenvolver uma sensação de estar no *mainstream* do fluxo de informações na rede, podendo influenciar o grupo retendo ou transmitindo informações.

Para Krackhardt (2010) o grau de centralidade local reflete a posição e o papel do nodo em termos de popularidade e atividade do nodo. Além disso, nodos com alto grau de centralidade podem ser identificados como líderes informais do grupo.

A medida de centralidade local não quer dizer que exista um único ponto central na rede. Na Figura 9, por exemplo, os pontos A, B e C podem ser vistos como centros locais. O grau, portanto, é uma medida de centralidade local, e uma comparação dos graus dos vários pontos em um grafo pode mostrar o quão bem os pontos são conectados com seus ambientes locais.

Figura 9 - Centralidade local e global



Fonte: Adaptado de Scott (2000).

O grau de centralidade local de um ponto (*degree centrality*) é dado por $C_D(\mathbf{n}_i) = \mathbf{d}(\mathbf{n}_i)$. Para um grafo direcionado a centralidade de um ponto é dada pelo seu *outdegree*.

A centralidade local relativa de cada ponto também está exibida na Figura 9. O cálculo da centralidade local relativa foi proposto por Freeman (1979) para normalizar o valor da centralidade local em relação ao número de pontos do grafo.

A medida de centralidade local tem uma limitação importante. Comparações de centralidade podem apenas ser feitas entre os membros de um mesmo grafo, ou entre grafos do mesmo tamanho.

b) Centralidade global (centralidade de proximidade e centralidade de intermediação)

Freeman (1979) propôs uma medida de centralidade global baseada no que ele denomina de "proximidade" dos pontos. A medida de centralidade global de Freeman é expressa em termos das distâncias entre os vários pontos. Um ponto é globalmente central se ele está a curta distância de muitos outros pontos, ou seja, o ponto é próximo a

muitos dos outros pontos no gráfico. Essa medida é conhecida como centralidade de proximidade (*closeness centrality*).

Para calcular a centralidade de proximidade se soma a distância geodésica do nó em relação a todos os demais nós do grafo e depois se inverte, uma vez que quanto maior a distância menor a proximidade, chegando-se à seguinte equação:

$$C_C(n_i) = \left[\sum_{j=1}^{\xi} d(n_i, n_j) \right]^{-1} \quad (1)$$

Como pode ser visualizado na Figura 9, B é o nodo mais globalmente central, seguido por G e M.

Para Krackhardt (2010) um nodo com uma posição mais próxima a todos os outros nodos pode alcançar os nodos mais rapidamente, obtendo e disseminando informação rapidamente através da rede. Assim, quando a transferência de informação for crítica para a rede, ter uma boa centralidade de proximidade é um importante atributo para o nodo que o detém.

Freeman (1979) também apresenta um conceito adicional de centralidade denominado de intermediação (*betweenness*). Este conceito mede o quanto um determinado ponto se encontra "entre" os vários outros pontos no grafo. A abordagem de intermediação de Freeman é construída em torno do conceito de "dependência local". Um ponto é dependente de outro se os caminhos que o ligam a outros pontos passam por este ponto. Um ponto com um grau alto pode ter um papel intermediário importante e ser muito central para a rede (SCOTT, 2000).

A centralidade de intermediação (*betweenness centrality*) é a mais complexa das medidas de centralidade de ponto. A intermediação de um ponto Y para um determinado par de pontos X e Z é definida como a proporção de geodésicos que ligam X e Z e passam por Y. A intermediação mede em que grau Y está "entre" X e Z. A centralidade de intermediação de um nó é denotada pela seguinte fórmula:

$$C_B(n_i) = \sum_{j < k} \frac{g_{jk}(n_i)}{g_{jk}} \quad (2)$$

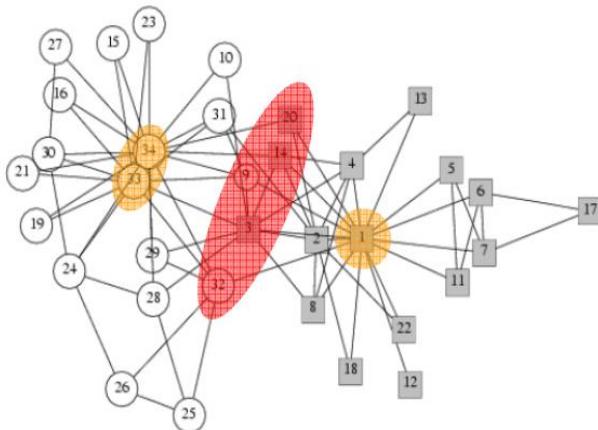
Seja g_{jk} a distância geodésica que liga os nós j e k e $g_{jk}(n_i)$ a distância geodésica que passa pelo nó n_i , o somatório dessas proporções

determina o grau em que um nó está no caminho de todos os demais nós do grafo.

No trabalho de Freeman pode ser encontrada a base para toda família de medidas de centralidade de ponto: a centralidade local (centralidade de grau) e centralidade global (centralidade de intermediação e centralidade de proximidade).

Na Figura 10 a rede representa um clube de karatê, resultado de uma pesquisa de Zachary (1974), onde os membros tiveram conflitos internos e se dividiram em dois grupos. Essa rede apresenta todos os conceitos de centralidade. Os nodos 1, 33 e 34 tem as maiores centralidades de grau e centralidade de proximidade. Os nodos 3, 9, 14, 20 e 32 tem as maiores centralidades de intermediação.

Figura 10 - Atores importantes da rede do clube de karatê de Zachary.



Fonte: Adaptado Zachary (1974)

2.2.6 Subgrupos

Uma das maiores preocupações de analistas de redes sociais é a identificação de subgrupos de atores dentro de uma rede. Subgrupos coesos são subconjuntos de atores entre os quais existem laços fortes, diretos, intensos, frequentes ou positivos.

A detecção de subgrupos dentro de uma rede pode ser importante para a compreensão de como a rede como um todo se comporta (HANNEMAN; RIDLEY, 2005).

Suponha que os atores de uma rede formem dois grupos que não se sobrepõem, e suponha-se que os atores em outra rede também

formem dois grupos, porém os membros se sobrepõem (algumas pessoas são membros de ambos os grupos). Onde os grupos se sobrepõem, pode-se esperar que o conflito entre eles seja menos provável do que quando os grupos não se sobrepõem. Onde os grupos se sobrepõem, a mobilização e a difusão podem espalhar-se rapidamente por toda a rede, onde os grupos não se sobrepõem, características podem ocorrer em um grupo e não se difundirem para o outro.

Saber como o ator está inserido na estrutura de grupos dentro de uma rede também pode ser crítico para compreender o seu comportamento. Por exemplo, alguns atores podem agir como "pontes" entre os grupos. Outros podem ter todas as suas relações dentro de um único grupo. Alguns atores podem estar fortemente ligados dentro do grupo enquanto que os outros podem possuir poucas ligações. As diversas maneiras que os atores estão embutidos na estrutura dos grupos dentro de uma rede impactam as formas que estes atores veem a sua "sociedade" e os comportamentos que eles podem praticar.

Hanneman e Ridley (2005) levantam algumas questões que podem ser respondidas a partir do estudo das subestruturas:

a) Quão separados são os subgrupos? Eles se sobrepõem e compartilham membros ou eles são divididos e particionam a rede?

b) Quão grandes são os subgrupos? Existem alguns grupos grandes ou um grande número de pequenos grupos?

c) Existem atores específicos que parecem desempenhar funções de rede? Por exemplo, atuam como nodos que ligam o grafo ou estão isolados dos grupos?

Segundo Wasserman e Faust (1994) as teorias de coesão social (Collins, 1988) forneceram a motivação para a definição de métodos de detecção e análise de subgrupos. Alguns pesquisadores de rede que desenvolveram métodos para detecção e análise de subgrupos são: Burt (1980) e Freeman (1984). Para esses autores o conceito de grupo social pode ser estudado olhando-se as propriedades dos subconjuntos de atores dentro da rede.

Hanneman e Ridley (2005) destacam que a detecção e análise de subgrupos pode acontecer a partir de uma perspectiva *top-down*, onde toda a rede é a referência inicial.

Para os autores, abordagens deste tipo tendem a olhar para a estrutura como um "todo" e identificar "subestruturas" como peças que são localmente mais densas. Em certo sentido, esta lente mais macro está à procura de "buracos", "vulnerabilidades" ou "pontos fracos" na estrutura global da rede. Estes buracos ou pontos fracos definem linhas de divisão no grupo maior e apontam como ele pode ser decomposto em

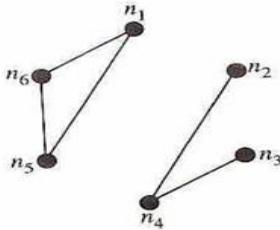
unidades menores. As análises representantes desse tipo de abordagem são: Componente, Blocos e *Cutpoints*. Neste trabalho somente a noção de Componente é apresentada.

Um componente é formalmente definido como um subgrafo conectado ao máximo. Um subgrafo é dito conectado quando todos os seus pontos se conectam através de um ou mais caminhos e não possuem outras conexões fora do subgrafo. Quando um subgrafo é máximo, é impossível adicionar novos membros sem destruir a qualidade de conectividade.

Se um grafo contém um ou mais pontos "isolados", esses pontos são chamados de componentes. Componentes de um grafo são subgrafos que estão conectados dentro do grafo, mas desconectados entre os subgrafos. A Figura 11 ilustra dois componentes em um grafo.

Segundo Scott (2000) a noção de um componente pode ser muito forte para encontrar todos os pontos fracos, buracos e subpartes localmente densas de um grafo. Existem algumas abordagens mais flexíveis, que não serão abordadas nessa revisão.

Figura 11 - Componentes de uma rede



Fonte: Adaptado de Wasserman e Faust (1994, p. 109)

2.2.7 ARS em Redes de Colaboração Científica

Newman (2001a, 2001b, 2001c, 2004a, 2004b) produziu uma série de artigos seminais de análise de rede de colaboração científica utilizando artigos em coautoria de bases de dados científicas, tais como MEDLINE, SPIRES e NCSTRL, as quais cobrem várias áreas de conhecimento. Entre vários resultados interessantes, os trabalhos relatam que as redes de coautoria formam "*small worlds*", no qual pares de pesquisadores estão separados por apenas um curto caminho de intermediários conhecidos e que a distância média que separa dois autores é 6, ou seja, existem 6 graus de separação na ciência. Além disso, ele demonstrou que as redes são altamente agrupadas, indicando

que dois autores tendem muito mais a colaborar se existe um terceiro autor que é comum a ambos.

Foram encontrados diversos trabalhos na literatura (Apêndice A) que analisam a colaboração científica através do método de análise de redes sociais. Nesses trabalhos, diversas métricas de ARS são aplicadas e contextualizadas em redes baseadas em coautoria.

Os trabalhos analisam a rede de colaboração de contextos diversos. Entre os trabalhos que analisam a rede de colaboração de um departamento universitário ou programa de pós-graduação, citam-se o de Bordin, Gonçalves e Todesco (2014), Bordin, et al (2015), Costa e Vanz (2012), Tamanini et al. (2013), Katerndahl (2012), Lima (2011), Rossoni e Filho (2009) e Silva et al. (2006).

Alguns trabalhos que analisam a rede de colaboração de áreas de conhecimento/pesquisa específicas, como os de Bordin e Gonçalves (2013), Yu et al. (2013), Lee et al. (2012), Wang et al. (2012) e González-Alcaide et al. (2012).

Outros trabalhos analisam a rede de colaboração de uma conferência ou revista (periódico) como os de Hou, Retschmer e Liu (2008), Fischbach, Putzke e Schoder (2011), Lil et al. (2012), Maia et al. (2012), Ye, Song e Li (2012) e Martins et al. (2012).

Os trabalhos de análise de rede de colaboração científica utilizam métricas de ARS variadas. Em nível de análise de toda a rede, densidade, distância média, diâmetro e grau médio são métricas muito utilizadas.

A densidade é utilizada pois ela indica o nível de conectividade ou colaboração da rede. Redes com alto grau de densidade indicam alta atividade de colaboração (KATERNDAHL, 2012). Ainda segundo Fischbach, Putzke e Schoder (2011), a densidade reflete o percentual do total da rede com o qual um ator foi coautor de um artigo.

A distância média entre nodos da rede é outra medida utilizada e indica o número médio de intermediários ou caminho de possibilidade de colaboração direta entre quaisquer dois autores.

O diâmetro, no contexto de uma rede de coautoria, indica o caminho de colaboração mais longo. O grau médio de uma rede também é um indicador do nível de colaboração da rede na medida em que determina o número médio de relacionamentos por autor (MARTINS et al., 2012).

Em nível de análise individual de cada nodo, métricas de centralidade, que levam em consideração as diferentes maneiras com que um ator interage e se comunica com o restante da rede, são comumente utilizadas.

Segundo Wasserman e Faust (1994), os atores mais importantes, ou mais proeminentes, estão normalmente localizados em posições estratégicas dentro da rede. Nodos com alto grau de centralidade podem ser identificados como líderes informais do grupo (KRACKHARDT, 2010). As métricas de centralidade conhecidas são: centralidade de grau, centralidade de intermediação e centralidade de proximidade.

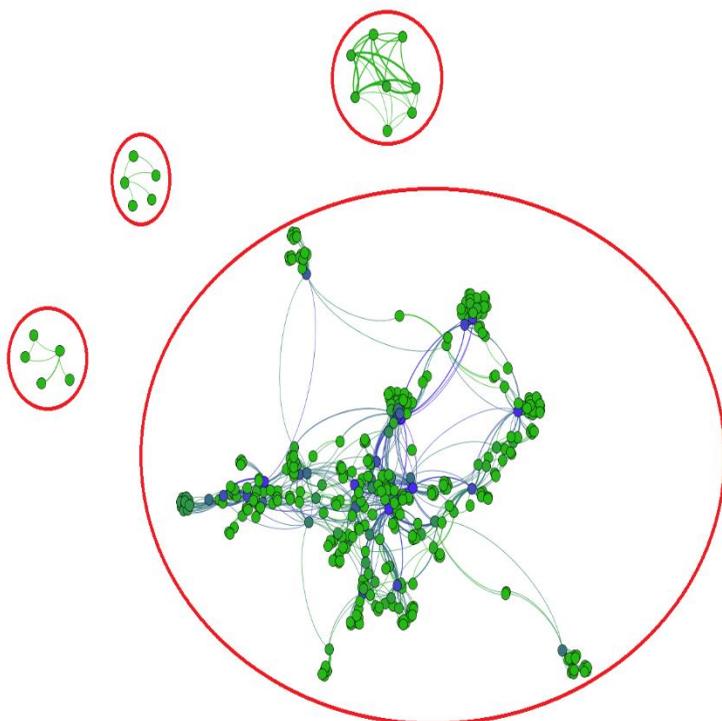
A centralidade de grau de um nodo em uma rede de coautoria indica o número de autores da rede que publicaram em conjunto com este nodo (autor), indicando, portanto, o nível de colaboração desse ator. Ela está associada com o nível de colaboração do ator, ou seja, quanto maior o grau maior será o nível de colaboração.

A centralidade de proximidade mede a proximidade de um determinado ator em relação aos demais atores da rede. Numa rede de coautoria, um autor com uma centralidade de proximidade alta pode indicar uma maior possibilidade de estabelecer parcerias de publicação na rede por estar mais próximo em relação a todos os outros autores (SOUZA; BARBASTEFANO; LIMA, 2012). Autores que tem um alto grau de centralidade de proximidade podem alcançar todos os outros autores através de uma cadeia menor de coautores em relação a autores com um baixo grau de centralidade (FISCHBACH; PUTZKE; SCHODER, 2011).

A centralidade de intermediação mede o quanto um determinado nodo se encontra "entre" os vários outros nodos da rede, atribuindo importância a um ator em função do fluxo que passa por ele para interligar outros dois atores da rede, através do menor caminho possível. Numa rede de coautoria, um autor com alto valor de centralidade de intermediação indica que um número significativo das parcerias estabelecidas na rede envolve, de forma direta ou indireta, as publicações relacionadas a esse ator (SOUZA; BARBASTEFANO; LIMA, 2012). Nesse contexto, um nodo ou autor que tem um alto grau de centralidade de intermediação pode ser considerado uma "ponte" entre distintos grupos de pesquisa, uma vez que está no caminho mais curto entre outros nodos (FISCHBACH; PUTZKE; SCHODER, 2011). Para Martins et al. (2012) a centralidade de intermediação indica o poder que os autores têm de intermediar ou conectar grupos de autores.

Em nível de grupo o número de **componentes** indica a presença de grupos que pesquisam isoladamente. Na Figura 12, que representa a rede de coautoria do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento - PPGEGC do ano de 2012, são encontrados quatro (4) componentes distintos.

Figura 12 - Rede de coautoria do PPGEGC no ano de 2012



Fonte: Bordin, Gonçalves e Todesco (2014).

2.3 SISTEMAS BASEADOS EM CONHECIMENTO

A área de Inteligência Artificial Simbólica possui um ramo conhecido como Representação e Raciocínio de Conhecimento, que objetiva a concepção de sistemas computacionais capazes de raciocinar sobre uma representação de máquina interpretável do mundo, de forma semelhante ao raciocínio humano.

Sistemas Baseados em Conhecimento (SBC) ou em inglês *Knowledge-Based Systems (KBS)*, têm um modelo computacional de algum domínio de interesse em que os símbolos servem como substitutos dos artefatos de domínio do mundo real, como objetos físicos, eventos, relacionamentos, etc. O domínio de interesse pode cobrir qualquer parte do mundo real ou qualquer sistema hipotético sobre o qual se deseja representar o conhecimento para fins computacionais (SOWA, 2000).

Um sistema baseado em conhecimento mantém uma base de conhecimento que armazena os símbolos do modelo computacional na forma de declarações sobre o domínio e executa raciocínio através da manipulação desses símbolos (GRIMM, HITZLER, ABECKER, 2007).

Chandrasekaran, Johnson e Smith (1992) oferecem outra definição adicional de SBC onde o processo de inferência é destacado. Para os autores um sistema baseado em conhecimento tem representações explícitas de conhecimento, bem como dos processos de inferência que operam nessas representações para atingir os objetivos. Um processo de inferência consiste em uma série de passos de inferência, onde cada passo cria um conhecimento adicional. O processo de aplicação dos passos de inferência é repetido até que as informações necessárias para preencher os requisitos do objetivo de solução de problema ou tarefa sejam geradas. Tipicamente, tanto o conhecimento de domínio quanto o processo de inferência devem ser modelados e representados.

Sistemas baseados em conhecimento usam uma representação computacional de tais conhecimentos em forma de declarações sobre o domínio de interesse. Exemplos de tais declarações no domínio de viagens de negócio, por exemplo, são "empresas reservam viagens para seus empregados", "voos e passeios de trem são tipos especiais de viagens" ou "os empregados são pessoas empregadas em alguma empresa". Este conhecimento pode ser usado para responder a perguntas sobre o domínio de interesse. A partir dessas declarações e por meio de dedução automática, um sistema baseado em conhecimento pode, por exemplo, inferir que "uma pessoa em um voo reservado por uma

empresa é um funcionário" ou "a empresa que reservou um voo para pessoa é empregador dessa pessoa".

Nesse contexto, é importante esclarecer as formas ou técnicas de representação do conhecimento que existem.

2.3.1 Representação do Conhecimento

A área de representação de conhecimento estuda a formalização do conhecimento e seu processamento dentro de máquinas. Existem muitas formas (técnicas) de representação do conhecimento, as quais serão apresentadas a seguir:

- a) Redes semânticas: os formalismos de rede semântica se concentram em expressar a taxonomia da estrutura de categorias de objetos e as relações entre eles. Uma rede semântica é um grafo cujos nodos representam conceitos e os arcos representam relações entre esses conceitos. Eles permitem uma representação estrutural das sentenças de um domínio de interesse. Segundo Mylopoulos (1980), a vantagem dessa forma é que a sua representação gráfica aumenta sua compreensibilidade. A maior desvantagem desse esquema em rede é a falta de uma semântica formal e uma terminologia padrão. Redes semânticas são muito relacionadas com outra forma de representação do conhecimento chamada *frames*. De acordo com Russel e Norvig (2009) redes semânticas e *frames* são idênticos em sua expressividade, mas possuem metáforas de representação diferentes. Na metáfora do sistema de *frames*, conceitos são visualizados como caixas (*frames*) e as relações como *slots* dentro de *frames* que podem ser preenchidos por outros *frames*.
- b) Regras: outra forma de expressar conhecimento é através de regras que refletem a noção de consequência. Regras possuem a forma IF-THEN e permitem expressar vários tipos de relacionamentos complexos. A maior vantagem dessa abordagem é a forma natural e simples de extrair, codificar conhecimento e entender o conhecimento codificado. Entretanto, um sistema baseado em regras pode crescer muito incorporando milhares de regras e exigindo esforço extra e ferramentas para manter sua consistência (VASSEV; HINCHEY, 2011).
- c) Lógica: Esquemas de representação do conhecimento baseados em lógica empregam as noções de constantes, variáveis, funções, predicados, conectores e quantificadores lógicos para representar fatos (MYLOPOULOS, 1980). De acordo com Vassev e Hinchey (2011), a lógica é relevante, pois fornece uma semântica precisa e possibilita o raciocínio (inferência de novo conhecimento a partir do

conhecimento existente), que por sua vez é relevante para a dedução. O formalismo lógico mais proeminente usado para representar conhecimento é lógica de primeira ordem (*First-Order Logic* - FOL). Outro importante formalismo é lógica descritiva (*Description Logic* - DL), que evoluiu de redes semânticas e permite representar um conhecimento de domínio primeiramente definindo conceitos relevantes em TBox e depois usando ABox para especificar as propriedades dos objetos. A DL é menos expressiva que FOL, mas tem uma sintaxe mais compacta, melhores características computacionais e no contexto da web semântica se tornou o formalismo de representação de conhecimento mais importante. Tanto redes semânticas como regras tem sido formalizadas usando lógica para dar a elas uma semântica precisa. Sem essa formalização precisa tais representações são vagas e ambíguas, e portanto problemáticas para propósitos computacionais (GRIMM, HITZLER, ABECKER, 2007).

2.3.2 SBC's e a Engenharia do Conhecimento

Em seus primórdios a pesquisa em Inteligência Artificial (IA) esteve focada no desenvolvimento de formalismos, mecanismos de inferência e ferramentas para operacionalizar Sistemas Baseados em Conhecimento.

Studer, Benjamins e Fensel (1998) relatam que estes estudos ofereceram resultados bastante promissores, mas que a aplicação desta tecnologia na construção de SBC's para uso comercial não se mostrou adequada. Ou seja, os meios utilizados para desenvolver pequenos protótipos acadêmicos não escalavam até o *design* e manutenção de grandes sistemas comerciais.

No desenvolvimento dos primeiros SBC's o conhecimento era extraído dos especialistas através de entrevistas e posteriormente era codificado através de regras heurísticas. O conhecimento era representado de maneira uniforme, num mesmo nível de abstração, dificultando o desenvolvimento de grandes sistemas no âmbito comercial, já que a manutenção dos mesmos se tornava difícil e custosa. MYCIN (Shortlife, 1976) e INTERNIST (MILLER; POPLÉ; MYERS, 1982) são exemplos dos primeiros SBCs.

Essa situação foi comparada a crise do software do final dos anos sessenta e da mesma forma que esta originou a disciplina de Engenharia de Software (ES), a disciplina de Engenharia do Conhecimento também nasceu da necessidade de metodologias mais adequadas, com o objetivo

de transformar o processo de construção de um SBC de uma arte para um processo de engenharia, trazendo consigo uma série de metodologias, técnicas, linguagens e ferramentas específicas para esse fim (STUDER; BENJAMINS; FENSEL, 1998).

Em uma revisão com perspectiva histórica, Studer, Benjamins e Fensel (1998) destacam a mudança de paradigma na construção de SBCs, de uma abordagem de transferência de conhecimento para uma abordagem de modelagem de conhecimento.

No início dos anos oitenta o desenvolvimento de um SBC era visto como um processo de transferência do conhecimento humano para uma base de conhecimento implementada. Esta transferência era baseada no pressuposto de que o conhecimento requerido pelo SBC já existia e tinha de ser apenas coletado e implementado. Na maioria das vezes, o conhecimento necessário era obtido através de entrevistas com especialistas sobre como eles resolviam tarefas específicas. Normalmente, o conhecimento era implementado em regras de produção que por sua vez eram executadas por um interpretador de regras (*shell*).

Ainda nessa mesma década (80), a comunidade de IA reuniu-se em um esforço chamado de "segunda geração de sistemas especialistas" (David et al., 1993). O trabalho desse grupo envolveu a análise de muitos sistemas especialistas, construídos na década de 1970, com domínios diferentes, da medicina à física, que abrangiam uma variedade de tarefas de manipulação de sistema (resolução de problemas, *design*, controle), com o objetivo de abstrair sua concepção e operação.

Segundo Clansey, Sierhuis e Seah (2009) entre as motivações desse grupo estavam a produção de *frameworks* de alto nível para a construção de sistemas especialistas mais eficientes, mais robustos e poderosos através da incorporação de princípios gerais, em vez de fatos e heurísticas isoladas, a facilitação da explicação do raciocínio para os usuários e facilitação do reuso no desenvolvimento de outros sistemas especialistas através do desenvolvimento de linguagens de modelagem.

O resultado desse trabalho, fruto da análise das várias bases de conhecimento de regras mostrou que o formalismo de representação de regras de produção não era suficiente para apoiar uma adequada representação de diferentes tipos de conhecimento (CLANSEY, 1983). Ou seja, concluiu-se que as regras representavam os diferentes tipos de conhecimento de uma maneira muito uniforme e que poderiam ser categorizadas de acordo com os papéis que elas desempenhavam na solução do problema.

Essa situação pode ser exemplificada através do clássico sistema especialista Mycin (SHORTLIFE, 1976), composto por uma

base de regras sobre sintomas, condições médicas, resultados de testes de laboratórios e suas possíveis associações com infecções. Tais axiomas e regras eram específicos para esse domínio e de maneira mais problemática, as regras muitas vezes incorporavam conhecimento processual implícito, como a ordenação das regras. Este conhecimento de resolução de problemas implícito tornava a base de regras pouco confiável e difícil de manter enquanto o conhecimento de domínio evoluía.

As lições aprendidas com as primeiras experiências com sistemas baseados em regras não foram apenas que conhecimento de domínio devia ser representado de maneira explícita, mas também que o comportamento de resolução de problemas de um sistema devia ser representado em um componente separado do sistema (CRUBÉZY; MUSEN, 2004).

A abordagem de transferência de conhecimento só era viável para o desenvolvimento de pequenos protótipos, e não conseguia produzir bases de conhecimento grandes, confiáveis e sustentáveis. Além disso, reconheceu-se que a abordagem de transferência, onde a aquisição de conhecimento se baseia no conjunto de conhecimentos já existentes, era inadequada diante do papel importante do conhecimento tácito do especialista na resolução de problemas. Essas deficiências resultaram em uma mudança de paradigma da abordagem de transferência para a abordagem de modelagem do conhecimento.

Ainda segundo Studer, Benjamins e Fensel (1998) se verificou mais tarde a existência de um consenso geral de que o processo de construção de um SBC deveria ser visto como uma atividade de modelagem, ou seja, construir um SBC deveria significar a construção de um modelo computacional com o objetivo de realizar as capacidades de resolução de problemas comparáveis à um especialista de domínio.

Na fase de aquisição do conhecimento o especialista pode conscientemente articular algumas partes do seu conhecimento, entretanto ele não está ciente de uma parte significativa desse conhecimento, uma vez que está escondido em suas habilidades. Este conhecimento não é diretamente acessível, tem que ser construído e estruturado durante a fase de aquisição de conhecimento. Portanto, nessa fase a aquisição de conhecimento não é mais vista como uma transferência de conhecimento em uma representação de computador, mas como um processo de construção de um modelo (CLANCEY, 1989).

Segundo Studer, Benjamins e Fensel (1998), a visão de modelagem na construção de um SBC tem as seguintes consequências:

- Como todo modelo, é apenas uma aproximação da realidade. Em princípio, o processo de modelagem é infinito, porque é uma atividade incessante, com o objetivo de aproximar o comportamento desejado.

- O processo de modelagem é um processo cíclico. Novas observações podem levar a um refinamento, modificação ou a conclusão do modelo já construído. Por outro lado, o modelo pode orientar a nova aquisição do conhecimento.

- O processo de modelagem é dependente das interpretações subjetivas do engenheiro do conhecimento. Logo, este processo é tipicamente defeituoso e uma avaliação do modelo com relação com a realidade é indispensável para a criação de um modelo adequado.

SBCs conhecidos também como Sistemas Especialistas, Sistemas de Informação Intensiva em Conhecimento ou simplesmente Sistemas de Conhecimento são a única e mais importante criação comercial e industrial da disciplina Inteligência Artificial e são usados na solução de problemas humanos em muitas áreas de aplicação (SCHREIBER et al., 2002). Os mesmos autores destacam, a rapidez e a qualidade na tomada de decisão, assim como o aumento na produtividade como os principais benefícios advindos da utilização de SBCs.

2.3.3 Método de Resolução de Problema (*Problem Solving Method – PSM*)

Ao analisar uma primeira geração de sistemas baseados em conhecimento (sistemas especialistas) que realizavam diferentes tarefas, Clansy (1985) descobriu um comportamento comum de solução de problema. Ele abstraiu esse comportamento comum em um padrão de inferência genérico chamado Classificação Heurística (*Heuristic Classification*).

Esse padrão consiste em identificar a classe de uma situação de domínio desconhecida e envolve três passos de inferência principais: primeiro, abstrair os fatos sobre a situação de domínio (por exemplo, sintomas e condições do paciente) em características de alto nível; depois combinar o conjunto de características abstratas para possíveis explicações (paciente tem uma infecção gram-negativa) usando um conjunto de heurísticas e finalmente refinar a explicação (o micro-organismo responsável pela infecção do paciente é “E Coli”). O conjunto de passos pode ser conferido na Figura 13.

Clansey (1985) descreveu esse comportamento de solução em um nível abstrato, conhecido como Nível de Conhecimento (*Knowledge Level*). O nível de conhecimento foi primeiramente apresentado por Newell (1982) com o argumento de que é possível descrever o conhecimento em um nível superior ao nível de símbolos comumente utilizados em sistemas de representação de conhecimento.

O nível de conhecimento permite descrever o raciocínio em termos de objetivos ou tarefas a serem alcançadas, as ações necessárias para atingir esses objetivos e os conhecimentos necessários para executar essas ações. Desde então, o nível de conhecimento se tornou o princípio fundamental da Engenharia do Conhecimento (SCHREIBER, 2007).

A descrição do processo de solução de um problema em nível de conhecimento abstraiu os detalhes relacionados com a implementação do processo de raciocínio e resultou na noção de Método de Resolução de Problema (*Problem-Solving Method* ou *PSM*) (BENJAMINS e FENSEL, 1998).

Métodos de resolução de problemas (*PSMs*) foram introduzidos como um paradigma de Engenharia do Conhecimento para codificar sequências de passos de inferência independente de domínio, sistemáticos e reutilizáveis envolvidos no processo de resolução de certos tipos de tarefas de aplicação com o conhecimento de domínio (CLANSEY, 1985; CHANDRASEKARAN, 1986 e MCDERMOTT, 1988).

PSMs abstraem e isolam conhecimento procedural do conhecimento de domínio, tornando tanto o comportamento de inferência do sistema como o conhecimento de domínio nesse sistema explícito e fácil de adaptar (CRUBÉZY; MUSEN, 2004).

De acordo com Studer, Benjamins e Fensel (1998), um *PSM* possui as seguintes características:

- Um *PSM* especifica quais ações de inferência precisam ser realizadas para resolver uma dada tarefa.
- Um *PSM* determina a sequência em que essas ações devem ser ativadas.
- Em um *PSM*, os chamados papéis de conhecimento (*knowledge roles*) determinam qual o papel que o conhecimento de domínio desempenha em cada ação de inferência. Estes papéis de conhecimento definem uma terminologia genérica independente de domínio.

No processo de engenharia do conhecimento, *PSMs* podem ser explorados de diferentes formas (Fensel, 1997):

1. *PSMs* contêm ações de inferência que necessitam de conhecimentos específicos, para cumprirem a sua missão. Assim, um *PSM* pode ser usado como um guia para aquisição de conhecimento de domínio estático, ou seja, pode guiar o processo de elicitação de conhecimento.
2. *PSMs* permitem descrever a linha principal do processo de raciocínio de um SBC, o que suporta a validação do SBC, porque o especialista é capaz de entender o processo de resolução do problema. Além disso, esta descrição abstrata pode ser usada durante o próprio processo de resolução de problema, com o objetivo de facilitar a explicação.
3. *PSMs* podem ser reutilizados no desenvolvimento de diferentes SBCs, logo a construção de SBCs pode ser feita a partir de uma biblioteca de componentes reutilizáveis.

Chandrasekaran (1986) caracteriza esse último ponto como um dos principais objetivos de pesquisa em *PSM*. A ideia é fornecer uma biblioteca de algoritmos pré-definidos e implementados para que os desenvolvedores possam usar para construir sistemas baseado em conhecimento. Essa ideia traz a promessa de reutilização de software para estes sistemas, permitindo o rápido desenvolvimento de aplicações baseados em conhecimento.

2.3.4 Abordagens de Modelagem de Conhecimento

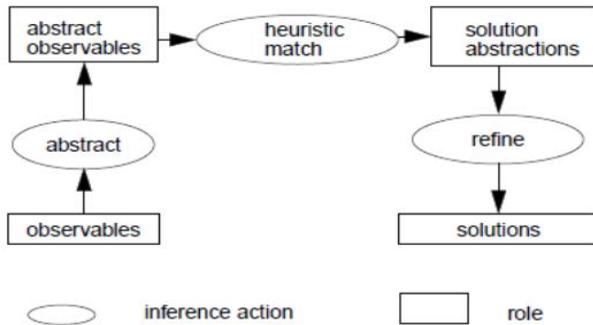
Três abordagens principais surgiram na década de 80 e tiveram influência nas abordagens de modelagem em engenharia do conhecimento: *Role-Limiting Methods* (MCDERMOTT, 1988), *Generic Tasks* (CHANDRASEKARAN, 1986) e *Task-Structure* (CHANDRASEKARAN; JOHNSON; SMITH, 1992).

a) *Role-Limiting Methods (RLM)*

RLM foi uma das primeiras tentativas de suportar o desenvolvimento de SBCs através da proposição de um método de resolução de problema reutilizável. A abordagem RLM pode ser caracterizada como uma abordagem do tipo *shell*, que possui uma implementação de um *PSM* específico que pode ser usado somente para resolver um tipo específico de tarefa para o qual foi desenvolvido. Esse *PSM* também define papéis (*roles*) genéricos que o conhecimento pode ter durante o processo de resolução de problema e ele fixa completamente a representação do conhecimento para esses papéis, de tal forma que o especialista somente tenha que instanciar os conceitos e relacionamentos genéricos. Como pode ser observado na Figura 13 o *PSM* de Classificação Heurística oferece um papel chamado

“observables” para o especialista. Usando esse papel o especialista tem que verificar qual conceito específico de domínio corresponde a esse papel, por exemplo, “dados de paciente” e tem que fornecer instâncias de domínio para esse conceito, no caso, fatos concretos sobre pacientes.

Figura 13 - Método de resolução de problema Classificação Heurística



Fonte: Adaptado de Studer, Benjamins e Fensel (1998).

b) Generic Taks (GT)

GT foi uma abordagem utilizada na construção de muitos SBCs na década de oitenta. Uma GT, tal como “Classificação Hierárquica” (*Hierarchical Classification*) é um bloco de construção que pode ser reutilizada em muitos SBCs. Nessa abordagem cada GT está associada com uma descrição genérica das suas entradas e saídas. Além disso, uma GT vem com esquema fixo de tipos de conhecimento especificando a estrutura do conhecimento de domínio necessário para resolver uma tarefa. Por fim, uma GT inclui uma estratégia de resolução de problema fixa especificando os passos de inferência da estratégia.

Entretanto, Chandrasekaran, Johnson e Smith (1992) analisaram essa abordagem com mais detalhes e identificaram duas principais desvantagens. Primeiro, a noção de tarefa entra em conflito com a noção de PSM usado para resolver uma tarefa, uma vez que cada GT inclui uma estratégia de resolução de problema determinada. Segundo, a complexidade das GTs que estavam sendo propostas era muito diferente, ou seja, não havia uma definição de qual nível de granularidade era o mais adequado para cada GT. Nesse contexto, os autores propuseram a abordagem “Estrutura de Tarefa” (*task-structure*).

c) Task-structure

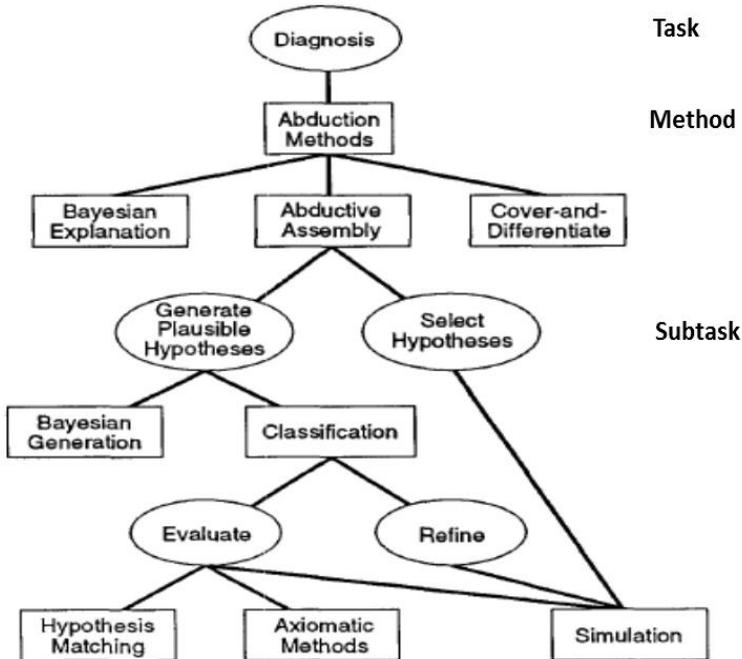
Task-structure é um *framework* em nível de tarefa para descrever sistemas. Ele consiste em uma árvore de tarefas, métodos e subtarefas aplicadas de forma recursiva até que as tarefas de nível mais baixo sejam resolvidas por um conhecimento diretamente disponível.

Nessa abordagem uma tarefa refere-se a um tipo de problema, por exemplo a tarefa de diagnóstico, e um método é uma forma de cumprir uma tarefa. Dessa maneira uma estrutura de tarefa pode ser definida como na Figura 14, onde uma tarefa, representada por um círculo, está associada a um conjunto de métodos, representados por retângulos, adequados para realizar esta tarefa. Cada método pode ser decomposto em várias subtarefas. A estrutura de decomposição é refinada a um nível onde são introduzidas subtarefas elementares que podem ser diretamente resolvidas através da utilização do conhecimento disponível.

Em outras palavras na metodologia analítica de tarefa, uma tarefa define "o que tem que ser feito". As tarefas são realizadas pela aplicação de um método, que define "como executar uma tarefa".

Esta metodologia tem sido utilizada com sucesso para modelar processos complexos como diagnóstico médico (BENJAMINS; JANSWEIJER, 1994), projetos de engenharia de software (HACKOS; REDISH, 1998), projetos de monitoramento de dados de saúde pública (BUCKERIDGE et al., 2004, 2008) e no entendimento de processos de proveniência da informação (GOMÉZ-PÉREZ; CORCHO, 2008).

Figura 14 - Estrutura de tarefa para a tarefa de Diagnóstico.



Fonte: Adaptado de Chandrasekaran, Johnson e Smith (1992).

Métodos são maneiras de cumprir tarefas e podem ser do tipo computacional ou situado. Por exemplo, a tarefa de prever o comportamento de um dispositivo pode ser executada por um método computacional que realiza uma simulação ou pode ser resolvida através da inspeção do modelo físico do dispositivo.

A estrutura de tarefa facilita a modelagem do conhecimento porque associa tarefas com métodos que cumprem essas tarefas e os conhecimentos necessários para usar os métodos. Os vários níveis da estrutura de tarefa mostram como o conhecimento pode ser decomposto em corpos de conhecimento que estão associados às tarefas específicas (CHANDRASEKARAN; JOHNSON; SMITH, 1992).

A abordagem de estrutura de tarefa, conhecida também como decomposição de tarefa, permite que desenvolvedores selecionem ou criem *PSMs* (métodos) para modelar o conhecimento de controle

necessário para resolver uma tarefa. Como mencionado anteriormente cada *PSM* pode implicar em subtarefas, que por sua vez são modeladas usando *PSMs* que por sua vez podem acarretar em novas subtarefas. No mínimo, cada *PSM* fornece relações de entrada e saída, que descrevem os dados da tarefa, requisitos de conhecimento e seus resultados. Essa abordagem modular facilita a montagem e a reutilização de *PSMs*.

Seguindo a mesma linha de modelar conhecimento usando a noção de tarefa, outras abordagens de modelagem, tais como KADS (WIELINGA; SCHREIBER; BREUKER, 1992) surgiram na mesma época. Entretanto, Chandrasekaran, Johnson e Smith (1992), argumentam que nessa abordagem a tarefa é entendida como uma estratégia fixa para resolver um problema, ou seja, o termo é sinônimo do método ou especificação do procedimento. Enquanto que na abordagem de estrutura de tarefa (*task-structure*) dos autores, eles separam a noção de tarefa do método que executada a tarefa.

KADS evoluiu posteriormente para CommonKADS, um conhecido *framework* de engenharia do conhecimento proposto por Schreiber (2000). A principal característica dessa metodologia é a construção de uma coleção de modelos como um conjunto estruturado de conhecimento que reflete todos os aspectos importantes para que um SBC tenha êxito dentro de um contexto organizacional. Esses modelos são: Modelo da Organização, Modelo de Tarefa, Modelos de Agente, Modelo de Comunicação, Modelo de Conhecimento e Modelo de Projeto. Os três primeiros objetivam modelar o ambiente onde o SBC irá operar. Os Modelos de Conhecimento e Agentes representam os requisitos de entrada que guiarão a implementação do sistema através do Modelo de Projeto.

A principal contribuição dessa metodologia é a proposta do Modelo de Conhecimento, o qual descreve o conhecimento que determinado agente possui e que é relevante para a execução de determinada tarefa, além de descrever a estrutura do mesmo em função do seu uso. Cada modelo é composto de três tipos ou níveis de conhecimento: nível de domínio, nível de inferência e nível de tarefa (SCHREIBER et al., 2002).

2.3.5 Ontologias

O conceito de ontologia foi adaptado da Filosofia, onde é empregado para descrever e classificar as coisas do mundo, para a Inteligência Artificial, área que lida com raciocínio sobre modelos do mundo, com o objetivo de descrever conceitos do mundo e suas relações (STUDER et al., 1998). Ontologias são modelos conceituais do que

existe em algum domínio e podem ser levadas para uma forma interpretável por máquina através de técnicas de representação de conhecimento (GRIMM; HITZLER; ABECKER, 2007).

Ao longo do tempo várias definições de ontologia foram dadas. Uma definição bastante conhecida foi dada por Gruber (1993): "Uma ontologia é uma especificação explícita de uma conceitualização". A definição de Gruber foi estendida por Borst (1997): "Ontologia é uma especificação formal de uma conceitualização compartilhada". Posteriormente Studer et. al. (1998) basearam-se nas definições acima e criaram a sua definição: "Uma especificação formal e explícita de uma conceitualização compartilhada". Formal porque uma ontologia deve ser legível por máquina. Explícita porque os tipos de conceitos utilizados e as restrições ao seu uso são explicitamente definidos. Conceitualização porque refere-se a um modelo abstrato que identifica os conceitos relevantes de algum fenômeno do mundo e compartilhado porque a ontologia captura o conhecimento consensual aceito por um grupo.

Segundo Chandrasekaran, Josephson e Benjamins (1998) ontologias são importantes por duas razões:

- a) Ontologias esclarecem a estrutura de conhecimento: ontologias são o coração de qualquer sistema de representação do conhecimento. Se não se tem as conceitualizações que fundamentam o conhecimento, então não se tem um vocabulário para a representação do conhecimento. Logo, o primeiro passo para a representação do conhecimento é a análise ontológica efetiva de algum campo do conhecimento.
- b) Ontologias permitem o compartilhamento de conhecimento: elas fornecem um meio para o compartilhamento de conhecimento entre pessoas com necessidades semelhantes de representação do conhecimento no mesmo, evitando a necessidade de replicar a análise do conhecimento.

Existem diferentes formalismos de representação e linguagens associadas para a criação de ontologias. Entretanto, todos eles compartilham um conjunto mínimo de componentes: classes, relações, funções, axiomas e instâncias (CORCHO et al., 2006)

- Classes representam conceitos usados em um amplo sentido, podendo ser abstratos ou concretos, elementares ou compostos, reais ou fictícios. Por exemplo, no domínio de viagem, conceitos podem ser locais (cidades), acomodações (hotel, camping), etc. Eles podem ser a descrição de qualquer coisa que é dita, logo, podem ser a descrição de uma tarefa, função, ação, processo de raciocínio,

etc. Classes normalmente são organizadas como uma taxonomia, onde mecanismos de herança podem ser aplicados.

- Relações representam os tipos de interações entre os conceitos de um domínio. Ontologias expressam relações binárias entre conceitos onde o primeiro argumento é conhecido como *domain* e o segundo argumento é conhecido como *range*. Por exemplo, a relação `temLugarDeChegada` tem como *domain* o conceito `Voo` e como *range* o conceito `Localizacao`. Relações podem ser instanciadas com o conhecimento do domínio, como por exemplo, `voo FLN5678 temLugarDeChegada Florianópolis`. Este tipo de relação também é conhecida como propriedade de objetos (*object property*). Relações binárias são também utilizadas para representar atributos de um conceito. Atributos são distinguidos de relações porque o seu *range* é sempre um tipo de dado, tal como *string*, *number*, etc., enquanto que o *domain* da relação é sempre um conceito.
- Axiomas são utilizados para representar conhecimento que não pode ser representado com outros componentes. Além disso, são usadas para verificar a própria ontologia e a consistência do conhecimento armazenado na base de conhecimento. Axiomas também são úteis para inferir novos conhecimentos.
- Instâncias são usadas para representar elementos ou indivíduos de uma ontologia.

As primeiras linguagens para representação de ontologias foram criadas no início da década de 90 como uma evolução das linguagens de representação do conhecimento. Corcho e Gomez-Perez (2000) apresentam uma comparação de diferentes linguagens ontológicas em relação às suas capacidades de representação do conhecimento e inferência. São exemplos de primeiras linguagens ontológicas: Ontolingua (FARQUHAR, 1997), OCML (MOTTA, 1998) e F-Logic (KIFER; LAUSER; WU, 1995).

Os paradigmas que deram suporte a essas linguagens são diversos: baseado em frame (*frame-based*), lógica descritiva (*description logic*), cálculo de predicado de primeira e segunda ordem e orientação a objetos.

Com o advento da *Internet*, linguagens para criação de ontologias que exploram as características da *Web* foram criadas. Essas linguagens são chamadas de linguagens de marcação de ontologias ou linguagens de ontologias baseada na *Web* porque sua sintaxe é baseada em linguagens de marcação como HTML (*HyperText Markup Language*) e

XML (*Extensible Markup Language*). São primeiros exemplos dessa categoria de linguagens: XOL (KARP; CHAUDHRI; THOMERE, 1999), SHOE (*Simple HTML Ontology Extensions*) (LUKE; HEFLIN, 2000) e OIL (*Ontology Inference Layer* ou *Ontology Interchange Language*) (HORROCKS et al., 2000).

A representante mais atual desse grupo de linguagens é OWL (*Ontology Web Language*)¹¹, uma linguagem de marcação de ontologia para a Web Semântica, recomendada pelo W3C¹².

Existem diferentes níveis de ontologias. Guarino (1998) propõe uma classificação de ontologias em quatro categorias distintas: Ontologias de Alto-Nível, Ontologias de Domínio, Ontologias de Tarefas e Ontologias de Aplicação.

- Ontologias de Alto Nível: descrevem conceitos gerais, como por exemplo, os conceitos básicos do espaço.
- Ontologias de Domínio: descrevem um vocabulário relacionado a um domínio específico, tal como Medicina.
- Ontologias de Tarefas: descrevem uma tarefa ou uma atividade, como a avaliação de poluição sonora em ambientes urbanos;
- Ontologias de Aplicação: descrevem conceitos que dependem tanto de um domínio específico como de uma tarefa específica.

Gómez-Pérez, López e Corcho (2002) apresentam outra forma de classificar ontologias, que se dá pela expressividade e formalismo:

- *Lightweight*: são ontologias que modelam informação de um determinado domínio (conceitos e sua taxonomia), sem incluir axiomas e restrições. Neste sentido, este tipo de ontologia não requer um nível de expressividade elevado, o que, por outro lado, dificulta o processo de raciocínio.

- *Heavyweigh*: são ontologias que requerem um alto nível de expressividade para incorporar axiomas e restrições, facilitando os processos de inferência neste tipo de ontologia.

O processo de construção de ontologia até metade dos anos noventa era mais uma arte do que um processo de engenharia. Cada equipe de desenvolvimento seguia seus próprios princípios e fases nessa construção, o que dificultava o reuso e a extensão de tais ontologias. A partir de então diversas metodologias foram desenvolvidas no intuito de

¹¹ <http://www.w3.org/2004/OWL/>

¹² World Wide Web Consortium. www.w3.org/

sistematizar a construção e a manipulação de ontologias (GÓMEZ-PEREZ; LÓPEZ; CORCHO, 2002).

Methontology (LÓPEZ, 1999) e *On-to-Knowledge* (SURE et al., 2003) segundo Suárez-Figueroa (2007) são consideradas as metodologias mais completas porque apresentam um guia para o desenvolvimento de ontologias a partir do zero.

Como contraponto a essas duas abordagens metodológicas citadas acima, Noy e McGuinness (2001) dizem que não existe uma maneira ou metodologia correta para desenvolver uma ontologia e propõe uma série de passos como um possível processo iterativo de construção.

O processo de construção de uma ontologia visa um conhecimento de domínio consensual, portanto, seu desenvolvimento é muitas vezes um processo de cooperação envolvendo diferentes pessoas ou organizações. No que diz respeito às pessoas é dito que elas comprometem-se com a ontologia se elas concordam em aceitar tal ontologia.

Para o desenvolvimento de ontologias existem ferramentas que são dependentes de linguagens ou outras que são independentes. Nessa última categoria se enquadra Protégé¹³, uma ferramenta (*framework*) bastante consolidada. Protégé foi criada no Grupo de Modelagem de Conhecimento da *Stanford Medical Informatics* e é definida pelos membros desse grupo como um ambiente de desenvolvimento de sistemas baseados em conhecimento (GENNARI et al., 2003).

O objetivo original de Protégé foi reduzir o gargalo na aquisição do conhecimento, minimizando o papel do engenheiro de conhecimento na construção de bases de conhecimento. Ao longo de mais de uma década, o grupo de modelagem de *Stanford* trabalhou para transformar Protégé em uma ferramenta de modelagem de conhecimento de propósito geral.

O funcionamento de Protégé está apoiado na ideia de que a construção da base de conhecimento deve ser dividida em (1) construção da ontologia por um engenheiro de conhecimento e, em seguida, (2) preenchimento da base de conhecimento por um especialista de domínio.

Além disso, Protégé implementa um conjunto de estruturas de modelagem de conhecimento e ações que suportam a criação, a visualização e a manipulação de ontologias representadas em diversos formatos.

¹³ <http://protege.stanford.edu/>

2.3.6 Ontologias e Métodos de Resolução de Problemas (PSM)

Qualquer SBC compreende pelo menos duas partes fundamentais: conhecimento de domínio e conhecimento de resolução de problemas. O conhecimento de domínio é o conhecimento sobre a realidade no domínio de interesse (objetos, relações, eventos, estados, relações causais, etc.) e o conhecimento de resolução de problemas é o conhecimento sobre a forma de resolver problemas (CHANDRASEKARAN; JOSEPHSON; BENJAMINS, 1998).

