



**Pós-Graduação em Ciência da Computação**

**Relacionamentos Ocultos de Confiança em Ambientes Virtuais Sociais como Apoio aos Sistemas de Recomendação de Especialistas**

***Edeilson Milhomem da Silva***

**TESE DE DOUTORADO**



Universidade Federal de Pernambuco  
posgraduacao@cin.ufpe.br  
www.cin.ufpe.br/~posgraduacao

RECIFE, NOVEMBRO/2013



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE INFORMÁTICA  
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

EDEILSON MILHOMEM DA SILVA

RELACIONAMENTOS OCULTOS DE CONFIANÇA EM  
AMBIENTES VIRTUAIS SOCIAIS COMO APOIO AOS SISTEMAS  
DE RECOMENDAÇÃO DE ESPECIALISTAS

TRABALHO APRESENTADO AO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIA DA COMPUTAÇÃO.

ORIENTADOR(A): Prof. Silvio Romero de Lemos Meira  
CO-ORIENTADOR (A): Prof. Ana Carolina Salgado

RECIFE, NOVEMBRO/2013

Catálogo na fonte  
Bibliotecário Jefferson Luiz Alves Nazareno, CRB 4-1758

Silva, Edeilson Milhomem da.

Relacionamentos ocultos de confiança em ambientes virtuais sociais como apoio aos sistemas de recomendação de especialistas./ Edeilson Milhomem da Silva. – Recife: O Autor, 2013.

XIV, 163f.: fig.

Orientador: Silvio Romero de Lemos Meira.

Tese (Doutorado) - Universidade Federal de Pernambuco. Cin. Ciência da Computação , 2013.

Inclui referências bibliográficas.

1. Engenharia de software. 2. Inteligência artificial.  
I. Meira, Silvio Romero de Lemos. (orientador). II. Título.

005.1

(22. ed.)

MEI 2014-1

Tese de Doutorado apresentada por Edeilson Milhomem da Silva à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título "**Relacionamentos Ocultos de Confiança em Ambientes Virtuais Sociais como Apoio aos Sistemas de Recomendação de Especialistas**" orientada pelo **Prof. Silvio Romero de Lemos Meira** e aprovada pela Banca Examinadora formada pelos professores:

\_\_\_\_\_  
Prof. Patricia Cabral de Azevedo Restelli Tedesco  
Centro de Informática / UFPE

\_\_\_\_\_  
Prof. Ricardo Bastos Cavalcante Prudêncio  
Centro de Informática / UFPE

\_\_\_\_\_  
Prof. Sérgio Castelo Branco Soares  
Centro de Informática / UFPE

\_\_\_\_\_  
Prof. Marcos Roberto da Silva Borges  
Núcleo de Computação Eletrônica / UFRJ

\_\_\_\_\_  
Prof. Flávia Maria Santoro  
Departamento de Informática Aplicada / UFRJ

Visto e permitida a impressão,  
Recife, 8 de novembro de 2013.

\_\_\_\_\_  
**Profa. Edna Natividade da Silva Barros**  
Coordenadora da Pós-Graduação em Ciência da Computação do  
Centro de Informática da Universidade Federal de Pernambuco.

*Dedico este trabalho à minha mãe querida pelo amor, paciência e dedicação constantes.*

## Agradecimentos

Inexplicavelmente, pelo menos para mim, muitas pessoas vêm a este imenso mundo e passam por grandes sofrimentos: ora por alguma limitação física ou mental, ora por doenças crônicas que adquiriram ao longo da vida ou simplesmente já nasceram com elas, ora por situações adversas em que estes vêm ao mundo (e.g. miséria ou fome), e por aí vai. Infelizmente existem muitas condições irreversíveis, já outras apenas impõem barreiras que fazem com que as lutas cotidianas sejam mais desafiadoras. E estas condições desafiadoras dependem exclusivamente das escolhas que são feitas. As condições sociais e genéticas que me foram impostas fizeram com que eu enxergasse a vida com uma perspectiva de crescimento, como um grande desafio: o moleque “diferente” sem “Pedigree” (né, Irenides? rs) que nascera no interior do interior, que desafiava as altas temperaturas da região correndo de cueca no meio da rua (né, Parcilene? rs).