Ontologias desempenham um papel principal na análise, modelagem e implementação do conhecimento de domínio, mas elas também no conhecimento de resolução de problemas (STUDER et al. 1998; CHANDRASEKARAN; JOSEPHSON; BENJAMINS, 1998).

Ontologias são usadas na modelagem do domínio de conhecimento, utilizado pelos *PSMs*. No entanto, *PSMs* e ontologias de domínio são desenvolvidos de forma independente e necessitam ser conciliados para formar um sistema de conhecimento coerente. De forma a possibilitar essa conciliação é necessário a declaração do formato e semântica do conhecimento e dos dados que o PSM espera do domínio para realizar a sua tarefa. A representação de um vocabulário de termos e relações dos requisitos de dados dos *PSMs* através de ontologias (ontologias de tarefa e método) permite explicitar a interação entre o conhecimento de domínio e o método de resolução de problemas.

De acordo com Crubézy e Musen (2003) ontologias e *PSMs* tem uma longa e intrínseca história de relacionamento. *PSMs* e as ontologias que especificam o conhecimento de domínio são dois blocos de construção de aplicações de conhecimento: *PSMs* fornecem o componente de raciocínio do sistema, que conta com o conteúdo da ontologia domínio para resolver problemas nesse domínio.

PSMs agem como modelos de raciocínio que precisam ser instanciados com conhecimento de domínio para cada nova aplicação. No entanto, essa instanciação muitas vezes não é simples já que *PSMs* permitem que o conhecimento de domínio seja livre do conhecimento procedural de resolução de problemas, tornando-os também reutilizáveis para finalidades diferentes. De maneira inversa, para serem reutilizáveis, *PSMs* também devem ser livres de qualquer domínio (CRUBÉZY; MUSEN, 2003).

Depois da introdução da ideia de *PSM* a comunidade de engenharia conhecimento pensou em várias formas de maximizar o potencial de ontologias como base para a reutilização do conhecimento de solução de problemas.

Já no início dos anos noventa, Mizoguchi, van Welkenhuysen e Ikeda (1995) começaram a falar sobre ontologias de tarefa como instrumentos valiosos para ligar o vocabulário e vista de um usuário ao de um solucionador de problemas. Estas ontologias de tarefas compreendiam o vocabulário e os passos de raciocínio de tarefas específicas e, nesse sentido, eles se relacionam com as noções de objetivo de resolução de problemas e conhecimento de resolução de problemas.

Gennari et al. (1994) e Studer et al. (1996) apresentaram as primeiras propostas de modelagem de resolução de problemas baseada em ontologias. Essas propostas foram motivadas pela necessidade de localização e configuração de *PSMs* de forma a possibilitar o reuso dos mesmos. Gennari et al. (1994) apresentam uma ontologia de métodos onde os requisitos de dados dos métodos são representados. Studer et al. (1996) propõem uma ontologia de método que representa todos os conceitos genéricos e relacionamentos que são usados pelo *PSM* para fornecer sua funcionalidade e uma ontologia de subtarefa que descreve os passos de inferência que compõem um método.

Fensel et al. (1997a, 1997b) usam ontologias para descrever classes de problemas genéricos (ontologia de tarefas) e *PSMs* (ontologia de método).

Como exemplo de utilização de *PSMs* modelados pela metodologia de estrutura de tarefa e representados através de ontologias de métodos e de tarefas, citam-se os diversos trabalhos da equipe da *Stanford Medical Informatics* (CRUBÉZY et al., 2005; BUCKERIDGE et al., 2004, 2008; O'CONNOR et al., 2009). Os trabalhos apresentam um *framework* de conhecimento para vigilância de dados de saúde pública.

2.4 MINERAÇÃO DE TEXTO

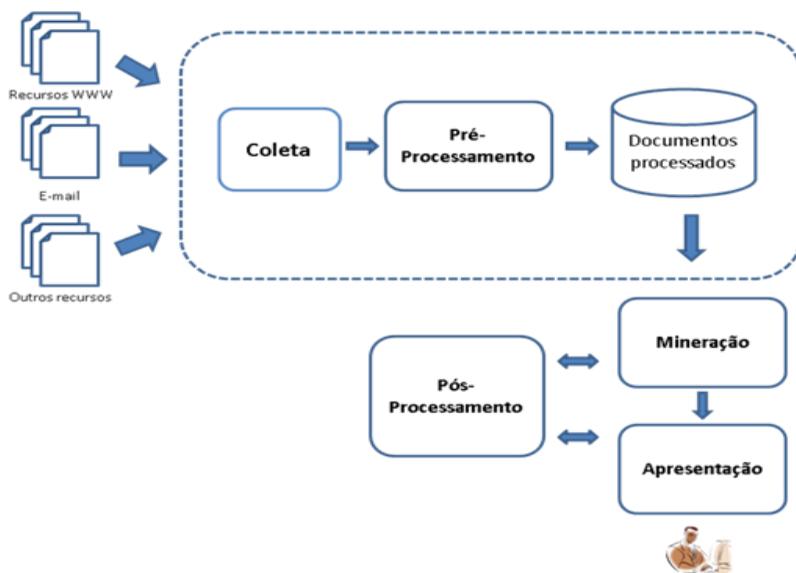
Mineração de texto ou *Text Mining* pode ser amplamente definido como um processo de conhecimento intensivo em que um usuário interage com uma coleção de documentos ao longo do tempo usando um conjunto de ferramentas de análise. De uma forma análoga à mineração de dados, a mineração de texto procura extrair informações úteis a partir de fontes de dados, neste caso coleções de documentos, através da identificação e exploração de padrões interessantes (SANGER; FELDMAN, 2007).

Esta é uma área de pesquisa que tenta resolver o problema da sobrecarga de informação, usando técnicas de mineração de dados, aprendizado de máquina, processamento de linguagem natural (NLP), recuperação de informação (IR) e gestão do conhecimento.

Em uma perspectiva que considera a mineração de texto como um processo que contempla um conjunto de etapas, Sanger e Feldman (2007) destacam que o processo envolve o pré-processamento de coleções de documentos (categorização de textos, extração de informação, extração de termos), o armazenamento das representações intermediárias, as técnicas para analisar estas representações intermediárias (tais como a análise de distribuição, agrupamento (*clustering*), análise de tendências e as regras de associação) e visualização de os resultados.

Os autores apresentam uma arquitetura geral de sistema de mineração de textos na qual cinco etapas emergem: Coleta, Pré-Processamento, Mineração, Apresentação e Pós-Processamento. A Figura 15 ilustra a integração dessas etapas. A seguir uma ênfase é dada na etapa de Pré-Processamento, cuja algumas tarefas foram utilizadas neste trabalho.

Figura 15 - Arquitetura de um sistema de mineração de texto



Fonte: Adaptado de Feldman e Sanger (2007).

2.4.1 Pré-Processamento

A etapa de pré-processamento objetiva preparar os dados para a etapa seguinte, a de mineração. É considerada a etapa mais onerosa do processo uma vez que não existe um único algoritmo que possa ser aplicado para gerar uma representação satisfatória para ser usada nas próximas etapas.

Tarefas de pré-processamento geralmente consistem em converter as informações originais de cada fonte de dados para um formato canônico antes de aplicar vários tipos de métodos de extração contra estes documentos para criar uma nova coleção de documentos totalmente representada por conceitos (SANGER; FELDMAN, 2007).

Entre as tarefas ou técnicas de propósito geral dessa etapa estão a tokenização, que tem a função de extrair unidades mínimas do texto, também conhecidas como *tokens*, e a redução do léxico, que objetiva diminuir o número de *tokens* gerados, realizado com a remoção de *stopwords* que são *tokens* que não possuem valor semântico, sendo úteis apenas para a compreensão geral do texto. Outras técnicas de redução podem ser utilizadas, tais como as técnicas de normalização *stemming*, onde cada palavra é reduzida até que seja obtida sua respectiva raiz, e

lemmatization, onde é substituído as diversas formas de representação da palavra pela sua forma primitiva, por exemplo, as formas “livro”, “livros” e “livraria” apontam todas para o lexema “livro”.

Categorização de Texto (CT), Extração da Informação (EI) e Clusterização são apresentadas por Feldman e Sanger (2007) como técnicas da etapa de pré-processamento. CT e EI são técnicas as vezes referidas como “*tagging*” (etiquetagem) pois ambas permitem a obtenção de representações formais e estruturadas de documentos. A clusterização, que tem a finalidade de agrupar documentos similares em unidades chamadas *clusters*, também é apresentada pelos autores na etapa de pós-processamento.

Durante a etapa de pré-processamento, os documentos são convertidos para uma representação mais gerenciável. Normalmente, os documentos são representados por vetores de características (*feature vector*). Uma característica é uma dimensão no espaço de características. Um documento é representado como um vetor - uma sequência de características e seus pesos - neste espaço (*vector space model*).

O modelo de vetor de características mais comum é o *bag-of-words*, que usa termos do documento como características (vetor de termos). Entre os métodos de pesos para as características da representação, o mais simples é o binário em que o peso da característica é 1 se o termo correspondente está presente no documento - ou zero caso contrário. Métodos mais complexos levam em conta a frequência do termo no documento, e ou em toda a coleção de documentos.

A Extração de Informação tem o objetivo de extrair informações relevantes de fontes de dados semiestruturadas e não-estruturadas. Essas informações consistem em entidades e relacionamentos que são suscetíveis de serem significativas. Para Feldman e Sanger (2007) a EI é talvez a mais proeminente técnica atualmente utilizada no texto de pré-processamento de mineração pois sem ela os sistemas de mineração de texto teriam suas capacidades de descoberta de conhecimento muito mais limitadas. O tipo mais simples de extração de informações é chamado de extração de termos. Não existe nenhum tipo de *frame*, há simplesmente apenas um tipo de entidade - “termo”.

A Clusterização tem a finalidade de agrupar documentos similares em unidades chamadas *clusters*. Difere da tarefa de classificação por não se conhecer a priori o número de *clusters* que serão gerados.

Para que a similaridade entre os documentos seja calculada é preciso que o mesmo seja representado na forma vetorial. No Modelo de Espaço Vetorial (*Vector Space Model - VSM*) (SALTON; WONG; YANG, 1975) cada documento é um vetor e pode ser comparado a outro vetor através de uma medida de similaridade.

Um bom agrupamento deve agrupar documentos semelhantes e separar os que não são semelhantes. Portanto, a qualidade do agrupamento é normalmente especificada em função da similaridade entre esses documentos.

Para o agrupamento de documentos a medida de similaridade mais comum é o cosseno (*cosine*). A equação do cosseno mede o ângulo entre dois vetores, variando de 1.0 ($\cos(0^\circ) = 1.0$) para vetores apontando na mesma direção, 0.0 ($\cos(90^\circ) = 0.0$) para vetores ortogonais¹⁴ e -1.0 ($\cos(180^\circ) = -1.0$) para vetores apontando em direções opostas, sendo definido como :

$$\cos \theta = \frac{\sum_{i=1}^n (t_i \times q_i)}{\sqrt{\sum_{k=1}^n (t_k)^2} \times \sqrt{\sum_{j=1}^n (q_j)^2}}$$

onde t_i e t_k são as frequências normalizadas dos i th e k th termos do vetor t , q_i e q_j são as frequências dos i th e j th termos do vetor q .

A seguir é apresentada uma abordagem de clusterização proposta por Osinski e Weiss (2005) denominada Lingo¹⁵ utilizada neste trabalho somente para a descoberta de termos significativos (rótulos) ou seja até uma determinada fase do algoritmo. A apresentação segue a proposição dos autores na íntegra.

2.4.2 Abordagem de clusterização Lingo

A abordagem de clusterização Lingo inverte a ordem tradicional de descoberta de *clusters*. Em vez de calcular a proximidade entre documentos e, em seguida, rotular os grupos descobertos, primeiramente tenta-se encontrar rótulos de agrupamentos (*clusters*) adequados e conceitualmente variados e, em seguida, atribuir documentos para os rótulos para formar grupos (OSINSKI; WEISS, 2005).

¹⁴ Que formam ângulos retos.

¹⁵ <http://carrotsearch.com/lingo3g-overview.html>

Na abordagem “descrição-vem-primeiro” de Lingo, a seleção cuidadosa dos candidatos a rótulo é crucial. O algoritmo deve assegurar que os rótulos são significativamente diferentes quando cobre a maioria dos tópicos nos *snippets* (fragmentos) de entrada. Para encontrar os candidatos a rótulo é usado Modelo de Espaço Vetorial (*Vector Space Model - VSM*) e Decomposição de Valores Singulares (*Single Value Decomposition -SVD*), sendo este último o construto matemático essencial que fundamenta a técnica *Latente Semantic Index (LSI)* (BERRY; DUMAIS; O'BRIEN, 1994).

VSM é um método de recuperação de informação que utiliza operações de álgebra linear para comparar dados textuais. VSM associa um único vetor multidimensional com cada documento da coleção, e cada componente desse vetor reflete uma determinada palavra-chave ou termo relacionado com o documento. Usa-se "termo" para uma única palavra e "frase" para se referir a uma sequência de termos. Isso permite a representação de um conjunto de documentos, organizando seus vetores em uma matriz “termo-documento”. O valor de um único componente da matriz “termo-documento” depende da força do relacionamento entre o seu termo associado e o respectivo documento.

Ao contrário do VSM, LSI objetiva representar a coleção de entrada usando conceitos encontrados nos documentos, em vez dos termos literais que aparecem neles. Para fazer isso, LSI aproxima a matriz “termo-documento” original utilizando um número limitado de fatores ortogonais.

Esses fatores representam um conjunto de conceitos abstratos, cada um transmitindo alguma ideia comum a um subconjunto do conjunto de entrada.

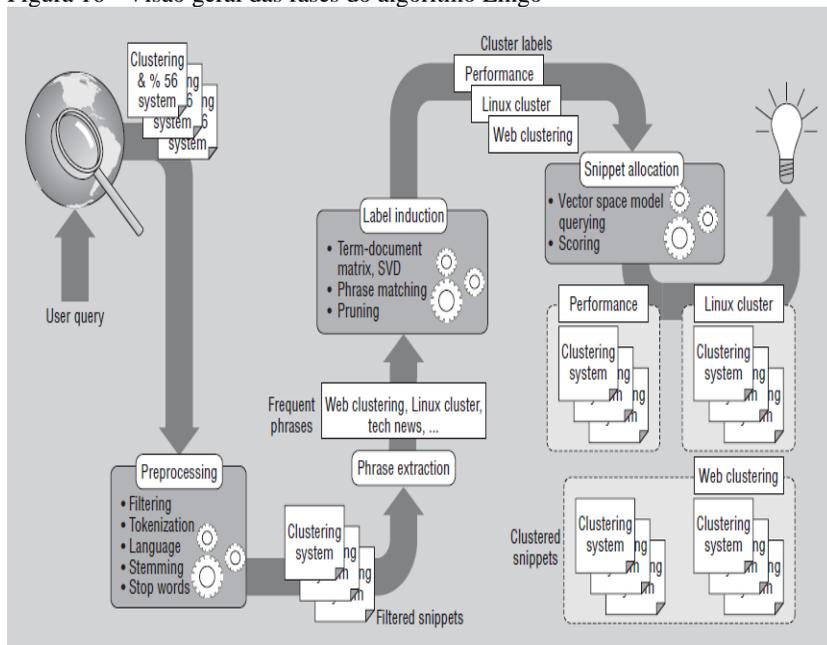
Para o Lingo, estes conceitos são candidatos adequados a rótulos de *clusters*. Entretanto, sua representação matricial é difícil para o entendimento humano. Para obter rótulos concisos e informativos, precisa-se extrair o significado verbal por trás da representação vetorial dos conceitos abstratos derivados da LSI.

Entre todas as coocorrências de termos em uma coleção de documentos, frases que aparecem pelo menos um certo número de vezes na coleção de entrada, parecem ser mais significativas para os usuários, e ao mesmo tempo são relativamente fáceis de se encontrar. Eles são muitas vezes colocações (“*collocations*”) ou nomes próprios, tornando-os ao mesmo tempo informativo e conciso. Logo, utiliza-se eles para aproximar o significado verbal dos conceitos abstratos derivados do SVD. Lingo usa essas frases frequentes como rótulos de *clusters*. Uma

vez que se tenha descoberto um conjunto diversificado de rótulos, atribuem-se os documentos mais relevantes a estas usando VSM padrão.

A Figura 16 apresenta uma visão geral das fases do algoritmo Lingo, que serão apresentadas na sequência.

Figura 16 - Visão geral das fases do algoritmo Lingo



Fonte: Osinski e Weiss (2005)

Pré-processamento (*Preprocessing*)

Nesta fase, usa-se uma combinação de três métodos comuns de pré-processamento de texto, sendo: a) *stemming*, uma técnica para encontrar uma representação semântica de uma palavra flexionada (geralmente um lema) para diminuir o impacto da sintaxe de uma língua; b) ignorar *stop-word*, para lidar com termos que ocorrem com frequência, mas que não têm nenhum significado (conjunções, artigos, etc.); e c) heurísticas de segmentação de texto, para dividir o texto em palavras e frases.

Lingo reconhece a linguagem de cada *snippet* separadamente e aplica o pré-processamento adequado a ele. A maioria dos algoritmos remove *stop-words* inteiramente. Lingo somente marca as *stop-words*,

porque elas podem ajudar os usuários a entender o significado de frases longas (por exemplo, "Câmara Comércio" versus "Câmara de Comércio").

Extração de frase (*Phrase extraction*)

A fase de extração de frase visa descobrir frases e termos individuais que poderiam potencialmente explicar os significados verbais por trás dos conceitos abstratos derivados da SVD.

Lingo requer que os rótulos de *cluster* apareçam no fragmento de entrada (*snippet*), pelo menos, um número especificado de vezes; não cruzem limites da sentença - marcadores de sentença indicam uma mudança de tópico, portanto, uma frase que se estende para além de uma sentença é improvável que seja significativa; ser uma frase frequente completa - mais longa possível, frases completas permitem descrições mais claras de *clusters*; não começar e nem terminar com uma *stop-word*.

Indução de rótulo de *cluster* (*Cluster-label induction*)

Durante a fase de indução de rótulo de *cluster*, Lingo identifica os conceitos abstratos que melhor descrevem a coleção de *snippets* de entrada e usa frases frequentes para construir uma representação legível desses conceitos. Isto produz um conjunto de rótulos, cada um dos quais determinará o conteúdo e descrição de um *cluster*.

Para ilustrar os conceitos apresentados nesta fase, exemplifica-se como Lingo manipula uma coleção exemplo de $d = 7$ títulos de documentos (Figura 17a), em que $t = 5$ termos (Figura 17b) e $p = 2$ frases (Figura 17c) aparecem mais do que uma vez e, portanto, são tratados como frequentes.

Primeiramente, Lingo constrói uma matriz termo-documento a partir dos *snippets* (fragmentos) de entrada. Em tal matriz, um vetor coluna representa cada trecho, e vetores de linha denotam termos selecionados para representar características dos documentos. No exemplo mostrado na Figura 18, a primeira linha representa a palavra "information", a segunda linha representa "singular", e assim por diante através dos termos na Figura 17b. Do mesmo modo, a coluna 1 representa D1 na figura Figura 17a, "large-scale singular value computations", a coluna 2 representa D2, "software for the sparse singular value decomposition", e assim por diante.

Nesta fase, desconsidera-se termos marcados como *stop-words* e termos que não aparecem mais do que um certo número de vezes (como

"software" na Figura 17a). Isto aumenta não só significativamente a eficiência de tempo do algoritmo, mas também reduz ruído entre os rótulos de *clusters*.

Usa-se o esquema de ponderação de peso de termos *tfidf* de Salton, Wong e Yang (1975) para eliminar vieses fortes. Calcula-se os valores na matriz A (Figura 17) usando *tfidf* e depois normaliza-se cada comprimento do vetor coluna.

Figura 17 - Dados de entrada na fase de indução de rótulo: (a) documentos, $d = 7$; (b) termos, $t = 5$; e (c) frases, $p = 2$

D1:	Large-scale <i>singular value computations</i>
D2:	Software for the sparse <i>singular value decomposition</i>
D3:	Introduction to modern <i>information retrieval</i>
D4:	Linear algebra for intelligent <i>information retrieval</i>
D5:	<i>Matrix computations</i>
D6:	<i>Singular value cryptogram analysis</i>
(a) D7:	Automatic <i>information organization</i>
T1:	Information
T2:	Singular
T3:	Value
T4:	Computations
(b) T5:	Retrieval
P1:	Singular value
(c) P2:	Information retrieval

Fonte: Osinski e Weiss (2005)

Figura 18 - Matriz termo-documento na fase de indução de rótulo

$$A = \begin{bmatrix} 0.00 & 0.00 & 0.56 & 0.56 & 0.00 & 0.00 & 1.00 \\ 0.49 & 0.71 & 0.00 & 0.00 & 0.00 & 0.71 & 0.00 \\ 0.49 & 0.71 & 0.00 & 0.00 & 0.00 & 0.71 & 0.00 \\ 0.72 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.83 & 0.83 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

Fonte: Osinski e Weiss (2005)

O processo de descobrir conceitos abstratos baseia-se no SVD da matriz termo-documento A , que a divide em três matrizes (U , S e V) tal que $A = USV^T$.

Uma das propriedades da *SVD* é que as primeiras r colunas de U (onde r é o ranking de A) formam uma base ortogonal para o espaço de termos da matriz de entrada. Em álgebra linear, vetores de base de espaço linear podem servir como blocos de construção para a criação de vetores sobre aquele espaço. Seguindo esta intuição, cada bloco de construção deve carregar uma das ideias referidas na coleção de entrada. Assim, para o Lingo, os vetores de base (isto é, as colunas do vetor em U) são exatamente o que ele configurou encontrar: uma representação vetorial dos conceitos abstratos dos *snippets*.

Na maioria das situações, considerando todos os r vetores base como conceitos abstratos resultaria em um número não gerenciável de *clusters* - normalmente próximo ao número de *snippets*. Portanto, Lingo utiliza os valores singulares da matriz A (encontram-se na diagonal da matriz S da SVD) para calcular quantas colunas de U devem avançar para a próxima fase do algoritmo.

Suponha que o cálculo resulte em $k = 2$, sendo definido como o número desejado de *clusters* do exemplo. Consequentemente, nos processamentos seguintes utiliza-se U_k que consiste nas primeiras k colunas de U (mostrada na Figura 19a).

Figura 19 - Matrizes na fase de indução de rótulo: (a) matriz de conceitos abstratos U , (b) matriz dos candidatos a rótulo de cluster P , e (c) matriz dos candidatos a rótulo de cluster (conceito abstrato) M .

$$\begin{array}{l}
 \text{(a)} \quad U = \begin{bmatrix} 0.00 & 0.75 & 0.00 & -0.66 & 0.00 \\ 0.65 & 0.00 & -0.28 & 0.00 & -0.71 \\ 0.65 & 0.00 & -0.28 & 0.00 & 0.71 \\ 0.39 & 0.00 & 0.92 & 0.00 & 0.00 \\ 0.00 & 0.66 & 0.00 & 0.75 & 0.00 \end{bmatrix} \\
 \\
 \text{(b)} \quad P = \begin{bmatrix} 0.00 & 0.56 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.71 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.71 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.83 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \\
 \\
 \text{(c)} \quad M = \bar{U}_k P = \begin{bmatrix} 0.92 & 0.00 & 0.00 & 0.65 & 0.65 & 0.39 & 0.00 \\ 0.00 & 0.97 & 0.75 & 0.00 & 0.00 & 0.00 & 0.66 \end{bmatrix}
 \end{array}$$

Fonte: Osinski e Weiss (2005)

Duas características dos vetores base são importantes aqui. Em primeiro lugar, os vetores são pares ortogonais, o que deve resultar em uma rica diversidade entre os conceitos abstratos descobertos. Em segundo lugar, os vetores base são expressos no espaço de termos da matriz A e assim, podem ser frases frequentes descobertas na fase anterior. Lingo usa a medida de distância clássica do cosseno para comparar os vetores e calcular qual frase ou palavra será a melhor representação verbal do conceito abstrato.

Considera-se termos individuais, nesta fase, porque é possível que nenhuma das frases frequentes descreva um conceito abstrato melhor do que um único termo.

Para cada frase frequente e um único termo frequente, Lingo cria um vetor coluna sobre o espaço de termos da matriz A. Quando montados conjuntamente, estes vetores formam uma matriz termo-documento P de frases e termos frequentes (Figura 19b). No exemplo, a coluna 1 corresponde a frase "*singular value*," a coluna dois corresponde a "*information retrieval*", a coluna 3 corresponde a "*information*" (termo único), assim por diante.

Assumindo vetores de coluna de ambas U e P normalizados pelo comprimento, como no exemplo, o problema do calcular a distância *cosseno* entre cada par conceito abstrato – frase-ou-termo fica reduzido a uma simples multiplicação de matriz.

As linhas da matriz resultante M (Figura 19c) representam conceitos abstratos, suas colunas representam frases e termos individuais, e os valores individuais são as similaridades calculadas através do cosseno. Assim, em uma única linha, o máximo componente

indica a frase ou termo único que melhor se aproxima do conceito abstrato correspondente.

No exemplo, o primeiro conceito abstrato está relacionado com a "*singular value*", enquanto o segundo está relacionado com "*information retrieval*". Como os valores na matriz M variam de 0,0 (nenhuma relação) a 1,0 (combinação perfeita), o Lingo também usa os componentes máximos componentes como pontuações rótulo de cluster.

Alocação do conteúdo do *cluster* (*Cluster-content allocation*)

O processo de alocação de conteúdo ao *cluster* assemelha-se a recuperação de documentos baseada em VSM, só que em vez de uma consulta, Lingo combina cada *snippet* (trecho) de entrada contra uma série de queries, cada uma das quais é um único rótulo de *cluster*. Assim, para um determinado rótulo de consulta, se a similaridade entre um *snippet* e o rótulo exceder um limiar pré-definido, Lingo aloca o *snippet* ao *cluster* correspondente.

As atribuições de valores limiares variam entre 0,0 e 1,0. Valores limiares mais elevados resultam em mais documentos sendo colocados em *clusters*, o que pode diminuir a precisão da atribuição. Valores menores conduzem a grupos menores com maior precisão de atribuição, mas com *recall* menor. A atribuição do valor limiar é uma questão de preferência do usuário. Verifica-se que os limiares entre 0,15 e 0,30 produzem os melhores resultados.

A última operação da fase de alocação é calcular a pontuação do grupo como um produto da pontuação do rótulo e do número de *snippets* do grupo. A Figura 20 mostra os resultados para o exemplo.

Figura 20 - Resultado final da alocação de conteúdo ao cluster¹⁶

¹⁶ Os números nos parênteses são os *scores* dos *clusters*.

Information retrieval (1.0)

D3: Introduction to modern *information retrieval*

D4: Linear algebra for intelligent *information retrieval*

(a) D7: Automatic *information organization*

Singular value (0.95)

D2: Software for the sparse *singular value* decomposition

D6: *Singular value* cryptograms analysis

(b) D1: Large-scale *singular value computations*

Other topics

(c) D5: *Matrix computations*

Fonte: Osinski e Weiss (2005)

2.5 TRABALHOS RELACIONADOS

Os trabalhos relacionados, que evidenciaram as lacunas e ajudaram a formar os pilares da problemática desta pesquisa, também buscaram soluções para esses problemas no contexto onde foram apresentados e segundo algum viés. Nesta seção, esses trabalhos são primeiramente classificados no Quadro 1 pelo tipo de problema que abordam, na sequência uma síntese de cada trabalho é apresentada, bem como são pontuadas algumas diferenças entre eles e o que é proposto nesta tese. Por fim, são apresentados trabalhos relacionados à proposição de um *framework* de modelagem de conhecimento.

Quadro 1 - Trabalhos relacionados classificados por tipo de problema

Trabalhos	Semântica da representação da rede	Conhecimento de ARS	Análise do conteúdo produzido pela rede
Mika (2005)	X		
Mika, Elfring e Groenewegen (2006)	X		
Erétéo et al. (2008, 2009a, 2009b e 2011)	X		
Abazi-Bexheti e Kadriu (2011)	X		
Balancieri (2010)		X	
Yang et al. (2012)			X
Xu. et al. (2012)			X

Fonte: Autor.

Mika (2005) e Mika, Elfring e Groenewegen (2006), apesar de não argumentarem especificamente sobre a falta de semântica na representação de uma rede, desenvolveram trabalhos pioneiros na área de inteligência social na web, envolvendo representação semântica integrada de dados de redes sociais conjuntamente com a utilização de métodos de ARS.

Para isso eles propuseram um *framework* para coleta, mineração e representação desses dados em uma estrutura unificada utilizando *Resource Description Framework* (RDF) e a ontologia *Friend-Of-A-Friend* (FOAF) (BRICKLEY e MILLER, 2014). Posteriormente, esses dados eram utilizados por bibliotecas e softwares específicos para representação visual e cálculo de medidas de ARS.

Erétéo et al. (2008, 2009a, 2009b e 2011) apresentaram uma proposição de representação de redes sociais utilizando *Resource Description Framework Schema* (RDFS), uma representação dos conceitos de ARS utilizando uma ontologia (SemSNA) em *Ontology Web Language* (OWL) e posterior análise da rede social através de algoritmos, implementados na linguagem de consulta SPARQL, que calculam as principais métricas de ARS.

Abazi-Bexheti e Kadriu (2011) apresentaram um *framework* para representar uma rede de coautoria através da extensão da ontologia FOAF em uma ontologia chamada CO-AUTHORONTO e a utilização de um sistema de regras em *Semantic Web Rule Language* (SWRL).

O *framework* desta tese se apropria do argumento da falta de semântica, mas apresenta uma proposição de representação de uma rede social mais específica relacionada a colaboração científica, através de uma ontologia de domínio em OWL onde grande parte dos conceitos relacionados ao contexto da colaboração científica são expressos se diferenciando assim das soluções de representação genéricas de Mika (2005) e Erétéo et al. (2008, 2009a, 2009b e 2011) e também da proposição de rede de colaboração científica de Abazi-Bexheti e Kadriu (2011).

Ao representar o domínio (conceitos e relacionamentos) de uma rede de colaboração científica por meio de uma ontologia de domínio com um maior número de conceitos e relacionamentos, proporciona-se uma representação semântica capaz de fornecer respostas mais completas e consequentemente um entendimento maior acerca do contexto modelado.

No trabalho de Balancieri (2010) foi proposto um método para explicitação de conhecimento advindo da análise de rede social de um domínio de aplicação. O método é composto por fases onde são construídas três ontologias para que os resultados da ARS sejam apresentados na linguagem do domínio de sua aplicação. A primeira fase culmina com a construção de uma ontologia para codificar o problema a ser tratado, denominada de ontologia de domínio. Na segunda fase é construída uma ontologia que codifica as medidas de ARS, denominada de ontologia de tarefa e na terceira fase é construída

uma ontologia que codifica as deduções que o especialista em ARS realiza quando descreve o significado das medidas no âmbito do domínio estudado, denominada de ontologia de aplicação.

No *framework* desta tese as tarefas de análise de colaboração, assim como os passos de inferência ou raciocínio (métodos) necessários para executar essas tarefas são representados por meio de ontologias de tarefa e de método, respectivamente. Essa abordagem difere do trabalho de Balancieri (2010), onde o modelo de aplicação proposto pelo autor armazena a descrição ou explicação dessas medidas, ou seja, não representa os passos de inferência ou raciocínio necessários para executar as tarefas de análise. Além disso, o que o autor denomina de modelo de tarefa e inclui uma ontologia de tarefa que armazena instâncias das medidas de ARS calculadas, é representado neste *framework* em uma ontologia de domínio de colaboração científica.

Em relação à análise do conteúdo ou dos artefatos originados da colaboração, como fonte adicional no processo de análise de rede de colaboração, tem-se os trabalhos de Yang et al. (2012) e Xu. et al. (2012) que analisaram a estrutura de uma rede conjuntamente com o conteúdo originado por seus membros, sendo essa última parte denominada pelos autores de análise semântica de conteúdo.

Em ambos os trabalhos, os autores utilizam o conteúdo produzido pelos atores para criar uma rede semântica de conceitos representada através de uma ontologia. Essa representação ontológica permite verificar a similaridade entre os perfis dos atores. Verificou-se nos trabalhos pesquisados que essa é a única representação ontológica utilizada pelos autores, ou seja, que a rede de colaboração em si não é representada com formalismos que permitam explorar a sua semântica.

Em relação a proposição de modelagem e desenvolvimento de artefatos de conhecimento combinado em um *framework*, que é o objetivo principal desta tese, além dos trabalhos já referenciados na fundamentação teórica que abordam partes do trabalho como o uso da modelagem de decomposição de tarefa e o uso de ontologias para representar os diferentes tipos de conhecimento, citam-se aqui dois trabalhos atuais que combinam esses conceitos na composição de um *framework*.

O'Connor et al (2009) desenvolveram um *framework* baseado em *PSM (Biological Spatio-Temporal Outbreak Reasoning Module - BioSTORM)* para detectar o aparecimento súbito de doenças infecciosas usando dados de vigilância de saúde pública de fontes diversas.

Este trabalho utilizou ontologias para modelar dados e conhecimento síndrômico e usou *PSMs* para definir métodos analíticos necessários para analisar os dados de vigilância. Os esforços de desenvolvimento foram divididos em processos de modelagem conceitual e desenvolvimento de software.

BioSTORM foi construído com quatro componentes principais: uma ontologia para descrever e integrar várias fontes de dados que podem servir como entrada para o sistema; uma biblioteca de *PSMs* baseada em conhecimento e estatística para analisar os dados de vigilância síndrômica; um componente de mediação com um *data broker* e um interpretador de mapeamento para traduzir os dados de entrada em uma forma usável pelos *PSM's* e um controlador para instalar configurações de *PSM* para analisar *streams* de dados. O *data broker* tem a função de integrar múltiplas fontes de dados descritas pela ontologia de fonte de dados e o interpretador de mapeamento alimenta os dados integrados para os *PSMs* apropriados.

O *framework* foi criado através da decomposição de tarefas de Chandrasekaran, na qual as tarefas gerais de vigilância foram modeladas como um conjunto de tarefas específicas e essas foram decompostas em subtarefas. Foi criada uma ontologia de *PSMs* de vigilância, que modela e classifica os métodos de vigilância que podem automatizar cada subtarefa de monitoramento. Com esta ontologia, a biblioteca de *PSMs* foi disponibilizada como um repositório de *PSMs* processáveis por computador que *BioSTORM* pode indexar, pesquisar e invocar. Além disso, cada *PSM* na biblioteca é associado com uma ontologia de métodos que define as classes de dados e conhecimento no qual o método opera. A ontologia de métodos torna explícito os requisitos de dados dos *PSMs*.

No trabalho de Gómez-Pérez e Corcho (2008) e Gómez-Pérez (2010) a modelagem e construção de *PSMs* é abordada no contexto da proveniência dos dados, que consiste em descrever e compreender onde e como dados são produzidos, os atores envolvidos na produção e os processos aplicados antes da chegada a coleção a partir da qual são acessados. Os autores propõem o uso de *PSMs* como um meio de reforçar a compreensão de especialistas no domínio na execução de processos em um ambiente de proveniência de dados orientado a conhecimento (*Knowledge-Oriented Provenance Environment - KOPE*).

Neste trabalho métodos de resolução de problemas (*PSMs*) foram usados como sobreposições semânticas que representam proveniência em múltiplos níveis de abstração. Essa abordagem facilita a compreensão do usuário de como informações de proveniência

relacionam-se a execução de um processo, simplifica a análise da execução do processo, mostrando a decomposição do processo em subprocessos em nível de domínio e oferece visualizações de execução do processo em vários níveis de detalhe.

Para isso KOPE utiliza os seguintes recursos de conhecimento para apoiar as interpretações por parte do usuário das informações de proveniência em diferentes níveis de abstração: um metamodelo de PSM descrevendo seus constructos e como eles se relacionam uns com os outros, uma biblioteca de PSMs contendo uma hierarquia de métodos e instâncias do metamodelo de PSM e uma ontologia de domínio que descreve o domínio da aplicação dos processos a serem analisados.

O trabalho apresentado nesta tese, buscou inspiração, principalmente na abordagem do trabalho de O'Connor et al (2009) acima citado, no sentido de que foi utilizada a mesma técnica de modelagem (*task-structure*) e as representações de tarefas e métodos se dão através de ontologias, mas ele apresenta uma proposição de modelagem que não pode ser comparada aos trabalhos por tratar de um domínio diferente que é a colaboração científica em rede.

3 METODOLOGIA

Este capítulo apresenta a metodologia que norteou o desenvolvimento do trabalho de pesquisa, mas inicialmente apresenta alguns conceitos fundamentais. Pesquisa, segundo Gil (2007, p. 17), é definida como:

“O procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa desenvolve-se por um processo constituído de várias fases, desde a formulação do problema até a apresentação e discussão dos resultados.”

É importante destacar a definição de metodologia proposta por Minayo (2007, p. 44):

“a) A apresentação adequada e justificada dos métodos, técnicas e dos instrumentos operativos que devem ser utilizados para as buscas relativas às indagações da investigação; b) a “criatividade do pesquisador”, ou seja, a sua marca pessoal e específica na forma de articular teoria, métodos, achados experimentais, observacionais ou de qualquer outro tipo específico de resposta às indagações específicas.”

Lacerda et al. (2013) complementa essa definição ao dizer que a metodologia se interessa pela validade do caminho escolhido para se chegar ao fim proposto pela pesquisa; e que não deve ser confundida com o conteúdo (teoria) nem com os procedimentos (métodos e técnicas).

Logo, em um trabalho de pesquisa, é necessário realizar e explicitar o enquadramento metodológico da pesquisa. Ainda segundo Lacerda et al. (2013), o enquadramento metodológico consiste em escolher e justificar um método de pesquisa que permita, principalmente:

“a) responder ao problema de pesquisa formulado; b) ser avaliado pela comunidade científica; c) evidenciar procedimentos que robusteçam os resultados da pesquisa, e, de forma tão importante quanto, não deve ser compreendido como um ato burocrático.”

Entretanto, antes de proceder ao enquadramento metodológico desta pesquisa, julga-se necessário apresentar um embasamento teórico que explicita a distinção entre conhecimento científico e conhecimento tecnológico e apresente a abordagem de pesquisa *Design Science*,

utilizada na condução dessa pesquisa. Optou-se por abordar os conceitos relacionados a *Design Science* com certo nível de detalhe porque isso facilitará o entendimento das escolhas metodológicas. A próxima seção apresenta esse embasamento e na sequência o enquadramento metodológico.

3.1 EMBASAMENTO TEÓRICO

As características do conhecimento tecnológico são apresentadas por Cupani (2006) que argumenta, com base em livro de Carl Mitcham, que a tecnologia pode ser abordada através de quatro perspectivas diferentes: a) como certo tipo de objeto (os artefatos); b) como uma classe específica de conhecimento (o saber tecnológico); c) como um conjunto de atividades para a produção e uso de artefatos; d) como uma manifestação de determinada vontade do ser humano em relação ao mundo.

Em relação a definição de tecnologia, Cupani (2006) explicita a de Mario Bunge em “*el estudio científico de lo artificial*”: “A tecnologia pode ser vista como o campo do conhecimento relativo ao projeto de artefatos e a planificação de sua realização, operação, ajuste, manutenção e monitoração, à luz do conhecimento científico.” (Bunge, 1985, p. 231).

Cupani (2006) argumenta que para compreender o conhecimento tecnológico, a primeira medida parece ser deixar de considerá-lo como uma aplicação do conhecimento científico. O autor pontua que, ao longo da história do desenvolvimento do conhecimento humano, muitas conquistas foram oriundas de descobrimentos científicos, com fortes bases teóricas elaboradas. Por outro lado, observaram-se inúmeros descobrimentos que não partiram de formulações analíticas, de bases teóricas, mas, de meras experimentações em que seus resultados positivos contribuíram para a formalização de certo tipo de conhecimento, como é o caso da consolidação do conhecimento em termodinâmica, que originou-se de uma invenção puramente tecnológica, a máquina a vapor.

A tecnologia é uma atividade dirigida a **produção** de algo novo e não ao descobrimento de algo existente. Assim o conhecimento tecnológico se configura como uma área *sui generis*, a das “ciências do artificial”, explorada por Herbert A. Simon em seu livro clássico “As Ciências do Artificial” (CUPANI, 2006).

O artificial constitui um sistema adaptado ao ambiente em função de determinado propósito humano, um objeto (artefato) com

propriedades desejadas, planejado e fabricado conforme um projeto (*design*). Por causa disso, pode ser chamado de “conhecimento prescritivo” em contraste com o conhecimento descritivo buscado pela ciência (SIMON, 1969, 1996)

Simon (1996) evidencia esse conceito com a distinção entre a ciência natural e artificial. A ciência natural se refere a um conjunto de conhecimentos sobre uma classe de objetos e/ou fenômenos do mundo (suas características, como se comportam e como interagem). Nesse sentido, é tarefa das disciplinas científicas naturais pesquisarem e ensinarem como as coisas são e como elas funcionam. Esse raciocínio pode ser aplicado para os fenômenos naturais (biologia, química, física) e sociais (economia, sociologia).

No contraponto, a “ciência do artificial” se ocupa da “concepção de artefatos que realizem objetivos” (SIMON, 1996, p. 198). Ou seja, as ciências do artificial dizem respeito a como as coisas devem ser para funcionar e atingir determinados objetivos. O autor argumenta que tem sido, justamente, a tarefa das escolas de engenharia, por exemplo, ensinar o que se refere ao artificial: como criar e projetar artefatos que tenham as propriedades desejadas e alcancem objetivos definidos.

Artificial é algo que foi produzido ou inventado pelo homem, sofrendo intervenções deste. Por consequência, máquinas, organizações, economia e aspectos da própria sociedade (como suas instituições) podem ser classificados como instâncias do artificial.

Simon (1996) também advoga pela necessidade de criar uma ciência (i.e., um corpo de conhecimento rigoroso e validado) que se dedique a propor como construir artefatos que possuam certas propriedades desejadas, isto é, como projetar esses artefatos e a denomina “Ciência do Projeto” ou *Design Science*. A autor argumenta que ao projeto interessa o quê e como as coisas devem ser, a concepção de artefatos que realizem objetivos.

Van Aken (2004) é outro proeminente autor que advoga a favor dessa abordagem. Para ele a missão principal da *Design Science* é desenvolver conhecimento para a concepção e desenvolvimento de artefatos. Em Costa e Silva (2014) pode ser encontrada uma compilação de autores que conceituam *Design Science* ou *Design Research*. Dessa compilação cabe destacar a de March e Smith (1995): “*Design Research* tenta criar coisas que servem a propósitos humanos e é orientado para a tecnologia e seus produtos são avaliados de acordo com critérios de valor ou de utilidade”.

Simon (1996) destaca ainda que, diferente das ciências naturais onde é comum descrever, explorar e explicar um fenômeno, artefatos

podem ser discutidos em termos descritivos, no que tange a comunicação e detalhamento dos principais componentes e informações sobre o artefato em si, e em termos imperativos no sentido de determinar as questões normativas que envolvem a construção a aplicação desse artefato.

Chakrabarti (2010) distingue *Design Science* e *Design Science Research*. Para o autor *Design Science* é a base epistemológica e *Design Science Research* é o método que operacionaliza a construção do conhecimento nesse contexto. Para Çağdaş e Stubkjær (2011) a *Design Science Research* se constitui em um processo rigoroso de projetar artefatos para resolver problemas, avaliar o que foi projetado ou o que está funcionando, e comunicar os resultados obtidos.

Segundo Lacerda et al (2013) a *Design Science/Design Science Research* tem amadurecido como abordagem principalmente na área de Tecnologia e Gestão da Informação (MARCH; SMITH, 2005; PEFFERS et al., 2008).

Na sequência julga-se importante entender os principais elementos constituintes da *Design Science* e *Design Science Research*: “classes de problemas” e os “artefatos”.

De acordo com Van Aken (2004) as classes de problemas permitem que os artefatos e, por consequência, suas soluções, não sejam apenas uma resposta pontual a certo problema em determinado contexto. A *Design Science* não se preocupa com a ação em si, mas com o conhecimento que pode ser utilizado para projetar as soluções. Assim, o conhecimento na *Design Science* é generalizável quando válido para uma dada classe de casos.

Lacerda et al (2013) destacam que não há uma definição conceitual de “classe de problemas” ou uma proposição para sua construção. Os autores então definem “classe de problemas” como: “A organização de um conjunto de problemas, práticos ou teóricos, que contenha artefatos avaliados, ou não, úteis para a ação nas organizações”.

Os autores propõem uma lógica sequencial para se chegar até a classe de problemas:

- 1) Conscientização: a partir de um problema, teórico ou prático, identificado, é necessário identificar quais objetivos ou metas seriam necessários para que o problema seja considerado satisfatoriamente resolvido;

- 2) Revisão sistemática na literatura: depois da conscientização é preciso estabelecer o quadro de soluções empíricas conhecidas. A

revisão consiste na busca e identificação dos artefatos que propõem soluções ao problema. É necessário buscar publicações que consolidassem as classes de problemas, os artefatos testados e suas soluções.

Depois de caracterizar classe de problemas, é necessário caracterizar os artefatos associados. Para SIMON, (1996, p. 28) artefatos são “objetos artificiais que podem ser caracterizados em termos de objetivos, funções e adaptações. São normalmente discutidos, particularmente durante a concepção, tanto em termos imperativos como descritivos”.

Uma vez definidos os artefatos, pode-se tipificá-los. Artefatos podem ser definidos como: Constructo, Modelo, Método e Instanciação (MARCH; SMITH, 1995). O Quadro 2 apresenta a definição de cada tipo de artefato.

Quadro 2 - Tipos de Artefatos

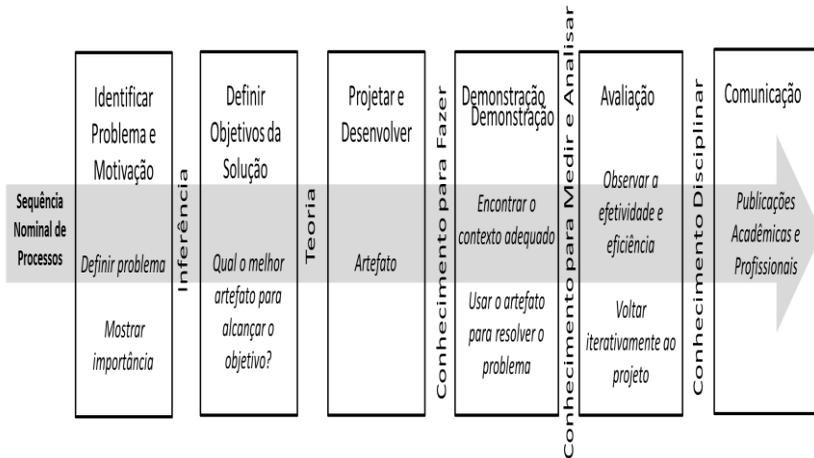
Tipos de Artefato	Constructo	Constructos ou conceitos formam o vocabulário de um domínio. Eles constituem uma conceituação utilizada para descrever os problemas dentro do domínio e para especificar as respectivas soluções. Conceituações são extremamente importantes em ambas as ciências, natural e de <i>design</i> . Eles definem os termos usados para descrever e pensar sobre as tarefas.
	Modelo	Um modelo é um conjunto de proposições ou declarações que expressam relacionamentos entre constructos. Modelos podem ser vistos como uma descrição ou uma representação de como as coisas são. Em atividades de design, modelos representam situações como problema e solução. Na <i>Design Science</i> , a preocupação é a utilidade de modelos. A relevância de um modelo é expressa pela sua utilidade e as suas imprecisões e abstrações são irrelevantes para

		esse propósito.
	Métodos	Um método é um conjunto de passos (um algoritmo ou orientação) usado para executar uma tarefa. Métodos baseiam-se em um conjunto de constructos subjacentes (linguagem) e uma representação (modelo) em um espaço de solução. Os métodos podem ser ligados aos modelos, nos quais as etapas do método podem utilizar partes do modelo como uma entrada que o compõe. Além disso, os métodos são, muitas vezes, utilizados para traduzir um modelo ou representação em um curso para resolução de um problema. Os métodos são criações típicas das pesquisas em <i>Design Science</i> .
	Instanciações	Uma instanciação é a concretização de um artefato em seu ambiente. Instanciações operacionalizam constructos, modelos e métodos. As instanciações informam como implementar ou utilizar determinado artefato e seus possíveis resultados. As instanciações podem se referir a um determinado artefato ou à articulação de diversos artefatos para a produção de um resultado em um contexto. Instanciações demonstram a viabilidade e a eficácia dos modelos e métodos que elas contemplam.

Fonte: Extraído de March e Smith (1995, p. 257-258).

Em relação à condução da *Design Science Research*, existem várias proposições na literatura e uma síntese gráfica delas pode ser encontrada em Lacerda et al (2013). Aqui apresenta-se a proposição de Peffers et al. (2007), utilizada como suporte metodológico deste trabalho de pesquisa, composta de seis passos ou atividades que são exibidos na Figura 21 e descritos a seguir:

Figura 21 - Processo de *Design Science Research Methodology (DSRM)*



Fonte: Adaptado de Peffers et al. (2007)

- **Identificar o problema e sua motivação:** define-se o problema de pesquisa específico e justifica-se o valor de sua solução;
- **Definir os objetivos para uma solução:** inferem-se os objetivos da solução proposta a partir da definição do problema e do conhecimento do que é viável e exequível;
- **Projetar e Desenvolver:** cria-se o artefato, que pode ser um constructo, modelo, método ou uma instanciação. Nesta etapa diferentes abordagens podem ser utilizadas, tais como: algoritmos computacionais, representações gráficas, protótipos, maquetes em escala, entre outros. De acordo com Lacerda et al (2013) o principal resultado desta etapa é o artefato em estado funcional, o conhecimento gerado, embora aplicado pontualmente na solução de problemas específicos, ou no desenvolvimento de novos artefatos, deve ser generalizável para o que foi definido como “Classe de Problemas”. Esta generalização permite a construção de um

conhecimento útil – no sentido pragmático – não pontualmente, mas que possa ser ampla e efetivamente aplicável pela/para a sociedade (VENABLE, 2006; MARCH; SMITH, 1995; VAISHNAVI; KUECHLER, 2009).

- **Demonstrar:** demonstra-se o uso do artefato, resolvendo uma ou mais instâncias do problema por meio de um experimento, simulação, estudo de caso, prova formal ou outra atividade apropriada;
- **Avaliar:** observa-se e mensura-se como o artefato atende à solução do problema, comparando-se, por meio de métricas e técnicas de análises, os objetivos que foram propostos para a solução, com os resultados observados na utilização do artefato, durante a sua etapa de demonstração. De acordo com Lacerda et al (2013) nesta etapa deve ser definida um processo rigoroso de verificação do comportamento do artefato no ambiente para o qual foi projetado, em relação às soluções que ele se propôs alcançar. Assim, é necessário: i) explicitar o ambiente interno, o ambiente externo e os objetivos clara e precisamente; ii) explicitar como o artefato pode ser testado; e iii) descrever os mecanismos que medem os resultados.
- **Comunicar:** divulga-se o problema e sua relevância, o artefato concebido, sua utilidade e ineditismo, o rigor do projeto e a sua efetividade, para outros pesquisadores e outras audiências quando isso for apropriado.

De acordo com March e Smith (1995) a *Design Science* é composta pela construção e pela avaliação. A construção é o processo de confecção de artefatos para um propósito específico, enquanto a avaliação é a verificação do desempenho dos artefatos como solução desejada.

Para estes mesmo autores a construção deve ser julgada com base no valor ou utilidade do artefato para uma comunidade de usuários. Os modelos devem ser avaliados em termos de sua fidelidade com os fenômenos do mundo real, completude, nível de detalhe, robustez e consistência interna. As instanciações devem ser avaliadas considerando a eficácia e eficiência do artefato e seus impactos sobre o ambiente e seus usuários.

Ao abordar a avaliação e validade na *Design Science*, Lacerda et al (2013) destacam que, para aumentar a confiabilidade nos resultados da pesquisa, é necessário um conjunto de cuidados e procedimentos

rigorosos que minimizem o viés nos resultados obtidos e/ou nas soluções geradas.

Os autores, apoiados em Hevner, March e Park (2004, p. 86), apresentam alguns métodos que podem ser utilizados para avaliar os artefatos, entre eles, a forma de avaliação “Descritiva”, com duas possibilidades de métodos: “Argumento informado”, que utiliza a informação das bases de conhecimento (por exemplo, das pesquisas relevantes) para construir um argumento convincente a respeito da utilidade do artefato; “Cenários”, que consiste em construir cenários detalhados em torno do artefato, para demonstrar sua utilidade.

Ainda segundo Lacerda et al (2013) outra forma de avaliação dos artefatos desenvolvidos poderia ser a utilização de *Focus Group* ou Grupo Focal. Grupos Focais podem ser utilizados como método de avaliação para a *Design Science Research*, uma vez que garantem uma discussão mais profunda e colaborativa em relação aos artefatos desenvolvidos pela pesquisa. Auxiliam, ainda, na realização da análise crítica dos resultados obtidos durante a pesquisa e que podem fazer brotar novas possibilidades com o intuito de encontrar melhores soluções para os problemas em estudo.

Dois tipos de Grupos Focais são apresentados por Tremblay, Hevner e Berndt (2010) e podem ser utilizados para a avaliação dos artefatos desenvolvidos pela *Design Science Research*. Estes tipos, bem como suas principais características estão explicitadas no Quadro 3.

Quadro 3 - Tipos de Grupos Focais em *Design Science Research*

Características	Grupo Focal Exploratório	Grupo Focal Confirmatório
Objetivo	Alcançar melhorias incrementais rápidas na criação de artefatos.	Demonstrar a utilidade dos artefatos desenvolvidos no campo de aplicação.
Papel do Grupo Focal	Fornecimento de informações que possam ser utilizadas para eventuais mudanças tanto no artefato, como no roteiro do Grupo Focal. Refinamento do roteiro do Grupo	O roteiro de entrevistas previamente definido para ser aplicado ao grupo de trabalho não deve ser modificado ao longo do tempo a fim de garantir a possibilidade de se fazer comparativos entre cada grupo focal participante.

	Focal e identificação de constructos a serem utilizados em outros grupos.	
--	---	--

Fonte: Adaptado de Tremblay, Hevner e Berndt (2010).

Em relação a validade, Chakrabarti (2010) afirma que é um fator determinante para amparar a pesquisa, com o intuito de facilitar a utilização ou aplicação da pesquisa no campo prático. De acordo com Lacerda et al (2013) em *Design Science Research*, se compreende como fonte de validade um conjunto de procedimentos para garantir que os resultados gerados pelo artefato provêm do ambiente interno projetado e o ambiente externo em que foi preparado para operar. Para isso é necessário: i) Explicitar o ambiente interno, o ambiente externo e os objetivos clara e precisamente; ii) Explicitar como o artefato pode ser testado; iii) Descrever os mecanismos que gerarão os resultados a serem controlados/acompanhados.

3.2 ENQUADRAMENTO METODOLÓGICO

De acordo com a explicitação das características do conhecimento/pesquisa científica e conhecimento/pesquisa tecnológico apresentados na seção anterior é possível classificar a pesquisa que constitui esta tese, como sendo uma pesquisa tecnológica. Esta classificação está embasada no objetivo principal deste trabalho que é disponibilizar um conjunto de artefatos combinados em um *framework* para análise de rede de colaboração científica.

Em função da ênfase na construção de artefatos, esta pesquisa teve sua condução apoiada na abordagem metodológica *Design Science Research* (DSR), mais especificamente na proposição de condução de Peffers et al. (2008), apresentada na seção anterior.

De acordo com a proposição da DSR, a primeira atividade consistiu em descobrir as classes de problemas da pesquisa e os tipos de artefatos associados para resolver esses problemas. Essa atividade está diluída nas duas primeiras etapas do processo geral de condução da pesquisa baseado na proposição de Peffers et al. (2008) e exibido na Figura 22.

Figura 22 - Processo de condução da pesquisa

<p>1. Identificar o problema e sua motivação</p> <ul style="list-style-type: none"> • Como representar o conhecimento de domínio e o conhecimento das tarefas de análise de uma rede colaboração científica de modo que esta representação possa auxiliar o processo de análise da referida rede?
<p>2. Definir os objetivos para uma solução</p> <ul style="list-style-type: none"> • O1: Propor e desenvolver um conjunto de representações dos conhecimentos envolvidos na análise de uma rede de colaboração científica, bem como um conjunto de artefatos para armazenar e recuperar dados dessas representações combinados em um <i>framework</i>. • O2: Avaliar a contribuição do <i>framework</i> no processo de análise de uma rede de colaboração científica, no que tange a relevância das tarefas de análise propostas, aos resultados fornecidos e a sua aplicabilidade como instrumento de gestão.
<p>3. Projetar e Desenvolver</p> <ul style="list-style-type: none"> • <i>Framework</i> de conhecimento (conhecimento de domínio, tarefa e solução de tarefa) para análise de rede de colaboração científica.
<p>4. Demonstrar</p> <ul style="list-style-type: none"> • Demonstração do <i>framework</i> em dois cenários distintos: Programa de Pós-Graduação interdisciplinar e disciplinar.
<p>5. Avaliação</p> <ul style="list-style-type: none"> • Avaliação do <i>framework através de grupos focais</i>
<p>6. Comunicar</p> <ul style="list-style-type: none"> • Apresentação dos resultados

Fonte: Autor.

O processo de descoberta das classes e artefatos começou com uma etapa denominada de Conscientização, que demandou uma revisão de literatura dos tópicos Colaboração Científica, Análise de Redes Sociais, Redes de Coautoria, Tecnologias Semânticas, e Sistemas Baseados em Conhecimento e possibilitou a definição inicial do problema de pesquisa e de algumas formas preliminares de solucionar esse problema.

Logo a seguir aconteceu a etapa de Revisão Sistemática de Literatura, que abordou com mais profundidade e afinou os resultados da etapa anterior, conduzindo efetivamente a descoberta das classes de problemas e dos tipos de artefatos associados que solucionariam os problemas. Para a revisão foram utilizadas bases de dados de artigos científicos, tais como *Scopus*, *Web of Science*, *ScienceDirect* e *SciELO*. No Apêndice 1 é apresentada uma síntese da revisão de literatura deste trabalho em cada um dos referidos tópicos.

Assim, chegou-se as classes de problemas e tipos de artefatos apresentados no Quadro 4. Cabe destacar que os modelos descritos são combinados nesta pesquisa no que denomina-se *framework*. O último tipo de artefato do Quadro 4 refere-se a instanciação dos modelos acima descritos.

Quadro 4 - Classes de Problemas e Tipos de Artefatos

Classe de Problemas	Tipo de Artefatos
Representação de uma rede colaboração científica	Modelo de Conhecimento de Domínio
Entendimento das métricas de Análise de Redes Sociais	Modelo de Conhecimento de Tarefa e Solução de Tarefa
Análise do conteúdo como complemento a análise estrutural da rede	Modelo de Conhecimento de Domínio, Tarefa e Solução de Tarefa
	Instanciação dos modelos

Fonte: Autor.

Com este processo, o problema e os objetivos foram definidos, conforme explicitado na Figura 22.

A terceira etapa da condução do processo refere-se ao Projeto e Desenvolvimento dos artefatos. A especificação detalhada desta etapa encontra-se no Capítulo 4 deste trabalho.

Na quarta etapa os artefatos (modelos combinados e instanciados em um *framework*) foram demonstrados em um ambiente com as características condizentes ao ambiente onde o problema se manifesta, ou seja, em um ambiente de P&D onde a análise da colaboração científica é importante.

O ambiente de P&D escolhido foi um Programa de Pós-Graduação (PPG) de uma Universidade Federal. Foram escolhidos dois Programas de Pós-Graduação através da técnica de amostragem não probabilística por julgamento. Esta técnica permite que o pesquisador

use seu julgamento para selecionar casos que habilitem uma melhor resposta para a questão de pesquisa. É usada frequentemente quando se trabalha com amostras pequenas, tais como o estudo de caso, e quando se deseja selecionar casos que são particularmente informativos (SAUNDERS; LEWIS; THORNHILL, 2009).

Neste caso, os critérios de julgamento levaram em consideração a natureza (disciplinar ou interdisciplinar) dos Programas, o interesse e disponibilidade dos gestores dos mesmos em responder o questionário e posterior entrevista de coleta de necessidades de análise de colaboração científica aplicados na etapa 3 de desenvolvimento dos artefatos (detalhada no Capítulo 4) e principalmente o sentimento de obter informações valiosas dos gestores desses programas, o que é, de acordo com Saunders, Lewis e Thornhill (2009), geralmente a tônica desse tipo de escolha.

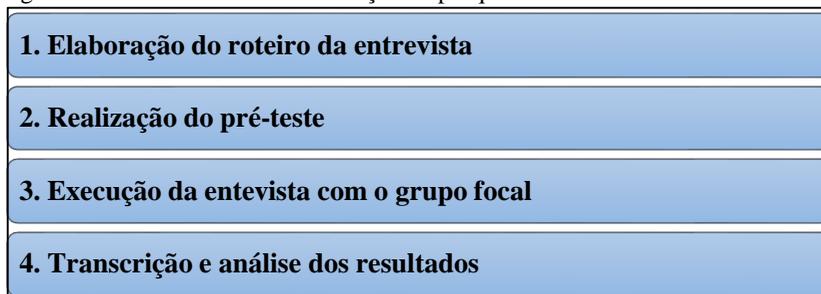
A etapa de demonstração consistiu em apresentar para os gestores (coordenador e subcoordenador) de cada PPG o *framework* em operação respondendo as questões de análise de colaboração científica do Programa. O detalhamento do processo de demonstração do *framework* é explicitado no Capítulo 5.

Para a etapa de avaliação, realizada de forma concomitante com a demonstração, foi utilizada a técnica de coleta de dados por meio de grupos focais, que consiste em uma entrevista com duas ou mais pessoas, onde um tópico é definido de forma precisa e clara e há um foco em habilitar e gravar a discussão entre os participantes (SAUNDERS; LEWIS; THORNHILL, 2009). Os procedimentos de avaliação da pesquisa são detalhados na subseção seguinte.

3.2.1 Procedimentos de Avaliação da Pesquisa

Esta seção explicita o objetivo da avaliação, bem como os procedimentos de avaliação da pesquisa. A pesquisa foi avaliada com o objetivo de verificar a contribuição do *framework* no processo de análise de uma rede de colaboração científica. A contribuição foi materializada em três dimensões denominadas: Colaboração Científica, Resultados do *framework* e Aplicabilidade do *framework* como instrumento de gestão. Os procedimentos de condução da avaliação são exibidos na Figura 23 e explicados na sequência.

Figura 23 - Procedimentos da avaliação da pesquisa.



Fonte: Autor.

Primeiramente foi desenvolvido um instrumento de avaliação ou roteiro de entrevista composto pelas três dimensões citadas anteriormente. O instrumento está disponível no Apêndice B.

A dimensão Colaboração Científica objetivou colher a percepção da importância da colaboração científica e da análise dessa colaboração como geradora de insumos para uma possível tomada de decisão no contexto analisado e a relevância das perguntas/tarefas de análise propostas no *framework*.

A dimensão Resultados do *framework* buscou avaliar o grau de confiança dos resultados das perguntas/tarefas de análise de colaboração científica do *framework*. Para isso foi proposto que os resultados de cada uma das 18 perguntas/tarefas que compõem o *framework* fossem analisados e após a análise, os resultados de cada pergunta/tarefa fossem avaliados em relação ao critério grau de confiança, escolhendo uma, entre três opções: alto, médio ou baixo.

A dimensão Aplicabilidade do *framework* como instrumento de gestão buscou avaliar a aplicabilidade do mesmo como instrumento de gestão. Para isso foram propostas possíveis ações de tomada de decisão em nível de gestão para cada uma das 18 perguntas/tarefas de análise que compõem o *framework*. Além disso, nesta dimensão o instrumento permitiu a proposição de novas ações para cada uma das perguntas/tarefas.

O segundo procedimento foi realizar o pré-teste do instrumento em uma entrevista. O pré-teste teve o objetivo de verificar o tempo de condução da entrevista, estimado em duas horas, bem como verificar o entendimento das questões propostas no roteiro da entrevista com o grupo focal. Foram escolhidos 2 entrevistados com experiência em

gestão de P&D, um deles na coordenação de Programa de Pós-Graduação e o outro na gestão de uma Fundação de Amparo à Pesquisa. Com o pré-teste, pôde-se confirmar o tempo estimado; reorganizar a forma de avaliar a dimensão Resultados e Aplicabilidade, de forma que as duas fossem avaliadas na sequência para cada pergunta/tarefa do *framework*; e em função da experiência em gestão dos entrevistados, coletar novas possíveis ações de gestão para a versão final do instrumento.

A execução da entrevista com os grupos focais, gestores de PPG, aconteceu em momentos diferentes, previamente agendados e devidamente consentidos (o modelo do termo de consentimento encontra-se no Apêndice C). As entrevistas foram gravadas e demandaram tempos diferenciados. Para o PPG de natureza interdisciplinar, o tempo demandado extrapolou o estimado, levando em torno de 3 horas de duração. Para o outro PPG, a execução ficou dentro do tempo estimado.

As respostas foram transcritas e analisadas na sequência e os resultados encontram-se no Capítulo 5.

4 **FRAMEWORK DE CONHECIMENTO**

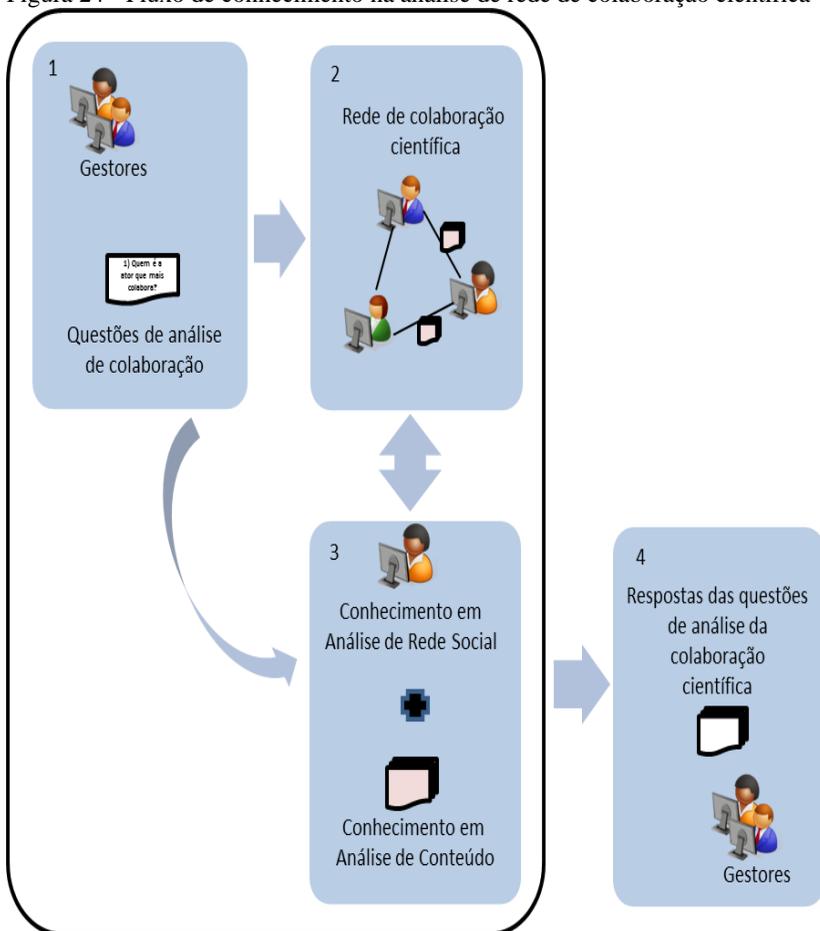
A definição de *framework* dada por Tomhave (2005) diz que “um *framework* é uma construção fundamental que define os pressupostos, conceitos, valores e práticas, e que inclui uma orientação para a implementação do mesmo”, o que indica uma orientação menos conceitual e abstrata e mais relacionada com um trabalho demonstrável. Ainda tem-se a definição encontrada no *Merriam Webster Dictionary* que define *framework* como “uma estrutura básica de alguma coisa: um conjunto de ideias ou fatos que provem suporte para alguma coisa” e a de *Dictionary.com* com “uma estrutura composta de partes encaixadas e unidas”.

Este trabalho de pesquisa vai ao encontro dessas definições e foi caracterizado como um *framework*, pois se vale de pressupostos, conceitos, práticas, inclui orientações para ser implementado e é composto por partes bem definidas que se complementam em prol do suporte a uma situação (colaboração científica), o que será demonstrado no decorrer deste capítulo.

O *framework* de conhecimento para análise da colaboração científica que constitui esta tese está alicerçado na modelagem e representação dos conhecimentos necessários para tal análise. A representação de conhecimento tem seus fundamentos na Inteligência Artificial e posterior evolução na área de Engenharia do Conhecimento, conjuntamente com a ideia de modelagem de conhecimento que busca separar o conhecimento de domínio do conhecimento de como resolver os problemas desse domínio.

Uma parte essencial do *framework* é o resultado de um processo de modelagem e representação de conhecimentos necessários para a análise de uma rede de colaboração científica através de métricas de análise de rede social e através da análise do conteúdo dos artefatos produzido pelas atividades de colaboração. A Figura 24 apresenta o fluxo de conhecimento existente nessa análise em um contexto de P&D, tal como uma universidade.

Figura 24 - Fluxo de conhecimento na análise de rede de colaboração científica



Fonte: Autor

O fluxo começa com os gestores ou comunidade acadêmica demandando questões ou tendo necessidades de análise de colaboração científica (quadro 1). Essas questões devem ser respondidas analisando-se uma rede de colaboração, formada por pesquisadores (professores e alunos) que originam artefatos frutos dessa colaboração (quadro 2). As questões demandam o conhecimento em análise de redes sociais, bem como o conhecimento em análise do conteúdo dos artefatos (produções bibliográficas, técnicas, artísticas, projetos, etc.) produzidos pelas

colaborações (quadro 3). Por sua vez, as respostas às questões devem ser fornecidas aos demandantes das mesmas (quadro 4).

A partir do fluxo de conhecimento explicitado na Figura 24 e dentro de uma perspectiva de modelagem de conhecimento é possível identificar o conhecimento de domínio, representado no quadro 2 pela rede de colaboração científica, o conhecimento de como resolver problemas nesse domínio, representado pelo conjunto de questões ou tarefas de análise (quadro 1) que exigem conhecimentos distintos em análise de rede social e análise do conteúdo dos artefatos (quadro 3) e que para serem operacionalizadas dependem do conhecimento de domínio.

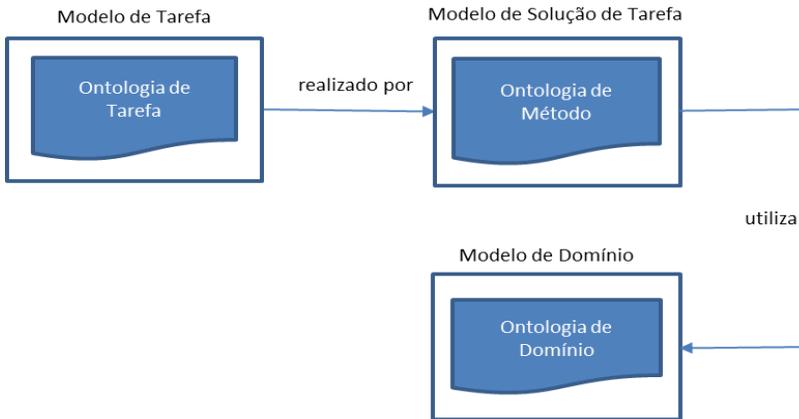
A partir dessa abstração, propõe-se neste trabalho, que o conhecimento de como resolver problemas seja desmembrado e modelado em dois tipos de conhecimentos distintos: o conhecimento que representa as necessidades ou tarefas de análise de colaboração e o conhecimento de como realizar essas tarefas.

Assim têm-se três tipos de conhecimentos distintos: conhecimento que representa as necessidades ou tarefas de análise de colaboração representado no quadro 1, o conhecimento de como realizar essas tarefas, representado no quadro 3 e o conhecimento do domínio no qual as análises serão feitas, representado no quadro 2.

O *framework* é composto por modelos que representam esses conhecimentos. No modelo de tarefa estão representadas as questões, necessidades ou tarefas de análise de colaboração científica, representando portanto as demandas de análise. O modelo de solução de tarefas representa os processos de inferência ou métodos necessários para realizar as tarefas de análise. O modelo de domínio representa os conceitos do domínio da colaboração científica, bem como os conceitos do domínio das métricas de ARS.

Todos os modelos são representados através de ontologias, respectivamente, ontologia de tarefa, ontologia de métodos e ontologia de domínio. A Figura 25 apresenta o relacionamento dos modelos de conhecimento do *framework*.

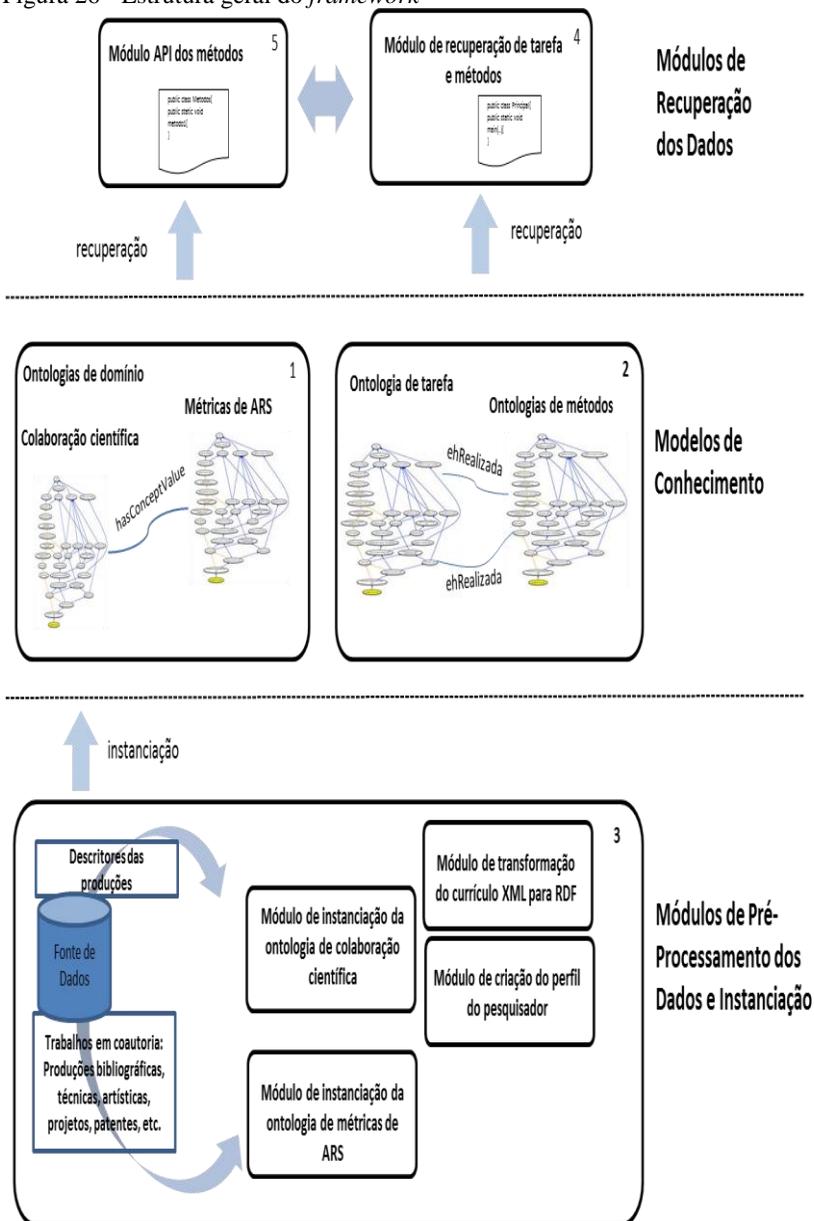
Figura 25 - Organização dos modelos de representação de conhecimento do *framework*



Fonte: Autor

Os modelos de conhecimento e suas representações caracterizam uma camada fundamental do *framework*. Entretanto, a sua estrutura é composta por mais duas camadas, compostas por elementos, que combinados caracterizam o *framework* na sua totalidade. A Figura 26 apresenta as três camadas onde todos os elementos do *framework* estão dispostos. Eles são brevemente explicados na sequência e detalhados no restante deste capítulo.

Figura 26 - Estrutura geral do *framework*



Fonte: Autor.

Na camada Modelos de Conhecimento tem-se todos os modelos de conhecimento representados através de ontologias.

O conhecimento de domínio (quadro 1) é representado por duas ontologias: a ontologia de Colaboração Científica que representa conceitos e relacionamentos desse domínio, tais como professores, alunos, publicações, etc. e a ontologia de métricas de Análise de Rede Social (ARS) que representa os conceitos das métricas de ARS. Essas ontologias são interligadas por propriedades de objetos que unem os conceitos (instâncias) de atores da Colaboração Científica, com conceitos (instâncias) de ARS.

O conhecimento de tarefa (quadro 2) é representado através de uma ontologia que define tarefas (questões ou necessidades) de análise de colaboração científica. O conhecimento de solução de tarefa é representado por uma ontologia de métodos que define os métodos que realizam as tarefas representadas na ontologia de tarefa. Tarefas da ontologia de tarefas são realizadas por métodos representados na ontologia de métodos, os quais são modelados observando-se a assinatura do método, isto é, os parâmetros e o controle da ordem destes parâmetros.

Os modelos de conhecimento, através de suas ontologias, precisam ser instanciados, ou seja, é necessário que sejam criadas instâncias ou indivíduos para os conceitos representados. A camada de Pré-processamento e instanciação (quadro 3) determina os requisitos de dados, orienta como fazer as instanciações e oferece alguns módulos desenvolvidos para realizar essa atividade.

Nos modelos de tarefa e solução de tarefa e suas respectivas ontologias de tarefa e método, as instâncias já fazem parte do *framework*, visto que o mesmo foi modelado até o momento para um conjunto específico de tarefas. Entretanto, as ontologias de tarefa e método podem ser estendidas e nesse caso, tanto a estrutura das ontologias quanto as instancias devem ser criadas por um engenheiro do conhecimento.

Por outro lado, as ontologias de Colaboração Científica e métricas de ARS, que representam o modelo de domínio, devem ser instanciadas com dados de um contexto de colaboração científica. Esses dados precisam ser coletados e processados antes da instanciação. Os módulos que compõem o processo de preparação dos dados e instanciação dessas ontologias estão caracterizados no quadro 3 da estrutura do *framework*.

Na camada Recuperação dos dados estão os módulos de recuperação dos dados das ontologias de domínio, tarefa e métodos.

No quadro 4 está a implementação do módulo responsável pela recuperação nas ontologias de tarefa e métodos, das instâncias relacionadas a tarefa de análise escolhida pelo usuário do *framework*. Para cada tarefa escolhida, existirá uma instância de tarefa e uma ou mais instâncias de métodos complexos, subtarefas, métodos primitivos e parâmetros associados. A configuração, que inclui a ordem, de todas essas instâncias de tipos diferentes determinam, conseqüentemente, quais métodos primitivos da API de métodos (quadro 5) serão invocados e em qual ordem.

Para que o usuário escolha uma tarefa de análise, recomenda-se o desenvolvimento uma aplicação *front-end*. Essa aplicação é independente do *framework*.

No quadro 5 está a implementação dos métodos primitivos do *framework*. Os métodos representados na ontologia de método estão implementados conforme a sua representação em termos de parâmetros e ordem de parâmetros. Essa implementação forma a API - *Application Programming Interface* - dos métodos do *framework*.

A seguir cada camada que compõem o *framework* é detalhada. Inicia-se pela camada dos Modelos de conhecimento, onde os aspectos metodológicos de construção dos modelos de Tarefa, Solução de Tarefa e Domínio são apresentados, logo seguem-se as camadas dos Módulos de Pré-Processamento e Instanciação das ontologias. A última camada apresentada refere-se aos módulos de Recuperação de Dados das ontologias.

4.1 MODELO DE TAREFA E MODELO DE SOLUÇÃO DE TAREFA

O modelo de tarefa representa as tarefas de análise da colaboração científica. Essas tarefas são consideradas intensivas em conhecimento porque demandam processos de raciocínio ou inferência por parte do especialista que as executa. Os processos ou passos de inferência necessários para cumprir as tarefas são representados pelo modelo de solução de tarefa. A separação do modelo de tarefa (representado pela ontologia de tarefas) do modelo de solução de tarefas (representado pela ontologia de métodos) aumenta as possibilidades de reuso de ambos.

Na metodologia de construção desses modelos, primeiramente foram identificadas as necessidades ou tarefas de análise de colaboração

científica que são respondidas através de métricas de análise de rede social e análise do conteúdo dos artefatos.

Posteriormente, as tarefas foram modeladas através de uma metodologia analítica de tarefa que permitiu a decomposição das tarefas em métodos e subtarefas e forneceu o insumo para a representação das tarefas e da solução das tarefas através da ontologia de tarefa e da ontologia de método, respectivamente. A seguir cada um desses passos é detalhado.

4.1.1 Identificação das Tarefas de Análise de Colaboração Científica

A metodologia de identificação das necessidades ou tarefas de análise de colaboração envolveu uma concepção inicial de 18 perguntas relacionadas a análise da colaboração científica em um contexto de P&D. Toda concepção das perguntas foi norteada pela revisão de literatura referente a aplicação do método de ARS na análise de redes de colaboração científica.

Essa concepção inicial levou em consideração alguns critérios:

- Primeiro, o método que responderia essas perguntas, ou seja, as perguntas deveriam ser respondidas através de métricas de ARS e através da análise do conteúdo dos artefatos produzidos pela colaboração;

- Segundo, a grau de relevância das questões para quem iria utilizá-las, por exemplo, gestores, pesquisadores do contexto de P&D, etc.;

- Terceiro, as questões foram elaboradas com objetivo de serem claras o suficiente para o usuário soubesse exatamente qual o resultado seria obtido com as mesmas.

Após a concepção inicial, foi realizado um procedimento para verificar a contribuição e grau de relevâncias das 18 questões concebidas junto a potenciais interessados em analisar um contexto de colaboração científica. Este procedimento de verificação ocorreu por meio de um instrumento de coleta de dados do tipo questionário eletrônico e de uma entrevista. No instrumento questionário, disponível no Apêndice D, para cada questão haviam três opções de resposta: muito relevante, relevante e pouco relevante, além de um espaço no final para inserir sugestões.

A amostra escolhida para esse procedimento foi constituída por coordenadores e subcoordenadores de Programas de Pós-Graduação (PPG), incluindo a Pró-reitora de Pós-Graduação, de uma Universidade Federal, totalizando 117 pessoas.

O procedimento de verificação foi feito em três etapas: primeiramente foi realizado um pré-teste com 5 respondentes com o objetivo de verificar o entendimento das questões e a adequação do próprio instrumento de coleta como um todo. Posteriormente foi enviado por e-mail, o questionário para o restante da amostra e finalmente, após o final do período de resposta, foram escolhidas 3 pessoas da amostra para uma entrevista, com o objetivo de verificar face a face o entendimento e relevância das questões.

Como resultado desse procedimento de verificação obteve-se as respostas de 23 gestores de PPG. Pode-se verificar a partir das respostas que a maioria das questões de análise de colaboração proposta é muito relevante ou relevante. A opção Não relevante obteve um total de 68 respostas. A compilação das respostas por pergunta encontra-se no Apêndice E.

O Quadro 5 apresenta o conjunto de demandas, perguntas ou tarefas de análise de colaboração científica que fundamentam este *framework*. Neste quadro, apresenta-se também o método e conseqüentemente, o conhecimento necessário para responder tais perguntas. As perguntas 1, 2, 3, 4, 5, 6 e 7 são respondidas através de métricas de ARS. Da pergunta 9 à pergunta 15 o método baseia-se na mineração do conteúdo dos artefatos produzidos pela colaboração e as perguntas 16, 17 e 18 baseiam-se nas métricas de ARS e na mineração do conteúdo.

Quadro 5 - Conjunto de perguntas/tarefas de análise de colaboração científica

Pergunta/Tarefa	Método que responde a pergunta/tarefa
1. Quem são os pesquisadores que mais colaboram?	Métrica de ARS: Centralidade de Grau
2. Quem são os pesquisadores que menos colaboram?	Métrica de ARS: Centralidade de Grau
3. Quem são os pesquisadores mais próximos aos demais pesquisadores?	Métrica de ARS: Centralidade de Proximidade
4. Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram?	Métrica de ARS: Centralidade de Intermediação
5. Quais os grupos de pesquisadores que trabalham isoladamente?	Métrica de ARS: Componente
6. Qual o nível geral de atividade de colaboração da rede?	Métrica de ARS: Densidade
7. Qual o nível médio de colaboração dos pesquisadores da rede?	Métrica de ARS: Grau médio
8. Quem são os colaboradores do pesquisador “x”?	Nodos adjacentes a um nodo.
9. Quem são os pesquisadores que colaboram no tópico “y”?	Palavras-chave das produções do pesquisador.
10. Quem é o pesquisador que mais colabora no tópico “y”?	Palavras-chave das produções do pesquisador
11. Quais grupos de pesquisadores que colaboram no tópico “y”?	Palavras-chave das produções do pesquisador

12. Qual o perfil das colaborações do pesquisador “x”?	Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador.
13. Quem são os pesquisadores que possuem perfis similares a um perfil específico?	Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador.
14. Qual o grau de similaridade dos perfis de colaboração de dois pesquisadores?	Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador.
15. Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador “x”?	Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador.
16. Quem são os pesquisadores que, com base em seus perfis e na proximidade com o pesquisador “x”, podem colaborar com esse pesquisador?	Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador. Proximidade refere-se a todos os nodos com distância = 2.

17. Qual o perfil dos pesquisadores que mais colaboram?	Métrica de ARS: Centralidade de Grau. Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador.
18. Qual o perfil dos grupos que trabalham isoladamente?	Métrica de ARS: Componente. Palavras-chave das produções do pesquisador, acrescido dos termos extraídos das produções do pesquisador.

Fonte: Autor.

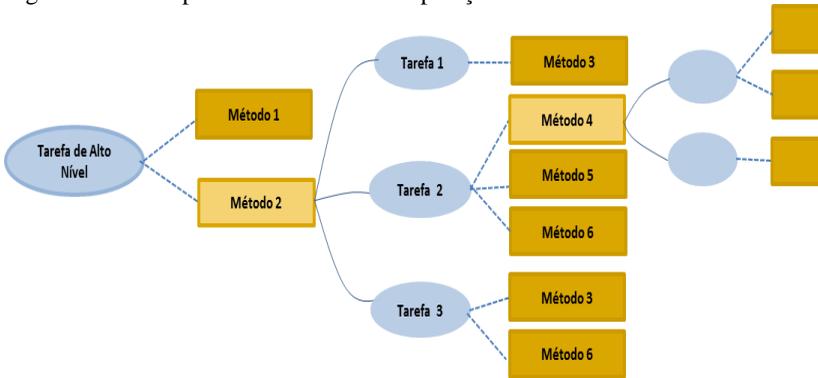
4.1.2 Modelagem das Tarefas de Análise de Colaboração Científica

A partir da identificação das tarefas de análise, foi utilizada uma abordagem de modelagem de conhecimento orientada a objetivo ou tarefa, proposta por Chandrasekaran, Johnson e Smith (1992) que baseia-se na noção da estrutura de tarefa. Essa abordagem consiste na decomposição de uma tarefa em passos (métodos) que cumprem a tarefa, permitindo assim que se identifique o que deve ser feito e como deve ser feito. Segundo os autores diversas abordagens de modelagem de conhecimento compartilham a visão de que um sistema deve ser modelado em relação a um objetivo ou tarefa.

Na metodologia analítica de tarefa, previamente apresentada no Capítulo 2, seção 2.3.4, uma tarefa define "o que tem que ser feito". Uma tarefa é realizada por um ou mais métodos, que definem "como executar uma tarefa". Um método pode executar uma tarefa diretamente (método primitivo) ou decompor uma tarefa em subtarefas (método complexo ou método de decomposição de tarefas ou *task-decomposition method - TDM*) com restrições à sua ordem de execução e delegando sua execução a outros métodos. Esse processo de decomposição recursiva cria uma estrutura de árvore que representa o processo modelado, como pode ser observado na Figura 27.

Tarefas (mostradas como elipses) são realizadas através da aplicação de métodos (retângulos) e pode existir mais de um método para cada tarefa. Os métodos podem ser primitivos (retângulos escuros) ou complexos (retângulos claros). Métodos complexos, também chamados de métodos de decomposição de tarefa, quebram uma tarefa em subtarefas. As linhas sólidas no gráfico são lidas como "método decompõe uma tarefa em" (relacionamento *AND*). As linhas tracejadas conectam tarefas com seus métodos (relacionamento *OR*).

Figura 27 - Exemplo de árvore de decomposição de tarefa



Fonte: Adaptado de Buckeridge et al. (2008)

Uma vez que é possível existir muitas maneiras de realizar uma tarefa, pode haver múltiplos métodos elegíveis para decompor a mesma tarefa em diferentes conjuntos de subtarefas. Além disso, um único método pode ser reutilizado para realizar várias tarefas.

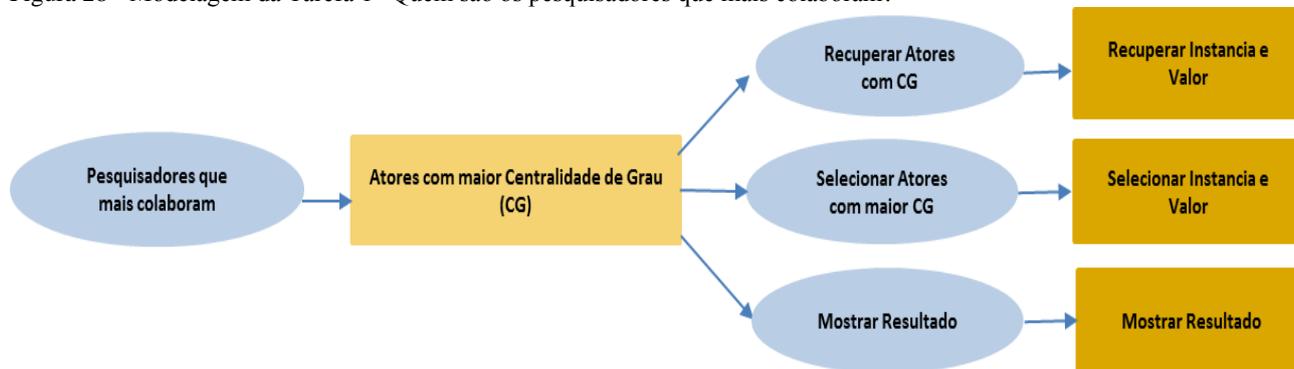
A metodologia analítica de tarefa é adequada para essa modelagem, primeiramente porque permite identificar claramente o “quê” e o “como” fazer, separando os papéis de conhecimento e possibilitando a criação dos modelos de conhecimento de tarefa (o quê) e de solução de tarefa (como). Além disso, ela esclarece os papéis dos diferentes métodos e explícita as semelhanças e diferenças estruturais entre os métodos (algoritmos). Identifica, por exemplo, quando um único método pode ser usado para várias tarefas.

A modelagem de algumas perguntas ou tarefas de análise de colaboração são apresentadas a seguir através da

Figura 28, Figura 29, Figura 30, Figura 31 e Figura 32 e seus respectivos quadros explicativos. O objetivo é apresentar algumas características da modelagem nesta seção. A modelagem de todas as demais perguntas encontra-se no Apêndice F.

A maioria das tarefas modeladas são realizadas por métodos complexos (retângulo claro), que por sua vez são decompostos em subtarefas (elipse) que são realizadas por métodos primitivos (retângulo escuro). Para cada tarefa modelada é apresentado um quadro que descreve textualmente seus métodos complexos, subtarefas e métodos primitivos. Os métodos primitivos são especificados em relação aos parâmetros que cada método espera.

Figura 28 - Modelagem da Tarefa 1 “Quem são os pesquisadores que mais colaboram? ”



Fonte: Autor.

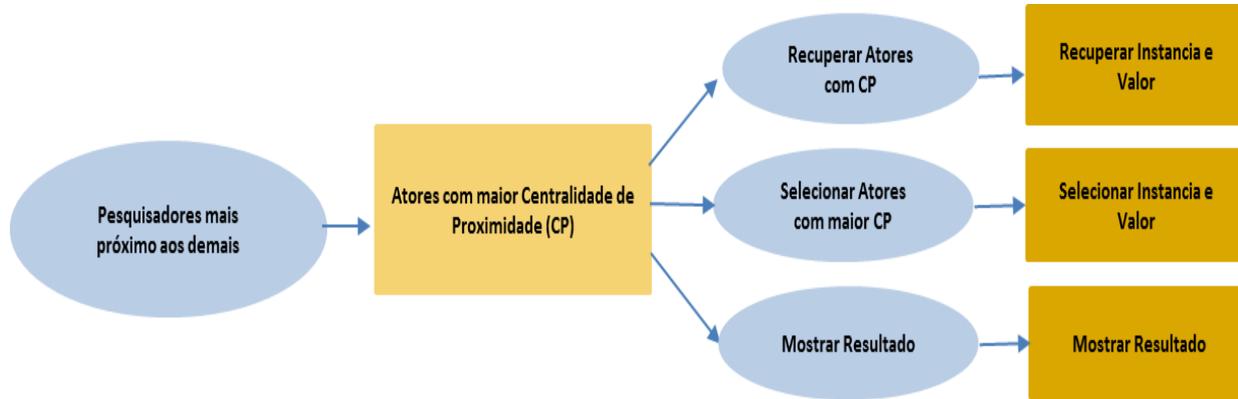
Quadro 6 - Quadro explicativo da modelagem da Tarefa 1

Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Atores com maior Centralidade de Grau (CG)	1- Recuperar atores com CG	<p>Recuperar Instância e Valor</p> <p>recuperarInstanciaValor (propriedade de objeto, propriedade de dados): lista de instâncias</p> <p>Descrição: Esse método recebe o nome de uma propriedade de objeto e de uma propriedade de</p>

		dados e retorna a lista de instâncias relacionadas com a propriedade de objeto, conjuntamente com o valor da propriedade de dados de cada instância.
	2- Selecionar atores com maior CG	<p>Selecionar Instancia e Valor</p> <p>selecionarInstanciaValor (lista de instâncias, filtro ordem, filtro limite): lista de instâncias</p> <p>Descrição: Esse método seleciona instâncias do parâmetro lista de instâncias através do parâmetro filtro ordem, que determinada se a ordem é ascendente ou descendente, e do parâmetro filtro limite que determinada o número de instâncias selecionadas.</p>
	3 – Mostrar resultado	<p>Mostrar resultado</p> <p>mostrarResultado (lista de instâncias)</p> <p>Descrição: Esse método exibe a propriedade nome das instâncias da lista de instâncias.</p>

Fonte: Autor.

Figura 29 - Modelagem da Tarefa 3 “Quem são os pesquisadores mais próximos aos demais pesquisadores?”



Fonte: Autor

Quadro 7 - Quadro explicativo da modelagem da Tarefa 3

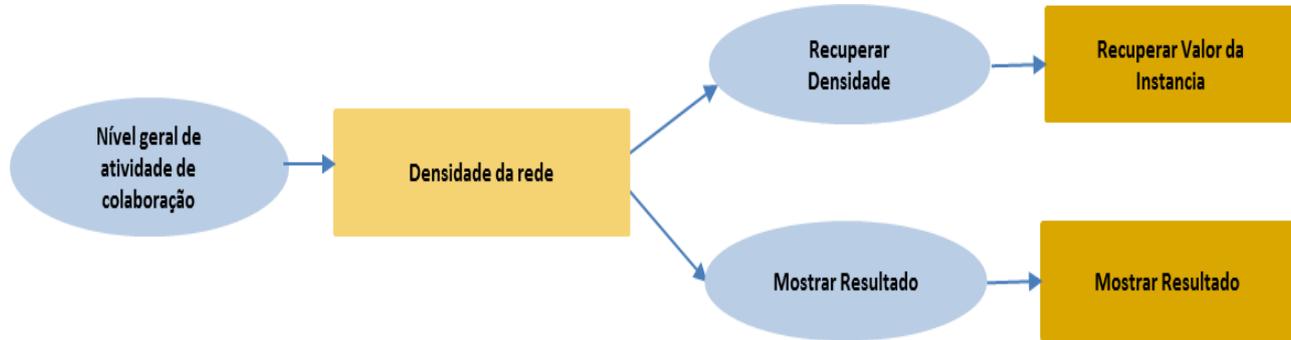
Método complexo	Subtarefas (passos de inferência)	Especificação de Método primitivo
Atores com maior Centralidade de Proximidade (CP)	1- Recuperar atores com CP	Recuperar Instância e Valor (Especificado anteriormente)

	2- Selecionar atores com maior CP	Selecionar Instancia e Valor (Especificado anteriormente)
	3 – Mostrar resultados	Mostrar resultado (Especificado anteriormente)

Fonte: Autor.

Na modelagem apresentada na Figura 28 e Figura 29 acima, é importante observar que as tarefas são diferentes, mas compartilham uma subtarefa e todos os métodos primitivos (Recuperar Instância Valor, Selecionar Instância Valor e Mostrar Resultado) e por isso eles não são especificados novamente. O compartilhamento de componentes da modelagem, como subtarefas e métodos primitivos é a essência dessa abordagem de modelagem. Isso fica mais evidente nas demais tarefas modeladas no Apêndice F.

Figura 30 - Modelagem da Tarefa 6 “Qual o nível geral de atividade de colaboração da rede?”



Fonte: Autor.

Quadro 8 - Quadro explicativo da modelagem da Tarefa 6

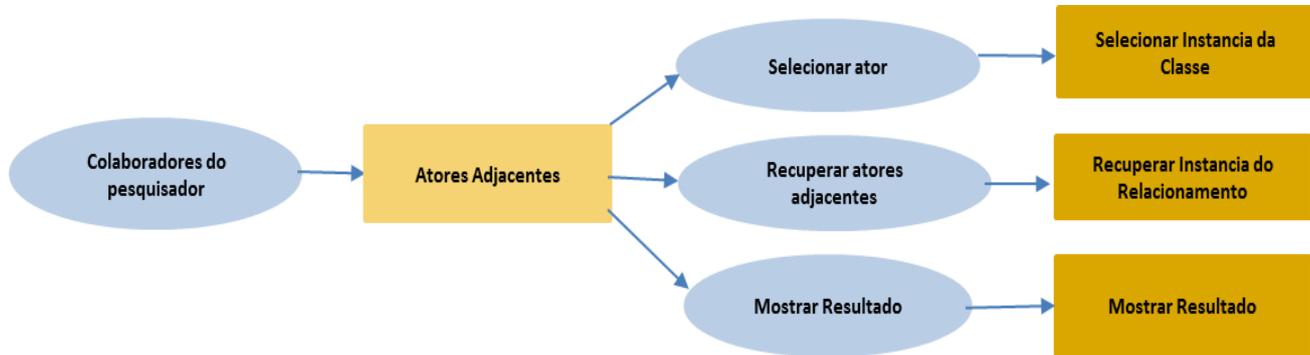
Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Densidade da rede	1 – Recuperar Densidade	<p>Recuperar Valor da Instancia</p> <p>recuperarValorInstancia (classe, propriedade de dados): lista de valores</p> <p>Descrição: Esse método recebe como</p>

		parâmetro o nome de uma classe e de uma propriedade de dados e retorna uma lista de valores das instâncias da classe.
	2 – Mostrar resultado	<p>Mostrar resultado</p> <p>mostrarResultado (lista de instâncias)</p> <p>Descrição: Esse método exibe a propriedade valor da instância da lista de instâncias.</p>

Fonte: Autor.

Na modelagem apresentada na Figura 30 acima e na Figura 31 abaixo, aparecem métodos primitivos diferentes dos que foram apresentados. Destaca-se que o nome dos métodos primitivos remete a ideia do que fazem. No caso do método Recuperar Valor da Instancia, este recupera valores de uma propriedade de dados de instâncias de uma classe e por causa disso o método necessita desses parâmetros. O método Selecionar Instancia da Classe, seleciona instâncias de uma classe com base no parâmetro de entrada recebido. Esse último método e o método Recuperar Instância do Relacionamento também estão presentes na modelagem apresentada na Figura 32.

Figura 31 - Modelagem da Tarefa 8 “Quem são os colaboradores do pesquisador “x”?”



Fonte: Autor.

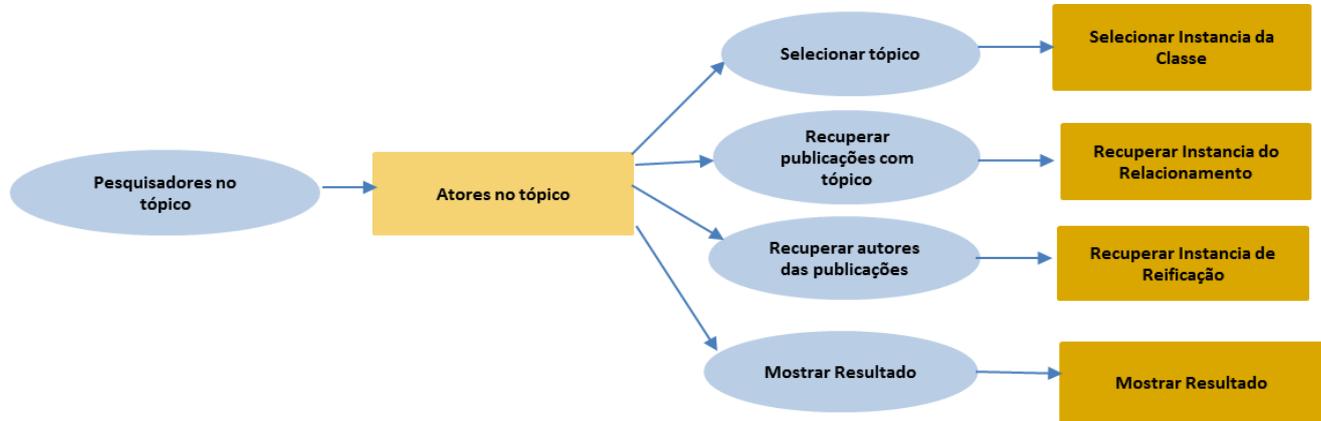
Quadro 9 - Quadro explicativo da modelagem da Tarefa 8

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Atores Adjacentes	1-Selecionar ator	<p>Selecionar Instancia da Classe</p> <p>selecionarInstanciadaClasseVivo (dado de entrada): lista de instâncias.</p> <p>Descrição: Esse método seleciona e retorna uma instancia correspondente ao dado de entrada (nome do pesquisador). Este método</p>

		é específico para a ontologia de domínio usada no <i>framework</i> .
	2- Recuperar atores adjacentes	<p>Recuperar Instancia do Relacionamento</p> <p>recuperarInstanciadoRelacionamento (Lista de instâncias, propriedade de objeto, valor lógico – V ou F): lista de instâncias.</p> <p>Descrição: Esse método retorna uma lista de instâncias que relacionam com o parâmetro lista de instâncias através do parâmetro propriedade de objeto. O parâmetro valor lógico indica se as instâncias da lista de parâmetro devem ser incluídas ou não na lista de instâncias de retorno.</p>
	3 – Mostrar resultados	<p>Mostrar resultados</p> <p>mostrarResultados (lista de instâncias)</p> <p>Descrição: Esse método exibe a propriedade nome das instâncias da lista de instâncias.</p>

Fonte: Autor.