Por isso, ó Deus, agradeço pela direção que tem me dado e que fizera com que eu chegasse até este importante momento, que se resume ao amor, ao carinho e à dedicação de pessoas especiais (né, mãe?); à mão e ao ombro amigo de pessoas chave em momentos cruciais (né, Lidiane e Parcilene?); à sabedoria para vencer os conflitos da mente que me foram e são impostos constantemente. Enfim, à direção que sempre faz com que eu encontre uma luz. Obrigado, ó Pai, por me guiar e pela certeza que tenho que continuarás me guiando.

Há muito que agradecer. Há muitas pessoas que em momentos oportunos contribuíram para que esta longa jornada continuasse sendo trilhada. A primeira, e a mais especial delas, é minha mãe querida, que em meio a uma turbulenta vida, sempre se manteve de pé, com a esperança e o sonho de tornar seus filhos pessoas melhores e com uma formação. Logo em seguida, minha vó querida, que sempre atuou como uma força propulsora que me movia, e ainda move. Agradeço também às minhas queridas irmãs, Lidiane e Suélen, e às sobrinhas Hanna Eduarda e Sofia, que sempre nos trazem um sorriso em momentos de agonia, pressão ou angústia.

Agradeço a tantos amigos que tive o privilégio de tê-los, que tantas vezes estiveram aqui ou ali para me estenderem uma mão, um ombro, ou uma companhia tanto em momentos alegres como também os tristes. Eu tive e tenho o privilégio de ter muitos amigos assim e é difícil enumerá-los, porque, com certeza, serei injusto deixando um ou

outro de fora. Mas vamos lá: Alison, André, Carlos, Cristina, Diêmy, Dayane, Douglas (pegador), Eloi, Fábio, Fabiano, Fernando, Jorge, Jackson, Kelienny, Leandro, Lucas, Madianita, Múcio, Parcilene, Poliana, Rafael, Ricardo, Roberta e Ronaldo. Embora cada um tenha um importante papel em cada momento, sempre há aqueles que são mais especiais.

Dentre estes, a Parcilene; uma pessoa com uma capacidade intelectual incomensurável, mas que, com certeza, também sempre será lembrada por causa da “imensidão” do seu coração (que romântico rs). A Parcilene ensinou-me muito, acreditou em mim, insistiu, orientou, guiou, cuidou e abriu meus olhos quando eles insistiam em se estreitar (Jack ?). É muito difícil pontuá-la com simples palavras. Valeu mesmo, Parcilene! As amizades que nasceram na graduação e perduram até hoje: André, Carlos, Leandro, Lucas e Jorge. E ao Leandro, em especial, pelo “terrível” trabalho da entrega final desta tese na biblioteca da Universidade. O Jackson, grande pessoa, com uma humildade proporcional à sua capacidade intelectual, que não é pouca, nem de longe. Obrigado Jack! Você foi muito importante em todo este processo e acompanhou sempre de perto o desenvolvimento do trabalho, onde trocamos ideias constantemente, discutimos, aprendemos um com o outro. A Keli, cuja admiração por mim sempre atuou como uma força motriz que me impulsionou. O seu carinho e a sua simplicidade sempre foram muito importantes para mim. Espero que esta simplicidade a acompanhe sempre.

Os amigos e companheiros de trabalho: Cristina, Diêmy, Douglas, Fabiano, Fernando, Irenides, Jackson, Madianita, Parcilene e Régia, por tantas vezes terem tido paciência para aguentar os estresses do dia-a-dia, aliados à pressão do desenvolvimento desta tese. Aproveito a oportunidade para pedir desculpas pelos estresses e agradecer pela paciência. E também agradecer à Mádia pelo trabalho de revisão, uma tarefa difícil aliada a seus tantos outros compromissos diários.

Agradeço ao Silvío por sempre ter me apoiado e dado o voto de confiança que precisava. E à Carol pelo companheirismo, dedicação e por ter compartilhado um pouco da sua experiência comigo. Aprendi muito com as vivências com a Carol. Obrigado pela oportunidade e privilégio!

Agradeço ao grupo de pesquisa do CEULP de Engenharia Inteligente de Dados, que teve forte contribuição para o desenvolvimento deste trabalho. Os membros atuais e os que já fizeram parte: Jackson, Parcilene, Diego, Charles e Flávio. Destes, vai um agradecimento especial ao Diego: dedicado, comprometido e responsável. Sei que você

tem uma longa jornada pela frente, que Deus o abençoe e que você sempre possa exercer a simplicidade e a humildade em sua vida: alguns dos principais fatores que nos tornam vencedores, ou pelo menos pessoas melhores, que no final é o que importa.

Enfim, a vida é repleta de batalhas e conquistas, e esta é mais uma importante conquista, mas com certeza, não mais importante do que as verdadeiras amizades. E esta certeza eu tenho: tive o privilégio de conquistar muitas com as quais poderei contar sempre! Por tudo isso agradeço a Ti, Pai querido, ó Deus Criador dos céus e da terra!



## Resumo

O volume de informação disponível na web aumenta a cada dia, muitas vezes por estímulo das plataformas sociais que vêm se destacando neste cenário, já que provêem facilidades para que as pessoas possam interagir e trocar experiências. A gestão eficiente deste conhecimento, embora não seja uma tarefa trivial, pode trazer diversos benefícios como, por exemplo, a recomendação personalizada para as pessoas, de acordo com as suas preferências. Neste sentido, a presente tese tem como propósito apresentar uma abordagem, intitulada SWEETS, que identifica automaticamente quais as áreas e nível (grau) de conhecimento das pessoas, isto é, os especialistas em determinadas áreas de conhecimento. Para isso, são usados os conhecimentos produzidos por estas pessoas e disponibilizados em diferentes plataformas virtuais. A identificação dos níveis de conhecimento destas pessoas em áreas específicas pode não ser suficiente, pois é interessante que sejam identificados os especialistas mais adequados para uma determinada pessoa, isto é, que tenham um relacionamento social mais próximo, aumentando assim a probabilidade de colaboração. As informações contextuais dos relacionamentos (e.g. distância social) entre pessoas oferecem *background* para a descoberta de quão estas pessoas confiam umas nas outras.

Assim, a presente tese defende que estes relacionamentos de confiança são fundamentais na decisão de colaboração entre indivíduos. Por isso, é apresentada uma abordagem, intitulada T-SWEETS, que se baseia nas informações contextuais dos relacionamentos entre indivíduos para inferir o grau de confiança entre eles, e a sua implantação em um cenário que comprovou a tendência de colaboração entre os indivíduos que possuem relações confiáveis. T-SWEETS baseia-se em 4 elementos: Similaridade entre Perfis, Relacionamento de Confiança, Nível de Maturidade e Reputação, que são oriundos do resultado de um experimento realizado com um grupo de pessoas. Outra constatação desta tese é que, embora haja um grande volume de conhecimento disponibilizado pelas pessoas nas plataformas virtuais, muitas vezes este conhecimento pode não ser suficiente para identificar e recomendar a pessoa especialista em assuntos específicos. Por isso, a descoberta (recomendação) de relacionamentos de confiança entre os indivíduos pode ser um elemento que encoraje as pessoas a trocarem experiências ou interajam, uma vez que as pessoas tendem a colaborar com pessoas que mais confiam, fornecendo assim, insumo para a inferência das especialidades dos indivíduos.