Figura 32 - Modelagem da Tarefa 9 “Quem são os pesquisadores que colaboram no tópicó “x”?”



Fonte: Autor.

Quadro 10 - Quadro explicativo da modelagem da Tarefa 9

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Atores no tópicó	1-Selecionar tópicó	Selecionar Instancia da Classe selecionarInstanciadaClasseVivo (dado de entrada): lista de instâncias.

		<p>Descrição: Esse método seleciona e retorna uma instancia correspondente ao dado de entrada (nome do pesquisador). Este método é específico para a ontologia de domínio usada no <i>framework</i>.</p>
	<p>2- Recuperar publicações com tópicos</p>	<p>Recuperar Instancia do Relacionamento</p> <p>recuperarInstanciadoRelacionamento (Lista de instâncias, propriedade de objeto, valor lógico – V ou F): lista de instâncias.</p> <p>Descrição: Esse método retorna uma lista de instâncias que relacionam com o parâmetro lista de instâncias através do parâmetro propriedade de objeto. O parâmetro valor lógico indica se as instâncias da lista de parâmetro devem ser incluídas ou não na lista de instâncias de retorno.</p>

	3 – Recuperar autores das publicações	<p>Recuperar Instancia da Reificação</p> <p>recuperarInstanciaReificacao (lista de instâncias, classe, propriedade de objeto, classe, propriedade de objeto): lista de instâncias</p> <p>Descrição: Esse método retorna uma lista de instâncias que se relacionam via reificação com o parâmetro lista de instâncias. Para isso são necessárias as classes e as propriedades de objeto dos relacionamentos.</p>
	4 – Mostrar resultados	<p>Mostrar resultados</p> <p>mostrarResultados (lista de instâncias)</p> <p>Descrição: Esse método exibe as propriedades, por exemplo, nome, das instâncias da lista de instâncias.</p>

4.1.3 Representação do Modelo de Tarefa e Solução de Tarefa

A modelagem analítica de tarefa ajuda a separar a noção de tarefa do método que executa a tarefa e com isso fornece o insumo para representar esses tipos de conhecimentos diferentes. Dessa forma, o modelo de tarefa é representado através de uma ontologia de Tarefas e o modelo de solução de tarefa através de ontologias de Métodos.

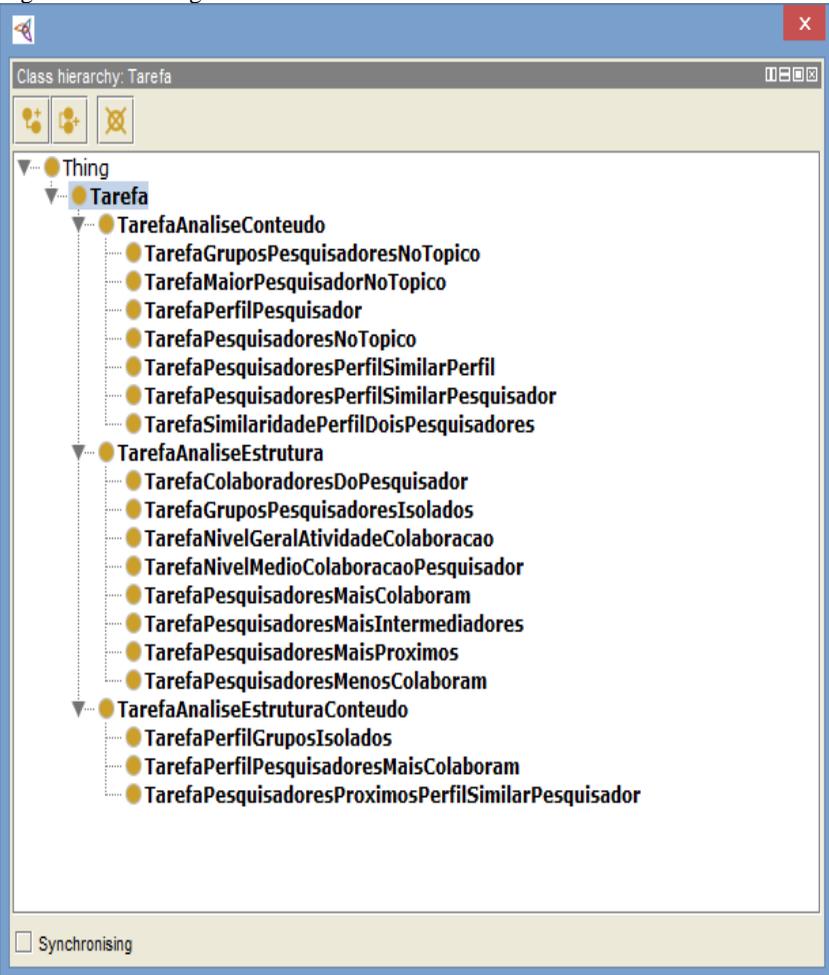
As ontologias foram criadas a partir da decomposição das tarefas de análise de colaboração científica em métodos complexos, que foram decompostos em subtarefas e foram realizadas por métodos primitivos (abordagem da modelagem analítica de tarefa), apresentada na seção anterior sendo representadas em OWL. A especificação dos parâmetros dos métodos primitivos apresentados nos quadros explicativos também serviram de insumo para a representação dos mesmos na ontologia de Métodos. Para a criação das ontologias foi utilizada a ferramenta Protégé®.

a) **Ontologia de Tarefas:** A hierarquia de classes da ontologia de Tarefas, apresentada na Figura 33 através de um recorte da ferramenta Protégé®, foi criada a partir das tarefas de nível mais alto (representadas na forma de elipse) de cada pergunta ou tarefa de análise de colaboração científica.

A hierarquia possui três níveis, o primeiro nível representa o conceito geral de Tarefa, o segundo nível representa o conceito da categoria de tarefa, que pode ser de três tipos:

- a. Tarefa de análise da estrutura da rede (**TarefaAnáliseEstrutura**): representam as tarefas que são respondidas somente por métricas de ARS. Sob este nível estão 8 (oito) tarefas do terceiro nível da hierarquia.
- b. Tarefa de análise do conteúdo dos artefatos da colaboração (**TarefaAnáliseConteúdo**): representam as tarefas que são respondidas somente pela análise do conteúdo dos artefatos da colaboração. Sob este nível estão 7 (sete) tarefas do terceiro nível da hierarquia.
- c. Tarefa de análise da estrutura e de conteúdo (**TarefaAnáliseEstruturaConteúdo**): representam as tarefas que são respondidas de forma mista, pela análise da estrutura e pela análise do conteúdo dos artefatos da colaboração. Sob este nível estão 3 (três) tarefas do terceiro nível da hierarquia.

Figura 33 - Ontologia de Tarefas



Fonte: Autor.

A Figura 34 exibe a única propriedade de objeto (*object property*) desta ontologia: **ehRealizada** (*isPerformed*). É através dessa propriedade que uma tarefa se liga a um método complexo que a realiza, representado na ontologia de Métodos.

Figura 34 – Propriedade de objeto da Ontologia de Tarefa



Fonte: Autor.

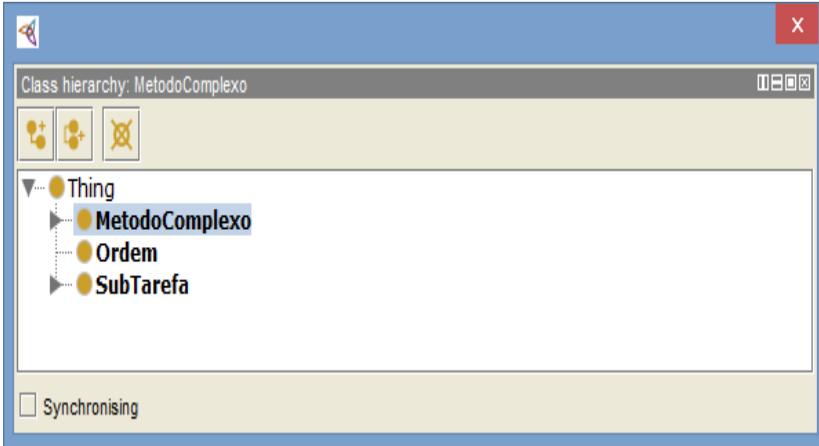
b) **Ontologia de Métodos:** A ontologia de métodos está dividida em duas ontologias, uma que representa os métodos complexos (retângulo claro) que se relacionam com subtarefas (elipse) e outra ontologia que representa os métodos primitivos (retângulo escuro) que se relacionam com parâmetros identificados na modelagem analítica de tarefa das perguntas ou tarefas de colaboração.

A ontologia de Métodos Complexos, apresentada na Figura 35,

Figura 36 e Figura 37, representa, além dos conceitos de **MetodoComplexo** e **Subtarefa**, o conceito de **Ordem**. Este conceito permite modelar a ordem de invocação de subtarefas.

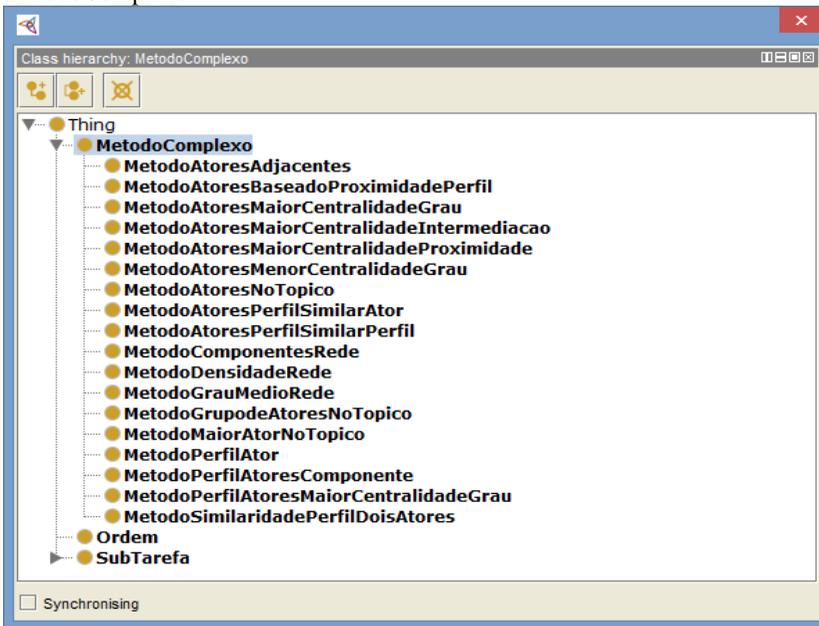
Nesta ontologia, instâncias de **MetodoComplexo** se relacionam com instâncias de **Subtarefa** e a ordem do relacionamento é estabelecida pelas instâncias do tipo **Ordem**. De maneira mais específica: uma instância de **MetodoComplexo** se relaciona através da propriedade de objeto (*object property*) **temOrdem** (*hasOrder*) com uma instância de **Ordem**, que por sua vez, se relaciona através da propriedade de objeto **ordemPara** (*orderFor*) com uma instância de **SubTarefa**. Instâncias de **Ordem** possuem a propriedade de dados (*data property*) **temValor** (*hasValue*) que armazena a ordem de invocação das instâncias de **SubTarefa**.

Figura 35 - Ontologia de Métodos Complexos – Visão Geral



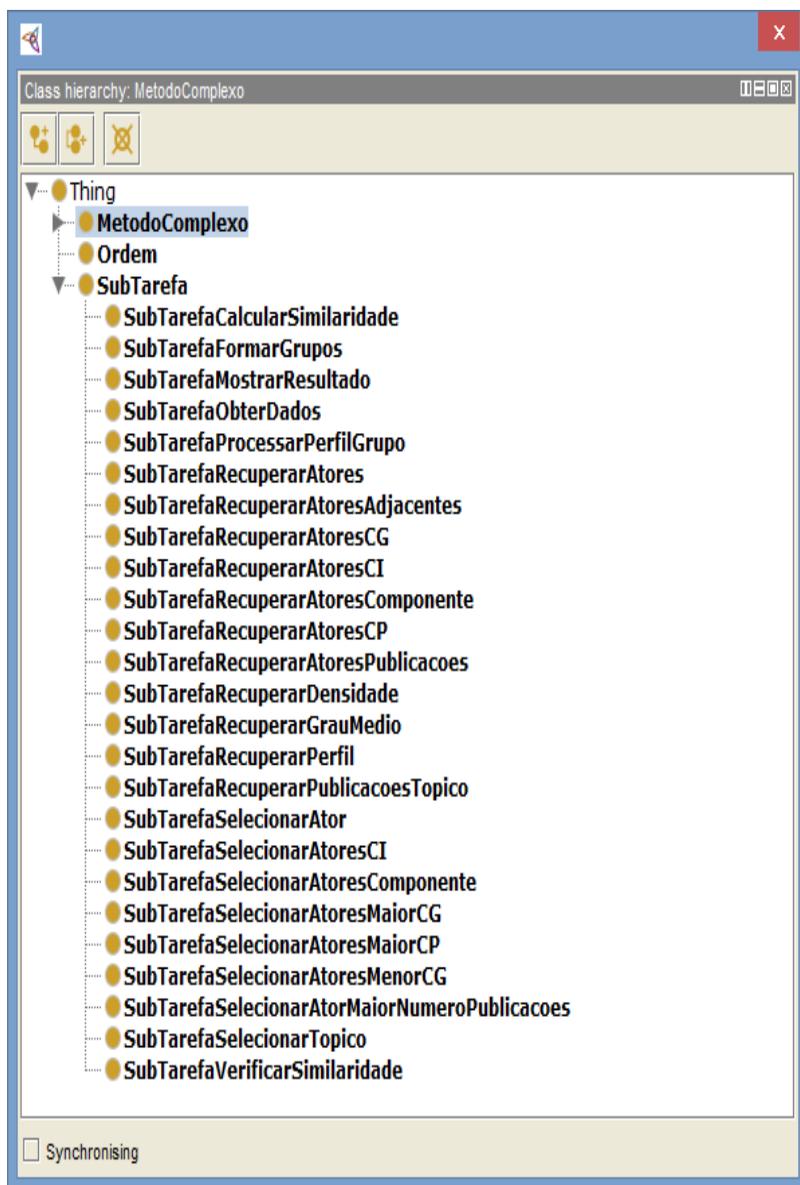
Fonte: Autor.

Figura 36 - Ontologia de Métodos Complexos – Subclasses de MetodoComplexo



Fonte: Autor.

Figura 37 - Ontologia de Métodos Complexos – Subclasses de Subtarefa

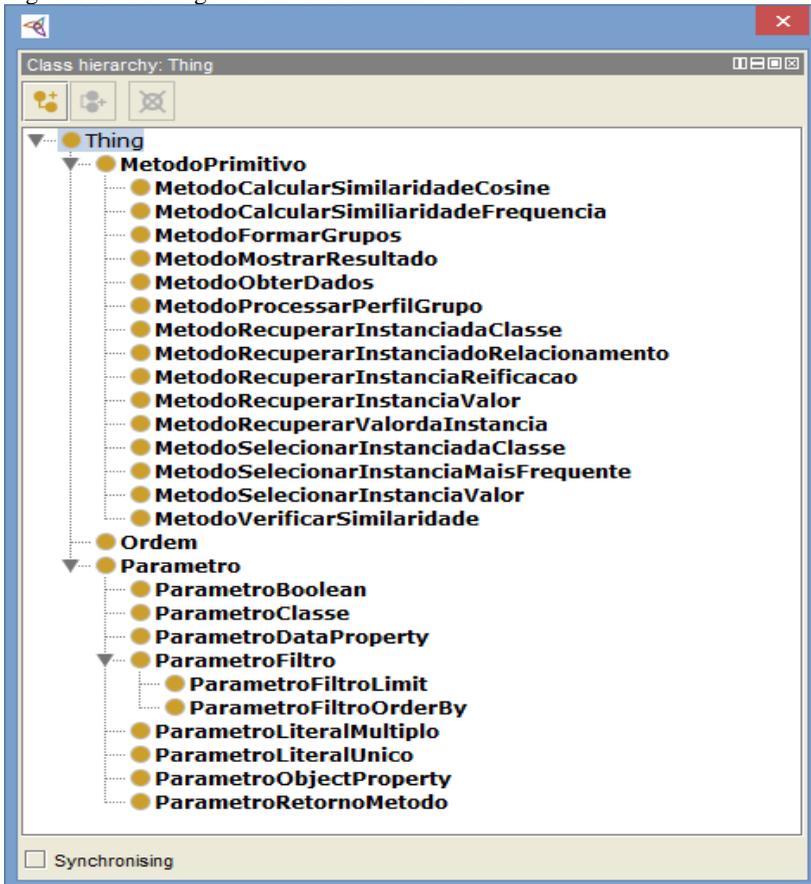


Fonte: Autor.

A ontologia de métodos primitivos, apresentada na Figura 38, representa os conceitos de **MetodoPrimitivo**, **Parametro** e **Ordem**.

MetodoPrimitivo representa os métodos que resolvem tarefas diretamente e que possuem uma implementação associada. **Parametro** representa os tipos de parâmetros dos métodos primitivos, através dos subtipos **ParametroClasse**, **ParametroDataProperty**, **ParametroObjectProperty**, **ParametroLiteral**, **Parametro Boolean**, **ParametroFiltro**, **ParametroRetornoMetodo**. O conceito de **Ordem** permite modelar a ordem de invocação dos parâmetros de um método.

Figura 38 - Ontologia de Métodos Primitivos



Fonte: Autor.

Nesta ontologia, instâncias de **MetodoPrimitivo** se relacionam com instâncias de **Parametro** e a ordem do relacionamento é

estabelecida pelas instâncias do tipo **Ordem**. De maneira mais específica: uma instância de **MetodoPrimitivo** se relaciona através da propriedade de objeto **temOrdem** (*hasOrder*) com uma instância de **Ordem**, que por sua vez, se relaciona através da propriedade de objeto **ordemPara** (*orderFor*) com uma instância de **Parametro**. Instâncias de **Ordem** possuem a propriedade de dados **temValor** (*hasValue*) que armazena a ordem de invocação das instâncias de **Parametro**.

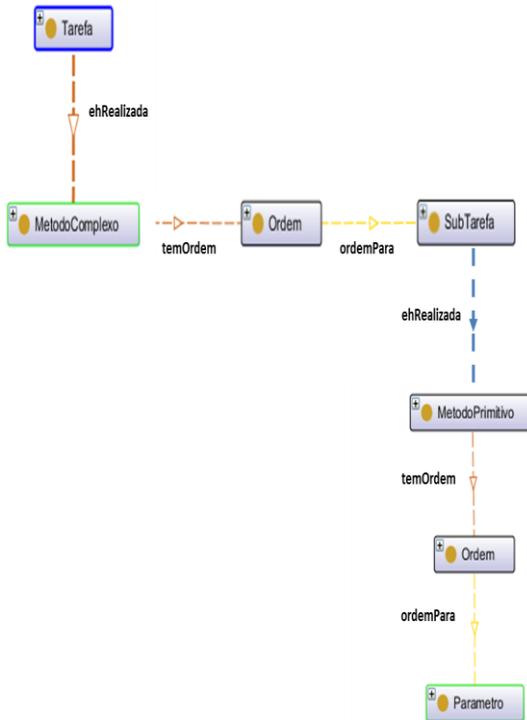
Cada instância de **MetodoPrimitivo** armazena através da propriedade de dados **temNome** (*hasName*) o nome do método implementado no API de métodos.

Cada instância das subclasses **ParametroClasse**, **ParametroDataProperty** e **ParametroObjectProperty** armazena ou está associada, através de uma propriedade de dados a uma descrição que identifica esse tipo nas ontologias de domínio. No caso do **ParametroClasse**, a instância estará associada, via propriedade de dados, ao nome de uma classe existente nas ontologias de domínio; no caso do **ParametroDataProperty**, a instância estará associada ao nome de uma propriedade de dados existente nas ontologias de domínio, assim por diante. Em um exemplo mais concreto, uma instância do tipo **ParametroClasse** pode estar associada através de uma propriedade de dados a uma classe chamada **FacultyMember** existente na ontologia de domínio.

As ontologias de Método Complexo e Método Primitivo se relacionam através da propriedade de objeto **ehRealizada** (*isPerformed*), que liga uma instância de **SubTarefa** a uma instância de **MetodoPrimitivo**.

Na Figura 39 são apresentadas todas as propriedades de objetos que relacionam instâncias dos conceitos da ontologia de Tarefas e instâncias dos conceitos das duas ontologias de Métodos. A explicação segue o sentido das setas da figura.

Figura 39 - Propriedades de objeto das ontologias de Tarefas e Métodos



Fonte: Autor.

Uma tarefa é realizada através de um método complexo. Sendo assim, a propriedade de objeto **ehRealizada** tem *domain* na classe **Tarefa** e *range* na classe **MetodoComplexo**.

Como um método complexo pode ser decomposto em várias subtarefas, a ordem de execução dessas subtarefas deve ser modelada. A propriedade de objeto **temOrdem** tem *domain* na classe **MetodoComplexo** e *range* na classe **Ordem**. A propriedade de objeto **ordemPara** tem *domain* na classe **Ordem** e *range* na classe **SubTarefa**.

Como uma sub tarefa é realizada através de método primitivo, a propriedade de objeto **ehRealizada** tem *domain* na classe **SubTarefa** e *range* na classe **MetodoPrimitivo**.

Um método primitivo pode possuir vários parâmetros e a ordem desses parâmetros também deve ser modelada. A propriedade de objeto

temOrdem tem *domain* na classe **MetodoPrimitivo** e *range* na classe **Ordem**. A propriedade de objeto **ordemPara** tem *domain* na classe **Ordem** e *range* na classe **Parametro**.

Para o estabelecimento da ordem com que os relacionamentos entre os pares de instâncias de **MetodoComplexo** e **SubTarefa** e os pares de instâncias de **MetodoPrimitivo** e **Parametro** aconteçam utilizou-se a noção de reificação, utilizada quando se quer armazenar informações sobre a relação binária entre duas instâncias. Neste caso, a informação que se quer armazenar é o valor correspondente a ordem, que pode ser 1, 2, 3 etc.

Sendo assim, cada instância de **Ordem** está relacionada a uma propriedade de dados chamada **temValor** (*hasValue*).

4.2 MODELO DE DOMÍNIO

O modelo de domínio representa o conhecimento de domínio, que é utilizado pelo modelo de solução de tarefa. Nesta tese o domínio a ser representado é o da Colaboração Científica, onde os conceitos relacionados a esse domínio, tais como Professores, Alunos, Publicações, estão presentes. Além disso, pertence ao conhecimento de domínio os conceitos relacionados as métricas de Análise de Rede Social (ARS).

O uso de ontologias é uma proposta consolidada na Engenharia do Conhecimento para representar o conhecimento de domínio de uma forma compartilhável e reutilizável. Nessa subseção será descrito o processo de construção das ontologias de domínio do *framework* proposto.

O processo de construção das ontologias buscou seguir o conceito de reutilização ou reuso de ontologias. Para D'aquin e Noy (2012) é menos custoso reutilizar ontologias bem estabelecidas e bem testadas do que desenvolver ontologias a partir do zero. Além disso, a reutilização de ontologias facilita a interoperabilidade de dados já que ao usar a mesma ontologia para descrever dados os mesmos podem ser integrados com muito mais facilidade.

Partindo do princípio de reuso, a primeira etapa do processo buscou ontologias do domínio da Colaboração Científica e ARS nas seguintes bibliotecas de ontologias, identificados por D'aquin e Noy (2012) como uma nova geração de bibliotecas de ontologias emergida recentemente: BioPortal¹⁷ (Noy et al. 2009), Cupboard¹⁸ (D'aquin e

¹⁷ <http://biportal.bioontology.org/>

Lewen, 2009), *oeGov*¹⁹ (Ontologies For E-Government), *OntologyDesignPatterns.org*²⁰ (Ontology Design Patterns.Org), *OntoSelect*²¹ (Buitelaar, Eigner e Declerck, 2004), The ONKI ontology server²² (Hyvonen et al., 2008), The TONES repository²³ e *Schema-Cache*²⁴.

Nessa pesquisa também foram utilizados os *search engines* de ontologias *Swoogle*²⁵ e *Watson*²⁶ (D'aquin e Motta, 2001). Além disso, foram utilizadas as bases de dados científicas *Scopus*, *WebofScience* e *ScienceDirect* para a busca de artigos com a proposição de ontologias de Colaboração Científica e ARS.

4.2.1 Ontologia de Colaboração Científica

Para a representação do domínio de Colaboração Científica foi utilizada a ontologia *VIVO-ISF (Integrated Semantic Framework)*²⁷ pertencente ao projeto *VIVO*²⁸.

O projeto *VIVO* é resultado de uma parceria entre sete universidades americanas (*University of Florida, Cornell University, Indiana University, Washington University School of Medicine in St. Louis, The Scripps Research Institute, Weill Cornell Medical College, and Ponce Medical College*) e financiado pelo *National Institute of Health (NIH)*. O projeto é materializado através de uma aplicação de código aberto, que representa comunidades acadêmicas e de pesquisa, desenvolvida com tecnologias de web semântica/*linked data*, como *RDF, OWL e SPARQL*.

A aplicação é preenchida com o perfil detalhado de professores, pesquisadores, alunos, etc., incluindo informações tais como publicações, atividades de ensino, afiliações profissionais, entre outras. Ela tem entre seus objetivos o fomento à colaboração, permitindo a descoberta de pesquisadores, projetos de pesquisas e bolsas de estudos

¹⁸ <http://cupboard.open.ac.uk:8081/cupboard-search/>

¹⁹ <http://oegov.org/>

²⁰ http://ontologydesignpatterns.org/wiki/Main_Page

²¹ <http://olp.dfki.de/OntoSelect/>

²² <https://onki.fi/>

²³ <http://owl.cs.manchester.ac.uk/repository>

²⁴ <http://schemacache.test.talis.com>

²⁵ <http://swoogle.umbc.edu/>

²⁶ <http://watson.kmi.open.ac.uk/WatsonWUI/>

²⁷ <http://ssepuntato.it/lode/owlapi/http://vivoweb.org/ontology/core#d4e3094>

²⁸ <http://vivoweb.org>

através de instituições, criando uma nuvem semântica de informações que podem ser pesquisadas.

Através de um modelo de dados em comum baseado em ontologias e a publicação de dados de acordo com os princípios dos *linked data*, as aplicações (instâncias) VIVO podem oferecer e consumir dados ligados, enriquecendo assim, os perfis dos pesquisadores. A instalação de várias instâncias institucionais da aplicação fomenta a colaboração inter-institucional.

A ontologia que suporta esse modelo de dados é chamada de VIVO-ISF. Ela é uma especificação unificada, formal e explícita de informações sobre pesquisadores, organizações e atividades envolvidas em pesquisa científica. VIVO-ISF integra classes e propriedades de várias outras ontologias existentes, além de definir suas próprias classes e propriedades. O Quadro 11 apresenta o conjunto de ontologias que compõem a ontologia principal.

Quadro 11 - Composição de ontologias da ontologia VIVO-ISF

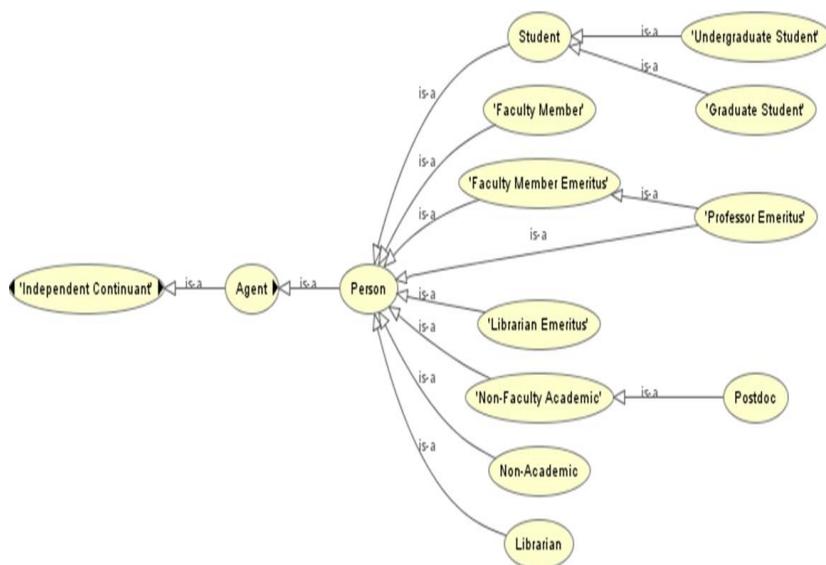
VIVO core	http://vivoweb.org/ontology/core
eagle-i - Resource Ontology (ERO)	http://code.google.com/p/eagle-i
Basic Formal Ontology (BFO)	http://www.ifomis.org/bfo
Bibliographic Ontology (BIBO)	https://github.com/structured-dynamics/Bibliographic-Ontology-BIBO
Cell Ontology (CL)	http://cellontology.org/?q=download
Event Ontology	http://motools.sourceforge.net/event/event.html
Friend of a Friend (FOAF)	http://www.foaf-project.org/
Gene Ontology (GO)	http://geneontology.sourceforge.net/#code
FAO's Geopolitical Ontology	http://www.fao.org/countryprofiles/geoinfo/en/Information
Artifact Ontology (IAO)	http://code.google.com/p/information-artifact-ontology/
Ontology for Biomedical Investigations (OBI)	http://obi.sourceforge.net/ontologyInformation/

Ontology of Clinical Research (OCRe)	http://code.google.com/p/ontology-of-clinical-research/
Reagent Ontology (ReO)	http://code.google.com/p/reagent-ontology/
Relations Ontology (RO)	http://obofoundry.org/ro/
Software Ontology (SWO)	http://theswo.sourceforge.net/
Sequence Ontology (SO)	http://www.sequenceontology.org/
SKOS (Simple Knowledge Organization System)	http://www.w3.org/2004/02/skos/
Uberon (Uber anatomy ontology)	http://obo.svn.sourceforge.net/viewvc/obo/uberon/releases/
vCard - Ontology for describing People and Organizations	http://www.w3.org/TR/vcard-rdf/

Fonte: Autor.

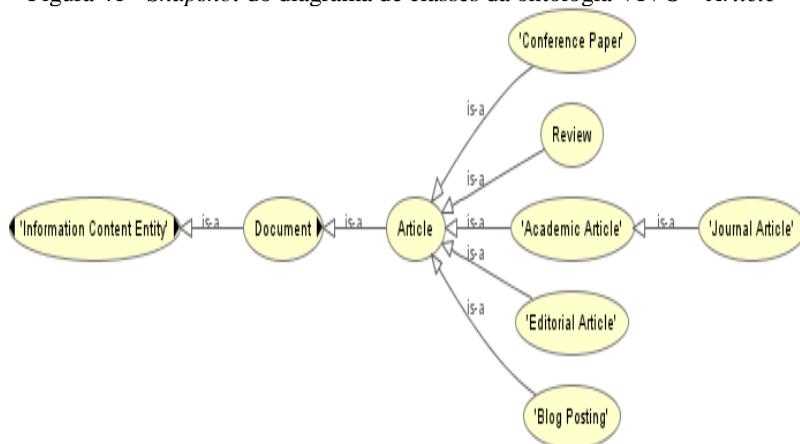
A ontologia VIVO-ISF na sua totalidade é composta por 200 classes e conseqüentemente, sua documentação é bastante extensa. No entanto, encontra-se no Anexo A dois diagramas fundamentais: o Diagrama de Classes e o Diagrama de Propriedade de Objetos, assim como um diagrama que ilustra como ocorre o relacionamento entre um autor (*Person*) e uma publicação (*Academic Article*) através do princípio de reificação usando a classe *Authorship*. Além desses diagramas, alguns *snapshots* do diagrama de classes da ontologia VIVO são apresentados na Figura 40 e Figura 41 porque exibem classes utilizadas por algumas tarefas de análise modeladas.

Figura 40 - *Snapshot* do diagrama de classes da ontologia VIVO – *Person*



Fonte: Autor.

Figura 41 - *Snapshot* do diagrama de classes da ontologia VIVO – *Article*



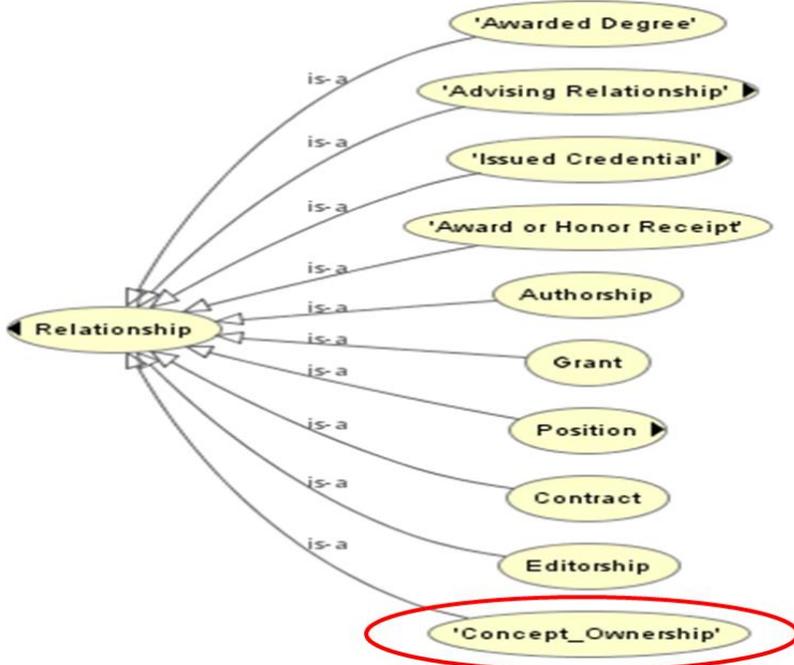
Fonte: Autor.

Apesar de ser uma ontologia bastante completa e essa ser a razão principal da sua escolha, ela precisou ser estendida para contemplar

algumas classes e relacionamentos que as tarefas de análise de colaboração identificadas e modeladas demandaram.

Na Figura 42 são ilustrados os vários tipos de classes de relacionamento que a ontologia VIVO representa e a extensão (destacada em vermelho) feita na mesma. Aqui um novo tipo de classe de relacionamento foi adicionado: “*Concept Ownership*”. Essa classe, assim como os outros tipos de *Relationship* já existentes na ontologia, foi criada para armazenar informações adicionais de um relacionamento entre outras duas classes. No caso de “*Concept Ownership*”, ela deve armazenar informações, como o peso ou frequência, fruto do relacionamento entre tipos de **Person** (Figura 40) e **SKOS:Concept** (quadro verde no Anexo A).

Figura 42 - *Snapshot* do diagrama de classes da ontologia Vivo – *Relationship*



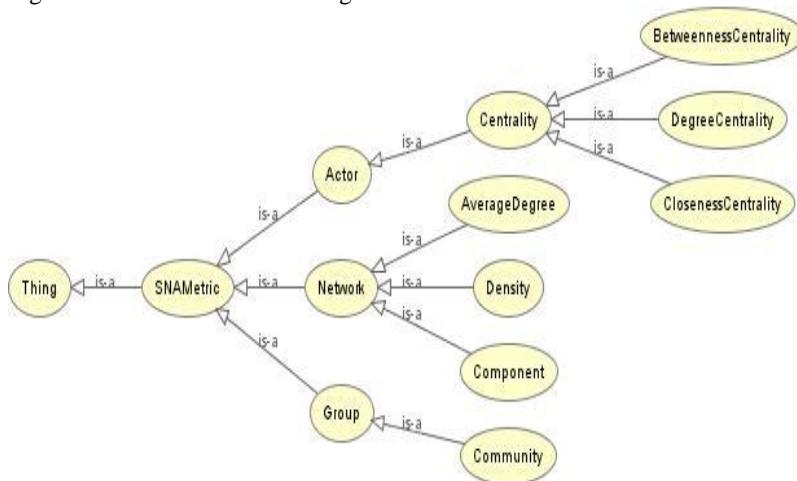
Fonte: Autor.

4.2.2 Ontologia de Métricas de Análise de Redes Sociais

Para a representação do domínio de análise de rede social foi encontrada a ontologia SemSNA (*Semantic Social Network Analysis*)²⁹ desenvolvida por Erétéo et al. (2008, 2009), onde estão representados conceitos relacionados a algumas métricas de ARS.

A ontologia original, que encontra-se no Anexo B, foi refatorada com a exclusão de alguns conceitos de ARS que não seriam utilizados nas tarefas de análise do *framework* e também com a organização e inclusão de outros conceitos, tais como, Densidade, Grau Médio além de algumas propriedades de objeto e dados. A Figura 43 exibe os conceitos e relacionamentos definidos nessa ontologia chamada de *SNAMetrics* (*Social Network Analysis Metrics*).

Figura 43 – *SNAMetrics*: Ontologia de Métricas de Análise de Rede Social



Fonte: Autor.

A classe principal **SNAMetric** é usada como superclasse de todos os conceitos relacionados às métricas de *Social Network Analysis* (SNA). As métricas foram agrupadas em conceitos relacionados com o nível de análise, ou seja, métricas do nível de análise do ator na rede (**Actor**), métricas do nível de análise da rede como um todo (**Network**) e métricas que analisam grupos formados por laços mais frequentes (**Group**).

²⁹ <http://ns.inria.fr/semsna/2009/06/21/voc>

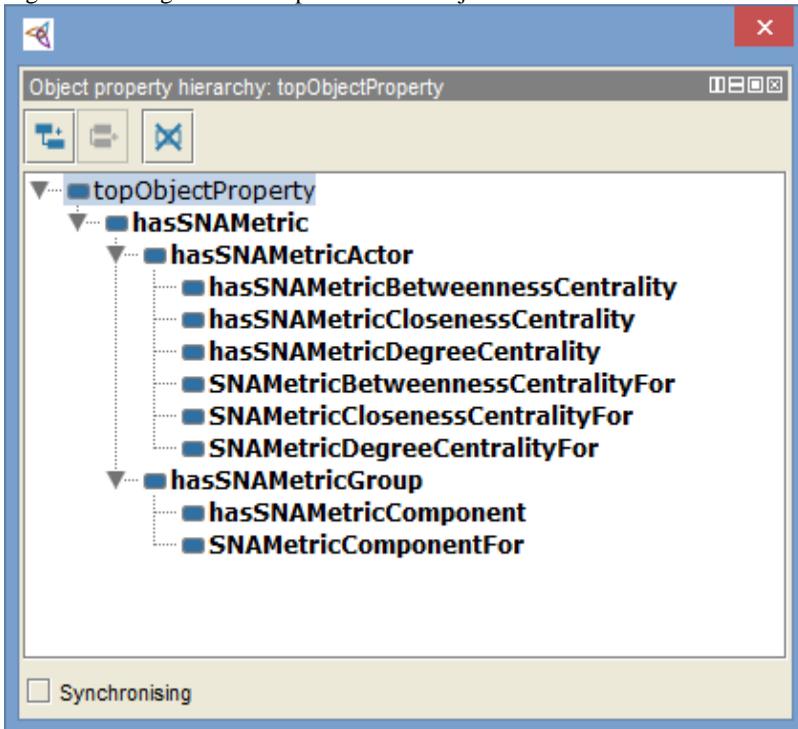
A classe **Actor** agrupa conceitos relacionados às métricas que analisam a posição do ator na rede, como os tipos de **Centrality: Degree Centrality, ClosenessCentrality e BetweennessCentrality**.

A classe **Group** contém o conceito de **Community** que pode ser utilizado para indicar em qual comunidade um ator pertence.

A classe **Network** agrupa conceitos relacionados a estrutura da rede, tais como **Component, AverageDegree e Density**.

A ontologia possui a propriedade de dados *hasValue*, que permite que as instâncias de *SNAMetrics* armazenem o valor de uma métrica. Ela também possui diversas propriedades de objeto, como exibido na Figura 44, que possibilitam o relacionamento entre instâncias da ontologia VIVO e instâncias que correspondem às métricas da ontologia *SNAMetrics*.

Figura 44 - Diagrama de Propriedades de Objeto da *SNAMetrics*



Fonte: Autor.

4.3 MÓDULOS DE PRÉ-PROCESSAMENTO DOS DADOS E INSTANCIACÃO DAS ONTOLOGIAS

O *framework* entrega as ontologias de Tarefas e de Métodos instanciadas para as tarefas identificadas, modeladas e representadas na seção 4.1. Contudo, as ontologias de domínio de Colaboração Científica (VIVO) e das métricas de ARS (*SNAMetrics*) precisam ser instanciadas.

Esta seção descreve os requisitos de dados necessários para a instanciação das ontologias de domínio do *framework*, os passos necessários para a instanciação, bem como os módulos desenvolvidos para realizar o pré-processamento dos dados e a instanciação dessas ontologias.

Em relação aos módulos que serão apresentados é importante destacar que eles foram concebidos em função de uma fonte de dados específica e que atende os requisitos de dados expostos a seguir, o currículo Lattes.

No cenário de C&T brasileiro, o currículo Lattes é uma fonte curricular de referência nacional e nele estão armazenadas informações relacionadas à formação, atuação profissional, produção científica, tecnológicas e artística de um pesquisador, além de projetos, orientações e participações em bancas. Logo existe uma possibilidade grande que essa fonte de dados seja escolhida para análises de colaboração científica, o que justifica a decisão dos módulos de pré-processamento e instanciação – que utilizam essa fonte – fazerem parte do *framework*.

Entretanto, o *framework*, é independente da fonte de dados nas suas camadas superiores: modelos de conhecimento e módulos de recuperação de dados.

4.3.1 Requisitos de Dados

Os dados para a instanciação das ontologias de domínio devem ser coletados de alguma fonte de dados e devem atender a alguns requisitos mínimos:

- a) Eles devem representar um contexto de colaboração científica através de evidências de trabalhos em coautoria tais como, produção bibliográfica, técnica e artística, projetos, patentes e registros, etc.
- b) Cada trabalho em coautoria devem conter dados dos coautores e também os descritores (palavras-chave) do trabalho. Os descritores servirão para criar o perfil de cada coautor (ator da rede) e também o perfil de cada trabalho

em coautoria. De cada coautor, é necessário minimamente o nome.

A partir dos trabalhos em coautoria, a rede de colaboração científica deve ser criada e o cálculo das medidas de ARS da rede deve ser efetuado por um software de análise de rede. As medidas de ARS necessárias são:

- a) Centralidade de grau, Centralidade de intermediação, Centralidade de proximidade;
- b) Densidade;
- c) Grau médio;
- d) Componente.

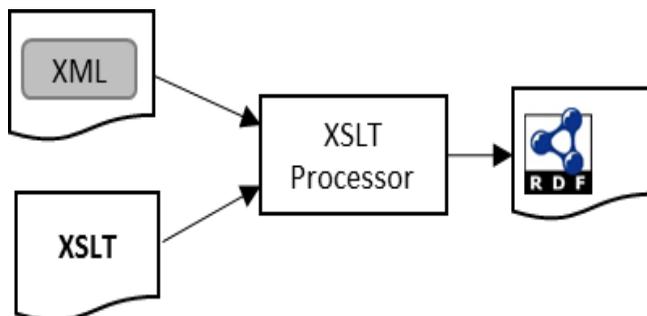
Destaca-se que os requisitos de dados estão diretamente relacionados com as perguntas ou tarefas de análise, ou seja, se as tarefas de análise forem estendidas e demandarem outros dados, eles deverão ser coletados. Além disso, as ontologias do modelo de domínio deverão suportar esses dados ou terão que ser estendidas.

4.3.2 Módulo de Instanciação da Ontologia de Colaboração Científica VIVO

O módulo de instanciação da ontologia VIVO-ISF tem como pré-requisitos dois módulos de pré-processamento dos dados: a) Módulo de transformação do currículo XML para RDF e b) Módulo de processamento do perfil expandido do pesquisador.

a. **Módulo de transformação do currículo XML para RDF:** este módulo é responsável por transformar cada currículo XML em um arquivo RDF contendo um conjunto de triplas anotadas com os mesmos tipos utilizados nas ontologias do projeto VIVO. Como pode ser visualizado na Figura 45 cada currículo é processado através de uma folha de estilo XSLT (*eXtensible Stylesheet Language for Transformation*) que transforma um Currículo Lattes XML em triplas RDF. Neste módulo isso é feito em um processo automatizado, onde para currículo XML é gerado um arquivo RDF correspondente.

Figura 45 - Transformação de XML em RDF



Fonte: Autor.

O XSLT (*lattes-vivo.xsl*) utilizado faz parte de um projeto *open source* chamado "*Semantic Lattes*"³⁰. O Quadro 12 abaixo exibe um *snapshot* deste XSLT que indica como cada *tag* TRABALHO-EM-EVENTOS do currículo XML deve ser transformada para triplas RDF. Observa-se na linha 536 deste *snapshot*, um padrão de código que permite a geração individualizada do URI (*Uniform Resource Identifier*) de um recurso que se repete para outros tipos de recursos. Neste caso, está sendo gerado um recurso (*resource*) do tipo **ConferencePaper** da ontologia VIVO_ISF.

³⁰ github.com/arademaker/slattes/

Quadro 12 - Snapshot do XSLT

```

535 <xsl:template match="TRABALHO-EM-EVENTOS">
536   <rdf:Description rdf:about="#P{@SEQUENCIA-PRODUCAO}-{/CURRICULO-VIT
537     <rdf:type rdf:resource="&vivo;ConferencePaper"/>
538     <dc:title>
539       <xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@TITULO-DO-
540     </dc:title>
541     <rdfs:label>
542       <xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@TITULO-DO-
543     </rdfs:label>
544     <dcterms:issued>
545       <xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@ANO-DO-TRA
546     </dcterms:issued>
547     <xsl:apply-templates select="DADOS-BASICOS-DO-TRABALHO/@IDIOMA"
548     <xsl:apply-templates select="DADOS-BASICOS-DO-TRABALHO/@HOME-PA
549     <xsl:apply-templates select="DADOS-BASICOS-DO-TRABALHO/@DOI"/>
550     <xsl:apply-templates select="AREAS-DO-CONHECIMENTO"/>
551     <xsl:apply-templates select="PALAVRAS-CHAVE"/>
552     <bibo:presentedAt>
553       <rdf:Description rdf:about="#conference-{generate-id(.)}">
554         <dc:title>
555           <xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@NOM
556         </dc:title>
557         <rdfs:label>
558           <xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@NOM
559         </rdfs:label>
560         <rdf:type rdf:resource="&bibo;Conference"/>

```

Fonte: Autor.

O XSLT original³⁰ foi modificado para permitir a geração não duplicada do URI de recursos, necessário quando ocorre o processamento de vários currículos XML e estendido para incluir a transformação das palavras-chave de cada tipo de produção bibliográfica.

b. **Módulo de criação do perfil expandido do pesquisador:** este módulo é responsável por criar o perfil expandido do pesquisador. Denomina-se perfil expandido, o conjunto ou vetor de termos oriundos dos títulos das produções bibliográficas do pesquisador conjuntamente com as respectivas frequências, que ajudam a formar, complementar ou qualificar o perfil inicial do pesquisador, formado pelas palavras-chave das produções bibliográficas cadastradas pelo pesquisador e suas respectivas frequências.

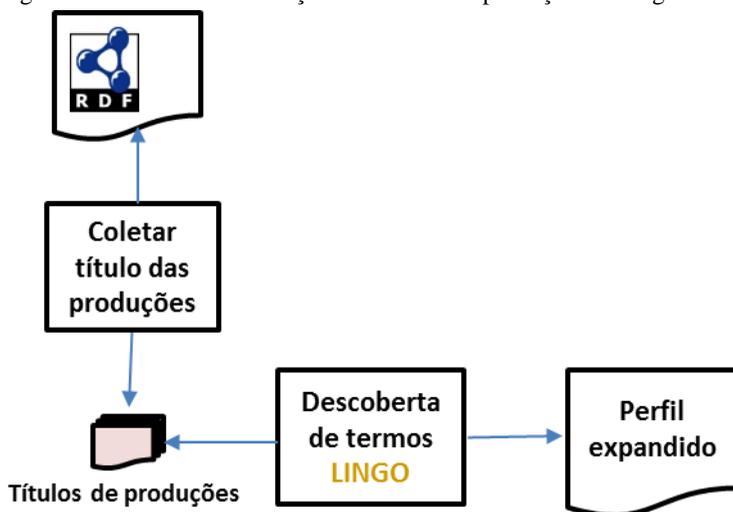
Neste trabalho, a criação do perfil expandido acontece através de um processo de mineração de dados textuais a partir dos títulos das

produções do pesquisador, no qual são extraídos os termos mais significativos das produções e que caracterizam o perfil do referido pesquisador. É através deste processo que o conteúdo produzido pelos pesquisadores da rede é minerado (coletado, pré-processado, representado) para posterior utilização em tarefas de análise de colaboração.

Verificou-se que aproximadamente 50% das produções bibliográficas dos currículos dos contextos coletados não tinham nenhuma palavra-chave cadastrada. Logo, a criação do perfil expandido em muitos casos, ajuda até mesmo a criar o perfil inicial do pesquisador.

A Figura 46 ilustra o processo de extração de termos e formação do perfil expandido.

Figura 46 - Processo de extração de termos das produções bibliográficas



Fonte: Autor.

O processo implementado neste módulo inicia com a coleta dos títulos das produções de cada currículo.rdf gerado no módulo anterior. A partir do conjunto de títulos de produção de cada pesquisador são identificados termos simples e compostos que aparecem com uma determinada frequência no conjunto de documentos. Para isso foi utilizado o algoritmo Lingo (OSINSKI; WEISS, 2005) até a sua fase de indução de rótulo (*label induction*) conforme explicado na seção 2.4.2.

Para a composição do perfil expandido, foram selecionados somente termos compostos com frequência maior ou igual a dois.

Finalmente, a instanciação da ontologia VIVO_ISF se dá através de um módulo específico que primeiramente popula a ontologia VIVO_ISF com as triplas dos arquivos .RDF gerados pelo primeiro módulo de pré-processamento (a) e posteriormente utiliza os resultados do segundo módulo de pré-processamento (b) para complementar a instanciação. Os módulos foram desenvolvidos na linguagem Java, utilizando Jena RDF e OWL API³¹ e a linguagem de consulta e manipulação de grafos RDF SPARQL³².

4.3.3 Módulo de Instanciação da Ontologia de Métricas de ARS

Este módulo demanda primeiramente que uma rede de colaboração científica seja criada e que a partir dela sejam calculadas as métricas de ARS que o *framework* requisita.

Sugere-se a utilização da ferramenta scriptLattes³³ para a coleta dos dados curriculares e para a criação da rede de colaboração baseada em coautoria. O scriptLattes é um sistema de código aberto, que permite coletar o currículo Lattes de um grupo de pesquisadores do site do CNPq, extrair informações de interesse, detectar produções científicas iguais ou similares e criar relatórios diversos através da compilação ou sumarização de produções intelectuais, projetos, orientações, etc., bem como o grafo de colaboração dos membros do grupo de pesquisadores (MENA-CHALCO; CESAR-JR, 2009).

O processo de detecção de produções acadêmicas iguais ou similares da ferramenta ScriptLattes visa detectar um relacionamento de colaboração entre os membros do grupo analisado. Assim, quando é encontrada uma produção considerada similar ou igual no conjunto das produções do grupo analisado conclui-se que exista um relacionamento de colaboração entre os mesmos. As explicações dos mecanismos de detecção de similaridade encontram-se em Mena-Chalco e Cesar-Jr (2013)

A rede de coautoria consiste em um arquivo .GDF contendo os nodos e as relações de coautoria juntamente com as frequências. As métricas de ARS podem ser calculadas a partir desse arquivo GDF com a utilização de um software de análise de rede.

³¹ <https://jena.apache.org/>

³² <http://www.w3.org/TR/sparql11-overview/>

Sugere-se a utilização do software de visualização e análise exploratória de dados de rede Gephi (BASTIAN; HEYMANN; JACOMY, 2009). Com esse software é possível calcular as métricas solicitadas e exportar os resultados em um arquivo .JSON.

O módulo de instanciação das métricas de ARS consiste em ler o arquivo .JSON e instanciar a ontologia *SNAMetrics* com os valores calculados. O módulo foi desenvolvido na linguagem Java.