**Palavras Chave:** Sistemas de Recomendação de Especialistas, confiança, Contexto Computacional, colaboração.

## Abstract

The volume of available information in internet increases every day, mostly in consequence of the motivation from social platforms, which are gaining prominence in this setting, because they provide easiness of interaction and experience exchange among persons. The efficient management of this knowledge, although is not a trivial task, can provide various benefits, such as the customized recommendation according to individuals' preferences. In this sense, this thesis aims to present an approach, titled T-SWEETS, which automatically identifies persons' knowledge areas and degrees, which is, the experts in specific knowledge areas. To achieve this, we use the knowledge persons produce and make available on different virtual platforms. The identification of persons' knowledge levels in specific areas may not be sufficient, because it is interesting to identify the most suitable experts for a given person, which is, those with a closer relationship, increasing the probability of collaboration. Relationships contextual information (i.e. the social distance) between persons offer a background for discovery of which of these person trust each other.

Therefore, this thesis states that these trust relationships are fundamental on the decision for collaboration between individuals. For this reason, we present an approach, titled T-SWEETS, which uses contextual information from relationships between individuals in order to infer their trust degree, in addition we present its deployment in a scenario that proves the trend for collaboration between individuals who pursue trustful relationships. T-SWEETS is based on four elements: User Profile Similarity; Trust Relationship; Maturity Level; and Reputation. Another conclusion this thesis presents is that, although there is a great volume of knowledge made available by users of virtual platforms, most of the time this knowledge may not be enough to identify and recommend the expert on specific subjects. For this reason, the discovery (recommendation) of trust relationships between individuals can be an element that encourages them to exchange experiences or to interact, because they tend to collaborate with those who they trust most, providing, finally, inputs for individuals' expertise inference.

**Keywords:** Expert recommender systems, trust, context, collaboration.

# Lista de Figuras

Figura 3.1: Conceito de transitividade.....	30
Figura 3.2: Conceito de agregabilidade.....	31
Figura 3.3: Conceito de assimetria .....	32
Figura 3.4: Grafo que representa as ligações de confiança de um nó $a$ em relação a $c$ (métricas $M1$ e $M2$ ). .....	35
Figura 3.5: Grafo que representa as ligações de confiança de um nó $u$ em relação a $v$ (métricas $M3$ e $M4$ ). .....	36
Figura 4.1: Estrutura do ambiente para implantação do algoritmo .....	50
Figura 4.2: Processo de análise da semelhança entre opiniões e conhecimentos dos usuários.....	52
Figura 4.3: Cenário para exemplificar o cálculo de semelhança dos perfis de opinião de dois usuários. ....	53
Figura 4.4: Grafo de Confiança.....	57
Figura 4.5: Grafo de relacionamentos de confiança: (A) grafo completo e (B) grafo com os menores caminhos. ....	58
Figura 4.6: Análise do nível de maturidade do usuário $a$ .....	61
Figura 4.7: Cenário para exemplificar o cálculo do Nível de Maturidade.....	62
Figura 4.8: Grafo de confiança para cálculo de reputação do usuário $A$ .....	64
Figura 5.1: Arquitetura do SWEETS .....	72
Figura 5.2: Indexação da base do conhecimento .....	73
Figura 5.3: Parte de uma ontologia $O_{ci}$ , hipotética.....	75
Figura 5.4: Exemplo de recomendação de especialistas em uma área $h$ para um usuário $U_{source}$ .....	90
Figura 5.5: Análise de Opiniões similares entre pessoas (A).....	91
Figura 5.6: Cenário para análise do nível de maturidade de um usuário. ....	91
Figura 5.7: (A) Grafo de Confiança entre os membros de uma rede social. ....	92
Figura 6.1: Arquitetura do Konnen .....	95
Figura 6.2: Tela principal pós-login do Konnen .....	96
Figura 6.3: Relações de confiança como apoio ao Modelo 3C .....	99
Figura 6.4: Widget para justificação das recomendações.....	101
Figura 6.5: Arquitetura do SWEETS implantado na plataforma Konnen.....	103
Figura 6.6: Integração dos perfis externos do LinkedIn e Lattes ao do Konnen .....	104
Figura 6.7: Procolo OAuth para conexão com a LinkedIn.....	105
Figura 6.8: Procolo de comunicação com a plataforma Lattes .....	106
Figura 7.1: Grafo dos Relacionamentos do Conecta durante as 16 Semanas do Experimento .....	124
Figura 7.2: Grafo dos Relacionamentos do Conecta com os nós identificados durante as 16 Semanas do Experimento .....	125

<i>Figura 7.3: Grafo organizado em camadas que apresenta a distância social entre os nós durante o experimento. ....</i>	<i>126</i>
<i>Figura 7.4: Usuários do experimento agrupados em grupos (turmas) que pertencem e seus relacionamentos.....</i>	<i>127</i>
<i>Figura 7.5: Estrutura de uma tríade .....</i>	<i>129</i>
<i>Figura 7.6: Possibilidades de tríades que podem ser usadas nas análises .....</i>	<i>131</i>
<i>Figura 7.7: Grafo de relacionamentos utilizado para analisar os algoritmos (Taherian, 2008).....</i>	<i>133</i>
<i>Figura 7.8: Dispersão dos valores obtidos para o grafo da Figura 7.7 com os algoritmos TidalTrust e T-SWEETS.....</i>	<i>135</i>
<i>Figura 7.9: Lightweight Ontology <math>O_{ch}</math> .....</i>	<i>139</i>
<i>Figura 7.10: Qualidade das recomendações na categoria Habilidade Calculada .....</i>	<i>142</i>

# Lista de Quadros

Quadro 2.1: Visão geral sobre características peculiares aos SREs prestados .....	22
Quadro 3.1: Principais características dos algoritmos de análise de confiança .....	45
Quadro 4.1: Cenário para exemplificar o cálculo de similaridade entre os perfis de conhecimento de dois usuários $U_1$ e $U_2$ . .....	55
Quadro 4.2: Demonstração da problemática do algoritmo TidalTrust em relação ao conceito de Transitividade de Confiança. ....	59
Quadro 4.3: Graus de confiança dos relacionamentos de um grupo de usuários $U$ em relação ao usuário $U_1$ . ....	65
Quadro 4.4: Resumo dos algoritmos que influenciaram na definição do T-SWEETS.....	67
Quadro 5.1: Cenário para exemplificação da Habilidade Demonstrada .....	79
Quadro 5.2: Visão sumarizada das habilidades de João e Pedro no assunto Ontologia .....	87
Quadro 6.1: Visão lógica do perfil indexado de um usuário.....	108
Quadro 6.2: Pesos para cada habilidade adotados no SWEETS .....	110
Quadro 6.3: Categorias para organização das publicações científicas na Plataforma Lattes e respectivos pesos adotados no SWEETS. ....	110
Quadro 7.1: Resultados das postagens da primeira fase do experimento .....	121
Quadro 7.2: Classificações e quantidades de tríades nas fontes de dados do Epinions, Slashdot e Wikipédia (Leskovec, 2010) .....	130
Quadro 7.3: Probabilidades de inferência de relacionamentos positivos e negativos para cada tríade. (Leskovec, 2010, adaptada).....	131
Quadro 7.4: Porcentagem da quantidade de relacionamentos positivos em cada tríade .....	131
Quadro 7.5: Resultados após a execução do algoritmo T-SWEETS aplicado no cenário da Figura 7.7. ....	133
Quadro 7.6: Resultados após a execução do algoritmo TidalTrust aplicado no cenário da Figura 7.7. (Taherian, 2008) .....	134
Quadro 7.7: Média e desvio padrão dos valores dos Quadros 7.5 e 7.6. ....	134