4.4 MÓDULOS DE RECUPERAÇÃO DOS DADOS

Existem dois módulos de recuperação de dados no *framework*, conforme apresentado nos quadros 4 e 5 na Figura 26, no início deste capítulo.

O módulo de Recuperação de Tarefa e Métodos (quadro 4) é responsável por recuperar da ontologia de Tarefas a instância correspondente a tarefa escolhida pelo usuário do *framework*. A partir dessa instância, todas as demais instâncias relacionadas na ontologia de Métodos Complexos e, por conseguinte, todas as instâncias na ontologia de Métodos Primitivos são recuperadas. Esse processo de recuperação consiste em empilhar a sequência de métodos primitivos e respectivos parâmetros a serem invocados.

Os métodos a serem invocados estão disponíveis no Módulo API dos métodos (quadro 5). Os módulos foram implementados utilizando a linguagem Java e utilizam Jena RDF e OWL API³¹ para recuperar instâncias (triplas ou *statements*) da estrutura de representação dos modelos de conhecimento.

4.5 PASSOS PARA OPERACIONALIZAÇÃO DO *FRAMEWORK*

Por fim, nesta seção apresenta-se o conjunto de passos necessários para operacionalizar o *framework* associados com um exemplo que ilustra os referidos passos.

4.5.1 Instanciação das ontologias de domínio de colaboração científica e métricas de análise de rede social

O processo de instanciação pode ser manual ou automatizado e neste último caso pode fazer uso dos módulos especificados anteriormente.

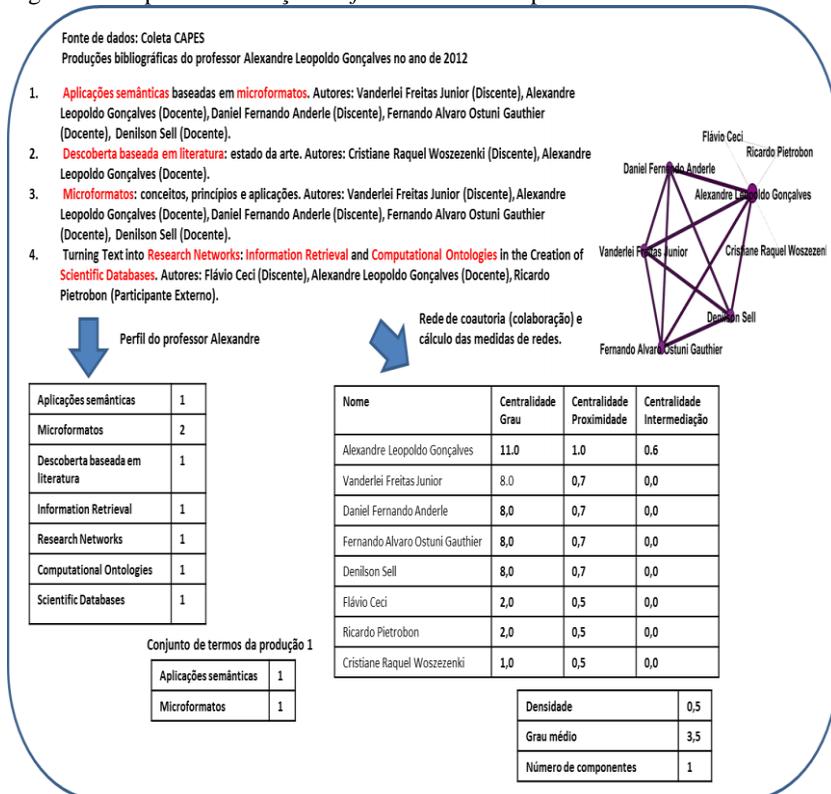
A Figura 47 demonstra, através de um exemplo, como acontece o processamento dos dados que servirão para criar as instâncias das ontologias. Nesse exemplo, primeiramente foi identificada uma fonte de dados com evidências de colaboração científica de um contexto de P&D

que apresentasse os requisitos mínimos de dados que o *framework* necessita.

Os relatórios gerados pelo Coleta Capes do PPGEGC atendem esses requisitos pois apresentam dados de produções bibliográficas que incluem o tipo da produção, o título, os coautores, o tipo dos coautores (docente ou discente), os dados do veículo de publicação, entre outros. A partir dessa fonte de dados foram coletadas, para fins de exemplo, as produções bibliográficas do ano de 2012 do professor Alexandre Leopoldo Gonçalves.

A partir desse conjunto de produções obteve-se o conjunto de termos (descritores ou palavras-chave) e frequência para compor o perfil do professor e para cada produção, foi obtido o conjunto de termos. No exemplo é exibido o perfil do professor Alexandre Leopoldo Gonçalves, a partir das suas produções no ano de 2012 e o conjunto de termos da produção 1.

A partir do conjunto de produções desse autor nesse ano obteve-se o conjunto de autores (nodos) e as relações de coautoria (arestas) necessárias. Através de um software de análise de rede como o Gephi efetuou-se o cálculo das medidas de ARS requeridas pelo *framework*.

Figura 47 - Operacionalização do *framework* – Pré-processamento dos dados

Fonte: Autor.

Os dados processados serviram de insumo para a criação das instâncias das ontologias de domínio VIVO-ISF e *SNAMetrics*, como pode ser visualizado na Figura 48.

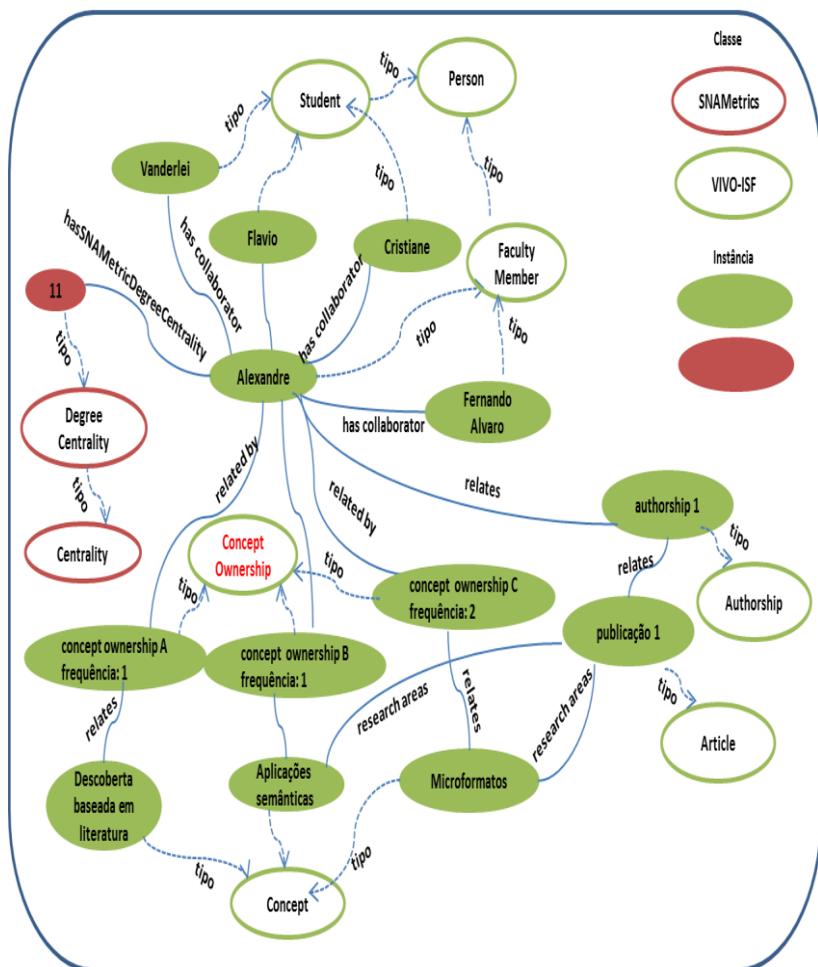
A partir dessa figura pode-se também perceber outros elementos dessas ontologias. A propriedade de dados *has collaborator* da ontologia VIVO relaciona as instâncias de *Faculty Member* (professor) e *Student* (aluno), ambos tipos de *Person*. A propriedade de dados *hasSNAMetricDegreeCentrality* da ontologia *SNAMetrics* relaciona a instância “Alexandre” do tipo *Faculty Member* à instância “11” do tipo *Degree Centrality*.

Instâncias do tipo *Person* (a instância Alexandre é do tipo *Faculty Member* que é um tipo de *Person*) podem se relacionar com instâncias de *Concept* e essa relação tem um valor associado que é o

peso ou frequência da relação. Isso demanda que as instâncias de **Person** se relacionem com instâncias de **Concept Ownership** (em vermelho para destacar que a classe foi adicionada à ontologia Vivo) através da propriedade de objeto *related by*. O valor do peso ou frequência é armazenado em uma *data property* associada a instância de **Concept Ownership**. No exemplo, a instância Alexandre está relacionada com três instâncias de conceitos (descritores) que fazem parte do seu perfil (calculado na etapa anterior) que são “Descoberta baseada em literatura”, “aplicações semânticas” e “microformatos”, cujas frequências são respectivamente 1, 1, e 2.

Instâncias do tipo **Article** são relacionadas a instâncias do tipo **Concept** através da propriedade de objeto *research areas*. No exemplo, a instância Publicação 1 está associada às instâncias “microformatos” e “aplicações semânticas”.

A autoria das publicações é representada pela classe **Authorship** que relaciona as instâncias de **Person** e **Article**. No exemplo, a instância Alexandre se relaciona com a instância authorship1 que por sua vez se relaciona com a instância publicação1.



Fonte: Autor.

4.5.2 Execução do *framework*

A execução é o último passo, que faz a estrutura proposta funcionar de maneira totalmente automatizada. Ela começa a partir da escolha, por parte do usuário, de alguma tarefa de análise de colaboração científica, através de uma aplicação com *front-end*. Essa aplicação com *front-end* não faz parte da estrutura do *framework*, ou seja, pode ser implementada com os requisitos de interface desejados

pelo utilizador do *framework*. Entretanto, ela deve somente permitir a escolha das tarefas modeladas e representadas nele.

A escolha de uma tarefa desencadeia o processo de recuperação de dados das ontologias explicitado na camada de recuperação de dados da estrutura do *framework* (Figura 26).

Um exemplo demonstra o processo de recuperação de dados das ontologias. Na Figura 49 o gestor escolhe a tarefa “Atores que mais colaboram”. A partir dessa escolha, o módulo de recuperação de tarefa e métodos do *framework* (quadro 4 na Figura 26) entra em execução. Isso é demonstrado na Figura 49 pelo “Pseudo-código da recuperação da Tarefa 1 “Atores que mais colaboram” nas ontologias“, onde o primeiro passo é carregar (ler) todas as ontologias de tarefa e métodos, o segundo passo é recuperar a instância correspondente à tarefa na classe apropriada e o terceiro passo é recuperar todas as instâncias relacionadas a essa instância. Todo o fluxo de busca das instâncias nas ontologias para posterior execução dos métodos implementados na API de métodos é exibido com detalhes na Figura 49.

A instância correspondente à tarefa “Atores que mais colaboram” está relacionada com a instância de **MetodoAtoresMaisColaboram** (tipo de **MetodoComplexo**) que está relacionada com duas instâncias da classe **Ordem** através da *object property* **temOrdem**, onde cada instância de **Ordem** se relaciona com a *data property* **temValor** que armazena a ordem com que as instâncias de **TarefaRecuperarAtoresCG** e **TarefaSelecionarAtorMaiorCG** devem ser executadas. A seguir a ordem de execução dessas tarefas será explicada.

Uma das instâncias de **Ordem**, citadas acima, armazena o número de ordem 1 na propriedade de dados **temValor** e está relacionada com a instância de **TarefaRecuperarAtoresCG** através da *object property* **ordemPara**, que por sua vez está relacionada através da *object property* **ehRealizada** com a instância de **MetodoRecuperarInstanciaValor**.

A instância do método **MetodoRecuperarInstanciaValor** armazena o nome do método, nesse caso “recuperarInstanciaValor”, através da *data property* **temNome** e relaciona-se com duas instâncias da classe **Ordem** através da *object property* **temOrdem**. Uma das instâncias de **Ordem** armazena o valor de ordem 1 e relaciona-se com a instância da classe **ParametroObjectProperty** através da *object property* **ordemPara**. A instância de **ParametroObjectProperty**

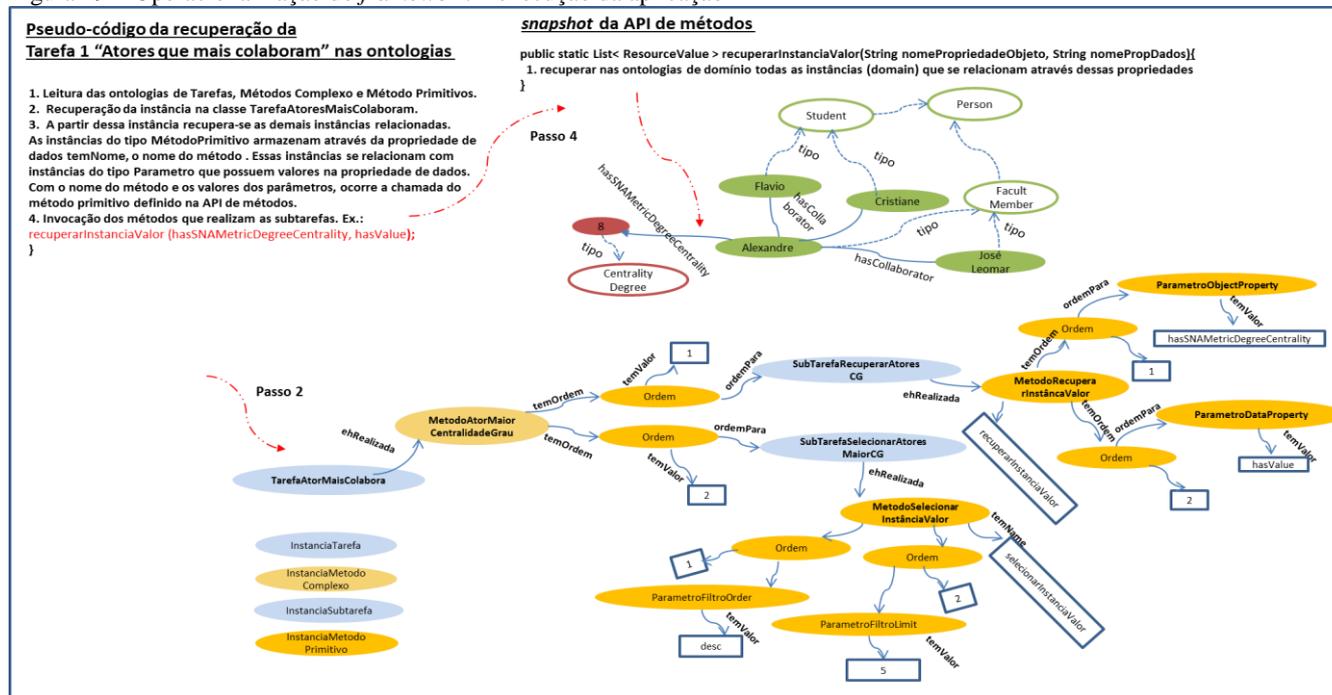
armazena através da *data property* **temValor** o valor do parâmetro, nesse caso “*hasSNAMetricDegreeCentrality*” (nome da propriedade de objeto na ontologia de métricas de ARS *SNAMetrics*). A outra instância de Ordem **Ordem** armazena o valor de ordem 2 e relaciona-se com a instância da classe **ParametroDataProperty** através da *object property* **ordemPara**. A instância de **ParametroDataProperty** armazena através da *data property* **temValor** o valor do parâmetro, nesse caso “*hasValue*” (nome da propriedade de dado na ontologia de domínio de métricas de ARS *SNAMetrics*).

No exemplo, é o método “recuperarInstanciaValor” e seus valores de parâmetro, “*hasSNAMetricDegreeCentrality*” e “*hasValue*”, que fazem com que o método **recuperarInstanciaValor** da API de métodos seja invocado (passo 4), executado e retorne um resultado. O resultado, neste caso, são todas as instâncias que possuem a *object property* “*hasSNAMetricDegreeCentrality*” (instâncias de **FacultyMember**) conjuntamente com os respectivos valores das instâncias de **DegreeCentrality** presentes na ontologia de domínio de colaboração científica (VIVO) e métricas de ARS (*SNAMetrics*) representadas em verde e vermelho na Figura 49. Esse retorno deste método é utilizado como entrada para o segundo método detalhado na sequência.

A outra instância de **Ordem**, que armazena o número de ordem 2, está relacionada com a instância de **TarefaSelecionarAtoresMaiorCG**, que por sua vez está relacionada com a instância de **MetodoSelecionarInstanciaValor**. Esse método se relaciona com instâncias de **ParametroFiltroOrder** e **ParametroFiltroLimit** através de instâncias do tipo **Ordem**, que armazenam a ordem dos parâmetros desse método.

Ao final da execução desse método, as instâncias de retorno do primeiro método (**recuperarInstanciaValor**), recebidas como parâmetro, são ordenadas de acordo com o valor do parâmetro **ParametroFiltroOrder** (“desc” de descendente) e limitadas em termos de quantidade de instâncias em função do valor do parâmetro **ParametroFiltroLimit** (5). Assim, as instâncias que correspondem aos atores que mais colaboram, conjuntamente com os graus de colaboração, são retornadas promovendo a resposta à tarefa “Atores que mais colaboram”.

Figura 49 – Operacionalização do *framework* – execução da aplicação



4.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou uma proposição composta por um conjunto de artefatos oriundos de processos de modelagem e representação dos conhecimentos utilizados na análise de colaboração científica de contextos de P&D e por um conjunto de artefatos que manipulam (armazenam e recuperam) os dados dessas representações. Este conjunto de artefatos foi apresentado em camadas com propósitos distintos e recebeu a denominação de *framework*.

A justificativa da escolha ou a metodologia de construção de cada artefato na sua respectiva camada foi apresentada. Primeiramente foram apresentados os artefatos da camada de modelos de conhecimento. Para cada modelo de conhecimento foram propostas representações ontológicas. Depois, foram apresentados os artefatos ou módulos implementados de armazenamento de dados nas representações e finalmente foram explicitados os módulos implementados de recuperação de dados das representações ontológicas.

Por fim, foram apresentados os passos para a operacionalização deste *framework*.

Acredita-se ser importante destacar que este é um *framework* computacional totalmente operacional para o conjunto de perguntas, tarefas ou demandas de análise de colaboração científica identificadas.

5 AVALIAÇÃO DO *FRAMEWORK*

Este capítulo tem o objetivo de apresentar o processo de operacionalização (instanciação), demonstração e avaliação do *framework*. Ele descreve o cenário onde estes processos aconteceram, os passos para a operacionalização do *framework* e o processo de avaliação em si, conjuntamente com as respostas. Por fim, a análise dos resultados da avaliação é apresentada.

5.1 CENÁRIO DE DEMONSTRAÇÃO E AVALIAÇÃO

Esta seção apresenta o cenário de operacionalização, demonstração e avaliação do *framework*, baseado nas escolhas metodológicas descritas na seção 3.2 do capítulo 3.

Os cenários de P&D escolhidos foram dois Programas de Pós-Graduação de uma Universidade Federal. Entendeu-se que um Programa de Pós-Graduação (PPG) se caracterizava como o melhor cenário de P&D para aplicação e avaliação do referido *framework*, pois a pesquisa e colaboração científica são elementos essenciais e inerentes a qualquer PPG.

Os dados ou insumos utilizados na análise da colaboração de cada PPG foram as produções bibliográficas (Artigos completos publicados em periódicos, Artigos aceitos para publicação, Livros e capítulos, Textos em jornais ou revistas, Trabalhos publicados em anais de congressos e Outras produções bibliográficas) produzidas pelos professores pesquisadores de cada PPG. Este escopo foi delimitado em função da factibilidade da condução da pesquisa frente a forma de acesso aos dados. A seguir os dois cenários são detalhados.

5.1.1 Cenário de Demonstração e Avaliação 1

O PPG deste cenário tem natureza interdisciplinar e possui um curso de Mestrado e Doutorado oferecidos desde 2004 (11 anos). Neste cenário foi analisada, através do *framework*, a rede de coautoria das produções bibliográficas de 37 professores pesquisadores do programa. Foram utilizadas em torno de 7000 produções bibliográficas deste grupo.

Os avaliadores do *framework* neste cenário foram os gestores – coordenador e subcoordenador - do PPG. A experiência destes gestores em cargos de gestão de P&D soma em torno de 34 anos.

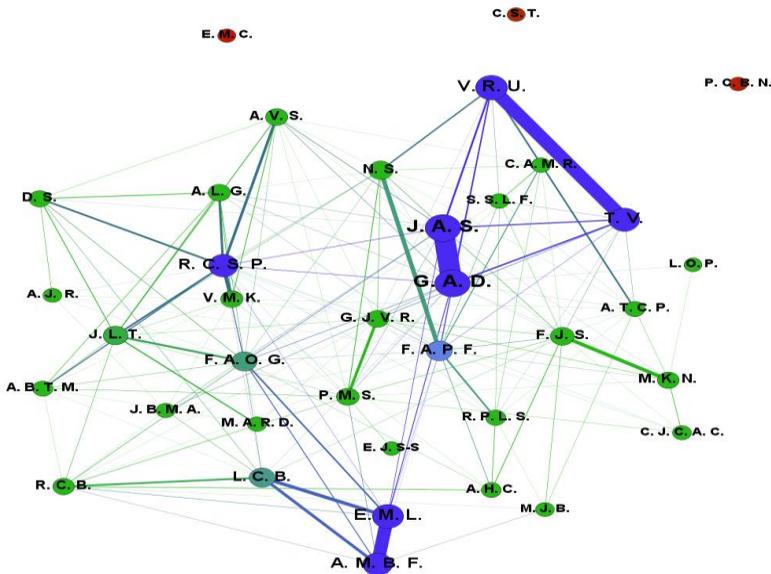
Como o objetivo de facilitar o entendimento das respostas da avaliação do *framework* é exibida na Figura 50 a rede de coautoria deste PPG e algumas análises básicas são feitas.

Nesta rede de coautoria existem 37 nodos e 175 arestas dispostas em 4 componentes distintos, sendo que 3 desses componentes são compostos por apenas 1 pesquisador. A densidade desta rede é 0,2.

A centralidade de grau com peso (*weight degree*), que leva em consideração o número de arestas de cada nodo conjuntamente com a frequência de cada aresta – neste caso o número de coautorias –, foi utilizada para destacar os nodos com maior centralidade na cor azul. Os nodos com maior centralidade de grau são J.A.S e G.A.D.

Os nodos com maior centralidade de proximidade são F.A.P.F. e F.A.O.G, com 0,70 e 0,68 respectivamente. Os nodos com maior centralidade de intermediação são F.A.P.F. e P.M.S. com 0,07 e 0,07 respectivamente.

Figura 50 - Rede de coautoria dos professores pesquisadores do PPG do Cenário 1



Fonte: Autor.

A centralidade de grau com peso (*weight degree*) também foi utilizada aqui para destacar os nodos na cor azul. Os nodos com maior centralidade de grau são C.B.W e C.M.W. Os nodos com maior centralidade de proximidade são R.F.C e A.v.W., com 0,48 e 0,42 respectivamente. Os nodos com maior centralidade de intermediação são R.F.C. e C.B.W. com 0,27 e 0,16 respectivamente.

5.2 OPERACIONALIZAÇÃO

A operacionalização ou instanciação do *framework* nos cenários descritos consistiu na execução dos passos descritos na seção 4.5. O primeiro passo - Instanciação das ontologias de domínio de colaboração científica e análise de rede social - demandou alguns procedimentos que são detalhados na sequência.

5.2.1 Coleta e Extração dos Dados

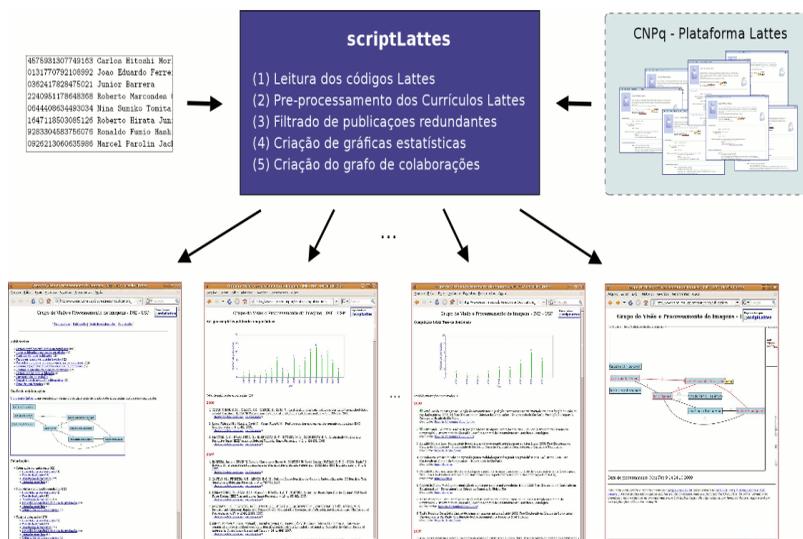
Os dados referentes a produção bibliográfica foram coletados do currículo Lattes dos pesquisadores. Foram utilizados dois procedimentos de coleta de dados, com objetivos e ferramentas distintas.

- a) **Coletar dados completos dos pesquisadores:** neste procedimento foram coletados manualmente os currículos XML dos pesquisadores de cada PPG com o objetivo de popular ou instanciar vários conceitos da ontologia de domínio VIVO-ISF.
- b) **Coletar dados da rede de coautoria dos pesquisadores:** neste procedimento foi utilizada a ferramenta scriptLattes³³ para a coleta automática dos currículos dos pesquisadores e consequente criação da rede de coautoria de cada PPG.

Para utilizar essa ferramenta é necessário informar a lista de identificadores (ids) de currículos Lattes que se deseja coletar e opcionalmente pode-se definir o período de coleta das produções. Nesta operacionalização os períodos não foram definidos. A Figura 52 exhibe as principais operações que o scriptLattes executa.

³³ <http://scriptlattes.sourceforge.net/>

Figura 52 - Principais operações do scriptLattes



Fonte: <http://scriptlattes.sourceforge.net/>

5.2.2 Pré-Processamento dos Dados

Os dados extraídos nos dois procedimentos de coleta foram pré-processados antes de serem inseridos nas ontologias de domínio *VIVO-ISF* e *SNAMetrics*.

Para o procedimento de coleta de dados A dois procedimentos de pré-processamento foram realizados:

- Transformar XML em RDF:** neste procedimento cada currículo no formato XML foi transformado em um conjunto de triplas RDF.
- Criar perfil expandido do pesquisador:** neste procedimento foi criado o perfil expandido do pesquisador.

Para o procedimento de coleta de dados B, o pré-processamento consistiu em calcular as métricas de análise de rede social. Neste procedimento a rede de coautoria foi criada e as métricas requisitadas pelo framework (centralidade de grau, centralidade de proximidade, centralidade de intermediação, densidade, grau médio e componentes) foram calculadas com o pelo software de visualização e análise

exploratória de dados de rede Gephi (BASTIAN; HEYMANN; JACOMY, 2009).

5.2.3 Instanciação das Ontologias

As ontologias que fazem parte do modelo de domínio do *framework*, mais especificamente, a ontologia de colaboração científica (*VIVO-ISF*) e a ontologia de ARS (*SNAMetrics*), foram instanciadas com os dados extraídos e processados através dos procedimentos anteriores. Para isso foram utilizados os dois módulos de instanciação descritos na seção 4.3.

5.2.4 Execução do *Framework*

Para a execução do *framework* foi desenvolvido uma aplicação *web* como *front-end*. A Figura 53 mostra a interface da aplicação onde a execução do *framework* começa a partir da escolha de alguma tarefa de análise de colaboração científica.

Figura 53 - Interface da aplicação *web-front-end* do *framework*



Fonte: Autor.

A partir da escolha da tarefa, as instâncias relacionadas a tarefa, métodos e parâmetros foram recuperadas das ontologias de tarefa e métodos pelo Módulo de recuperação de tarefa e métodos, que por sua vez invocou os métodos implementados no Módulo API de métodos, fornecendo a eles os dados ou parâmetros necessários para sua execução. Os métodos da API buscaram os dados de resposta da tarefa nas ontologias de domínio do *framework*.

5.3 DEMONSTRAÇÃO E AVALIAÇÃO

O *framework* foi demonstrado e posteriormente avaliado em momentos distintos e previamente agendados, nos dois cenários descritos no início deste capítulo. O áudio do processo foi gravado e teve consentimento firmado através de um Termo de Consentimento cujo modelo encontra-se no Apêndice C.

A demonstração consistiu em inicialmente apresentar os benefícios da colaboração científica, alguns conceitos de análise de rede, o potencial de explorar o conteúdo dos artefatos resultantes da colaboração (produções bibliográficas), bem como os objetivos do *framework*. Também foi apresentada a imagem da rede de coautoria do PPG (Figura 50 ou Figura 51), com o objetivo de facilitar o entendimento do *framework* e das respostas que ele fornece.

Logo após, cada uma das 18 tarefas de análise de colaboração científica foi explicada e executada *on the fly*. A demonstração foi feita para os gestores - coordenador e subcoordenador - de cada PPG.

Concomitante a demonstração dos resultados de cada tarefa foi realizada a avaliação do *framework* através de uma entrevista com o grupo focal, composto pelos 2 gestores do PPG.

A avaliação foi conduzida através de uma entrevista com um roteiro ou questões (Apêndice B) previamente definidas em três dimensões denominadas: Dimensão 1 - Colaboração Científica, Dimensão 2 - Resultados e Dimensão 3 - Aplicabilidade do *Framework* como Instrumento de Gestão, especificadas na seção 3.2.1

Inicialmente foram solicitados os dados de identificação, formação e o tempo de experiência em gestão de ambientes de P&D dos avaliadores. Logo a seguir foi solicitada a opinião dos gestores em relação a uma dimensão introdutória com questões relativas a importância da colaboração científica e dos resultados da sua análise para uma possível tomada de decisão.

A seguir, as respostas da avaliação com os dois grupos focais de gestores para as questões abordadas na entrevista serão explicitadas.

5.3.1 Perguntas e Respostas da Dimensão 1

Esta dimensão avaliou a importância da colaboração científica, da análise da colaboração e das perguntas/tarefas de análise propostas. Ela é composta de três questões, sendo a última questão realizada no final da entrevista e exposta no final da próxima seção.

1) Você considera importante o estabelecimento de relacionamentos de colaboração científica em um Programa de Pós-Graduação?

- Concordo plenamente Concordo parcialmente
 Indiferente
 Discordo parcialmente Discordo totalmente

Resposta: Os dois grupos concordaram.

2) Você considera que a análise da colaboração científica possa trazer subsídios para a tomada de decisão em vários níveis (gestor, pesquisador, comunidade acadêmica) em um Programa de Pós-Graduação?

- Concordo plenamente Concordo parcialmente
 Indiferente
 Discordo parcialmente Discordo totalmente

Resposta: Em ambos os grupos, a ideia de tomada de decisão causou um certo desconforto. Os gestores argumentaram que em nível de gestão esse termo é forte e remete a uma ação de cunho mais impositivo e que os gestores não podem tomar decisões no sentimento mais imperativo. Eles podem executar ações mais brandas, como por exemplo, orientar, induzir, incentivar, fazer um diagnóstico e tentar criar políticas de incentivo e de envolvimento das pessoas que são pouco colaborativas.

5.3.2 Perguntas e Respostas da Dimensão 2 e Dimensão 3

A dimensão 2 (D2) e a dimensão 3 (D3) da avaliação foram conduzidas uma após a outra para cada pergunta/tarefa de análise de colaboração científica do *framework*. Na sequência são apresentados todos os quadros com as perguntas e respostas.

A dimensão 2 avaliou o grau de confiança (alto, médio ou baixo) dos resultados das perguntas/tarefas. Já a dimensão 3 avaliou a aplicabilidade do *framework* como instrumento de gestão de análise de

colaboração científica. Nesta dimensão, para cada resultado de pergunta/tarefa foram propostas possíveis ações de tomada de decisão em nível de gestão. Essa proposição teve o objetivo de facilitar e reduzir o tempo da entrevista, oferecendo ao avaliador um ponto de partida, para dentro de uma escala que ia de “concordar totalmente” a “discordar totalmente”, decidir se a ação era possível, mas também propor novas ações, o que aconteceu na maioria das questões avaliadas.

Quadro 13 - Avaliação da tarefa 1

1. Quem são os pesquisadores que mais colaboram?
D2
<p>(X) Alto () Médio () Baixo</p> <p>- Ambos concordaram que o grau de confiança dos resultados com a métrica utilizada era alto. Entretanto, os dois grupos solicitaram explicações sobre a métrica utilizada. Nessa tarefa foi utilizada a Centralidade de Grau com Peso, que leva em consideração a quantidade de conexões do nodo pesquisador, mas também o peso das relações, que no caso da rede de coautoria é a quantidade de artigos publicados conjuntamente por dois nodos. Ambos os grupos concordaram no argumento que a melhor métrica a ser utilizada é a Centralidade de Grau pura, pois ela é mais fidedigna ao conceito colaboração.</p> <p>Considerações do grupo de gestores do cenário 1:</p> <p>- Abordaram a importância de avaliar o aspecto qualitativo da colaboração com os seguintes questionamentos: “Será que aquele pesquisador que colabora com muitos tem qualidade nas suas publicações? ”, “Ou tem mais qualidade aquele que colabora com poucos? ”</p> <p>- Perguntaram sobre a variável tempo e destacaram a sua importância nesse tipo de análise, pois os pesquisadores e o próprio Programa tem histórias diferentes ao longo do tempo.</p> <p>- Sugeriram mudar a pergunta para: “Quem são os pesquisadores que colaboram com mais pesquisadores? ”.</p> <p>- Salientaram que a colaboração se manifesta de diversas formas, além da coautoria, e que quando é baseada na coautoria, poderia se chamar coprodução. E nesse caso, o <i>framework</i> estaria avaliando os mais coprodutivos.</p> <p>- Em relação a forma de apresentação dos resultados dessa pergunta, sugeriram não restringir o número de pesquisadores em cinco, mas</p>

exibir todos ordenados pelo *ranking* de colaboração.

D3

Um alto grau de colaboração denota liderança, popularidade e intensa atividade científica (KRACKHARDT, 2010; FREEMAN, 1979). O nodo pode ser visto como um grande canal de informação, retendo ou transmitindo informações (FREEMAN, 1979).

A identificação desses pesquisadores permite que os gestores do PPG:

Ação A: Indiquem os mesmos para atuarem em atividades e ações do PPG, tais como, projetos de pesquisa, organização de eventos, etc.

Resposta do grupo de gestores do cenário 1

(X) Concordo parcialmente

- A resposta teve um sim e um não. Sim, “porque são pessoas conectadas e criaram um capital relacional bacana”. Não, essa ação acaba fortalecendo quem já é forte. Para esse grupo, a ação deve envolver os que menos colaboram.

Resposta do grupo de gestores do cenário 2:

(X) Concordo parcialmente

- Argumentaram que existem outros fatores que devem ser observados. Por exemplo, se algum pesquisador que aparece como mais colaborativo, for de “difícil acesso”, ou muito ocupado com suas pesquisas, talvez não tenha tempo para se envolver em outras ações.

- Destacaram que o foco da ação deveria ser ‘sugerir’ ao invés de indicar.

Ação B: Indiquem os mesmos para atuarem como disseminadores de informações na rede.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

- Citaram como exemplo, a utilização de um pesquisador que colabora com 20 dos 37 professores do Programa, para disseminar a informação de mudanças nos parâmetros da avaliação Qualis.

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

Ação C: Divulguem os resultados para os demais pesquisadores e comunidade acadêmica.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

- Sugeriram mostrar os resultados em uma reunião de colegiado, parabenizando os professores que trabalham com mais professores com frases do tipo “Gente, temos que trabalhar mais juntos”, “Professor, o senhor sozinho trabalha com meio programa”.

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

Você sugere novas sugestões de ações?

Sugestões do grupo de gestores do cenário 1:

Ação D: Realizar uma ação de diagnóstico dos resultados, o que pode conduzir, por exemplo, a) ao entendimento das “melhores práticas” dos pesquisadores mais colaborativos; b) a verificação da possibilidade de ocorrência de privilégios em determinadas posições, com professores ministrando disciplinas como as obrigatórias, que privilegiam um maior *networking* com os alunos e seus orientadores, consequentemente ligando esse professor colaborativo a todos os demais professores pesquisadores; c) a verificação das áreas de concentração dos professores que mais colaboram, ou seja, se são todos de uma mesma área de concentração ou se os mais colaborativos estão equilibrados nas áreas de concentração do programa. Se a maioria estiver em uma área de concentração pode ser inferido que existe uma “cultura de grupo” nessa área. E que o restante dos pesquisadores podem ter “clusters” mais fechados.

Quadro 14 - Avaliação da tarefa 2

2. Quem são os pesquisadores que menos colaboram?

D2

(X) Alto () Médio () Baixo

Ambos concordaram que o grau de confiança dos resultados foi alto.

Considerações do grupo de gestores do cenário 2:

- Comentaram que aproxima-se o período de avaliação da CAPES e que o programa é avaliado pela produção. Identificaram pesquisadores que colaboram pouco e que também não são tão produtivos, assim como pesquisadores que colaboram pouco e são altamente produtivos. Consideram importante divulgar resultados de forma anônima em reunião.

D3

No sentido contrário dos argumentos de Krackhardt (2010) e Freeman (1979) um baixo grau de colaboração denota pouca liderança, popularidade.

A identificação desses pesquisadores permite que os gestores do PPG:

Ação A: Orientem os mesmos a intensificarem sua atividade colaborativa com outros pesquisadores.

Resposta do grupo de gestores do cenário 1

(X) Concordo parcialmente

- Reiteraram que a ação deve ter uma conotação de envolvimento, por parte dos gestores, desses pesquisadores que menos colaboram com outros pesquisadores e ações do programa. Citam o caso de um professor que aparece na lista dos menos colaborativos em função da produção bibliográfica, mas que é altamente colaborativo em outras ações do programa. Logo, orientar esse professor a colaborar mais cientificamente pode causar desconforto.

Resposta do grupo de gestores do cenário 2

(X) Concordo plenamente

Você sugere outras ações de gestão?

Sugestões do grupo de gestores do cenário 1:

Ação B: Desenvolver mecanismos para envolver os pesquisadores em ações que propiciem a colaboração.

Sugestões do grupo de gestores do cenário 2:

Ação C: Divulgação dos resultados para os demais professores.

Quadro 15 - Avaliação da tarefa 3

3. Quem são os pesquisadores mais próximos a todos os outros pesquisadores?

Em uma perspectiva de rede, maior proximidade refere-se ao menor caminho entre dois nodos.

D2

() Alto (X) Médio () Baixo

Ambos sentiram dificuldade em entender a pergunta e também a resposta, que é dada em função da métrica de Centralidade Global de Proximidade. Em função disso, não se sentiram confortáveis para

<p>avaliar como Alto.</p> <p>Para um gestor do Grupo 1 o desconforto veio da ausência de um pesquisador que ele achou que deveriam estar entre os cinco apresentados. Na análise posterior a essa entrevista, foi verificado que esse pesquisador encontra-se na posição 7 do ranking.</p>
<p>D3</p>
<p>Um grau alto de proximidade denota maiores possibilidades de estabelecimento de parcerias de publicação por estar mais próximos em relação a todos os outros (SOUZA; BARBASTEFANO; LIMA, 2012). Posições mais próximas a todas as outras podem alcançar outras posições mais rapidamente, obtendo e disseminando informação, auxiliando a transferência de informação (KRACKHARDT, 2010).</p> <p>A identificação desses pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Indiquem os mesmos para tarefas que demandem a disseminação de informações na rede, tais como, divulgação de eventos, etc.</p> <p>Resposta do grupo de gestores do cenário 1 e do cenário 2: <input checked="" type="checkbox"/> Concordo plenamente - Consideraram que esses pesquisadores por estarem mais próximos aos demais, são disseminadores de informação e formadores de opinião que podem “levar aos demais algo que se precisa”, mas também podem ser considerados receptores, pois recebem informações do grande grupo.</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo plenamente</p>
<p>Você sugere outras ações de gestão?</p> <p>Sugestões do grupo de gestores do cenário 1: Ação B: Contatar esses pesquisadores, na situação de receptores de informações, para obter informações sobre situações específicas.</p>

<p>4. Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram?</p>
<p>D2</p> <p>() Alto (X) Médio () Baixo</p> <p>Ambos sentiram dificuldade em entender a pergunta e também a resposta, que é dada em função da métrica de Centralidade Global de Intermediação. Em função disso e dos valores muito baixos em ambos os cenários, não se sentiram confortáveis para avaliar como Alto.</p>
<p>D3</p> <p>Um grau alto de intermediação indica o poder de intermediação ou conexão de grupos (MARTINS et al., 2012). Pode ser considerado uma “ponte” entre distintos grupos de pesquisa, uma vez que está no caminho mais curto entre outros nodos (FISCHBACH; PUTZKE; SCHODER, 2011).</p> <p>Ação A: A identificação desses pesquisadores permite que os gestores do PPG indiquem os mesmos para atuarem e/ou orientarem pesquisas de cunho multi ou interdisciplinar.</p> <p>Resposta do grupo de gestores do cenário 1</p> <p>(X) Concordo parcialmente</p> <p>- Consideraram que é importante ter em mente que existem várias outras variáveis de análise, mas que podem ser pontos de coesão do grupo, formadores de opinião.</p> <p>Resposta do grupo de gestores do cenário 2</p> <p>(X) Concordo plenamente</p>
<p>Você sugere outras ações de gestão?</p> <p>Sugestões do grupo de gestores do cenário 1:</p> <p>Ação B: Sugerir a indicação dos mesmos para representar o programa em alguma ação, que exija o olhar do todo, tais como reuniões na reitoria, assumir cargos de coordenação de área ou de programa.</p>

Quadro 17 - Avaliação da tarefa 5

<p>5. Quais os grupos de pesquisadores que trabalham isoladamente?</p>

D2
<p>(X) Alto () Médio () Baixo</p> <p>Ambos consideraram alto o grau de confiança. Entretanto, ambos sentiram dificuldade em entender a pergunta e sugeriram modificá-la para “Quais os grupos de pesquisadores que trabalham desconectados entre si? ”. Também sentiram dificuldade em entender a numeração dos grupos, que começa em 0 porque utiliza a mesma numeração dos grupos (componentes) do software de análise de redes.</p>
D3
<p>A existência de grupos que trabalham isoladamente denota fragmentação da rede de pesquisadores. Quanto mais grupos, mais fragmentada a rede está.</p> <p>A identificação desses grupos permite que os gestores do PPG:</p> <p>Ação A: Orientem os grupos isolados a atuarem de forma mais colaborativa (quando a natureza do PPG permitir).</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo parcialmente - Destacam que os gestores precisam atuar mais através da criação de condições, políticas, estratégias para que isso aconteça e menos orientando o pesquisador a fazer isso.</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Ação B: Orientem esses grupos a ampliarem seu número de colaboradores.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo parcialmente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo plenamente</p>
Você sugere outras ações de gestão?

Sugestões do grupo de gestores do cenário 1:

Ação C: Analisar outras variáveis relativas aos grupos ou pesquisadores isolados para sugerir um possível desligamento do programa. Nesse caso, os gestores destacam que é muito importante verificar se a situação é justificável, pois existem muitas formas do pesquisador colaborar com o programa, não apenas através da produção científica.

Quadro 18 - Avaliação da tarefa 6

6. Qual o nível geral de atividade de colaboração da rede?

Em uma perspectiva de rede, onde os pesquisadores estão ligados pela coautoria de produções, o nível geral de atividade de colaboração é determinado pela quantidade de ligações existentes comparado a quantidade de ligações possíveis (quando todos estão ligados com todos).

D2

Ambos sentiram dificuldade em entender a pergunta.

Resposta do grupo de gestores do cenário 1:

(X) Alto

- Deram como sugestão a possibilidade de ver esse índice por área de concentração. Dessa forma, esperariam ver um valor alto por área de concentração e um valor menor quando toda a rede fosse analisada.

Resposta do grupo de gestores do cenário 2:

(X) Médio

- Este grupo considerou o valor muito baixo (0,07) e por isso não se sentiu confortável para avaliar como Alto. Apesar disso, justificaram o número em função dos critérios da CAPES de avaliação da área, que contabiliza meio ponto para cada coautor. Destacaram que esse critério é bastante evidenciado nas reuniões de colegiado.

- Sugeriram utilizar projetos como um recorte adicional para analisar a colaboração, além das produções bibliográficas, pois muitos pesquisadores colaboram em projetos, mas isso não necessariamente se materializa em produções.

D3

Um alto grau de densidade indica alta atividade de colaboração (KATERNDAHL, 2012).

A identificação desse nível permite que os gestores do PPG:

Ação A: Orientem os pesquisadores a ampliar/diversificar suas parcerias de colaboração, quando o nível for baixo:

Resposta do grupo de gestores do cenário 1:

(X) Concordo parcialmente

- Argumentaram a partir de um determinado ponto não é salutar que todos colaborem com todos porque entendem que quantidade de produções pode cair, pois três pesquisadores tendem a produzir 1 artigo ao invés de 3.

- Entretanto, destacaram que novos indicadores de avaliação para a área interdisciplinar estão sendo criados para avaliar de forma mais direta a colaboração nos programas.

Ação B: Orientem os pesquisadores a manter os níveis atuais de colaboração e/ou expandir as fronteiras de colaboração para o nível inter PPG, interinstitucional e internacional, quando o nível for alto.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Você sugere outras ações de gestão?

Sugestões do grupo de gestores do cenário 1:

Ação C: Promover ações que levem ao aumento do nível geral de atividades de colaboração, ao invés de orientar os pesquisadores a fazer isso.

Quadro 19 - Avaliação da tarefa 7

7. Qual o nível médio de colaboração dos atores da rede?

Medido pela média das colaborações de todos os pesquisadores.

D2

(X) Alto () Médio () Baixo

- Ambos consideraram o nível alto.

D3

O grau médio reflete a média de coautores de cada pesquisador. Um

valor baixo indica que, na média, os pesquisadores não produzem muito em coautoria.

A identificação desse nível permite que os gestores do PPG:

Ação A: Orientem os pesquisadores a colaborar mais, no caso do nível ser baixo, ou a incentivar a continuação da colaboração, caso o nível seja alto.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Resposta do grupo de gestores do cenário 2:

(X) Concordo parcialmente

- Consideram importante achar o índice médio de colaboração ideal, porque consideram um índice médio alto não seja tão adequado.

Quadro 20 - Avaliação da tarefa 8

8. Quem são os colaboradores do pesquisador “x”?
D2
() Alto (X) Médio () Baixo - Ambos ao testarem a tarefa consigo mesmo como pesquisador “x”, sentiram falta de um ou dois pesquisadores.
D3
A identificação desses colaboradores/pesquisadores permite que os gestores do PPG:
Ação A: Na presunção de que esses colaboradores pesquisem nos mesmos tópicos do pesquisador “x”, indiquem os mesmos para formação de futuras parcerias de pesquisa na área do pesquisador “x”.
Resposta do grupo de gestores do cenário 1:
(X) Discordo parcialmente.
- Discordou da premissa da pergunta porque num programa interdisciplinar os tópicos que os outros pesquisadores trabalham podem ser complementares e não tópicos da área do pesquisador “x”.
- Argumentaram que talvez a resposta explique a maneira como a interdisciplinaridade é construída no programa, ou seja, ajuda a enxergar

as conexões de um pesquisador de uma área com outras áreas, mas não no mesmo tópico.

- Concordaram que é possível propor que esses pesquisadores formem um grupo de pesquisa.

Resposta do grupo de gestores do cenário 2:

(X) Concordo parcialmente.

- Consideraram que é difícil fazer esse tipo de orientação.

- Concordaram que uma forma de colaboração possível seria indicar esses pesquisadores para participação em bancas.

Ação B: Orientem os colaboradores do pesquisador “x” a procurar o próprio pesquisador “x” para fomentar outros tipos de colaboração, tais como, participação em projetos.

Resposta do grupo de gestores do cenário 1:

(X) Concordo parcialmente

- Entenderam que a existência de laços de colaboração científica (via produção bibliográfica) já é uma consequência de outras ações desenvolvidas em conjunto dentro do programa.

Resposta do grupo de gestores do cenário 2:

(X) Concordo parcialmente

Você sugere outras ações de gestão?

Sugestões do grupo de gestores do cenário 1:

Ação C: Entender os laços do pesquisador, de forma a diagnosticar a sua atuação no programa. “Laços evidenciam preferências”.

Quadro 21 - Avaliação da tarefa 9

9. Quem são os pesquisadores que colaboram no tópico “y”?

D2

(X) Alto

- Ambos consideraram alto o grau de confiança.

- Ambos consideraram que a semântica da pergunta não estava adequada para a forma como os resultados foram recuperados (a partir das

palavras-chave das produções desses pesquisadores). Sugeriram mudar para “Quem são os pesquisadores que pesquisam/ trabalham no tópico “y”?”

Resposta do grupo de gestores do cenário 1:

- Sugeriram trabalhar com a dimensão tempo para evidenciar o surgimento e desaparecimento de tópicos ao longo do tempo.
- Destacaram que a resposta desta questão permite evidenciar qual a cobertura de um tópico dentro do programa.

D3

“A identificação de temáticas permite definir estratégias de atuação e posicionamento. Por exemplo, temas relevantes em nível nacional e internacional, mas pouco desenvolvidos no âmbito do PPG, podem ser desenvolvidos através da busca por fomento ou parcerias. Ao contrário, temas bem desenvolvidos por membros do programa, são muitas vezes latentes, ou seja, não são conhecidos da comunidade, e podem ser mais explorados através da agregação de mais pessoas e consequente criação de um grupo de pesquisa”.

A identificação desses pesquisadores permite que os gestores do PPG:

Ação A: Orientem os pesquisadores a formarem parcerias de colaborações de pesquisa no tópico “y”.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

Ação B: Direcionem as oportunidades de ações nesse tópico dentro do programa, tais como participação em projetos, editais de fomento à pesquisa, oportunidades de experiências no exterior (pós-doc, professor visitante, etc.).

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente
Você sugere outras ações de gestão?
Sugestões do grupo de gestores do cenário 1: Ação C: Sugerir a esses pesquisadores a orientação de alunos e projetos de pesquisa com essa temática (por exemplo, em um processo de seleção).

Quadro 22 - Avaliação da tarefa 10

10. Quem é o pesquisador que mais colabora no tópico “y”?
D2
Ambos consideram que a semântica da pergunta não estava adequada considerando a forma como os resultados foram recuperados (a partir das palavras-chave das produções desses pesquisadores). Sugeriram mudar para “Quem é o pesquisador que mais pesquisa/trabalha no tópico “y”?”, deixando mais claro que o foco da pergunta está no nó.
Resposta do grupo de gestores do cenário 1: (X) Alto - Sugeriram mostrar mais pesquisadores na resposta e não somente o que mais colabora.
Resposta do grupo de gestores do cenário 2: (X) Baixo - Consideraram baixo porque testaram alguns tópicos (termos) e os resultados não pareceram confiáveis: - tópico “ontologia”: não resultou no pesquisador que mais publica nessa área. Entretanto, verificações e testes posteriores a entrevista identificaram que esse pesquisador utiliza o termo “ontologias” nas palavras-chave das suas produções. - tópico “ <i>images</i> ”: não resultou no pesquisador que mais publica nessa área. Entretanto, verificações e testes posteriores a entrevista identificaram que esse pesquisador não utilizou esse termo nas palavras-chave das suas produções.
D3

A identificação desse pesquisador permite que os gestores do PPG:

Ação A: indiquem o pesquisador mais qualificado (que produziu mais no referido tópico) para formação de futura parceria de pesquisa no tópico “y”.

Resposta do grupo de gestores do cenário 1:

(X) Concordo parcialmente

- Argumentaram que o termo “qualificado” é forte, apesar de estar explicado na descrição da ação, e sugeriram usar “ativo”.

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

Você sugere outras ações de gestão?

Sugestões do grupo de gestores do cenário 1:

Ação B: Direcionem as oportunidades de ações nesse tópico, tais como participação em projetos, editais de fomento à pesquisa, para este pesquisador.

Ação C: Sugerem que o foco da ação saia da pessoa/pesquisador e vá para o entendimento ou diagnóstico de porque as pessoas estão ou não estão trabalhando com esse tema. A partir deste entendimento, novas estratégias podem ser propostas.

Quadro 23 - Avaliação da tarefa 11

11. Quais grupos de pesquisadores colaboram no tópico “y”?

D2

() Alto (X) Médio () Baixo

Ambos consideraram a análise dos resultados um pouco difícil e por isso sentiram-se mais confortáveis em informar o grau de confiança médio.

Resposta do grupo de gestores do cenário 1:

- Acharam a tarefa muito interessante.

- Ao analisar o tópico “conhecimento”, que conduziu a descoberta de oito grupos distintos, inferiram que o tópico é bem transversal no programa. A partir desse teste, também foi inferido que a situação ideal para o programa interdisciplinar, pode ser a existência de poucos grupos

concentrados no tópico.
D3
A identificação desses grupos de pesquisadores permite que os gestores do PPG:
Ação A: Orientem grupos distintos a unirem forças para participar de editais de fomento à pesquisa no referido tópico.
Resposta do grupo de gestores do cenário 1: (X) Concordo plenamente
Resposta do grupo de gestores do cenário 2: (X) Concordo plenamente
Ação B: Orientem os distintos grupos de pesquisadores que colaboram no referido tópico a unirem forças para produzir de forma mais qualitativa.
Resposta do grupo de gestores do cenário 1: (X) Concordo parcialmente - Considera o termo qualitativa forte. Sugere alterar para integrada .
Resposta do grupo de gestores do cenário 1: (X) Concordo parcialmente - Considera o uso do termo qualitativo “complicado” porque a união dos esforços não é indicativa que terá qualidade.

Quadro 24 - Avaliação da tarefa 12

12. Qual o perfil das colaborações do pesquisador “x”?
D2
Resposta do grupo de gestores do cenário 1: (X) Alto - Sugeriram modificar a pergunta para “Qual o perfil de pesquisa do pesquisador “x”? ”, pois o foco de análise é o nodo pesquisador e não a sua rede de colaboração. - Ao analisar o seu perfil (composto de termos com a respectiva frequência), um dos gestores se surpreende e percebe como sua história

passada, ou seja, os tópicos que pesquisou há muitos anos ainda é forte.

Resposta do grupo de gestores do cenário 2:

(X) Médio

- Ao testar o seu perfil, um dos gestores estranhou a existência do termo “biblioteca digital” com frequência 6, apesar de confirmar que já tinha trabalhado com este tópico há alguns anos. Verificações e análises posteriores confirmaram a existência deste termo em títulos de produções deste pesquisador nos anos de 2007 e 2009.

D3

A identificação desse perfil permite que os gestores do PPG:

Ação A: Indiquem o pesquisador para colaborar nos tópicos que compõem o referido perfil. Por exemplo:

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

- Entenderam que partindo de uma intuição ou escolha inicial de um pesquisador, é possível confirmar e posteriormente indicar o pesquisador para participação em projeto de pesquisa, participação em banca, onde o colegiado ou alguém que o represente, precise aprovar a participação.

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

- Foi dado como exemplo pela entrevistadora, a situação onde um aluno tenha interesse em trabalhar com um pesquisador “x”, mas desconheça os seus tópicos de pesquisa. Os gestores concordaram com o exemplo dado.

Ação B: Indiquem o pesquisador para orientar alunos com interesse nos tópicos do perfil.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

Quadro 25 - Avaliação da tarefa 13

13. **Quem são os pesquisadores que possuem perfis similares a um perfil específico*?** *O perfil específico deve ser composto por um (1) ou mais termos definidos e inseridos pelo usuário do *framework*.

D2**Resposta do grupo de gestores do cenário 1:**

(X) Alto

- Destacaram que apesar da análise ter sido feita através de uma dimensão (produção bibliográfica), o resultado permite fazer um diagnóstico do pesquisador e por consequência do programa.

Resposta do grupo de gestores do cenário 2:

(X) Médio

- Questionaram o termo “similar” e sugeriram usar “idêntico”. Foi respondido pela entrevistadora que “idêntico” sugere que somente os termos inseridos no perfil específico devem ser encontrados e que o perfil do pesquisador sempre será mais amplo que os termos inseridos. Argumentaram então, que por outro lado, a não existência no perfil do pesquisador de algum dos termos inseridos no perfil específico não poderia deixar de retornar esse pesquisador.

D3

A identificação desse perfil permite que os gestores do PPG:

Ação A: Indiquem os pesquisadores para colaborar e/ou orientar nos tópicos que compõem o referido perfil.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

Ação B: Sugiram pesquisadores para a composição de bancas.

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

<p>Resposta do grupo de gestores do cenário 2: (X) Concordo plenamente</p>
<p>Você sugere outras ações de gestão?</p> <p>Sugestões do grupo de gestores do cenário 1: Ação C: Realizar um diagnóstico de cada pesquisador, e por consequência do programa como um todo, que produza elementos para serem utilizados no planejamento estratégico do programa. Ação D: Convidar o pesquisador para ser avaliador de artigo.</p>

Quadro 26 - Avaliação da tarefa 14

<p>14. Qual o grau de similaridade (<i>matching</i>) dos perfis de colaboração de dois pesquisadores? *Medido pela similaridade dos perfis dos pesquisadores.</p>
<p>D2</p>
<p>Resposta do grupo de gestores do cenário 1: (X) Alto</p> <p>Resposta do grupo de gestores do cenário 2: (X) Médio</p>
<p>D3</p> <p>A identificação desse grau permite que os gestores do PPG:</p> <p>Ação A: Orientem pesquisadores com alto grau de similaridade, que não possuem interação, a unirem forças e colaborarem em prol da quantidade e qualidade das produções.</p> <p>Resposta do grupo de gestores do cenário 1: (X) Concordo plenamente - Consideraram “orientar” invasivo, sugeriram modificar para “promover”.</p> <p>Resposta do grupo de gestores do cenário 2: (X) Concordo plenamente</p> <p>Ação B: Indiquem pesquisadores com alto grau de similaridade para orientação e coorientação de pesquisa de mestrado ou doutorado (quando a natureza do programa assim permitir).</p>

<p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo parcialmente. -Argumentaram que é salutar ter orientadores de áreas diferentes.</p>
<p>Ação C: Indiquem ou busquem pesquisadores com alto grau de similaridade para a composição de bancas.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo plenamente</p>

Quadro 27 - Avaliação da tarefa 15

<p>15. Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador “x”?</p>
<p>D2</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Alto</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Alto</p>
<p>D3</p> <p>A identificação desse grau permite que os gestores do PPG:</p> <p>Ação A: Orientem os pesquisadores que possuem perfis semelhantes para colaborar.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo parcialmente. - Não teve justificativa.</p>

<p>Você sugere outras ações de gestão?</p> <p>Sugestões do grupo de gestores do cenário 1: Ação B: Verificar a aderência de perfis e promover a integração de pesquisadores novos no programa.</p> <p>Ação C: Verificar a aderência de perfis no processo de credenciamento e recredenciamento de novos pesquisadores no programa.</p>

Quadro 28 - Avaliação da tarefa 16

<p>16. Quem são os pesquisadores que, com base na proximidade e na similaridade de seus perfis com o pesquisador “x”, podem colaborar com esse pesquisador?</p>
<p>D2</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Alto - Sugeriram fazer uma ponderação da variável proximidade. Aqueles que estão a um (1) grau de distância não trazem surpresas, aqueles que estão há dois graus podem interessar mais.</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Alto</p>
<p>D3</p> <p>A identificação desses pesquisadores permite que os gestores do PPG:</p> <p>Ação A: indiquem os mesmos ao pesquisador “x”, fomentando as possibilidades de colaboração.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo parcialmente.</p>

Quadro 29 - Avaliação da tarefa 17

17. Qual o perfil dos pesquisadores que mais colaboram?
D2
<p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Alto</p> <ul style="list-style-type: none"> - Destacaram que com esse resultado é possível verificar se esse grupo mais colaborativo pertence a uma das áreas de concentração do programa ou se está distribuído nas outras áreas. - Concordaram que é possível indicá-los para as ações de uma forma mais orientada (através do perfil) do que somente pela resposta da tarefa 1 (só exibe os nomes). <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Alto</p>
D3
<p>A identificação desses pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Indiquem, de forma mais qualificada (baseada no perfil), os mesmos para atuarem em projetos de pesquisa, organização de eventos, etc.</p> <ul style="list-style-type: none"> - As respostas e justificativas seguem a avaliação da tarefa 1. <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo parcialmente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo parcialmente</p>
<p>Ação B: Indiquem, de forma mais qualificada (baseada no perfil), os mesmos para atuarem como disseminadores de ideias, eventos, etc.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo parcialmente.</p>

Quadro 30 - Avaliação da tarefa 18

18. Qual o perfil dos grupos que trabalham isoladamente?
D2
<p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Alto - Foi observado que o termo “conhecimento”, objeto de estudo deste programa, tinha uma frequência muito baixa, o que conduziu a proposição de uma nova ação.</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Alto</p>
D3
<p>A identificação do perfil desses grupos permite que os gestores do PPG:</p> <p>Ação A: orientem os mesmos para atuarem de forma mais colaborativa.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Discordo totalmente. - Argumentaram que o perfil do maior grupo (subgrafo desconectado com mais nodos) fica muito grande e confuso para analisar. Não acham que seja possível analisar e realizar alguma ação em cima do maior grupo. - Entretanto, concordaram que para os grupos menores (subgrafo desconectado com um ou mais nodos) isso é possível. - Destacaram que a qualificação com o perfil não faz diferença para esse tipo de ação e que a resposta da tarefa 1 já seria suficiente para essa ação.</p> <p>Ação B: Orientem os grupos pequenos a ampliarem seu número de colaboradores.</p> <p>Resposta do grupo de gestores do cenário 1: <input checked="" type="checkbox"/> Concordo plenamente</p> <p>Resposta do grupo de gestores do cenário 2: <input checked="" type="checkbox"/> Concordo plenamente</p>

<p>Você sugere outras ações de gestão?</p> <p>Sugestões do grupo de gestores do cenário 1: Ação C: Realizem um diagnóstico dos tópicos que estão sendo pesquisados nos grupos e proponham ações estratégicas. Por exemplo: Notar que um determinado tópico de pesquisa é muito importante no programa, ou seja, é pesquisado, mas não aparece com frequência no grupo, pode gerar uma ação de conscientização da importância do uso e preenchimento deste tópico no currículo dos pesquisadores.</p>
--

A última pergunta, referente a Dimensão 1 da avaliação, foi realizada no final da entrevista com os grupos focais.

3) As perguntas ou tarefas de análise de colaboração científica propostas são relevantes para a análise da colaboração científica?

Resposta do grupo de gestores do cenário 1:

(X) Concordo plenamente

Resposta do grupo de gestores do cenário 2:

(X) Concordo plenamente

5.4 ANÁLISE DOS RESULTADOS

O objetivo desta seção é sintetizar e analisar os resultados da avaliação descritos na seção anterior. Para isso, os resultados foram sintetizados nos gráficos exibidos na Figura 54, Figura 55 e Figura 56 e no mapa mental exibido na Figura 57. A partir dessa síntese, que colocou as respostas da avaliação em evidência, ficou mais claro realizar a análise e chegar a algumas conclusões.

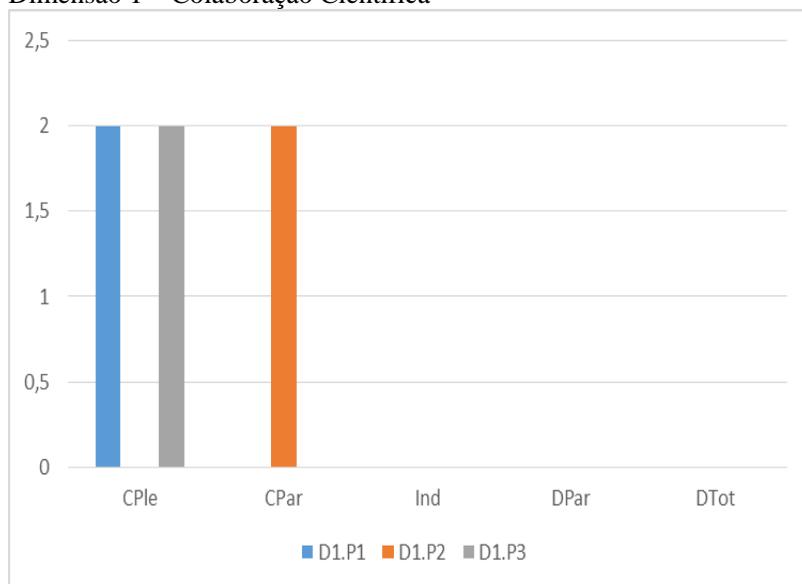
Contudo, antes de proceder a análise dos resultados, julga-se necessário destacar os diversos problemas encontrados na fonte de dados principal utilizada na operacionalização do framework nos dois cenários: o currículo Lattes. Os problemas encontrados foram de naturezas diversas, tais como: a) currículos desatualizados por parte dos pesquisadores³⁴; b) não-preenchimento das palavras-chave das

³⁴ Os currículos utilizados na demonstração foram baixados pela última vez no período de 04 a 09 de junho de 2015.

produções bibliográficas³⁵; c) preenchimento incorreto da produção (duplicidade, título da produção, local do evento, ano, etc.); d) não identificação da coautoria³⁶, etc. e impactaram diretamente nas respostas fornecidas pelo *framework*.

Para contornar algumas dessas limitações foram implementados alguns mecanismos. No caso da opção b, foi implementado um Módulo para extração dos termos mais frequentes a partir do título das produções dos pesquisadores. No caso da opção d, foi implementado um conjunto de regras para verificação da coautoria no Módulo de instanciação da ontologia de colaboração científica.

Figura 54 - Síntese das respostas da avaliação do *framework* na Dimensão 1 – Colaboração Científica



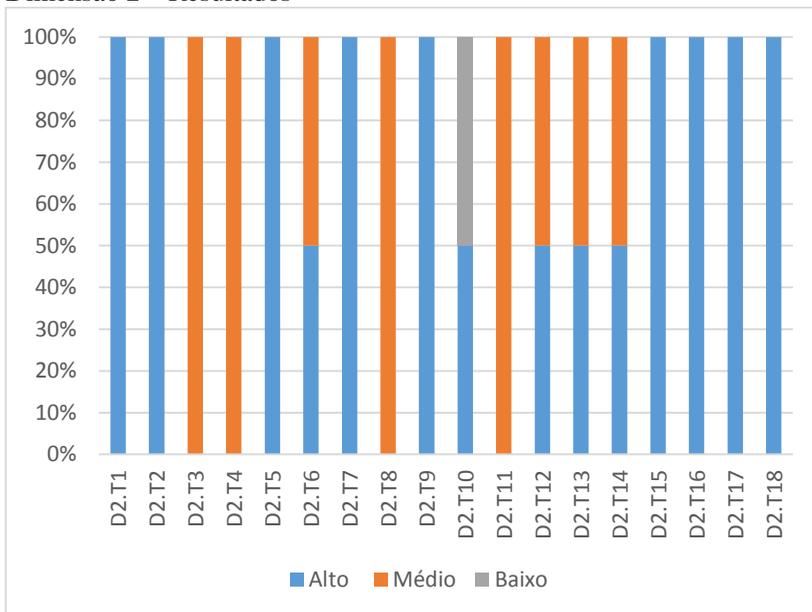
Fonte: Autor.

³⁵ Em ambos cenários em torno de 40% a 50% das produções não tinham nenhuma palavra-chave cadastrada.

³⁶ A identificação de coautoria é um módulo oferecido pelo sistema de preenchimento do currículo Lattes que permite confirmar o coautor de uma produção. Com isso, o identificador correto do coautor (NRO-ID-CNPQ) fica vinculado a produção intelectual no currículo.

Em relação a Dimensão 1 (D1) e a pergunta que aborda a importância do estabelecimento dos relacionamentos de colaboração (P1), bem como a pergunta que avalia a relevância do conjunto de perguntas propostas no *framework* (P3) observa-se que os dois grupos responderam CPl (Concordo Plenamente). Em relação a pergunta sobre o uso dos resultados de análise da colaboração como insumo para a tomada de decisão (P2) a resposta foi CPar (Concordo Parcialmente) pois ambos os grupos de gestores consideraram não ser possível tomar muitas decisões em suas posições.

Figura 55 - Síntese das respostas da avaliação do *framework* na Dimensão 2 – Resultados



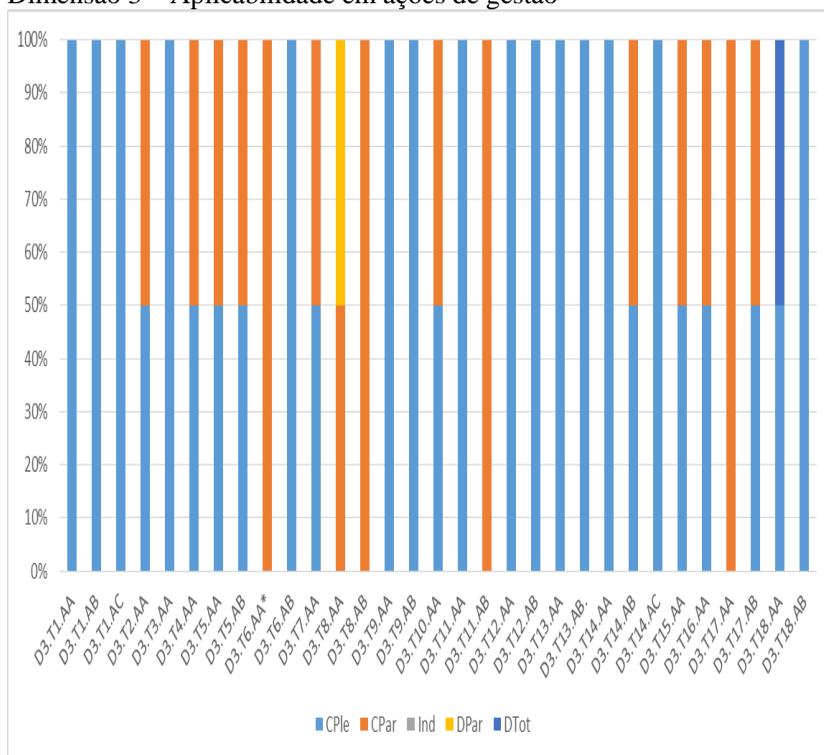
Fonte: autor.

Em relação a Dimensão 2 (D2), que mediu o grau de confiança dos resultados, é importante destacar que o julgamento baseou-se na percepção pessoal de cada grupo de gestores e essa percepção é fruto da experiência dos gestores, mas principalmente do conhecimento dos mesmos sobre os pesquisadores (seus colegas no PPG) da rede.

Posto isso, percebe-se que nove (9) das dezoito (18) perguntas obtiveram o total de respostas (2 respostas) no nível Alto e três (3) perguntas obtiveram o total de respostas no nível Médio. O grau de confiança Baixo foi dado em somente uma (1) pergunta e correspondeu a 1 dos grupos respondentes.

De uma outra forma é possível dizer que o grau de confiança Alto corresponde a 63% do total, e o grau de confiança Médio a 33%, o que é considerado bastante adequado diante do problema da qualidade da fonte de dados utilizada. Ainda assim, para muitas respostas onde o grau de confiança foi considerado Médio e até mesmo Baixo (1 resposta), análises posteriores revelaram a fonte do problema. A justificativa das respostas da maior parte das perguntas está descrita na subsecção 5.3.2

Figura 56 - Síntese das respostas da avaliação do *framework* na Dimensão 3 – Aplicabilidade em ações de gestão

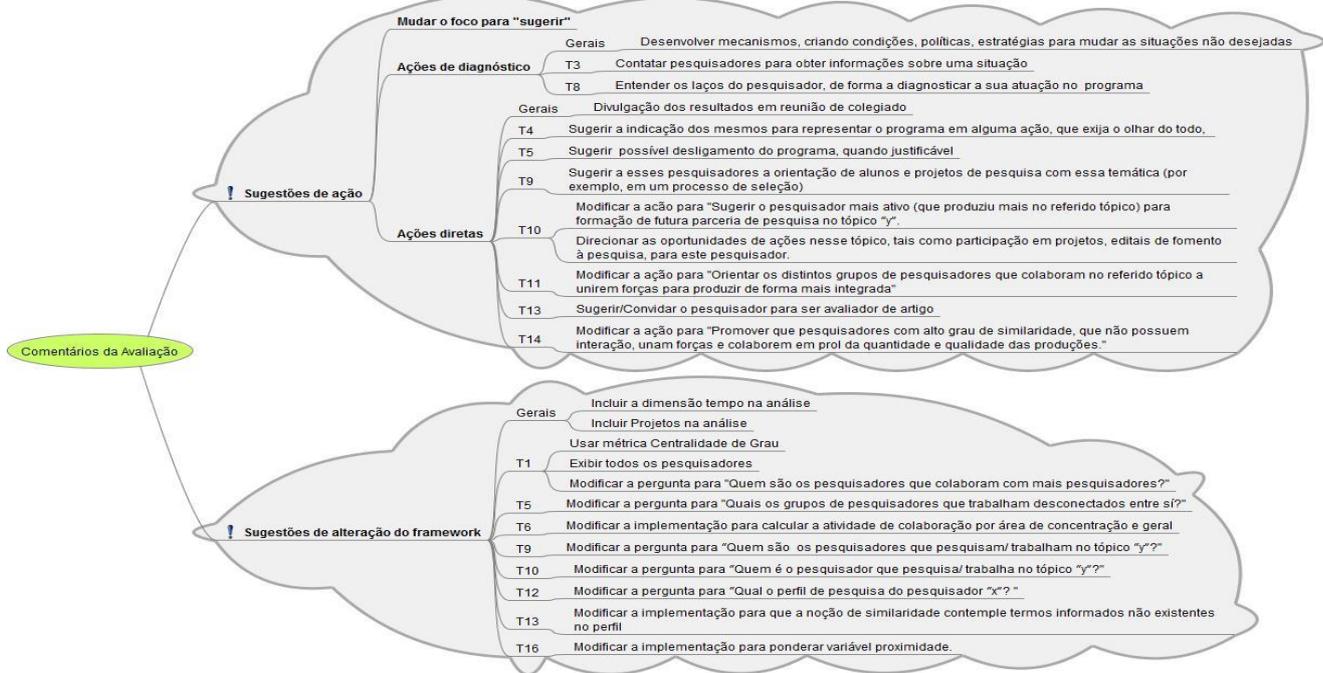


Fonte: autor.

Em relação a Dimensão 3 (D3), aplicabilidade do *framework* como instrumento de gestão, percebeu-se que as ações de gestão propostas tiveram uma aceitação plena, com os dois grupos de respondentes escolhendo CPlé (Concordo Plenamente), em 15 das 31 ações de gestão propostas, seguido de respostas divididas em CPlé e CPar em 10 ações de gestão. As opções DPar (Discordo Parcialmente) e DTot (Discordo Totalmente) apareceram somente em duas ações de gestão. Além das ações avaliadas, várias novas ações foram propostas pelos gestores como mostra o mapa mental da Figura 57.

O mapa mental foi desenvolvido para organizar os comentários referentes as sugestões relacionadas diretamente ao *framework* e as sugestões relacionadas as novas ações de gestão propostas.

Figura 57 - Mapa mental com as sugestões da avaliação



Fonte: Autor.

Em relação as sugestões de novas ações de gestão, percebe-se a existência de ações de cunho mais direto, ou orientadas a ação propriamente dita e sugestões orientadas ao diagnóstico, ao entendimento de uma situação, que gere subsídios para um futuro estabelecimento de novas ações, tanto de cunho mais operacional, quanto de cunho mais estratégico.

Todas elas são frutos da conversa frutífera que foi desenvolvida com os experientes gestores dos PPGs avaliados. Talvez seja importante destacar que elas denotam visões diferentes, e isso fica mais explícito na seção anterior, de gestores de PPG onde em um caso a natureza do conhecimento é disciplinar e em outro é interdisciplinar. Acima de tudo, elas denotam o tempo de experiência em gestão dos mesmos.

Finalmente, em relação as sugestões de mudanças no *framework*, entende-se que apesar de algumas terem vindo de um ou de outro grupo, todas são pertinentes, factíveis e tornarão o *framework* mais adequado ao seu propósito principal, que é analisar a colaboração científica de um determinado contexto.

6 CONSIDERAÇÕES FINAIS

Este trabalho foi concebido com base em uma problemática de pesquisa que se construiu a partir de três oportunidades de pesquisa não exclusivas do contexto da colaboração científica, mas certamente relacionadas a ela, encontradas na literatura.

A primeira delas está relacionada à forma de representação de uma rede, que normalmente é realizada através de grafo ou matrizes, e com isso perde parte da semântica relacionada aos elementos dessa rede (nodos e relacionamentos), pois não permite especificar mais informações acerca desses elementos. A segunda está relacionada ao conteúdo dos artefatos produzidos pelas relações de uma rede, que não são explorados, mas tem um potencial para complementar a análise da rede. Por fim, o terceiro relaciona-se com o problema do conhecimento especializado necessário para realizar tarefas de análise de colaboração científica. Entendeu-se, portanto, que o problema situa-se essencialmente na representação de conhecimentos distintos (domínio, tarefa e solução de tarefa) envolvidos em um processo de análise de rede de colaboração científica.

A representação de conhecimento de uma rede de colaboração científica, com os elementos que representam o seu domínio e elementos que representam a forma de raciocinar sobre esse domínio, fornecem uma estrutura que permite análises elaboradas sobre a referida rede, possibilitando com isso um amplo entendimento a respeito do contexto de colaboração.

Assim, foi proposto e desenvolvido um *framework* composto por um conjunto de representações dos conhecimentos envolvidos na análise de uma rede de colaboração científica, bem como um conjunto de artefatos (módulos) para armazenar e recuperar dados dessas representações, com o objetivo de auxiliar o processo de análise dessa rede.

Esse conjunto de artefatos e sua aplicabilidade neste e em outros domínios caracterizam a contribuição científica deste trabalho. A aplicabilidade do *framework* como suporte a ações gestão, verificada no processo de avaliação, caracteriza a contribuição social.

6.1 CONTRIBUIÇÕES

Como contribuições primeiramente destacam-se as ontologias resultantes da modelagem do conhecimento de domínio e do conhecimento das tarefas de análise.

O conhecimento do domínio da colaboração científica e das métricas de ARS foram representadas através da ontologia *VIVO-ISF* (*vivo-isf-public-1.6.owl*) e *SNAMetrics* (*SNAMetrics.owl*), respectivamente. As ontologias de domínio deste trabalho não foram criadas a partir do zero. Baseando-se no princípio da reutilização, foram utilizadas ontologias já existentes no mercado e na literatura. No entanto, elas foram refatoradas e essa refatoração indica uma contribuição de melhoria das mesmas.

Na ontologia de colaboração científica *VIVO-ISF* foi incluído um novo tipo do conceito *Relationship* denominado *ConceptOwnership*, que permite o estabelecimento de relacionamentos entre conceitos do tipo *Person* com conceitos do tipo *Concept*, possibilitando assim a representação e armazenamento do perfil de um pesquisador.

A ontologia de métricas de ARS *SNAMetrics* foi totalmente refatorada a partir de outra ontologia encontrada na literatura, com a remoção e criação de novos conceitos e relacionamentos, o que a caracteriza como uma nova ontologia independente, que pode ser utilizada em aplicações diversas.

O conhecimento das tarefas intensivas em conhecimento foi modelado através de uma técnica de decomposição de tarefas, a partir da identificação inicial das necessidades de análise de colaboração científica mais relevantes. Com esta técnica, a estrutura das tarefas de colaboração foi desmembrada, permitindo tanto a identificação dos passos de inferência necessários para cumprir a tarefas (fornecer a resposta), como a identificação do conhecimento de domínio utilizado nas tarefas. A partir desse processo de modelagem foram criadas a ontologia de tarefas de colaboração científica (*TarefaColaboracao.owl*), e as ontologias de solução das tarefas através de duas ontologias de métodos (*MetodoComplexo.owl* e *MetodoPrimitivo.owl*).

A ontologia *TarefaColaboracao.owl* representa um conjunto relevante de tarefas de análise de colaboração, classificadas de acordo com o método que as responde (ARS, análise de conteúdo ou ambos). Ela foi criada de forma independente das ontologias de métodos e com isso, possibilita a sua associação com outros métodos modelados e representados de formas diferentes nas ontologias de métodos, ou seja, uma determinada tarefa dessa ontologia pode ser associada (configurada) com outras formas de realizá-la.

A ontologia de métodos foi dividida para representar os métodos complexos de maneira independente dos métodos primitivos, gerando assim, as ontologias *MetodoComplexo.owl* e *MetodoPrimitivo.owl*.

A ontologia *MetodoComplexo.owl* contempla os tipos de métodos que realizam as tarefas de colaboração representadas na ontologia *TarefaColaboracao.owl* (classes do tipo *MetodoComplexo*), bem como os tipos que representam a decomposição desses métodos em subtarefas (classes do tipo *SubTarefa*). Os tipos de *MetodoComplexo* correspondem aos métodos de ARS, análise de conteúdo e métodos que usam as duas abordagens, e nessa ontologia suas instâncias já estão relacionadas com as instâncias do tipo *SubTarefa*. Entretanto, essa configuração pode ser modificada e/ou estendida. Em relação aos tipos de *SubTarefa*, muitos são reaproveitados em métodos complexos diferentes.

A ontologia *MetodoPrimitivo.owl* contempla os tipos de métodos que efetivamente realizam as tarefas de colaboração, ou de maneira mais específica, as subtarefas especificadas na ontologia *MetodoComplexo*. A representação destes métodos também pode ser utilizada em aplicações de outros domínios.

Outra contribuição relacionada aos artefatos produzidos, refere-se aos módulos implementados que fazem parte das camadas superior e inferior da estrutura do *framework* (Figura 26). Sem esses módulos, existiria apenas o conjunto de representações ontológicas expostas anteriormente. Logo, suas instâncias deveriam ser inseridas manualmente ou por outros processos automatizados que teriam que ser desenvolvidos. Neste trabalho, esses módulos foram propostos e implementados.

Os módulos da camada inferior, de pré-processamento de dados e instanciação das ontologias, são específicos para uma fonte de dados, o currículo Lattes. Todavia, justifica-se a sua inclusão na estrutura do *framework* por tratarem dados curriculares de referência nacional que expressam evidências de colaboração científica.

Os módulos da camada superior, de recuperação de dados das ontologias, também caracterizam-se como contribuições. No Módulo API dos Métodos Primitivos estão implementados todos os métodos que efetivamente recuperam os dados das ontologias de domínio, possibilitando o fornecimento da resposta das tarefas de colaboração. Os métodos implementados na API são genéricos e podem ser utilizados para a recuperação de dados em qualquer representação em OWL ou RDF. Neste trabalho um conjunto de vinte e quatro (24) métodos

consegue realizar todas as setenta e três (73) subtarefas de análises representadas.

O Módulo de Recuperação de Tarefa e Métodos também é genérico e é capaz de recuperar instâncias de qualquer ontologia de tarefas e métodos representada de maneira similar à que foi utilizada neste trabalho.

Entende-se que além das contribuições individuais dos artefatos produzidos, existe ainda a contribuição do conjunto de todos eles, que combinados constituem um *framework* que potencializa, com recursos de representação semântica as possibilidades de análise de uma rede de colaboração científica. Ou seja, existe uma estrutura que permite o armazenamento e recuperação de informações que vai além das tarefas modeladas e representadas no *framework*, seja através da modelagem e representação de novas tarefas de análise ou simplesmente através da recuperação de dados diretamente das representações.

Além disso, no que tange a contribuição social, relacionada à utilidade do mesmo no contexto para o qual foi construído, utiliza-se primeiramente a fala de um dos gestores no processo de avaliação, “tem-se um instrumento que mostra com lentes nunca vistas, a situação real, ou seja, o que está acontecendo no PPG”.

Para além do contexto onde o *framework* foi demonstrado, ele pode ser utilizado em qualquer cenário que tenha evidências de trabalho em coautoria. E nesse cenário, pode ser utilizado em níveis diferentes, do operacional (pesquisador) até o estratégico (gestor), bem como pela comunidade interna ou externa ao cenário, que tenha interesse nas respostas que o mesmo fornece. Como instrumento em nível de gestão, ele permite uma série de ações que foram geradas no processo de avaliação, caracterizando assim a aplicabilidade do mesmo como instrumento de gestão.

6.2 TRABALHOS FUTUROS

Como trabalhos futuros pretende-se primeiramente implementar as sugestões de modificações relacionadas ao *framework* propriamente dito e expostas na seção de Análise dos Resultados. Na seqüência destacam-se todas as possibilidades de extensão deste trabalho.

A primeira delas trata a utilização do *framework* com outras unidades de análise. Neste trabalho ele foi concebido, desenvolvido e avaliado para a unidade de análise pesquisador, sendo este o nodo da rede de colaboração. Entretanto, entende-se que exista o interesse em

analisar outros “grãos”, como por exemplo, o programa de pós-graduação (PPG), que neste caso permitiria a avaliação da colaboração entre PPGs distintos. Para isso, ao se utilizar a mesma fonte de dados (Lattes), os módulos de pré-processamento e instanciação das ontologias devem ser modificados e possivelmente novos módulos necessitem ser desenvolvidos para processar o conjunto de produções dos pesquisadores do PPG como uma unidade única e verificar as relações de colaboração com outros PPGs.

A segunda refere-se a modelagem de tarefas que envolvam outras métricas ou algoritmo de ARS, tais como, a detecção de comunidades também está prevista, assim como, tarefas que explorem mais a combinação das métricas de ARS com processos de mineração de conteúdo.

A terceira aborda a extensão da estrutura de representação do *framework* para comportar o armazenamento das respostas das tarefas de colaboração na própria estrutura. Por exemplo, a resposta da tarefa “Quem são os pesquisadores que mais colaboram” consiste em instância do tipo *FacultyMember* já armazenada no próprio *framework*. Essas instâncias podem ser vinculadas, por exemplo, através de uma *object property* a instância da tarefa *TarefaPesquisadoresMaisColaboram*, permitindo com isso que as respostas às tarefas não tenham que ser processadas a cada vez que forem demandadas.

Outra possibilidade refere-se ao uso da linguagem SPARQL para calcular as métricas de ARS na própria estrutura de armazenamento em grafo RDF, evitando assim o uso de um *software* externo que calcule as métricas separadamente.

Por fim, percebeu-se a importância da dimensão temporal nas análises realizadas. Essa dimensão temporal permitiria que forem tirados *snapshots* de fatias de tempo desejados para análise que conduziram a análises mais precisas e comparativas. Para viabilizar essa característica, entende-se que alterações substanciais dos artefatos de representação e dos módulos de coleta e recuperação de dados seriam necessárias.

REFERÊNCIAS

- ABAZI-BEXHETI, L.; KADRIU, A. A Uniform Semantic Web Framework for Co-authorship Networks. IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing. **Anais...**, p. 958-965, 2011. Sydney, Australia.
- ABBASI, A.; HOSSAIN, L.; LEYDESDORFF, L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. **Journal of Informetrics**, v. 6, n. 3, July 2012, p. 403-412.
- ADAMS, L. A. Delphi forecasting: future issues in grievance arbitration. **Technological Forecasting and Social Change**, v. 18, p. 151-160, 1980.
- ALMACK, John C. The influence of intelligence on the selection of associates. **School and Society**, 16:529-530, 1922.
- BALANCIERI, R. **Um Método Baseado em Ontologias para Explicitação de Conhecimento Derivado da Análise de Redes Sociais de um Domínio de Aplicação**. Tese apresentada ao PPEGC. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Programa de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, 2010.
- BARABÁSI, A. L. et al. Evolution of the social network of scientific collaborations. **Physica A: Statistical Mechanics and its Applications**, 311(3-4), 590-614, 2002.
- BASTIAN, M.; HEYMANN, S.; JACOMY, M. **Gephi: an open source software for exploring and manipulating networks**. In: International AAAI Conference on Weblogs and Social Media, 2009.
- BATAGELJ, V.; MRVAR, A. Pajek - Analysis and Visualization of Large Networks. In: Jünger, M., Mutzel, P., (Eds.). **Graph Drawing Software**. Springer, Berlin, 2003. p. 77-103.
- BAVELAS, A. Communications Patterns in Task Oriented Groups. **Journal of the Acoustical Society of America**, v.22, 1950.
- BEAVER, D. d. ROSEN, R. Studies in Scientific Collaboration: Part II - Scientific Coauthorship, Research Productivity and Visibility in the French Scientific Elite, 1799-1830. **Scientometrics**, v.1, n. 2, p.133-49, 1979.

- BEAVER, D. d. Reflections on scientific collaboration (and its study): past, present, and future. **Scientometrics**, v.52, n.3, p.365-377, 2001.
- BEAVER, D. d.; ROSEN, R. Studies in Scientific Collaboration: Part I – The Professional Origins of Scientific Co-authorship. **Scientometrics**, v.1, p. 65-84, 1978.
- BENJAMINS, R., JANSWEIJER, W. Toward a competence theory of diagnosis. **IEEE Expert: Intelligent Systems and Their Applications**, v.9, p. 43–52, 1994.
- BENJAMINS, V. R; FENSEL, D. Editorial: problem-solving methods. **International Journal of Human-Computer Studies**, v. 49, n.4, p. 305–313, 1998.
- BERKOWITZ, Stephen D. **An Introduction to Structural Analysis: The Network Approach to Social Research**. Toronto: Butterworths, 1982.
- BERRY, M.W.; DUMAIS, S. T. ;O'BRIEN, G.W. Using Linear Algebra for Intelligent Information Retrieval, tech. report UT-CS-94-270, Univ. of Tennessee, 1994.
- BORDIN, A.S; GONÇALVES, A.L.; TODESCO, J.L. Análise da colaboração científica departamental através de redes de coautoria. **Perspectivas em Ciência da Informação**, v.9, n.2, p. 37-52, 2014.
- BORDIN, A. S. et al. Identificação e Análise de Comunidades de Colaboração Científica: Estudo de Caso em um Programa de Pós-Graduação Interdisciplinar. **Perspectivas em Gestão e Conhecimento**. No prelo. Outubro, 2015.
- BORGATTI, S.P.; EVERETT, M.G.; FREEMAN, L.C. **Ucinet for Windows: Software for Social Network Analysis**. Harvard, MA: Analytic Technologies, 2002.
- BORST, W. N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse, PhD Thesis, University of Twente – Centre for Telematica and Information Technology, Enschede, Nederland, 1997.
- BOTT, Helen. Observation of play activities in a nursery school. **Genetic Psychology Monographs**, v.4, p.44–88, 1928.

BRAUN, T.; GOMEZ, I.; MENDEZ, A.; SCHUBERT, A. International Co-authorship Patterns in Physics and its Subfields, 1981-85. **Scientometrics**, v.24, p. 181-200, 1992.

BRICKLEY, D.; MILLER, L. FOAF Vocabulary Specification 0.99. Disponível em: <http://xmlns.com/foaf/spec/>. Acessado em: 01/08/2014.

BRODER, A. et al. Graph structure in the web. **Computer Networks**, v.33, p. 309–320, 2000.

BUCKERIDGE, D.L et al. **A knowledge-based framework for deploying surveillance problem solvers**. Int. Conf. Information and Knowledge Engineering (IKE '04); Las Vegas, 2004.

BUCKERIDGE, D.L et al. Understanding detection performance in public health surveillance: modeling aberrancy-detection algorithms. **Journal of the American Medical Informatics Association**, v. 15, n. 6, p 760–769, 2008.

BUITELAAR, P.; EIGNER, T.; DECLERCK, T. **OntoSelect: A dynamic ontology library with support for ontology selection**. In: Proceedings of the Demo Session at the ISWC 04, Hiroshima, Japan, Nov, 2004.

BUNGE, M. Treatise on Basic Philosophy. Volume 7. **Philosophy of Science and Technology. Part II: Life Science, Social Science and Technology**. Springer, 1985.

BURT, R.S. Models of network structure. **Annual Review of Sociology**, v.6, p. 79-141, 1980.

ÇAĞDAŞ, V.; STUBKJÆR, E. Design research for cadastral systems. **Computers, Environment and Urban Systems**, v. 35, p. 77-87, 2011.

CHAKRABARTI, A. A course for teaching design research methodology. **Artificial Intelligence for Engineering Design, Analysis and Manufacturing**, v. 24, p. 317-334, 2010.

CHANDRASEKARAN, B.; JOHNSON, T.R.; SMITH, J.W. Task-structure analysis for knowledge modeling. **Communications of the ACM**, v.35, p. 124 –37, 1992.

CHANDRASEKARAN, B. Generic Tasks in Knowledge-based Reasoning: High-level Building Blocks for Expert System Design. **IEEE Expert**, v.3, p. 23-30, 1986.

- CHANDRASEKARAN, B., JOSEPHSON, J. R., BENJAMINS, R. **Ontology of Tasks and Methods**. In: Workshop on Knowledge Acquisition, Modeling and Management (KAW'98), Banff, Canada, 1998.
- CHELMIS, C.; PRASANNA, V. K. Social Networking Analysis: A State of the Art and the Effect of Semantics. In: IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 2011.
- CHEN, L., WEI, S., QINGPU, Z. **Semantic Description of Social Network Based on Ontology**. In: IEEE International Conference on E-Business and E-Government – 2010. Anais... 2010. p. 1936–1939.
- CHOMPALOV, I.; GENUTH, J.; SHRUM, W. The organization of scientific collaborations. **Research Policy**, v.31, n.5, 749–767, 2002.
- CLANCEY, W. J. The Knowledge Level Reinterpreted : Modeling How Systems Interact. **Machine Learning**, v. 291, n.4, p. 285–291, 1989.
- CLANCEY, W.J. Heuristic Classification. **Artificial Intelligence**, 27, p. 289-350, 1985
- CLANCEY, W.J. The Epistemology of a Rule-Based Expert System - a Framework for Explanation. **Artificial Intelligence**, 20, p. 215-251, 1983.
- CLANCEY, W.J., SIERHUIS, M., SEAH, C. Workflow agents versus expert systems: Problem solving methods in work systems design. **Artificial Intelligence for Engineering Design, Analysis and Manufacturing**, v. 23, pp 357-371, 2009.
- COLLINS, R. **Theoretical Sociology**. San Diego: Harcourt Brace Jovanovich. 1988.
- CORCHO, O.; FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A. Ontological Engineering: Principles, Methods, Tools and Languages. In: .Calero, F. Ruiz, M. Piattini (eds). **Ontologies for Software Engineering and Software Technology**. Springer-Verlag, 2006.
- CORCHO, O.; GÓMEZ-PÉREZ, A. **A Roadmap to Ontology Specification Languages**. In: International Conference on Knowledge Acquisition, Modeling and Management, 2000, pp. 80-96.

COSTA, J. G.; VANZ, S. A. S. Indicadores da Produção Científica e Coautoria: Análise do Departamento de Ciências da Informação da UFRGS. **Encontros Bibli**, Florianópolis, v. 17, n. 33, p. 97-115, 2012.

COSTA, R. M.; SILVA, J.V. M.. **Design Research é uma metodologia de aplicação prática?**. In: III Congresso Internacional de Pós-Graduações e Pesquisas Latinos em Administração e Estudos Organizacionais A administração e o Estudos Organizacionais no contexto Latino-Americano, 2014, Porto Alegre.

CRUBÉZY, M. et al. Ontological for Bioterrorism. **IEEE Intelligent Systems**. Set-oct, 26- 35, 2005.

CRUBÉZY, M.; MUSEN, M.A. Ontologies in support of problem solving. In: Staab, S.; Studer, R. (Eds). **Handbook on Ontologies**. Berlin: Springer; 2004. p. 321-342.

CUPANI, A. **La peculiaridad del conocimiento tecnológico**. Scientia Studia, São Paulo, v.4, n.3, p.353-71, 2006.

D'AQUIN, M.; MOTTA, E. Watson, more than a semantic web search engine. **Semantic Web Journal**, v.2, n. 1, p. 55-63, 2011.

D'AQUIN, M.; NOY, N. F. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. **Journal of Web Semantics**, v. 11, n. August, p. 96-111, 2012.

DAVID, J.-M.; KRIVINE, J.-P; SIMMONS, R. **Second Generation Expert Systems**. Berlin: Springer-Verlag, 1993.

DE HAAN, J. Authorship patterns in dutch sociology. **Scientometrics**, v.39, n.2, p. 197-208-208, 1997.

DRESCH, A.; LACERDA, D. P.; ANTUNES JÚNIOR, J. A. V. Design Science Research: Método de Pesquisa para Avanço da Ciência e Tecnologia. Porto Alegre: Bookman. 2015.

ERÉTEO, G. et al.: **A State of the Art on Social Network Analysis and its Applications on a Semantic Web**. In: Social Data on Web (SDoW), Workshop at 7th International Semantic Web Conference (ISWC'08). 2008

ERÉTEO, G. et al. **Analysis of a Real Online Social Network using Semantic Web Frameworks**. In: 8th International Semantic Web Conference (ISWC'09). Anais... Wiley Publishing, 2009a.

ÉRÉTÉO, G.; GANDON, F.; BUFFA, M. et al. **Semantic social network analysis**. In: Web Science 2009. 2009b.

ÉRÉTÉO, G. et al. Semantic Social Network Analysis: a concrete case. In: Ben Daniel. **Handbook of Research on Methods and Techniques for Studying Virtual Communities**. Hershey, PA: IGI Global Ed., 2011. p. 122-156.

FALOUTSOS, M. P.; FALOUTSOS, C. On power-law relationships of the internet topology. **Computer Communications Review**, 29, 251–262 (1999).

FARQUHAR, A.; FIKES, R.; RICE, J. The Ontolingua Server: A Tool for Collaborative Ontology Construction. **International Journal of Human-Computer Studies**, vol. 46, n. 6, 1997, pp. 707-727.

FELDMAN, R.; SANGER, J. **The Text Mining Handbook -- Advanced Approaches in Analyzing Unstructured Data**. São Paulo: Cambridge University Press, 2007.

FELL, D. A.; WAGNER, A. The small world of metabolism. **Nature Biotechnology**, v.18, p. 1121–1122, 2000.

FENSEL, D. et al. **The Use of Ontologies for Specifying Tasks and Problem-Solving Methods : A Case Study**. In: 10th European Knowledge Acquisition, Modeling, and Management Workshop, EKAW '97. Proceedings... 1997a. p. 1–16.

FENSEL, D. **An Ontology-based Broker : Making Problem-Solving Method Reuse Work**. In: Proc. of the IJCAI-97 Workshop on Problem-Solving Methods for Knowledge-Based Systems (pp. 23–29). 1997b.

FERNÁNDEZ-LÓPEZ, M. Overview of methodologies for building ontologies. 1999. Disponível em:
<<http://www.ontology.org/main/presentations/madrid/analysis.pdf>>.

FISCHBACH, K.; PUTZKE, J.; SCHODER, D. Co-authorship networks in electronic markets research. **Electronic Markets**, v. 21, n. 1, p. 19-40, 2011.

FREEMAN, L. **The Development of Social Network Analysis: A Study in the Sociology of Science**. Vancouver, BC: Empirical Press, 2004.

FREEMAN, L. Centrality in Social Networks: Conceptual Clarification. **Social Networks**, v.1, n.3, p. 215–239, 1979.

FREEMAN, L. Turning a profit from mathematics: the case of social networks. **Journal of Mathematical Sociology**, 10, p. 343-360, 1984.

FREEMAN, L. Some antecedents of social network analysis. **Connections**, v. 19, n. 1, p. 39-42, 1996.

FREEMAN, L. Visualizing social networks. **Journal of Social Structure**, v.1, 2000.

GENNARI, J. H. et al. The Evolution of Protégé: An Environment for Knowledge-Based Systems Development. **International Journal of Human-Computer Studies**, vol. 58, n. 1, pp. 89-123, 2003.

GENNARI, J.H. et al. Mappings Domains to Methods in Support or Reuse. **International Journal of Human-Computer Studies**, 41, p. 399-424, 1994.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GLÄNZEL, W. Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. **Library Trends**, v.50, n.3, p. 461-473, 2002

GLÄNZEL, W.; SCHUBERT, A. Analysing scientific networks through co-authorship. In: H. F. Moed, W. Glänzel & U. Schmoch (Eds.). **Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems**. Dordrecht: Kluwer Academic Publishers, 2004. p. 257-276.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, O. **Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web**. Springer, 2002.

GOMÉZ-PÉREZ, J. M.; CORCHO, O. Problem-Solving Methods for Understanding Process Executions. **Computing in Science & Engineering**, v.10, n.3, p.47–52, 2008.

GOMÉZ-PÉREZ, J. M. **Aquisition and Undertanding of Process of Knowledge Using Problem Solving Methods**. Heidelberg: GmbH, 2010.

GONZÁLEZ-ALCAIDE, G. et al. Scientific authorships and collaboration network analysis on Chagas disease: papers indexed in PubMed (1940-2009). **Revista Do Instituto de Medicina Tropical de São Paulo**, v. 54, n.4, p. 219–228, 2012.

GRIMM, S., HITZLER, P., ABECKER, A. Knowledge Representation and Ontologies: Logic, Ontologies and Semantic Web Languages. In: Rudi Studer, Stephan Grimm, Andreas Abecker (Eds). **Semantic Web Services: Concepts, Technologies, and Applications**. Springer, 2007.

GROSSMAN, J.W. The evolution of the mathematical research collaboration graph. **Congressus numeratium**, v. 158, p. 202-212, 2002.

GRUBER, T. R. **Toward principles for the design of ontologies used for knowledge sharing**. In: Guarino, N. and Poli, R. (Eds.), International Workshop on Formal Ontology, Padova, Italy, 1993.

GUARINO, N. **Formal Ontology in Information Systems**. Proceeding of FOIS'98, Trento, Italy, IOS Press, June 1998.

HACKOS, J.T.; REDISH, J.C. **User and task analysis for interface design**: John Wiley & Sons; 1998.

HANNEMAN, R. A.; RIDDLE, M. **Introduction to social network methods**. Riverside, California, 2005.. Disponível em: <http://faculty.ucr.edu/~hanneman/nettext/>. Acesso em: 01/05/2013.

HARA, N., et al. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. **Journal of the American Society for Information Science and Technology**, v. 54, n. 10, p. 952–965, 2003.

HARARY, F.; NORMAN, R. Z.; CARTWRIGHT, D. **Structural Models: An Introduction to the theory of Directed Graphs**. New York: John Wiley and Sons. 1965.

HEVNER, A. R.; MARCH, S.T.; PARK, J. Design Science in Information Systems Research. **MIS Quarterly**, v. 28, n. 1, p. 75-105, 2004.

HORROCKS, I. et al. **OIL in a Nutshell**. In: Proceedings of the ECAI'00 Workshop on Application of Ontologies and PSMs. Berlin. Germany. August, 2000.

HOU, H. A. H.; RETSCHMER, H. I. K.; LIU, Z. E. L. The Structure of Scientific Collaboration Networks in Scientometrics. **Scientometrics**, v.75, n.2, p. 189–202, 2008.

HSU, C.; SANDFORD, B. A. The Delphi Technique: Making Sense of Consensus. **Practical Assessment, Research & Evaluation**, v. 12, n. 10, p. 1–8, 2007.

HUISMAN, M.; VAN DUIJN, M.A.J. **Software for statistical analysis of social networks**. In: 6th International Conference on Logic and Methodology (RC33), Amsterdam, August 16-20, 2004.

HUMMON, N. P.; CARLEY, K. M. Social networks as normal science. **Social Networks**, v.15, p.71–106, 1993.

HYVONEN, E. et al. **Building a national semantic web ontology and ontology service infrastructure — The FinnONTO approach**. In: The Semantic Web: Research and Applications: Proceedings of the European Semantic Web Conference (ESWC 2008), pages 95–109, 2008.

INSNA (International Network for Social Network Analysis).
Disponível em: <http://www.insna.org>. Acesso em: 01/07/2013.

JIANG, Y. Locating active actors in the scientific collaboration communities based on interaction topology analyses. **Scientometrics**, v.74, n.3, p. 471–482, 2008.

KARP, R.; CHAUDHRI, V.; THOMERE, J. XOL: An XML-Based Ontology Exchange Language. July, 1999.

KATERNDAHL, D. Evolution of the research collaboration network in a productive department. **Journal of Evaluation in Clinical Practice**, v. 18, n. 1, p. 195-201, 2012.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research Policy**, v. 26, p. 1–18, 1997.

KELLY, J. A. et al. Prevention of HIV and sexually transmitted diseases in high risk social networks of young Roma (Gypsy) men in Bulgaria: randomised controlled trial. **BMJ**, v. 333, p.1098-1101, 2006.

KIFER, M.; LAUSEN, G.; WU, J. Logical foundations of object-oriented and frame-based languages. **Journal of the ACM**, v. 42, n.4, p. 741-843, 1995.

- KNOKE, D.; BURT, R. Prominence. In R. Burt & M. Minor (Eds.). **Applied network analysis: a methodological introduction**. California: Sage Publications, 1983. p. 196-222.
- KRACKHARDT, D. Social networks. In: Levine, J. M., Hogg, M. A. (Eds). **Encyclopedia of Group Processes and Intergroup Relations**. Los Angeles: SAGE, 2010. p. 817-821.
- KRAFFT, D. et al. VIVO: Enabling National Networking of Scientists. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC: US.ZIMAN, J. M.
- KRETSCHMER, H.; LIMING, L.; KUNDRA, R. Foundation of a global interdisciplinary research network (COLLNET) with Berlin as the virtual center. **Scientometrics**, v. 52, n.3, p. 531-537, 2001.
- LACERDA, D. P. et al. Design Science Research: método de pesquisa para a engenharia de produção. **Gest. Prod.**, São Carlos, v. 20, n.4, p. 741-761, 2013.
- LAUDEL, G. Collaboration and reward. What do we measure by co-authorship? **Research Evaluation**, v.11, n.1, p. 3-15, 2002.
- LEE, D. H. et al. Collaboration network patterns and research performance: The case of Korean public research institutions. **Scientometrics**, v.91, n.3, 925-942, 2012.
- LI, T. et al. **A Social Network Analysis Methods based on Ontology**. In: Proceedings of the 3rd International Symposium on Knowledge Acquisition and Modeling, p. 258-261, 2010.
- LIAO, C. H. et al. Quantifying the degree of research collaboration: A comparative study of collaborative measures. **Journal of Informetrics**, v.6, n.1, p. 27-33, 2012.
- LIL, D. et al. **The Structure Analysis of the CSCWD Conference's Collaboration Network**. In: Proceedings of the IEEE 16th International Conference on Computer Supported Cooperative Work in Design, p. 713-718, 2012.
- LIMA, M. Y. Coautoria na produção científica do PPGGeo/UFRGS: uma análise de redes sociais. **Ci. Inf**, Brasília, v. 40, n. 1, p. 38-51, jan./abr. 2011.

LIU, X., et al. Coauthorship networks in the digital library research community. **Inf. Process. Manage.**, v. 41, n. 6, p. 1462-1480, 2005.

LUKE, S.; HEFLIN, J. SHOE 1.01. Proposed Specification. SHOE Project. February, 2000.
<http://www.cs.umd.edu/projects/plus/SHOE/spec101.htm>. Acesso em: 01/06/2013.

MAIA, G. et al. **Análise da Rede de Colaboração do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos: As primeiras 30 edições**. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, pp. 14–27, 2012.

MAIA, M.; CAEGNATO, S. Coautoria como indicador de colaboração científica. **Perspectivas em Ciência da Informação**, v.13, n.2, p. 18-31, 2008.

MALI, F. et al. Dynamic Scientific Co-Authorship Networks. In A. Scharnhorst, K. Borner & P. van den Besselaar (Eds.). **Models of Science Dynamics. Understanding Complex Systems**. Heidelberg: Springer, 2012. p. 195-232.

MARCH, S. T.; SMITH, G. F. Design and natural science research on information technology. **Decision Support Systems**, v. 15, n. 4, p. 251–266, 1995.

MARTINS, M. E. et al. Service's scientific community: A social network analysis (1995-2010). **Journal of Service Management**, 23(3), 455–469, 2012.

MCDERMOTT, J. Preliminary Steps toward a Taxonomy of Problem-solving Methods. In: S. Marcus (Ed.). **Automating Knowledge Acquisition for Experts Systems**. Boston: Kluwer Academic Publisher, 1988.

MELIN, G.; PERSSON, O. Studying research collaboration using co-authorship. **Scientometrics**, v.36, n.3, p.363-377, 1996.

MENA-CHALCO, J. P.; Cesar-Jr, R. M. scriptLattes: An open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n.4, p. 31-39, 2009.

MENA-CHALCO, J. P.; CESAR-JR, R. M. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. In: Hayashi, M.

C. P. I.; Leta, J (Org.). **Bibliometria e Cientometria: reflexões teóricas e interfaces**. São Carlos: Pedro & João Editores. 2013.

MICKENBERG, R.; DUGAN, J. **Taxi Driver Wisdom**. San Francisco: Chronicle, 1995.

MIKA, P. Social Networks and the Semantic Web. In: Proceedings of the International Conference on Web Intelligence (WIC 04), 2004.

MIKA, P.; ELFRING, T.; GROENEWEGEN, P. Application of semantic technology for social network in sciences. **Scientometrics**, v. 68, n. 1, p. 3–27, 2006.

MILLER, L. E. Determining what could/should be: The Delphi technique and its application. In: Annual Meeting of the Mid-Western Educational Research Association, Columbus, Ohio. 2006.

MILLER, R.A., POPLER, H.E., MYERS, J. D. INTERNIST-i, an experimental computer-based diagnosis consultant for general internal medicine. New England. **Journal of Medicine**, v.37, n.8, p.468-476, 1982.

MILOJEVIĆ, S. Modes of Collaboration in Modern Science – Beyond Power Laws and Preferential Attachment. **Journal of the American Society for Information Science and Technology**, v. 61, n. 7, p. 1410-1423, 2010.

MINAYO, M. C. S. **O desafio do conhecimento. Pesquisa qualitativa em saúde**. São Paulo: HUCITEC, 2007.

MIZOGUCHI, R.; VAN WELKENHUYSEN, J.; IKEDA, M. Task ontology for reuse of problem solving knowledge. In: Mars, N. J. I., (Ed.). **Towards very large knowledge bases**. IOS Press. 1995.

MOED; H. F.; BRUIN, R. E. de; NEDERHOF, A. J.; TIJSSEN, R. J. W. International Scientific Cooperation and Awareness within the European Community: Problems and Perspectives. **Scientometrics**, v. 21, p. 291-311, 1991.

MORENO, J. Who Shall Survive? Washington, DC: **Nervous and Mental Disease**. Publishing Company. 1934.

MOTTA, E. An Overview of the OCML Modelling Language. In: Workshop on Knowledge Engineering Methods & Languages, p. 21-22, 1998.

MYLOPOULOS, J. An Overview of Knowledge Representation. Proceedings of the 1980 workshop on Data abstraction, databases and conceptual modeling. **Anais...** p.5–12, 1980. New York, NY: ACM.

NEWELL, A. The knowledge level. **Artificial Intelligence**, v.18, p. 87-127, 1982.

NEWMAN, M E J; GIRVAN, M, Finding and evaluating community structure in networks. **Physical Review E**, v. 69, n.026113, p.1-16, 2004b.

NEWMAN, M. E. Coauthorship networks and patterns of scientific collaboration. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. suppl 1, p. 5200-5205, april 2004a.

NEWMAN, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. **Physical Review E**, 64(1), 016131, 2001b.

NEWMAN, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. **Physical Review E**, 64(1), 016132, 2001c.

NEWMAN, M. E. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 2, p. 404, jan. 2001a.

NOY, F. N.; MCGUINNESS, D. L. Ontology development 101: a guide to create your first ontology. 2001. Disponível em: <<http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc>>. Acesso em: 01/02/2014.

NOY, N. F. et al. Biportal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 10.1093/nar/gkp440, 2009.

O'CONNOR, M.J. et al. Software Engineering Challenges of building and deploying reusable problem solvers. **Artif Intell Eng Des Anal Manuf**, v. 23, n.4, p. 339–356, 2009.

ONTOLOGIES FOR E-GOVERNMENT. Disponível em:

<<http://oegov.org/>, 2010>. Acesso em: 01/05/2013.

ONTOLOGY DESIGN PATTERNS.ORG (ODP). Disponível em:

<<http://ontologydesignpatterns.org>, 2010>. Acesso em: 01/05/2013.

OWL WEB ONTOLOGY LANGUAGE OVERVIEW. World Wide

Web Consortium. Disponível em: <<http://www.w3.org/TR/owl2-overview/>>. Acesso em: 01/05/2013.

OSINSKI, S.; WEISS, D. A Concept-Driven Algorithm for Clustering Search Results. **IEEE Intelligent Systems**. 1541-1672, 2005.

PACHECO, R. C. S.; TOSTA, K. C. B. T.; FREIRE, P. S.

Interdisciplinaridade vista como um processo complexo de construção de conhecimento: uma análise do Programa de Pós-Graduação EGC/UFSC. **RBPB-Revista Brasileira de Pós-Graduação**, 2010, vol. 7, n. 12, p. 136-159.

PASSMORE, D. L. **Social Network Analysis - Theory and**

Applications. Free Book. 2011. Disponível em:

http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf. Acesso em: 01/02/2014.

PEFFERS, K. et al. A Design Science Research Methodology for

Information Systems Research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45-77, 2007.

PERSSON, O.; GLÄNZEL, W.; DANELL, R. Inflationary bibliometric

values: The role of scientific collaboration and the need for relative indicators in evaluative studies. **Scientometrics**, v.60, n.3, p. 421-432, 2004.

PRICE, D. deSolla, BEAVER, D. deB. Collaboration in an invisible

college. **American Psychologist**, v. 21, p. 1011–1018, 1966.

ROSSONI, L., FILHO, E. G., Cooperação entre Programas de Pós -

Graduação em Administração no Brasil: evidências estruturais em quatro áreas. **RAC**, Curitiba, v. 13, n. 2, p. 366–390, jul./ago. 2009.

ROWE, G.; WRIGHT, G. The Delphi technique as a forecasting tool:

issues and analysis. **International Journal of Forecasting**, v. 15, p. 353–375, 1999.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence – A Modern Approach**. Prentice-Hall, 3ed. 2009.

SALTON, G.; WONG, A.; YANG, C. A Vector Space Model for Automatic Indexing. **Communications of the ACM**, v.18, n.11, p. 613-620, 1975.

SAUNDERS, M.; LEWIS, P.; THORNHILL, A. **Research Methods for Business Students**. England: Prentice Hall, 2009.

SCHREIBER, G. et al. **Knowledge engineering and management: the commonKADS methodology**. MIT Press, 2002.

SCHREIBER, G. Knowledge engineering. In V. Lifschitz, F. van Harmelen, B. Porter (Eds). **Handbook of Knowledge Representation**. Elsevier, 2007. p. 929-946.

SCOTT, J. **Social Network Analysis. A Handbook**. 2nd edition. SAGE Publications: London, 2000.

SEIN, M. K. et al. Action Design Research. **MIS Quaterly**, v. 35, n. 1, p. 37-56, 2011.

SHORTLIFE, E. H. **Computer-Based Medical Consultations: Mycin**. American Elsevier, New York, 1976.

SILVA, A. B. O. et al. Estudo da rede de coautoria e da interdisciplinaridade na produção científica com base nos métodos de análise de redes sociais: avaliação do caso do Programa de Pós-Graduação em Ciência da Informação - PPGCI / UFMG. **Encontros Bibli**, Florianópolis, n. esp., p. 179-194, 2006.

SIMON, H. A. **The Sciences of the Artificial**. 3rd ed. Cambridge/Massachussets: MIT Press, 1996 [1961].

SIMONIN, B.L. The importance of collaborative know-how: an empirical test of the learning organization. **Academy of Management Journal**, v.40, p.1150–1174, 1997.

SMITH, M. The Trend Toward Multiple Authorship in Psychology. **American Psychologist**, n.13, p. 596-99, 1958.

SONNENWALD, D. H. Scientific collaboration: Challenges and solutions. In B. Cronin (Ed.), **Annual Review of Information Science & Technology**. New York, v. 41, n. 1, p. 643-681, 2007.

SONNENWALD, D.H. Challenges in corporate and university R&D collaboration. INFORMS (Institute for Operations Research and Management Science). In: Annual Conference, Cincinnati, OH, 1999.

SOUZA, C. G.; BARBASTEFANO, R. G.; LIMA, L. S. Redes de colaboração científica na área de química no Brasil: um estudo baseado nas coautorias dos artigos da revista *Química Nova*. **Química Nova**, São Paulo, v. 35, n. 4, p. 671-676, 2012.

SOWA, J. **Knowledge Representation: Logical, Philosophical and Computational Foundations**. Pacific Grove: Brooke Cole Publishing Co, 2000.

STEVENS, M.J.; CAMPION, M.A. The knowledge, skill, and ability requirement for teamwork: implications for human resource management. **Journal of Management**, 20, 503–530. Stokes, D.E.1997.

STRUSS, P. Model-based Problem Solving. In: Harmelen, F., Lifschitz, V.; Porter, B. (eds.) **Handbook of Knowledge Representation**, chapter 10, p. 395–465. Amsterdam, Elsevier B.V., 2008.

STUDER, R. et al. Ontologies and the Configuration of Problem-Solving Methods. In: Proc. of the 10th Knowledge Acquisition for Knowledge based Systems Workshop, Banff, 1996.

STUDER, R.; BENJAMINS, V.; FENSEL, D. Knowledge Engineering: Principles e Methods. **Data and Knowledge Engineering**, v.25, n.1, p. 161-197, 1998.

SUÁREZ-FIGUEROA, M. C. et al. NeOn: Lifecycle Support for Networked Ontologies: D5.3.1 NeOn Development Process and Ontology Life Cycle. Disponível em: http://www.neon-project.org/web-content/images/Publications/neon_2007_d5.3.1.pdf. Acesso em: 01/05/2013.

SUBRAMANYAM, K. Bibliometric Studies of Research Collaboration: A Review. **Journal of Information Science**, v.6, n.35, 1983.

SURE, Y.; STAAB, S.; STUDER, R. On-To-Knowledge methodology. In: Staab S., Studer R. (eds.) **Handbook on Ontologies**, chapter 6, pp. 117–132. Berlin: Springer-Verlag, 2003.

TAMANINI, M. T. A. et al. Análise Contextualizada de Redes Sociais: a colaboração em Comunidades Científicas. In: Brazilian Workshop on Social Network Analysis and Mining. 2013.

TOMHAVE, B. L. **Alphabet Soup : Making Sense of Models, Frameworks, and Methodologies**. 2005.

TONES ONTOLOGY REPOSITORY. Disponível em: <http://owl.cs.manchester.ac.uk/repository/>, 2010. Acesso em: 01/05/2013.

TREMBLAY, M. C.; HERVNER, A. R.; BERNDT, D. J, Focus Groups for Artifact Refinement and Evaluation in Design Research. **Communications of the Association for Information Systems**, v. 26, n. 27, p. 599-618, 2010.

VAISHNAVI, V.; KUECHLER, W. Design Research in Information Systems. 2009. Disponível em: <[http:// desrist.org/design-research-in-information-systems](http://desrist.org/design-research-in-information-systems)>. Acesso em: 15/03/2015.

VAN AKEN, J. E. Management Research Based on the Paradigm of the Design Sciences: The Quest for Field- Tested and Grounded Technological Rules. **Journal of Management Studies**, v. 41, n. 2, p. 219-246, 2004.

VASSEV, E.; HINCHEY, M. Knowledge Representation and Reasoning for Intelligent Software Systems. **Computer**, v.44, n.8, p. 96–99, 2011.

VENABLE, J. R. The Role of Theory and Theorising in Design Science Research. **DESRIST**, v. 24-25, p. 1-18, 2006.

WAGNER, C.; LEYESDORFF, L. Mapping the network of global science: Comparing international co-authorships from 1990 to 2000. **International Journal of Technology and Globalization**, 1(2), 185-208, 2005.

WAGNER, C.; LEYESDORFF, L. Network structure, self-organization, and the growth of international collaboration in science. **Research Policy**, v.34, n.10, p.1608-1618, 2005.

WANG, T. et al. On social computing research collaboration patterns: A social network perspective. **Frontiers of Computer Science in China**, v.6, n.1, 122–130, 2012.

WASSERMAN, Stanley; Faust, Katherine. **Social Network Analysis: methods and applications**. Cambridge University Press. Structural analysis in social the social sciences series, v. 8, 1994.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **Nature**, v. 393, p. 440–442, 1998.

WELLMAN, Beth. BOTT. **Journal of Educational Research**, v.14, p.126–132, 1926.

WHITE, H. C. **Chains of Opportunity; System Models of Mobility in Organizations**. Cambridge, Mass.: Harvard University. 1970.

WHITE, H.C., BOORMAN, S.A.; BREIGER, R. Social structure from multiple networks I: Blockmodels of roles and positions. **American Journal of Sociology**, v. 81, p.730–781, 1976.

WHITLEY, R. **The Intellectual and Social Organization of the Sciences**. Oxford: Oxford University Press, 2000.

WIELINGA, B.J.; SCHREIBER, A.T.; BREUKER, J. KADS: a modelling approach to knowledge engineering. **Knowledge Acquisition**, v.4, p. 5–53, 1992.

WILL, U.K.; GNIADEK, J.; MEMON, N. Measuring Link Importance in Terrorist Networks. Social Network Analysis. In: International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2010), 2010, Odense, Denmark. **Anais...**IEEE Computer Society, 2010. p. 225-232.

XU, Bo; LIU, Lu. Information diffusion through online social networks. In: IEEE International Conference on Emergency Management and Management Sciences (ICEMMS), p.53-56, 8-10 Aug. 2010.

XU, Y. et al. Combining social network and semantic concept analysis for personalized academic researcher recommendation. **Decision Support Systems**, v.54, n.1, p.564–573, 2012.

YANG, A. et al. The similar scholar recommendation in Schol@t. In: Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2012, p. 666–670, 2012.

YE, Q.; SONG, H.; LI, T. Cross-institutional collaboration networks in tourism and hospitality research. **Tourism Management Perspectives**, v.2, n.3, p.55–64, 2012.

YU, Q. et al. World scientific collaboration in coronary heart disease research. **International Journal of Cardiology**, v.167, n.3, p.631–9, 2013.

ZACHARY, W. W. An Information Flow Model for Conflict and Fission in Small Groups. **Journal of Anthropological Research**, v. 33, n. 4, p. 452-473, 1977.

ZIMAN, J. M. Real Science: **What it is, and What it Means**. Cambridge, UK: Cambridge University Press, 2000.

ZIMAN, J. M. **Conhecimento público**. Belo Horizonte: Itatiaia; São Paulo: Ed. da USP, 1979. 164 p.

GLOSSÁRIO

Conhecimento de domínio

Representação declarativa, decomposicional do comportamento dos constituintes elementares do sistema no domínio. (STRUSS, 2008)

Conhecimento de tarefa

Conhecimento de como resolver um problema. (MIZOGUCHI; VAN WELKENHUYSEN; IKEDA, 1995).

Mecanismos de resolução de problemas que realizam inferências em um modelo de domínio. (STRUSS, 2008)

Framework

“*Framework* é uma construção fundamental que define os pressupostos, conceitos, valores e práticas, e que inclui uma orientação para a implementação do mesmo” (TOMHAVE, 2005).

Método

Conceito similar a tarefa que consiste na estrutura da tarefa com controle. (MIZOGUCHI; VAN WELKENHUYSEN; IKEDA, 1995).

Método de Resolução de Problemas

PSM é uma especificação em nível de conhecimento de um padrão de raciocínio que pode ser usado para executar uma tarefa intensiva em conhecimento. (SCHREIBER, G., 2007).

De uma perspectiva mais programática, são componentes de software que representam e codificam algoritmos reutilizáveis. Eles podem ser combinados com representações do conhecimento de domínio para produzir sistemas inteligentes. (O’CONNOR, 2009)

Definem estratégias que permitem que tarefas sejam cumpridas. Estratégias descrevem como tarefas podem ser recursivamente decompostas em tarefas de nível mais simples. (GOMÉZ-PÉREZ, 2010)

Modelo de domínio

Descreve um domínio particular sob o qual tarefas e métodos são aplicados. (GOMÉZ-PÉREZ, 2010)

Ontologia de tarefa

Especificação para quais objetos e relações entre eles são necessárias para realizar uma tarefa. Provê um vocabulário necessário e suficiente

para construir um modelo de processo de resolução de problema humano. (MIZOGUCHI; VAN WELKENHUYSEN; IKEDA, 1995).

Semântica

Esta definição apoia-se na definição de Web Semântica dada por Berners-Lee (2001), afirmando que a “web semântica é uma extensão da web atual, na qual é atribuído à informação um significado bem definido, habilitando pessoas e computadores a trabalhar em melhor cooperação”. Logo, semântica refere-se a dar significado a algo.

Tarefa (*Task*)

Sequência de passos para resolver um problema. (MIZOGUCHI; VAN WELKENHUYSEN; IKEDA, 1995).

Sinônimo de tipos (famílias) de objetivos de resolução de problemas. Existe uma separação entre a tarefa e o método que executa essa tarefa. (CHANDRASEKARAN; JOHNSON; SMITH, 1992)

Descrevem o que vai ser alcançado, através da execução de um processo, sendo fortemente relacionada com o objetivo da execução. (GOMÉZ-PÉREZ, 2010)

APÊNDICE A – Revisão Sistemática

A Tabela 1, Tabela 2, Tabela 3, Tabela 4 e Tabela 5 apresentam os resultados de uma busca sistemática na literatura sobre os tópicos que permeiam esta pesquisa, apresentando o número de trabalhos revisados, os meios de publicação e os autores com duas ou mais publicações no tópico.

Tabela 1 - Revisão do tópico “Colaboração científica”

Tópico de pesquisa: Colaboração científica	
Principais termos utilizados: colaboração científica OR <i>scientific collaboration</i> OR <i>research collaboration</i>	
Número de documentos: 28	
Meios de publicação (periódicos, conferências, livros, etc.)	Nú. publicações
<i>American Psychologist</i>	1
<i>Annual Review of Information Science and Technology</i>	1
<i>Scientometrics</i>	8
<i>Research Policy</i>	3
<i>Research Evaluation</i>	1
Perspectivas em Ciência da Informação	2
<i>Library Trends</i>	1
<i>Journal of the American Society for Information Science and Technology</i>	3
<i>Journal of Informetrics</i>	1
<i>Journal of Information Science</i>	1
<i>International Journal of Technology and Globalisation</i>	1
<i>Handbook of Quantitative Science and Tecnology Research</i>	1
<i>International Journal of Business, Management and Social Sciences</i>	1
<i>Book</i>	1
<i>Report</i>	2
Principais autores	Número de publicações
Beaver, D. deB.	2
Glänzel, W.	3

Katz, J.	2
Schubert, A.	2
Sonnenwald, D.	2

Fonte: Autor.

Tabela 2 - Revisão do tópico “Análise de redes sociais”

Tópico de pesquisa: Análise de redes sociais	
Principais termos utilizados: análise de redes sociais OR <i>social network analysis</i>	
Número de documentos: 54	
Meios de publicação (periódicos, conferências, livros, etc.)	Nú. publicações
<i>Asian Journal Of Social Psychology</i>	1
<i>BMJ</i>	1
<i>Book Networks, Crowds, and Markets : Reasoning about a Highly Connected World</i>	1
<i>Book Social Network Analysis Theory and Applications</i>	1
<i>Book Social Network Analysis: A handbook</i>	1
<i>Book Social Network Analysis: methods and applications</i>	1
<i>Book The Development of Social Network Analysis</i>	1
<i>Chapter in Atualizações em Informática</i>	1
<i>Chapter in Encyclopedia of Group Processes & Intergroup Relations</i>	1
<i>Chapter in Handbook of Graph Drawing and Visualization</i>	1
<i>Chapter in Handbook of Social Economics</i>	1
<i>Chapter in The SAGE Handbook of Social Network Analysis</i>	2
<i>Computer Networks</i>	1
<i>Computer Science Review</i>	1
<i>Connections</i>	1
<i>Decision Support Systems</i>	1
<i>Human Resource Development Review</i>	1
<i>International Conference on Advances in</i>	1

<i>Social Networks Analysis and Mining</i>	
<i>International Conference on Emergency Management and Management Sciences</i>	1
<i>International Conference on Logic and Methodology</i>	1
<i>International Conference on Web Intelligence, Mining and Semantics</i>	1
<i>Journal of Anthropological Research</i>	1
<i>Journal of Statistical Mechanics: Theory and Experiment</i>	1
<i>Journal of the Information Science</i>	1
<i>Linkoping Eletronic Articles in Computer and Information Science</i>	1
<i>Nature</i>	1
<i>Parallel Computing</i>	1
<i>Physica A: Statistical Mechanics and its Applications</i>	1
<i>Physical Review E.</i>	3
<i>Proceedings of the IEEE</i>	1
<i>Proceedings of the National Academy of Sciences of the United States of America</i>	4
<i>Report</i>	3
<i>Revista de Administração de Empresas</i>	1
<i>Revista Informação & Informação</i>	1
<i>Social Network Analysis and Mining</i>	1
<i>Social Networks</i>	5
<i>Sociological Methodology</i>	1
<i>The American Journal of Sociology</i>	3
<i>The Journal of Planning Literature</i>	1
<i>Transboundary and Emerging Diseases</i>	1
Principais autores	Número de publicações
Freeman, Linton C.	6
Girvan, M.	2
Newman, M. E. J.	6
Matheus, R. F.	3
Scott, John	2

Fonte: Autor.

Tabela 3 - Revisão do tópico “Rede de coautoria”

Tópico de pesquisa: Rede de coautoria	
Principais termos utilizados: rede de coautoria OR <i>coauthorship network</i> OR " <i>research network</i> " OR " <i>collaboration network</i> "	
Número de documentos: 74	
Meios de publicação (periódicos, conferências, livros, etc.)	Nú. publicações
<i>American Journal of Physics</i>	1
<i>Ansiedad y Estrés</i>	1
<i>BMJ Open</i>	1
<i>Brazilian Workshop on Social Network Analysis and Mining</i>	2
<i>Chapter in Models of Science Dynamics</i>	1
Ciência da informação	4
<i>Decision Support Systems</i>	2
<i>Electronic Markets</i>	1
Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação	2
<i>Escola Regional de Banco de Dados</i>	1
<i>European conference on Technology enhanced learning</i>	1
<i>Frontiers of Computer Science in China</i>	1
<i>Hawaii International Conference on System Sciences</i>	2
Informação e Sociedade	1
<i>Information Processing and Management</i>	3
<i>International Conference on Computer Supported Cooperative Work in Design</i>	1
<i>International Conference on Information and Knowledge Management, Proceedings</i>	1
<i>International Conference on Web Intelligence, Mining and Semantics</i>	1
<i>International Journal of Cardiology</i>	2
<i>Journal of Engineering and Technology Management</i>	1
<i>Journal of Evaluation in Clinical Practice</i>	2
<i>Journal of Informetrics</i>	6

<i>Journal of Service Management</i>	1
<i>Journal of the American Medical Informatics Association</i>	1
<i>Journal of Universal Computer Science</i>	1
<i>Lecture Notes in Computer Science</i>	4
<i>Pacific Asia Conference on Information Systems</i>	1
<i>Physica A</i>	2
<i>Physical Review E</i>	1
<i>PLoS Neglected Tropical Diseases</i>	1
<i>Proceedings of the American Society for Information Science and Technology</i>	1
<i>Proceedings of the ASIST Annual Meeting</i>	1
<i>Proceedings of the National Academy of Sciences of the United States of America</i>	2
<i>Quality & Quantity</i>	1
Quimica Nova	1
RAC	1
<i>Research Policy</i>	6
Revista do Instituto de Medicina Tropical de São Paulo	1
<i>Revista Española de Cardiología</i>	1
<i>Scientometrics</i>	6

Principais autores	Número de publicações
Abbasi, A.	5
Altmann, J.	3
Balconi, M.	2
Duan, Z.	2
Freire, V. P.	2
González-Alcaide, G.	2
Hossain, L.	3
Katerndahl, D.	2
Kretschmer, H.	3
Leydesdorff, L.	2
Liu, Z.	2
Ma, J.	2
Ma, Jian	2
Newman, M. E. J.	3
Shao, H.	2

Fonte: Autor.

Tabela 4 - Revisão dos tópicos “Análise de redes sociais” e “Tecnologias semânticas”

Tópicos de pesquisa: Análise de redes sociais e Tecnologias semânticas	
Principais termos utilizados: “social network analysis” OR “análise de redes sociais” AND (ontolog* OR "semantic technologies" OR "semantic web")	
Número de documentos: 47	
Meios de publicação (periódicos, conferências, livros, etc.)	Nú. publicações
<i>ACM WebScience</i>	1
<i>Chapter in Exploratory Analysis in Dynamic Social Networks – Theoretical and Practical Applications</i>	1
<i>Chapter in Handbook of Research on Methods and Techniques for Studying Virtual Communities</i>	1
<i>Computers & Education</i>	1
<i>Decision Support Systems</i>	1
<i>Expert Systems with Applications</i>	2
<i>IEEE Conference IEEE on Open Systems</i>	1
<i>IEEE International Conference on Social Computing</i>	1
<i>International Conference on Advances in Social Networks Analysis and Mining (ASONAM)</i>	3
<i>International Conference on Computer Supported Cooperative Work in Desig</i>	1
<i>International Conference on Dependable, Autonomic and Secure Computing</i>	1
<i>International Conference on E-Business and E-Government</i>	1
<i>International Conference on Information Integration and Web-Based Applications and Services</i>	1

<i>International Conference on Information Visualisation</i>	1
<i>International Conference on Knowledge and Smart Technology</i>	1
<i>International Conference on Learning Analytics & Knowledge</i>	1
<i>International Conference on Management of Innovation and Technology</i>	1
<i>International Conference on Network-Based Information Systems</i>	1
<i>International Conference on Ubiquitous Information Management and Communication</i>	2
<i>International Conference on Web Intelligence</i>	4
<i>International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support</i>	1
<i>International Semantic Web Conference</i>	1
<i>International Symposium on Knowledge Acquisition and Modeling</i>	1
<i>International Workshop on Cooperative Intelligent Agents</i>	1
<i>Journal of Software</i>	1
<i>Knowledge Science, Engineering and Management</i>	1
<i>Knowledge-Based Systems</i>	1
<i>Learning Organization</i>	1
<i>Procedia - Social and Behavioral Sciences</i>	1
<i>Report</i>	1
<i>Scientometrics</i>	3
<i>Social Network Analysis and Mining</i>	1
<i>The Semantic Web: Research and Applications. Lecture Notes in Computer Science</i>	2
<i>Thesis</i>	2
<i>Working Conference on Virtual Enterprises</i>	1
<i>Workshop on Social Data on the Web</i>	1
Principais autores	Número de publicações
Buffa, Michel	6
Chelmis, C.	2
Corby, Olivier	3

Erétéo, Guillaume	7
Gandon, Fabien	5
Grohan, Patrick	2
Mansur, A. B. F.	2
Mika, Peter	2
Prasanna, V. K.	2
Sander, Peter	2
Thovex, Christophe	4
Trichet, Francky	4
Yusof, N.	2

Fonte: Autor.

Tabela 5 - Revisão do tópico “Sistemas baseados em conhecimento”

Tópico de pesquisa: Sistemas baseados em conhecimento	
Principais termos utilizados: “sistemas baseados em conhecimento” OR “ <i>knowledge-Based Systems</i> ”	
Número de documentos: 69	
Meios de publicação (periódicos, conferências, livros, etc.)	Nú. publicações
<i>ACM Symposium On Applied Computing</i>	1
<i>American Medical Informatics Association Symposium (AMIA)</i>	1
<i>Artificial Intelligence</i>	2
<i>Artificial Intelligence for Engineering Design, Analysis and Manufacturing</i>	4
<i>Artificial intelligence in Medicine</i>	1
<i>Artificial Intelligence Magazine</i>	2
<i>Book Knowledge engineering and management: the commonKADS methodology</i>	1
<i>Book Ontological engineering : with examples from the areas of knowledge management, e-commerce and the semantic web</i>	1
<i>Chapter in Handbook of Knowledge Representation</i>	1
<i>Chapter in Handbook on Ontologies</i>	1

<i>Chapter in Ontologies for Software Engineering and Software Technology</i>	1
<i>Chapter in Ontologies in Urban Development Projects</i>	1
<i>Chapter in Semantic Web Services: Concepts, Technologies and Applications</i>	1
<i>Chapter in Towards Very LArge Knowledges Bases</i>	1
<i>Communications of the ACM</i>	1
<i>Computing in Science & Engineering</i>	1
<i>Data & Knowledge Engineering</i>	1
<i>ENECOMP</i>	1
<i>European Conference on Artificial Intelligence</i>	2
<i>European Knowledge Acquisition, Modeling, and Management Workshop</i>	2
<i>Expert Update</i>	1
<i>Hawaii International Conference on System Sciences</i>	1
<i>IEEE Computer</i>	1
<i>IEEE Expert</i>	1
<i>IEEE Intelligent Systems</i>	1
<i>IEEE International Conference on Web Services</i>	1
<i>IEEE Transactions on Knowledge and Data Engineering</i>	1
<i>International Conference on Formal Ontology in Information Systems</i>	1
<i>International Conference on Information and Knowledge Engineering</i>	1
<i>International Conference on Knowledge Capture</i>	3
<i>International Conference on Knowledge Engineering and Knowledge Management</i>	1
<i>International Conference on Software Engineering & Knowledge Engineering</i>	1
<i>International Conference on Web Intelligence and Intelligent Agent Technology</i>	1
<i>International Journal of Human-Computer Studies</i>	5
<i>International Journal of u- and e- Service, Science and Technology</i>	1
<i>International Workshop on Metamodels,</i>	1

<i>Ontologies and Semantic Technologies</i>	
<i>Journal of the American Medical Informatics Association</i>	1
<i>Knowledge Acquisition for Knowledge based Systems Workshop</i>	1
<i>Knowledge and Information Systems</i>	1
<i>Knowledge Engineering: Practice and Patterns</i>	1
<i>Machine Learning</i>	1
<i>Report</i>	6
<i>Semantic Web</i>	1
<i>Simpósio Brasileiro de Sistemas de Informação</i>	1
<i>Thesis</i>	2
<i>The Electronic Journal of Information Systems Evaluation</i>	1
<i>The Knowledge Engineering Review</i>	1
<i>Web semantics (Online)</i>	1
<i>Workshop on Data Abstraction, Databases and Conceptual Modelling</i>	1
<i>Workshop on Ontologies and Problem-Solving Methods</i>	2
<i>Workshop on Ontologies and their Applications</i>	1

Principais autores	Número de publicações
Benjamins, V.	8
Buckeridge, D.	5
Chandrasekaran, B.	4
Clancey, W.	5
Corcho, O.	5
Crubézy, M.	6
Eriksson, H.	2
Falbo, R.	2
Fensel, D.	9
Fernández-López, M.	2
Gennari, J.	3
Gómez-pérez, A.	4
Gómez-Pérez, A.	7
Gómez-Pérez, J.M.	2
Groenboom, R.	3

Harmelen, F.	2
Lopes, E.C.	2
Martins, A. F.	2
Motta, E.	5
Musen, M.	13
Noy, N.	3
O'Connor, M.	5
Okhmatovskaia, A.	3
Schreiber, G.	4
Studer, R.	4
Tu, S.	5
Wielinga, B.	2

Fonte: Autor.

APÊNDICE B - Roteiro da Entrevista do Grupo Focal

Nome completo:

E-mail:

Titulação:

Programa de Pós-Graduação:

Anos de experiência em gestão na universidade:

Dimensão 1: Colaboração Científica

Esta dimensão avalia a importância da colaboração científica, da análise da colaboração e das perguntas/tarefas de análise propostas.

1. Você considera importante o estabelecimento de relacionamentos de colaboração científica em um Programa de Pós-Graduação?
 Concordo plenamente Concordo parcialmente
 Indiferente
 Discordo parcialmente Discordo totalmente

Justifique: _____

2. Você considera que a análise da colaboração científica possa trazer subsídios para a **tomada de decisão** em vários níveis (gestor, pesquisador, comunidade acadêmica) em um Programa de Pós-Graduação?
 Concordo plenamente Concordo parcialmente
 Indiferente
 Discordo parcialmente Discordo totalmente

Justifique: _____

3. As perguntas ou tarefas de análise de colaboração científica propostas são **relevantes** para a análise da colaboração científica?

- () Concordo plenamente () Concordo parcialmente
 () Indiferente
 () Discordo parcialmente () Discordo totalmente

Justifique: _____

Dimensão 2 – D2: Resultados do Framework

Esta dimensão avalia o **grau de confiança** dos resultados das perguntas/tarefas de análise de colaboração científica do framework.

Prezado gestor, analise os resultados de cada uma das 18 perguntas/tarefas. Após a análise, avalie os resultados de cada pergunta/tarefa em relação ao critério **grau de confiança**, escolhendo uma, entre três opções: alto, médio ou baixo.

Dimensão 3 – D3: Aplicabilidade do *Framework* como Instrumento de Gestão

Esta dimensão avalia a aplicabilidade do *framework* como instrumento de gestão. Para cada resultado de pergunta/tarefa foram propostas possíveis ações de tomada de decisão em nível de gestão.

Prezado gestor, com base nos resultados de cada uma das 18 perguntas/tarefas do *framework*, confirme se as ações de gestão propostas são apropriadas. Baseado na sua experiência como gestor, identifique mais ações de gestão para cada uma das respostas.

1. Quem são os pesquisadores que mais colaboram?	
D2	() Alto () Médio () Baixo

D3	<p>Um alto grau de colaboração denota liderança, popularidade e intensa atividade científica. (KRACKHARDT, 2010; FREEMAN, 1979). O nodo pode ser visto como um grande canal de informação, retendo ou transmitindo informações (FREEMAN, 1979).</p> <p>A identificação desses pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Indiquem os mesmos para atuarem em atividades e ações do PPG, tais como, projetos de pesquisa, organização de eventos, etc.</p> <p>() Concordo plenamente () Concordo parcialmente () Indiferente () Discordo parcialmente () Discordo totalmente</p> <p>Ação B: Indiquem os mesmos para atuarem como disseminadores de informações na rede.</p> <p>() Concordo plenamente () Concordo parcialmente () Indiferente () Discordo parcialmente () Discordo totalmente</p> <p>Ação C: Divulguem os resultados para os demais pesquisadores e comunidade acadêmica, como uma ação de endomarketing.</p> <p>() Concordo plenamente () Concordo parcialmente () Indiferente () Discordo parcialmente () Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>
-----------	---

2. Quem são os pesquisadores que menos colaboram?	
D2	() Alto () Médio () Baixo
D3	<p>No sentido contrário dos argumentos de Krackhardt (2010) e Freeman (1979) um baixo grau de colaboração denota pouca liderança, popularidade.</p> <p>A identificação desses pesquisadores permite que os gestores do</p>

	<p>PPG:</p> <p>Ação A: Orientem os mesmos a intensificarem sua atividade colaborativa com outros pesquisadores.</p> <p>() Concordo plenamente () Concordo parcialmente</p> <p>() Indiferente</p> <p>() Discordo parcialmente () Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>
--	---

<p>3. Quem são os pesquisadores mais próximos a todos os outros pesquisadores?</p> <p>Em uma perspectiva de rede, maior proximidade refere-se ao menor caminho entre dois nodos.</p>	
D2	() Alto () Médio () Baixo
D3	<p>Um grau alto de proximidade denota maiores possibilidades de estabelecimento de parcerias de publicação por estar mais próximos em relação a todos os outros (SOUZA; BARBASTEFANO; LIMA, 2012). Posições mais próximas a todas as outras podem alcançar outras posições mais rapidamente, obtendo e disseminando informação, auxiliando a transferência de informação (KRACKHARDT, 2010).</p> <p>A identificação desses pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Indiquem os mesmos para tarefas que demandem a disseminação de informações na rede, tais como, divulgação de eventos, etc.</p> <p>() Concordo plenamente () Concordo parcialmente</p> <p>() Indiferente</p> <p>() Discordo parcialmente () Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

4. Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram?	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>Um grau alto de intermediação indica o poder de intermediação ou conexão de grupos (MARTINS et al., 2012). Pode ser considerado uma “ponte” entre distintos grupos de pesquisa, uma vez que está no caminho mais curto entre outros nodos (FISCHBACH; PUTZKE; SCHODER, 2011).</p> <p>Ação A: A identificação desses pesquisadores permite que os gestores do PPG indiquem os mesmos para atuarem e/ou orientarem pesquisas de cunho multi ou interdisciplinar.</p> <p><input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente</p> <p><input type="checkbox"/> Indiferente</p> <p><input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

5. Quais os grupos de pesquisadores que trabalham isoladamente?	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>A existência de grupos que trabalham isoladamente denota fragmentação da rede de pesquisadores. Quanto mais grupos, mais fragmentada a rede está.</p> <p>A identificação desses grupos permite que os gestores do PPG:</p> <p>Ação A: Orientem os grupos isolados a atuarem de forma mais colaborativa (quando a natureza do PPG permitir).</p> <p><input type="checkbox"/> Concordo plenamente <input checked="" type="checkbox"/> Concordo parcialmente</p> <p><input type="checkbox"/> Indiferente</p> <p><input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Ação B: Orientem esses grupos a ampliarem seu número de</p>

	colaboradores. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente Você sugere outras ações de gestão?
--	--

6. Qual o nível geral de atividade de colaboração da rede?	
Em uma perspectiva de rede, onde os pesquisadores estão ligados pela coautoria de produções, o nível geral de atividade de colaboração é determinado pela quantidade de ligações existentes comparado a quantidade de ligações possíveis (quando todos estão ligados com todos).	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>Um alto grau de densidade indica alta atividade de colaboração (KATERNDAHL, 2012).</p> <p>A identificação desse nível permite que os gestores do PPG:</p> <p>Ação A: Orientem os pesquisadores a ampliar/diversificar suas parcerias de colaboração, quando o nível for baixo:</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente <p>Ação B: Orientem os pesquisadores a manter os níveis atuais de colaboração e/ou expandir as fronteiras de colaboração para o nível inter PPG, interinstitucional e internacional, quando o nível for alto</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente <p>Você sugere outras ações de gestão?</p>

7. Qual o nível médio* de colaboração dos atores da rede?	
*Medido pela média das colaborações de todos os pesquisadores.	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>O grau médio reflete a média de coautores de cada pesquisador. Um valor baixo indica que, na média, os pesquisadores não produzem muito em coautoria.</p> <p>Ação A: A identificação desse nível permite que os gestores do PPG orientem os pesquisadores a colaborar mais, no caso do nível ser baixo, ou a incentivar a continuação da colaboração, caso o nível seja alto.</p> <p><input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

8. Quem são os colaboradores do pesquisador “x”?	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>A identificação desses colaboradores/pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Na presunção de que esses colaboradores pesquisem nos mesmos tópicos do pesquisador “x”, indiquem os mesmos para formação de futuras parcerias de pesquisa na área do pesquisador “x”.</p> <p><input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Ação B: Orientem os colaboradores do pesquisador “x” a procurar o próprio pesquisador “x” para fomentar outros tipos de colaboração, tais como, participação em projetos.</p> <p><input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p>

	Você sugere outras ações de gestão?
--	--

9. Quem são os pesquisadores que colaboram no tópico “y”?	
--	--

D2	() Alto () Médio () Baixo
-----------	------------------------------

D3	<p>A identificação de temáticas permite definir estratégias de atuação e posicionamento. Temas relevantes em nível nacional e internacional, mas pouco desenvolvidos no âmbito do PPG, podem ser desenvolvidos através da busca por fomento ou parcerias. Ao contrário, temas bem desenvolvidos dentro do programa, são muitas vezes latentes, e podem ser mais explorados, por exemplo, através da criação de grupos de pesquisa.</p>
-----------	--

A identificação desses pesquisadores permite que os gestores do PPG:

Ação A: Orientem os mesmos a **formarem** parcerias de colaborações de pesquisa no tópico “y”.

() Concordo plenamente () Concordo parcialmente

() Indiferente

() Discordo parcialmente () Discordo totalmente

Ação B: Direcionem as oportunidades de ações dentro do programa, tais como projetos, editais de fomento à pesquisa, oportunidades de experiências no exterior (pós-doc, professor visitante, etc.)

() Concordo plenamente () Concordo parcialmente

() Indiferente

() Discordo parcialmente () Discordo totalmente

Você sugere outras ações de gestão?

10. Quem é o pesquisador que mais colabora no tópico “y”?
--

D2	() Alto () Médio () Baixo
D3	<p>Ação A: A identificação desse pesquisador permite que os gestores do PPG indiquem o pesquisador mais qualificado (que produziu mais no referido tópico) para formação de futura parceria de pesquisa no tópico “y”.</p> <p>() Concordo plenamente () Concordo parcialmente () Indiferente () Discordo parcialmente () Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

11. Quais grupos de pesquisadores colaboram no tópico “y”?	
D2	() Alto () Médio () Baixo
D3	<p>A identificação desses grupos de pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Orientem grupos distintos a unirem forças para participar de editais de fomento à pesquisa no referido tópico</p> <p>() Concordo plenamente () Concordo parcialmente () Indiferente () Discordo parcialmente () Discordo totalmente</p> <p>Ação B: Orientem os distintos grupos de pesquisadores que colaboram no referido tópico a unirem forças para e produzir de forma mais qualitativa.</p> <p>() Concordo plenamente () Concordo parcialmente () Indiferente () Discordo parcialmente () Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

12. Qual o perfil das colaborações do pesquisador “x”?

D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>A identificação desse perfil permite que os gestores do PPG:</p> <p>Ação A: Indiquem o pesquisador para colaborar nos tópicos que compõem o referido perfil. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Ação B: Indiquem o pesquisador para orientar alunos com interesse nos tópicos do perfil. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

13. Quem são os pesquisadores que possuem perfis similares a um perfil específico*? *Perfil específico corresponde a um ou mais tópicos previamente estabelecidos.	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>A identificação desse perfil permite que os gestores do PPG:</p> <p>Ação A: Indiquem os pesquisadores para colaborar e/ou orientar nos tópicos que compõem o referido perfil. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Ação B: Sugiram pesquisadores para a composição de bancas. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

14. Qual o grau de similaridade* (matching) dos perfis de colaboração de dois pesquisadores? *Medido pela similaridade dos perfis dos pesquisadores.	
D2	() Alto () Médio () Baixo
D3	<p>A identificação desse grau permite que os gestores do PPG:</p> <p>Ação A: Orientem pesquisadores com alto grau de similaridade, que não possuem interação, a unirem forças e colaborarem em prol da quantidade e qualidade das produções. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Ação B: Indiquem pesquisadores com alto grau de similaridade para orientação e coorientação de pesquisa de mestrado ou doutorado (quando a natureza do programa assim permitir). <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Ação C: Indiquem ou busquem pesquisadores com alto grau de similaridade para a composição de bancas. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente</p> <p>Você sugere outras ações de gestão?</p>

15. Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador “x”?	
D2	() Alto () Médio () Baixo
D3	<p>A identificação desse grau permite que os gestores do PPG:</p> <p>Ação A: Orientem os pesquisadores que possuem perfis semelhantes para colaborar. <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente</p>

	<input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente Você sugere outras ações de gestão?
--	---

16. Quem são os pesquisadores que, com base na proximidade e na similaridade de seus perfis com o pesquisador “x”, podem colaborar com esse pesquisador?	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>Ação A: A identificação desses pesquisadores permite que os gestores do PPG indiquem os mesmos ao pesquisador “x”, fomentando as possibilidades de colaboração.</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente
	Você sugere outras ações de gestão?

17. Qual o perfil dos pesquisadores que mais colaboram?	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>A identificação desses pesquisadores permite que os gestores do PPG:</p> <p>Ação A: Indiquem, de forma mais qualificada (baseada no perfil), os mesmos para atuarem em projetos de pesquisa, organização de eventos, etc.</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente
	<p>Ação B: Indiquem, de forma mais qualificada (baseada no perfil), os mesmos para atuarem como disseminadores de ideias, eventos, etc.</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente

	<input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente Você sugere outras ações de gestão?
--	---

18. Qual o perfil dos grupos que trabalham isoladamente?	
D2	<input type="checkbox"/> Alto <input type="checkbox"/> Médio <input type="checkbox"/> Baixo
D3	<p>A identificação desses grupos permite que os gestores do PPG:</p> <p>Ação A: orientem os mesmos para atuarem de forma mais colaborativa.</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente <p>Ação B: Orientem os mesmos a ampliarem seu número de colaboradores.</p> <input type="checkbox"/> Concordo plenamente <input type="checkbox"/> Concordo parcialmente <input type="checkbox"/> Indiferente <input type="checkbox"/> Discordo parcialmente <input type="checkbox"/> Discordo totalmente <p>Você sugere outras ações de gestão?</p>

APÊNDICE C - Termo de Consentimento Livre e Esclarecido

Eu, _____, consinto em participar voluntariamente da pesquisa realizada pela doutoranda **Andréa Sabedra Bordin**, do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da UFSC, sob orientação do **Prof. Dr. Alexandre Leopoldo Gonçalves**, para fins de investigação científica sobre o título: ***Framework de conhecimento para análise da colaboração científica.***

Eu concordo que o material e informações obtidas com a entrevista e formulário de pesquisa possam ser publicados, de forma anônima, para a pesquisa científica de Doutorado e, posteriormente, possam ser produzidos materiais para aulas, congressos, eventos científicos, palestras ou periódicos científicos, além de concordar com a gravação.

Estou ciente de que minha participação na entrevista durará aproximadamente 120 minutos, onde serão abordados aspectos referentes a minha percepção como gestor de um Programa de Pós-Graduação, em relação aos resultados e as possíveis ações de gestão que esses resultados sugerem.

Estou ciente de que poderei solicitar informações durante qualquer fase da pesquisa, inclusive após a sua publicação.

Se eu tiver qualquer dúvida a respeito da pesquisa, poderei entrar em contato com a pesquisadora através do e-mail: andrea.bordin@posgrad.ufsc.br. Também poderei manter contato com o orientador do trabalho através do e-mail a.l.goncalves@ufsc.br

Nome por extenso:	
Cargo:	
Local e Data:	
Assinatura:	

APÊNDICE D - Instrumento de Verificação das Perguntas de Análise de Colaboração Científica³⁷

Verificação de demandas para análise de colaboração científica

Prezado(a) Coordenador(a),
Baseado na sua experiência de gestor(a), verifique a contribuição das seguintes questões para a análise da colaboração científica do seu Programa de Pós-Graduação.
Por favor, marque a opção que indica o grau de relevância de cada questão sugerida pela pesquisa e insira, caso deseje, mais questões no final do questionário.

Para analisar a colaboração científica é importante saber:

1. Quem são os pesquisadores que mais colaboram?

Medido pela quantidade de coautores e quantidade de artigos coautorados.

- Muito relevante
- Relevante
- Pouco relevante

2. Quem são os pesquisadores que menos colaboram?

- Muito relevante
- Relevante
- Pouco relevante

3. Quem são os pesquisadores mais próximos aos demais pesquisadores?

Em uma perspectiva de rede, onde os pesquisadores estão ligados pela coautoria de produções, pesquisadores que são mais próximos (menor caminho) aos demais pesquisadores, oferecem mais possibilidades de colaboração.

- Muito relevante
- Relevante
- Pouco relevante

4. Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram?

- Muito relevante
- Relevante
- Pouco relevante

5. Quais os grupos de pesquisadores que trabalham isoladamente?

- Muito relevante
- Relevante
- Pouco relevante

6. Qual o nível geral de atividade de colaboração dos pesquisadores?

Em uma perspectiva de rede, onde os pesquisadores estão ligados pela coautoria de produções, o nível geral de atividade de colaboração é determinado pela quantidade de ligações existentes comparado a quantidade de ligações possíveis (quando todos estão ligados com todos).

- Muito relevante
- Relevante
- Pouco relevante

7. Qual o nível médio de colaboração dos pesquisadores?

Medido pela média das colaborações de todos os pesquisadores.

- Muito relevante
- Relevante
- Pouco relevante

8. Quem são os colaboradores do pesquisador "x"?

"x" refere-se a um pesquisador específico.

- Muito relevante
- Relevante
- Pouco relevante

9. Quem são os pesquisadores que colaboram no tópico "y"?

- Muito relevante
- Relevante
- Pouco relevante

10. Quem é o pesquisador que mais colabora no tópico "y"?

- Muito relevante
- Relevante
- Pouco relevante

11. Quais grupos de pesquisadores colaboram no tópico "y"?

- Muito relevante
- Relevante
- Pouco relevante

12. Qual o perfil* das colaborações do pesquisador "x"?

*Perfil: conjunto de tópicos de interesse do pesquisador, criado a partir do conteúdo (título e palavras-chave) das produções C&T produzidas em coautoria pelo pesquisador.

- Muito relevante
- Relevante
- Pouco relevante

13. Quem são os pesquisadores que possuem perfis similares a um perfil específico?

*Perfil específico é composto por tópicos (termos ou palavras-chave) especificado pelo gestor.

- Muito relevante
- Relevante
- Pouco relevante

14. Qual o grau de similaridade* dos perfis de colaboração de dois pesquisadores?

*Medido pela similaridade dos perfis dos pesquisadores.

- Muito relevante
- Relevante
- Pouco relevante

15. Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador "x"?

- Muito relevante
- Relevante
- Pouco relevante

16. Quem são os pesquisadores que, com base em seus perfis e na proximidade com o pesquisador "x", podem colaborar com esse pesquisador?

- Muito relevante
- Relevante
- Pouco relevante

17. Qual o perfil dos pesquisadores que mais colaboram?

- Muito relevante
- Relevante
- Pouco relevante

18. Qual o perfil dos grupos que trabalham isoladamente?

- Muito relevante
- Relevante
- Pouco relevante

Sugestões de questões:

Submit



100%: You made it.

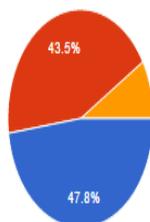
Never submit passwords through Google Forms.

APÊNDICE E - Sumário das Respostas do Instrumento de Verificação das Perguntas de Análise de Colaboração Científica³⁷

Summary

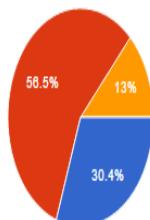
Para analisar a colaboração científica é importante saber:

1. Quem são os pesquisadores que mais colaboram?



Muito relevante	11	47.8%
Relevante	10	43.5%
Pouco relevante	2	8.7%

2. Quem são os pesquisadores que menos colaboram?

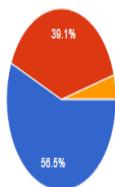


Muito relevante	7	30.4%
Relevante	13	56.5%
Pouco relevante	3	13%

³⁷ Link para o formulário:

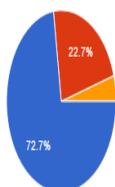
https://docs.google.com/forms/d/1Eyi_dj8258qhcSFMXzXR34h2HeQnydIgH12_OAFa7VY/viewform

3. Quem são os pesquisadores mais próximos aos demais pesquisadores?



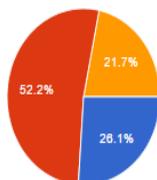
Muito relevante	13	56.5%
Relevante	9	39.1%
Pouco relevante	1	4.3%

4. Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram?



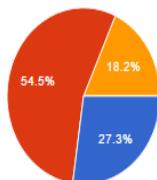
Muito relevante	16	72.7%
Relevante	5	22.7%
Pouco relevante	1	4.5%

5. Quais os grupos de pesquisadores que trabalham isoladamente?



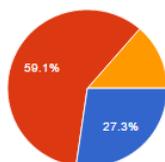
Muito relevante	6	26.1%
Relevante	12	52.2%
Pouco relevante	5	21.7%

6. Qual o nível geral de atividade de colaboração dos pesquisadores?



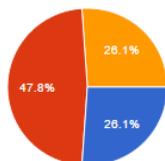
Muito relevante	6	27.3%
Relevante	12	54.5%
Pouco relevante	4	18.2%

7. Qual o nível médio de colaboração dos pesquisadores?



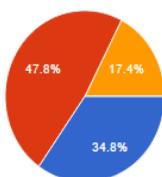
Muito relevante	6	27.3%
Relevante	13	59.1%
Pouco relevante	3	13.6%

8. Quem são os colaboradores do pesquisador "x"?



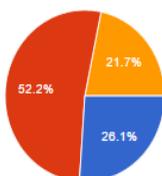
Muito relevante	6	26.1%
Relevante	11	47.8%
Pouco relevante	6	26.1%

9. Quem são os pesquisadores que colaboram no tópico "y"?



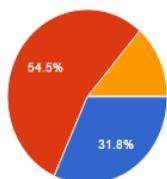
Muito relevante	8	34.8%
Relevante	11	47.8%
Pouco relevante	4	17.4%

10. Quem é o pesquisador que mais colabora no tópico "y"?



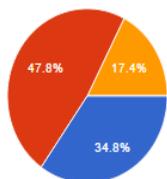
Muito relevante	6	26.1%
Relevante	12	52.2%
Pouco relevante	5	21.7%

11. Quais grupos de pesquisadores colaboram no tópico “y”?



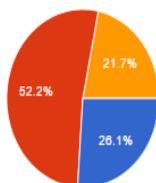
Muito relevante	7	31.8%
Relevante	12	54.5%
Pouco relevante	3	13.6%

12. Qual o perfil* das colaborações do pesquisador “x”?



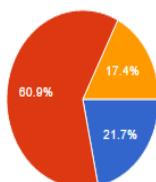
Muito relevante	8	34.8%
Relevante	11	47.8%
Pouco relevante	4	17.4%

13. Quem são os pesquisadores que possuem perfis similares a um perfil específico*?



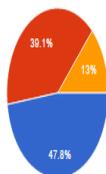
Muito relevante	6	26.1%
Relevante	12	52.2%
Pouco relevante	5	21.7%

14. Qual o grau de similaridade* dos perfis de colaboração de dois pesquisadores?



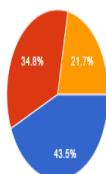
Muito relevante	5	21.7%
Relevante	14	60.9%
Pouco relevante	4	17.4%

15. Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador "x"?



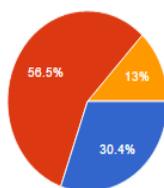
Muito relevante	11	47.8%
Relevante	9	39.1%
Pouco relevante	3	13%

16. Quem são os pesquisadores que, com base em seus perfis e na proximidade com o pesquisador "x", podem colaborar com esse pesquisador?



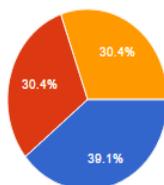
Muito relevante	10	43.5%
Relevante	8	34.8%
Pouco relevante	5	21.7%

17. Qual o perfil dos pesquisadores que mais colaboram?



Muito relevante	7	30.4%
Relevante	13	56.5%
Pouco relevante	3	13%

18. Qual o perfil dos grupos que trabalham isoladamente?



Muito relevante	9	39.1%
Relevante	7	30.4%
Pouco relevante	7	30.4%

Sugestões de questões:

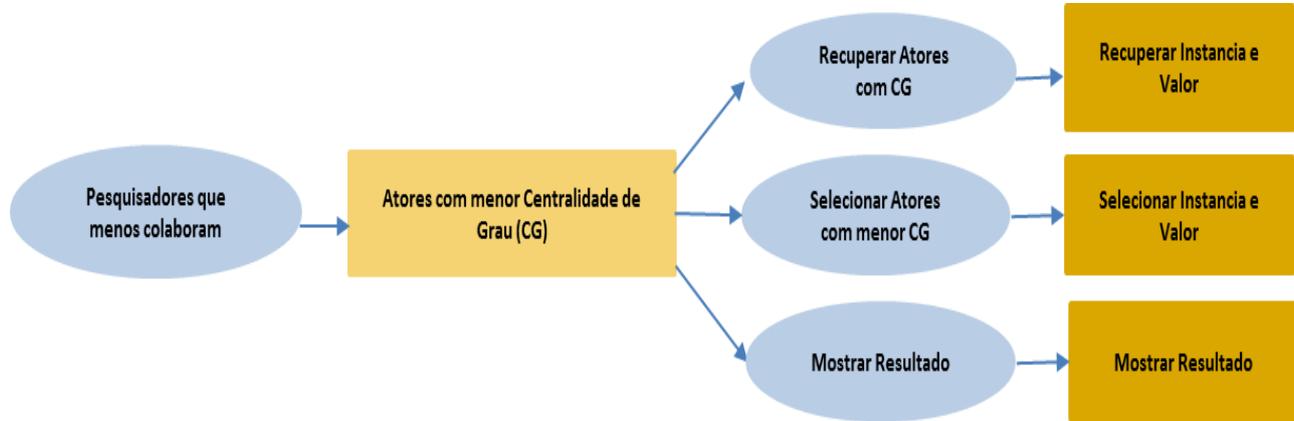
A colaboração científica pressupõe que cada um dos participantes deva ter uma participação efetiva no trabalho e não apenas ser a pessoa que cede o equipamento ou reagentes. Atualmente as boas revistas já exigem que cada um dos participantes de um estudo

descrevam qual sua participação naquele trabalho específico.

Todas as questões apresentadas são importantes para que a gestão possa identificar quem pode contribuir com quem, quem pode ajudar fazer meio campo para outros colaborarem, e assim por diante.

APÊNDICE F - Modelagem da Estrutura das Tarefas de Análise de Colaboração Científica

Figura 58 - Modelagem da Tarefa 2 “Quem são os pesquisadores que menos colaboram?”



Fonte: Autor

Quadro 31 - Quadro explicativo da modelagem da Tarefa 2

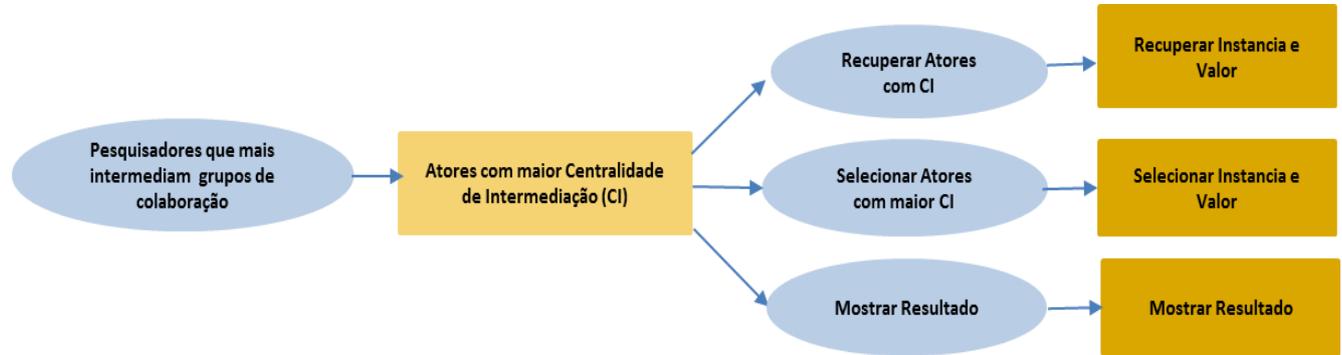
Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Atores com menor	1- Recuperar atores com CG	Recuperar Instância e Valor

Centralidade de Grau (CG)		<p>recuperarInstanciaValor (Propriedade de objeto, propriedade de dados): lista de instâncias</p> <p>Descrição: Esse método recebe o nome de uma propriedade de objeto e de uma propriedade de dados e retorna a lista de instâncias relacionadas com a propriedade de objeto, conjuntamente com o valor da propriedade de dados de cada instância.</p>
	2- Selecionar atores com menor CG	<p>Selecionar Instancia e Valor</p> <p>selecionarInstanciaValor (lista de instâncias, filtro ordem, filtro limite): lista de instâncias</p> <p>Descrição: Esse método seleciona instâncias do parâmetro lista de instâncias através do parâmetro filtro ordem, que determinada se a ordem é ascendente ou descendente, e do</p>

		parâmetro filtro limite que determinada o número de instâncias selecionadas.
	3 – Mostrar resultados	<p>Mostrar resultados</p> <p>mostrarResultados (lista de instâncias)</p> <p>Descrição: Esse método exibe as propriedades, por exemplo, nome, das instâncias da lista de instâncias.</p>

Fonte: Autor.

Figura 59 - Modelagem da Tarefa 4 “Quem são os pesquisadores que mais atuam como intermediador/facilitador entre diferentes grupos de pesquisadores que colaboram?”



Fonte:

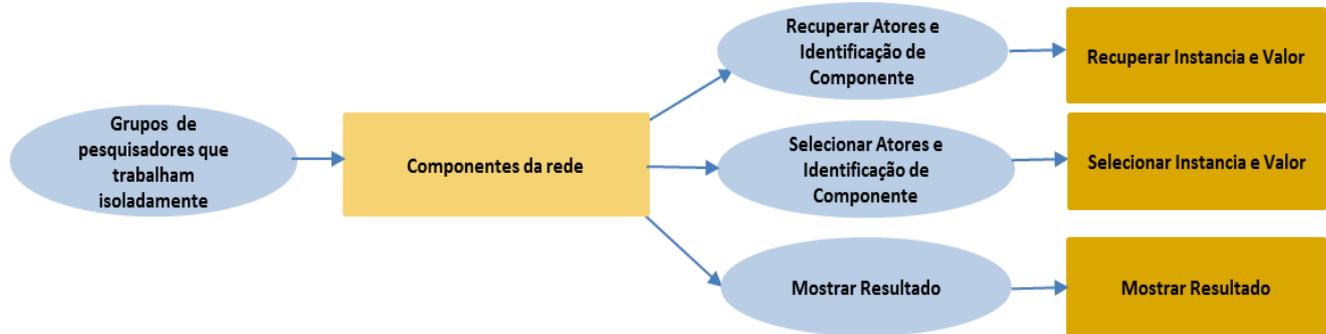
Quadro 32 - Quadro explicativo da modelagem da Tarefa 4

Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Atores com menor Centralidade de Intermediação (CI)	1- Recuperar atores com CI	Recuperar Instância e Valor (Especificado anteriormente)
	2- Selecionar atores com maior CI	Selecionar Instancia e Valor

		(Especificado anteriormente)
	3 – Mostrar resultados	Mostrar resultados (Especificado anteriormente)

Fonte: Autor.

Figura 60 - Modelagem da Tarefa 5 “Quais os grupos de pesquisadores que trabalham isoladamente?”



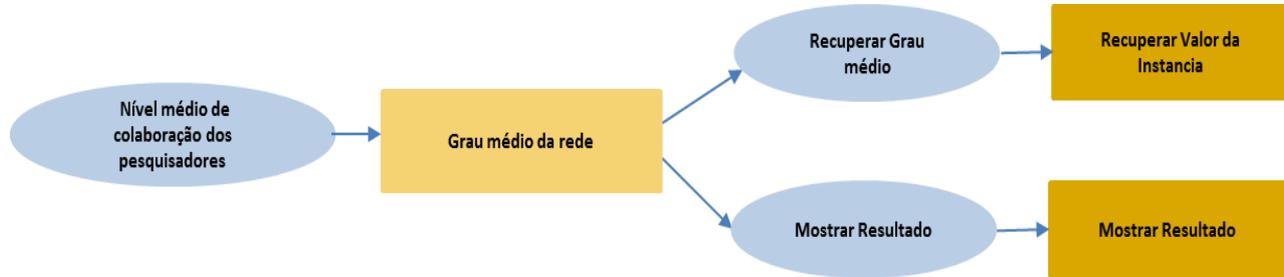
Fonte: Autor.

Quadro 33 - Quadro explicativo da modelagem da Tarefa 5

Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Componentes da rede	1- Recuperar atores e identificação de componente	Recuperar Instância e Valor (Especificado anteriormente)
	2- Selecionar atores e identificação de componente	Selecionar Instancia e Valor (Especificado anteriormente)
	3 – Mostrar resultados	Mostrar resultados (Especificado anteriormente)

Fonte: Autor.

Figura 61 - Modelagem da Tarefa 7 “Qual o nível médio de colaboração dos pesquisadores da rede?”



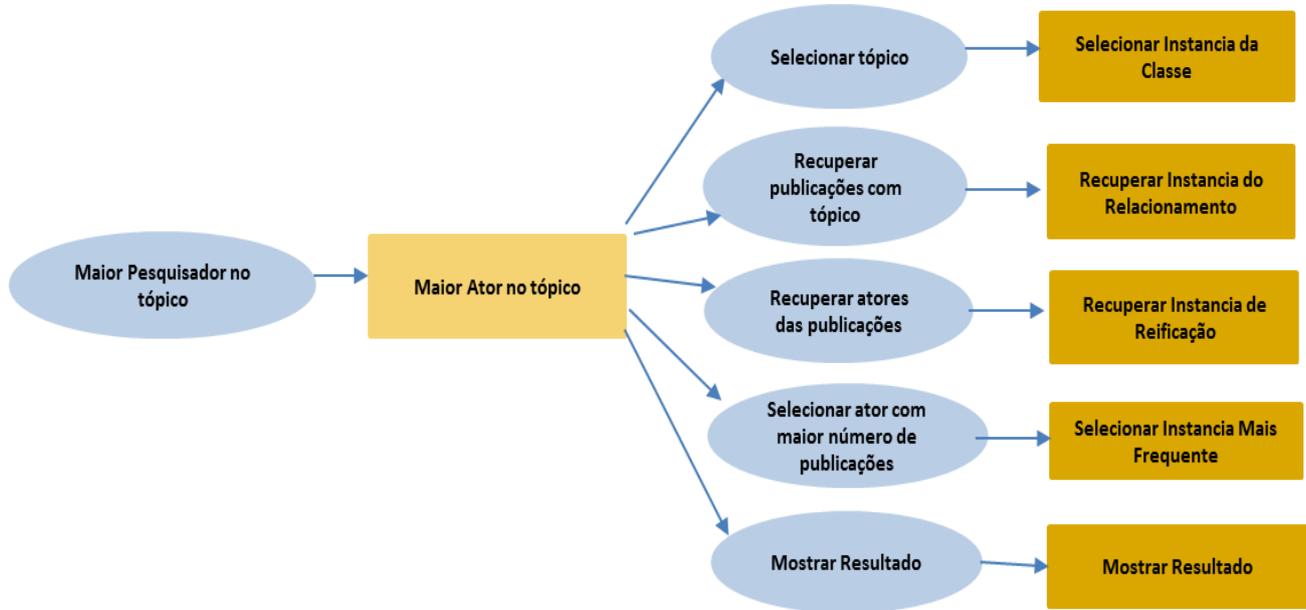
Fonte: Autor.

Quadro 34 - Quadro explicativo da modelagem da Tarefa 7

Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Nível médio de colaboração dos pesquisadores	1 – Recuperar Grau médio	Recuperar Valor da Instancia recuperarValorInstancia (classe, propriedade de dados): lista de valores

		<p>Descrição: Esse método recebe como parâmetro o nome de uma classe e de uma propriedade de dados e retorna uma lista de valores das instâncias da classe.</p>
	2 – Mostrar resultados	<p>Mostrar resultados</p> <p>mostrarResultados (lista de instâncias)</p> <p>Descrição: Esse método exibe as propriedades, por exemplo, nome, das instâncias da lista de instâncias.</p>

Figura 62 - Modelagem da Tarefa 10 “Quem é o pesquisador que mais colabora no tópico “x”?”



Fonte: Autor

Quadro 35 - Quadro explicativo da modelagem da Tarefa 10

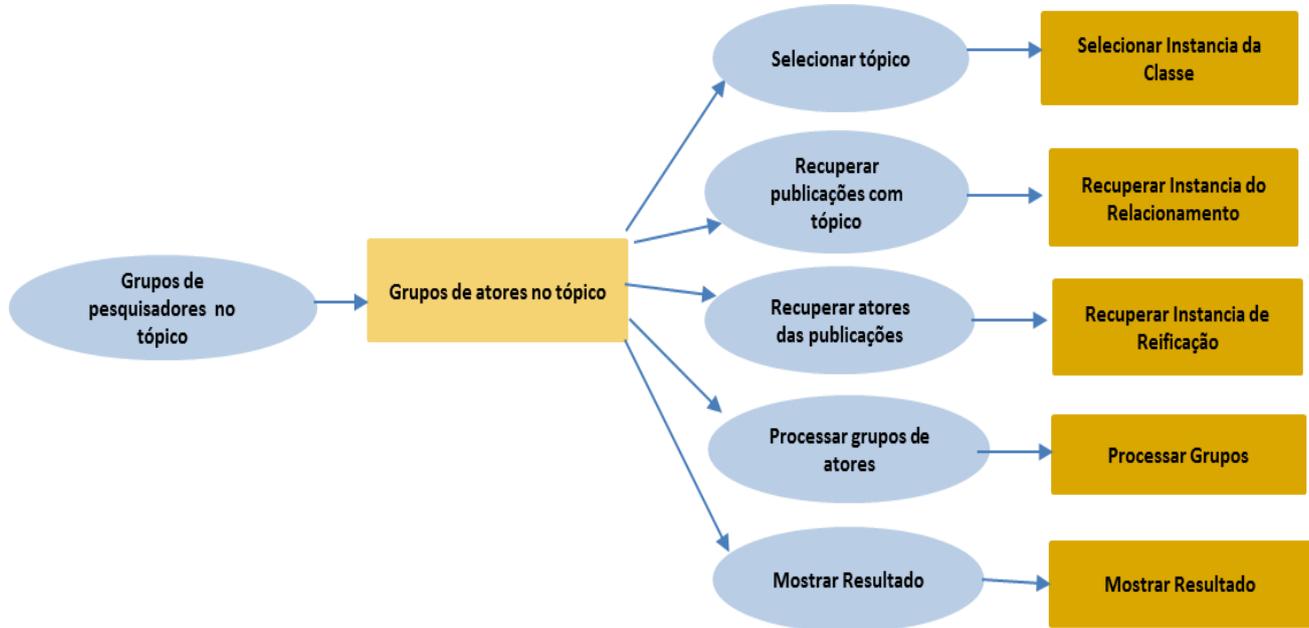
Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Maior ator no tópico	1-Selecionar tópico	<p>Selecionar Instancia da Classe</p> <p>selecionarInstanciadaClasseVivo (dado de entrada): lista de instâncias.</p> <p>Descrição: Esse método seleciona e retorna uma instancia correspondente ao dado de entrada (nome do pesquisador). Este método é específico para a ontologia de domínio usada no <i>framework</i>.</p>
	2- Recuperar publicações com tópicos	<p>Recuperar Instancia do Relacionamento</p> <p>recuperarInstanciadoRelacionamento (Lista de instâncias, propriedade de objeto, valor lógico – V ou F): lista de instâncias.</p> <p>Descrição: Esse método retorna uma lista de instâncias que relacionam com o parâmetro lista de instâncias através do parâmetro propriedade de objeto. O</p>

		parâmetro valor lógico indica se as instâncias da lista de parâmetro devem ser incluídas ou não na lista de instâncias de retorno.
	3 – Recuperar autores das publicações	<p>Recuperar Instancia da Reificação</p> <p>recuperarInstanciaReificacao (lista de instâncias, classe, propriedade de objeto, classe, propriedade de objeto): lista de instâncias</p> <p>Descrição: Esse método retorna uma lista de instâncias que se relacionam via reificação com o parâmetro lista de instâncias. Para isso são necessárias as classes e as propriedades de objeto dos relacionamentos.</p>
	4- Selecionar autor com maior número de publicações	<p>Selecionar Instancia Mais Frequente</p> <p>selecionarInstanciaMaisFrequente (Lista de instâncias): instancia</p> <p>Descrição: Esse método retorna a</p>

		instancia mais frequente na lista de instâncias recebida como parâmetro.
	5 – Mostrar resultados	Mostrar resultados mostrarResultados (lista de instâncias) Descrição: Esse método exhibe as propriedades, por exemplo, nome, das instâncias da lista de instâncias.

Fonte: Autor.

Figura 63 - Modelagem da Tarefa 11 “Quais grupos de pesquisadores colaboram no tópico “x”?”



Fonte: Autor.

Quadro 36 - Quadro explicativo da modelagem da Tarefa 11

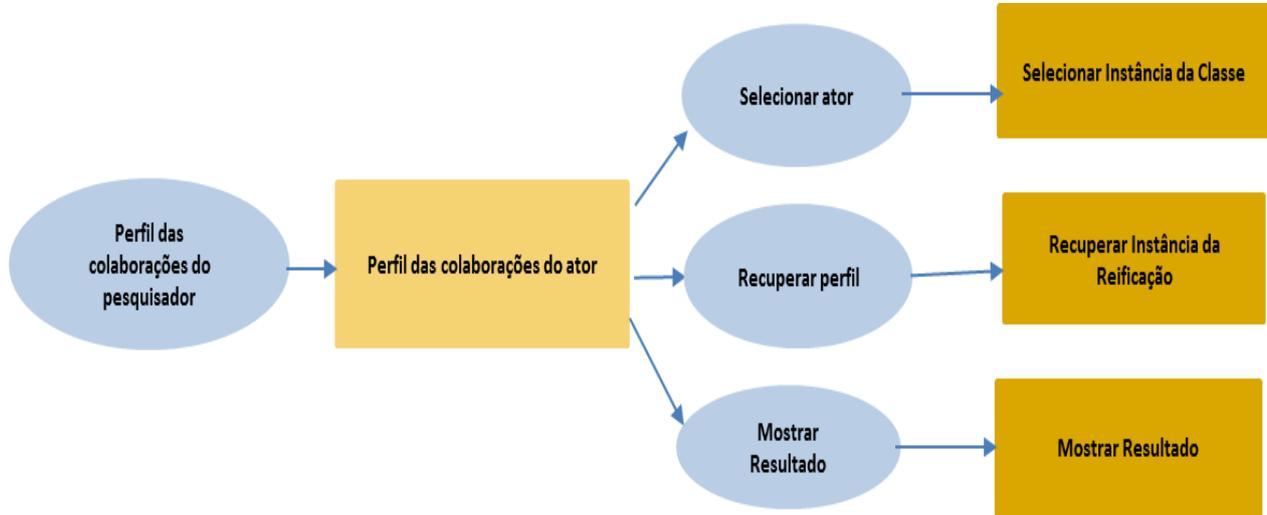
Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Grupos de atores no tópico	1-Selecionar tópico	Selecionar Instancia da Classe (Especificado anteriormente)
	2- Recuperar publicações com tópico	Recuperar Instancia do Relacionamento (Especificado anteriormente)
	3 – Recuperar autores das publicações	<p>Recuperar Instancia da Reificação</p> <p>recuperarInstanciaReificacao3 (lista de instâncias, classe, propriedade de objeto, classe, propriedade de objeto): mapa com as instâncias que iniciam a ligação como chave e a lista de instâncias associadas.</p> <p>Descrição: Esse método retorna um mapa com as instâncias que iniciam a ligação como chave e a lista de instâncias associadas via reificação. Para isso são necessárias as classes e as</p>

		propriedades de objeto dos relacionamentos.
	4 – Processar grupos de atores	<p>Processar Grupos</p> <p>processarGrupos (mapa com instâncias e lista de instâncias associadas): mapa com os identificadores dos grupos e as instancias associadas ao grupo.</p> <p>Descrição: Esse método retorna um mapa com os grupos de atores criados a partir do parâmetro mapa. Os grupos são criados a partir da coautoria das publicações. Estarão no mesmo grupo todos os autores de publicações que coautoraram diretamente ou indiretamente. Por exemplo, a publicação P1 tem como coautores A1 e A2, a publicação P2 tem como coautores A2 e A3 e a publicação</p>

		P3 tem como coautores A4 e A5. Serão formados dois grupos compostos pelos autores A1, A2 e A3 em um grupo e A4 e A5 no segundo grupo.
	5 – Mostrar resultados	<p>Mostrar resultados</p> <p>mostrarResultados (mapa com os identificadores dos grupos e as instâncias associadas ao grupo)</p> <p>Descrição: Esse método exibe o identificador do grupo e o nome dos pesquisadores vinculados ao grupo.</p>

Fonte: Autor.

Figura 64 - Modelagem da Tarefa 12 “Qual o perfil das colaborações do pesquisador “x”?”



Fonte: Autor.

Quadro 37 - Quadro explicativo da modelagem da Tarefa 12

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Perfil das colaborações do ator	1- Selecionar ator	Selecionar Instancias da Classe (Especificado anteriormente)
	2- Recuperar perfil	<p>Recuperar Instancia da Reificação</p> <p>recuperarInstanciaReificacao2 (lista de instância(s), classe, propriedade de objeto, classe, propriedade de dado, propriedade de objeto): mapa de instâncias chave-valor.</p> <p>Descrição: Esse método retorna um mapa com as instâncias que iniciam a ligação como chave e a lista de instâncias e valor associadas via reificação como valor. Para a recuperação das instâncias associadas são necessárias as classes e as propriedades de objeto dos</p>

		relacionamentos.
	3- Mostrar resultado	Mostrar resultado mostrarResultado (mapa de instâncias chave-valor) Descrição: Esse método exibe algumas propriedades das instâncias do mapa.

Fonte: Autor.

Figura 65 - Modelagem da Tarefa 13 “Quem são os pesquisadores que possuem perfis similares a um perfil específico?”



Fonte: Autor.

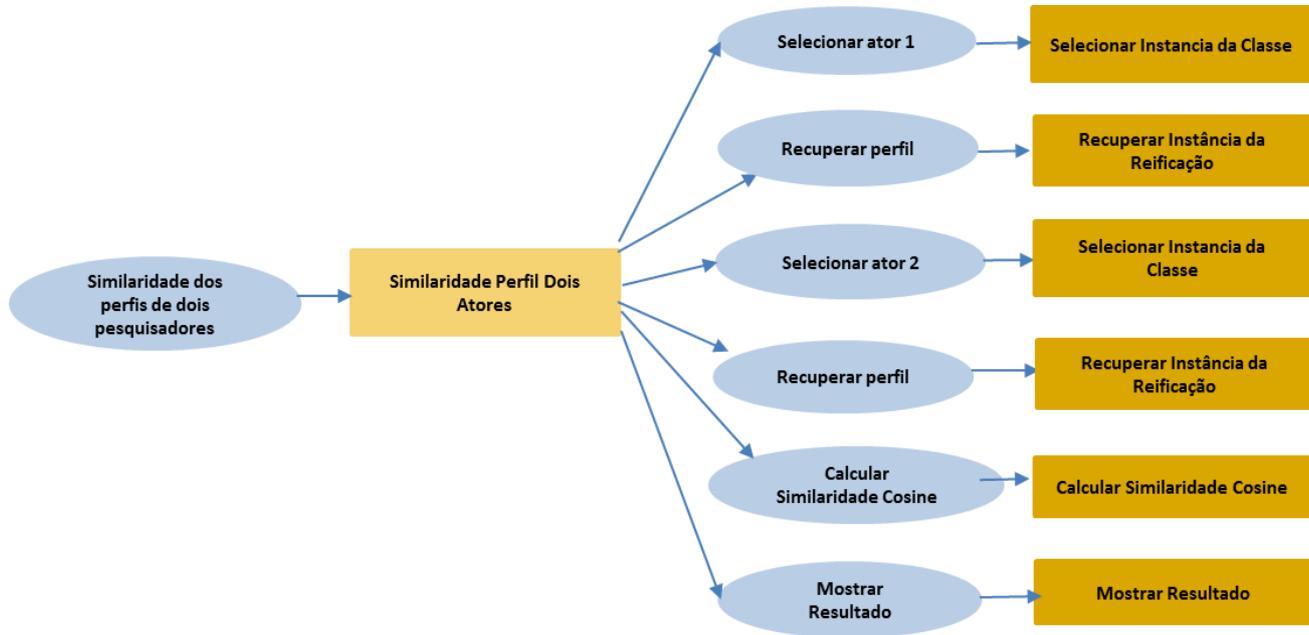
Quadro 38 - Quadro explicativo da modelagem da Tarefa 13

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Atores com perfis similares a um perfil específico	1- Recuperar atores	<p>Recuperar Instância da Classe</p> <p>recuperarInstanciadaClasse(classe): lista de instâncias.</p> <p>Descrição: Esse método recebe o nome da classe e retorna uma lista com todas as instâncias dessa classe.</p>
	2- Recuperar perfil	<p>Recuperar Instância da Reificação</p> <p>recuperarInstanciaReificacao2 (Especificado anteriormente)</p>
	3-Calcular Similaridade Freqüencia	<p>Calcular Similaridade Freqüencia</p> <p>calcularSimilaridadeFreqüencia (entrada de dados, mapa de instâncias chave-valor): lista de instâncias.</p> <p>Descrição: Esse método verifica a similaridade entre um vetor de termos origem e um conjunto de vetores</p>

		destino. A similaridade é calculada pela existência dos termos no vetor destino. A frequência absoluta dos termos no vetor destino estabelece o <i>ranking</i> . O método retorna uma lista de instâncias com valor (frequência) associados.
	4- Mostrar resultado	Mostrar resultado mostrarResultado (lista de instâncias) Descrição: Esse método exibe algumas propriedades das instâncias da lista.

Fonte: Autor.

Figura 66 - Modelagem da Tarefa 14 “Qual o grau de similaridade dos perfis de colaboração de dois pesquisadores?”



Fonte: Autor.

Quadro 39 - Quadro explicativo da modelagem da Tarefa 14

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
-----------------	-----------------------------------	-----------------------------------

Similaridade perfil dois atores	1- Selecionar Ator 1	Selecionar Instância da Classe (Especificado anteriormente)
	2- Recuperar perfil	Recuperar Instância da Reificação (Especificado anteriormente)
	3- Selecionar Ator 2	Selecionar Instância da Classe (Especificado anteriormente)
	4- Recuperar perfil	Recuperar Instância da Reificação (Especificado anteriormente)
	5- Calcular Similaridade Cosine	<p>Calcular Similaridade Cosine</p> <p>calcularSimilaridadeCosine (mapa de instâncias chave-valor 1, mapa de instâncias chave-valor 2): lista de instâncias.</p> <p>Descrição: Esse método verifica a similaridade entre o vetor termo-frequencia 1 e o vetor termo-frequencia 2. A similaridade é calculada através da equação de similaridade coseno (<i>cosine</i>).</p>
	4- Mostrar resultado	<p>Mostrar resultado</p> <p>mostrarResultado (lista de instâncias)</p>

		Descrição: Esse método exhibe o grau da similaridade.
--	--	--

Fonte: Autor.

Figura 67 - Modelagem da Tarefa 15 “Quem são os pesquisadores que, com base em seus perfis, podem colaborar com o pesquisador “x”?”



Fonte: Autor.

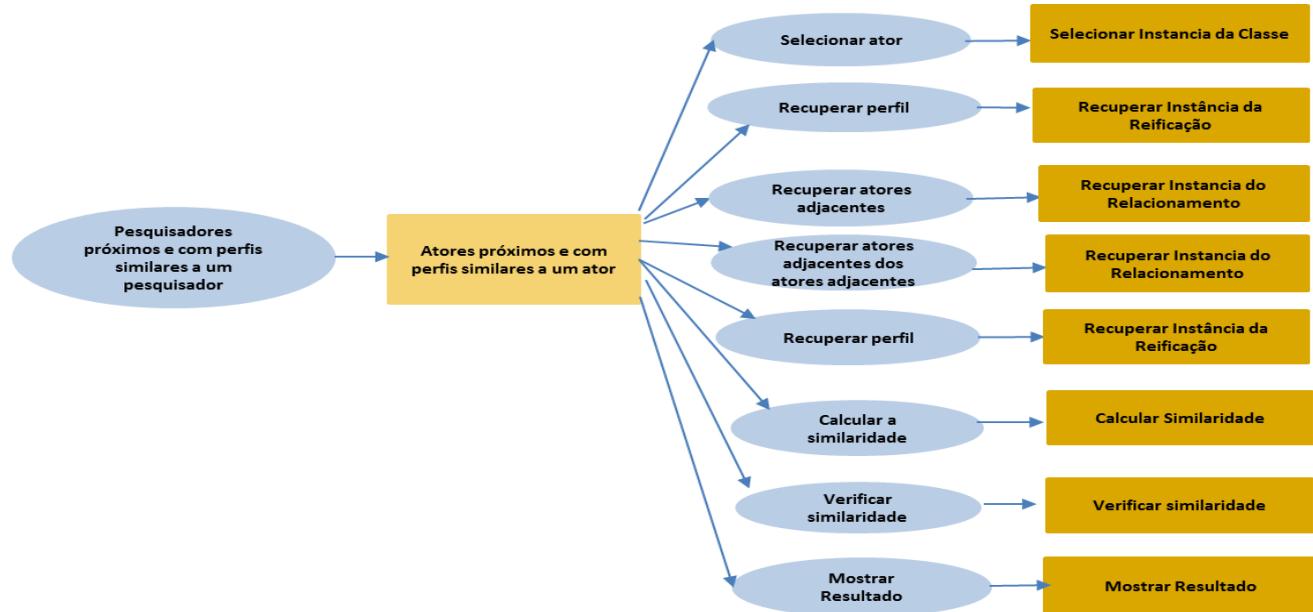
Quadro 40 - Quadro explicativo da modelagem da Tarefa 15

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Similaridade perfil dois atores	1- Selecionar Ator 1	Selecionar Instância da Classe (Especificado anteriormente)
	2- Recuperar perfil	Recuperar Instância da Reificação recuperarInstanciaReificacao2 (Especificado anteriormente)
	3- Recuperar Atores	Recuperar Instância da Classe (Especificado anteriormente)
	4- Recuperar perfil	Recuperar Instância da Reificação recuperarInstanciaReificacao2 (Especificado anteriormente)
	5- Calcular Similaridade Cosine	Calcular Similaridade Cosine calcularSimilaridadeCosine (mapa de instâncias chave-valor 1, mapa de instâncias chave-valor 2): lista de instâncias. Descrição: Esse método verifica a similaridade entre o vetor termo-frequencia 1 e o vetor termo-frequencia 2. A similaridade é calculada através da equação de similaridade coseno (<i>cosine</i>).

	6- Verificar Similaridade	<p>Verificar Similaridade</p> <p>verificarSimilaridade(lista de instâncias): lista de instâncias</p> <p>Descrição: Esse método recebe uma lista de instâncias termo-grau de similaridade e filtra as instâncias em função do grau de similaridade.</p>
	7- Mostrar resultado	<p>Mostrar resultado</p> <p>mostrarResultado (lista de instâncias)</p> <p>Descrição: Esse método exibe as propriedades nome do pesquisador e grau de similaridade da lista de instâncias.</p>

Fonte: Autor.

Figura 68 - Modelagem da Tarefa 16 “. Quem são os pesquisadores que, com base na proximidade e na similaridade de seus perfis com o pesquisador “x”, podem colaborar com esse pesquisador?”



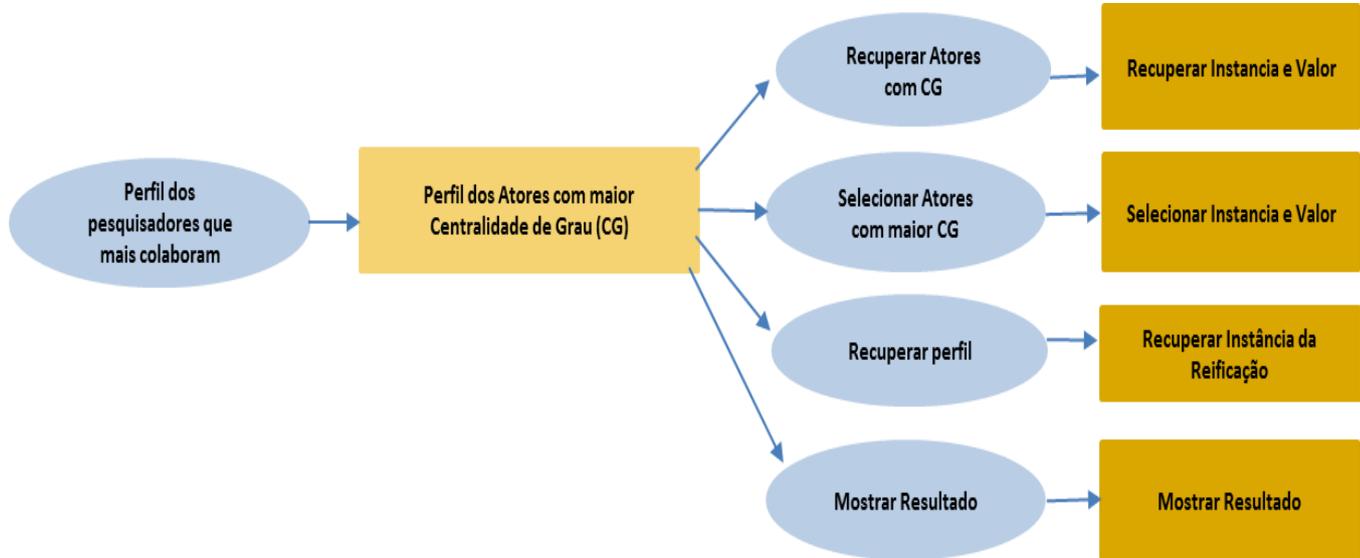
Fonte: Autor.

Quadro 41 - Quadro explicativo da modelagem da Tarefa 16

Método complexo	Subtarefas (passos de inferência)	Especificação do Método Primitivo
Atores próximos e com perfis similares a um ator	1- Selecionar Ator	Selecionar Instância da Classe (Especificado anteriormente)
	2- Recuperar perfil	Recuperar Instância da Reificação recuperarInstanciaReificacao2 (Especificado anteriormente)
	3- Recuperar Atores adjacentes	Recuperar Instância do Relacionamento (Especificado anteriormente)
	4- Recuperar Atores adjacentes dos Atores adjacentes	Recuperar Instância do Relacionamento (Especificado anteriormente)
	5- Recuperar Atores	Recuperar Instância da Classe (Especificado anteriormente)
	6- Recuperar perfil	Recuperar Instância da Reificação recuperarInstanciaReificacao2 (Especificado anteriormente)

	5- Calcular Similaridade Cosine	Calcular Similaridade Cosine (Especificado anteriormente)
	6- Verificar Similaridade	Verificar Similaridade (Especificado anteriormente)
	7- Mostrar resultado	Mostrar resultado mostrarResultado (lista de instâncias) Descrição: Esse método exibe as propriedades nome do pesquisador e grau de similaridade da lista de instâncias.

Figura 69 - Modelagem da Tarefa 17 “Qual o perfil dos pesquisadores que mais colaboram?”



Fonte: Autor.

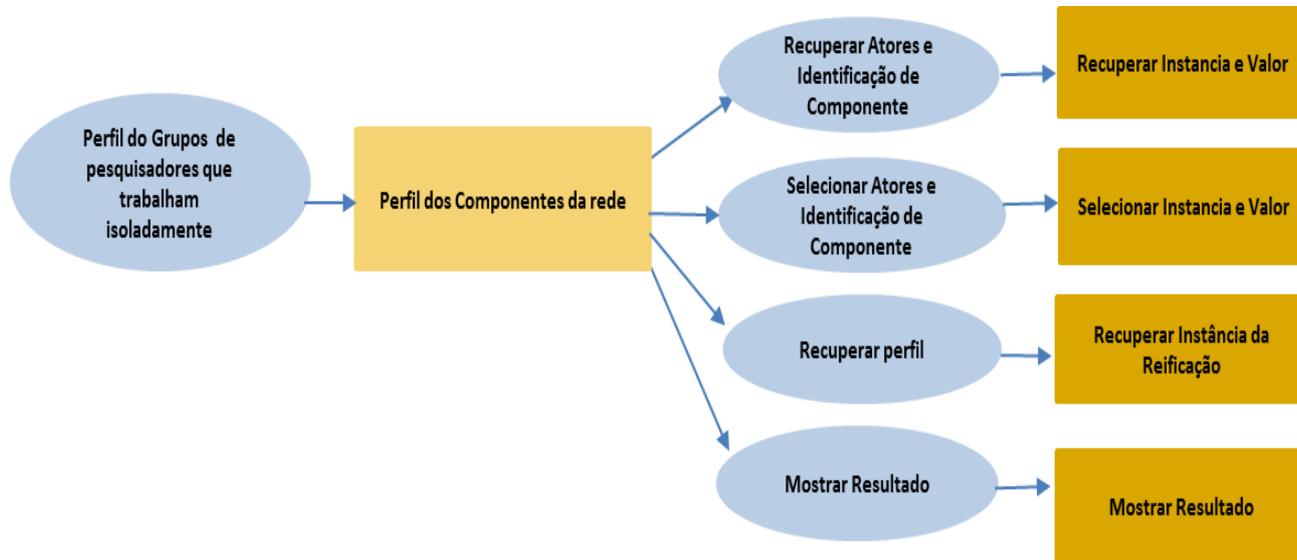
Quadro 42 - Quadro explicativo da modelagem da Tarefa 17

Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Perfil dos atores com maior Centralidade de Grau (CG)	1- Recuperar atores com CG	Recuperar Instância Valor (Especificado anteriormente)
	2- Selecionar atores com maior CG	Selecionar Instancia Valor (Especificado anteriormente)
	3- Recuperar perfil	<p>Recuperar Instância da Reificação</p> <p>recuperarInstanciaReificacao4((lista de instância(s), classe, propriedade de objeto, classe, propriedade de dado, propriedade de objeto): mapa de instâncias chave-valor.</p> <p>Descrição: Esse método retorna um mapa com as instâncias que iniciam a ligação como chave e a lista de instâncias e valor associadas via reificação como valor. Para a recuperação das instâncias associadas são necessárias as classes e as propriedades de objeto dos</p>

		relacionamentos
	4- Mostrar resultados	<p>Mostrar resultados</p> <p>mostrarResultados (lista de instâncias)</p> <p>Descrição: Esse método exibe as propriedades nome do pesquisador, termo e frequência das instâncias da lista de instâncias.</p>

Fonte: Autor.

Figura 70- Modelagem da Tarefa 18 “Qual o perfil dos grupos que trabalham isoladamente? ”



Fonte: Autor.

Quadro 43 - Quadro explicativo da modelagem da Tarefa 18

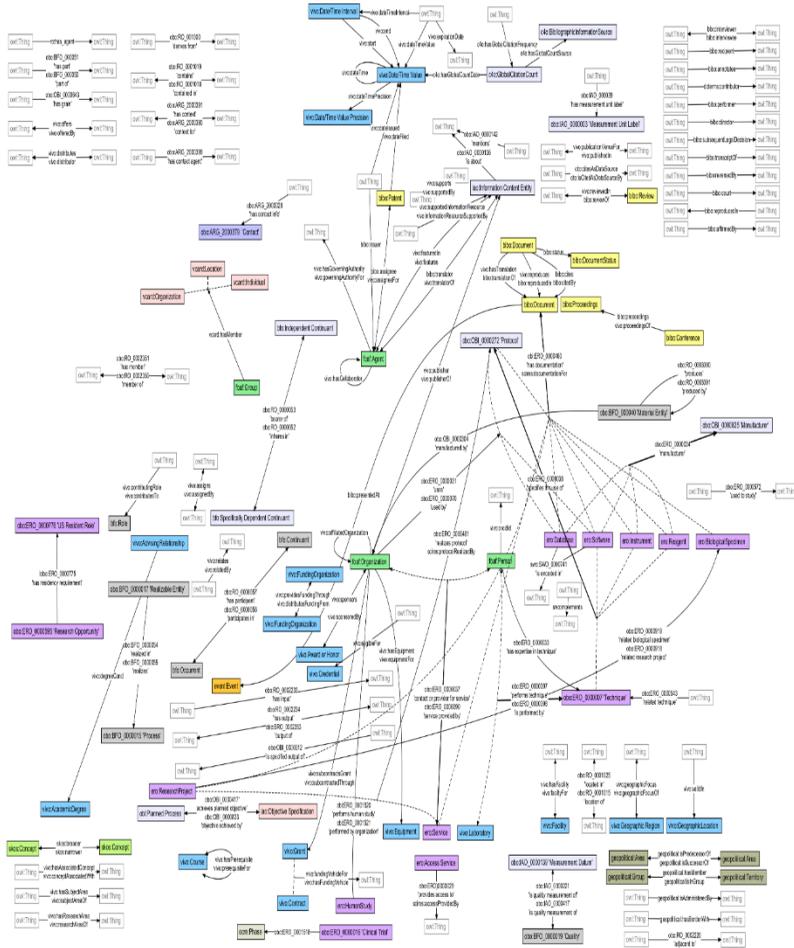
Método complexo	Subtarefas (passos de inferência)	Especificação do Método primitivo
Perfil dos Componentes da rede	1- Recuperar atores e identificação de componente	Recuperar Instância Valor (Especificado anteriormente)
	2- Selecionar atores e identificação de componente	Selecionar Instancia Valor (Especificado anteriormente)
	3- Recuperar perfil	Recuperar Instância da Reificação recuperarInstanciaReificacao4 (Especificado anteriormente)
	4 – Mostrar resultados	Mostrar resultados mostrarResultados (lista de instâncias) Descrição: Esse método exhibe as propriedades termo e frequência das instâncias da lista de instâncias.

Fonte: Autor.

2. Diagrama de Propriedades de Objetos da Ontologia VIVO³⁹

Updated 3/16/14

VIVO-ISF Ontology Object Properties | version 1.6

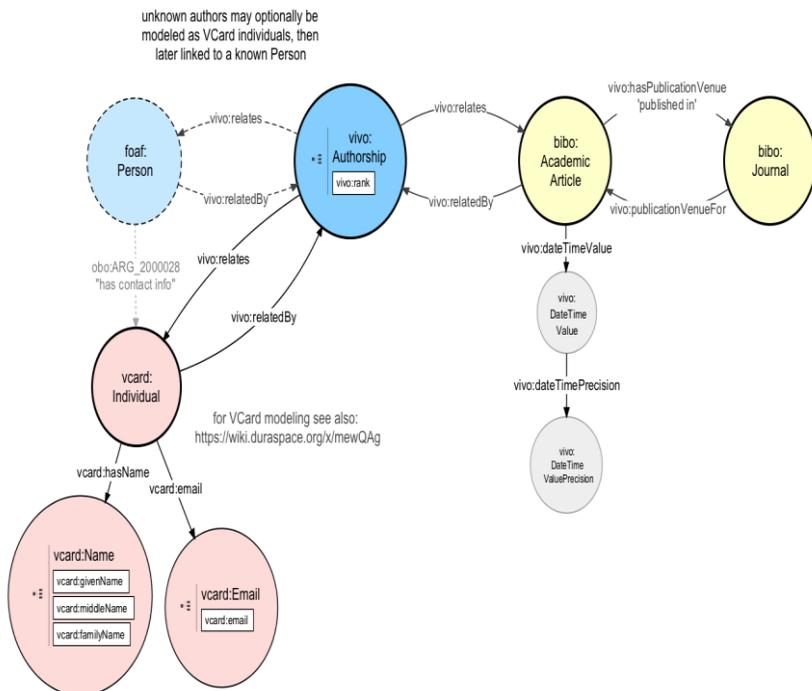


³⁹ <https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology+Overview+%3A+Object+Properties>

3. Diagrama do Relacionamento de Autor (*Person*) e Publicação (*Academic Article*) através da classe Coautoria (*Authorship*)

8/18/14

VIVO Authorship: Connecting
an Author with a Publication



ANEXO B – Ontologia SemSNA