# Sumário

<b>1</b>	<b><i>Introdução</i></b> .....	<b>1</b>
<b>1.1</b>	<b>Motivação</b> .....	<b>2</b>
<b>1.2</b>	<b>Objetivos</b> .....	<b>4</b>
<b>1.3</b>	<b>Organização da Tese</b> .....	<b>5</b>
<b>2</b>	<b><i>Sistemas de Recomendação de Especialistas com Contexto</i></b> .....	<b>7</b>
<b>2.1</b>	<b>Definição de Sistemas de Recomendação e Contexto Computacional</b> .....	<b>7</b>
<b>2.2</b>	<b>ReferralWeb</b> .....	<b>10</b>
2.2.1	Inferência de especialidades .....	11
2.2.2	Contexto no <i>ReferralWeb</i> .....	11
<b>2.3</b>	<b>ICARE: Intelligent Context Awareness for Recommending Experts</b> .....	<b>11</b>
2.3.1	Inferência de especialidades .....	13
2.3.2	Contexto no <i>ICARE</i> .....	14
<b>2.4</b>	<b>SmallBlue</b> .....	<b>15</b>
2.4.1	Inferência de especialidades .....	16
2.4.2	Contexto no <i>SmallBlue</i> .....	17
<b>2.5</b>	<b>Conscious</b> .....	<b>18</b>
2.5.1	Inferência de especialidades .....	19
2.5.2	Contexto no <i>Conscious</i> .....	20
2.5.3	Trabalhos Similares ao <i>Conscious</i> .....	21
<b>2.6</b>	<b>Estudo Comparativo dos SREs Discutidos</b> .....	<b>21</b>
<b>2.7</b>	<b>Considerações Finais</b> .....	<b>24</b>
<b>3</b>	<b><i>Confiança em Redes Sociais</i></b> .....	<b>26</b>
<b>3.1</b>	<b>Definição</b> .....	<b>26</b>
<b>3.2</b>	<b>Propriedades de Confiança</b> .....	<b>29</b>
<b>3.3</b>	<b>FlowTrust</b> .....	<b>33</b>
3.3.1	Inferência de confiança .....	34
3.3.2	Algoritmos semelhantes .....	37
<b>3.4</b>	<b>SUNNY</b> .....	<b>38</b>

3.4.1	Inferência de confiança.....	40
3.4.2	Algoritmos semelhantes .....	41
<b>3.5</b>	<b>Algoritmo para cálculo de reputação em redes sociais .....</b>	<b>42</b>
3.5.1	Inferência de confiança.....	42
<b>3.6</b>	<b>Considerações Finais .....</b>	<b>44</b>
<b>4</b>	<b><i>T-SWEETS</i>.....</b>	<b>47</b>
<b>4.1</b>	<b>Introdução .....</b>	<b>47</b>
<b>4.2</b>	<b>Pré-requisitos para utilização do T-SWEETS .....</b>	<b>49</b>
<b>4.3</b>	<b>Elementos para Inferência de Confiança.....</b>	<b>50</b>
4.3.1	Similaridade entre Perfis .....	51
4.3.2	Relacionamento de Confiança .....	55
4.3.3	Nível de Maturidade .....	60
4.3.4	Reputação .....	63
<b>4.4</b>	<b>Inferência de Confiança .....</b>	<b>65</b>
<b>4.5</b>	<b>Considerações Finais .....</b>	<b>66</b>
<b>5</b>	<b><i>SWEETS</i> .....</b>	<b>70</b>
<b>5.1</b>	<b>Introdução .....</b>	<b>70</b>
<b>5.2</b>	<b>Arquitetura do SWEETS .....</b>	<b>72</b>
<b>5.3</b>	<b>Módulo Fonte de Conhecimento.....</b>	<b>72</b>
<b>5.4</b>	<b>Módulo Indexação de Conhecimento.....</b>	<b>73</b>
<b>5.5</b>	<b>Módulo Geração de Ontologia.....</b>	<b>74</b>
<b>5.6</b>	<b>Módulo Identificação de Especialidades.....</b>	<b>75</b>
5.6.1	Processo de identificação de especialidades .....	76
<b>5.7</b>	<b>Módulo Recomendação de Especialistas.....</b>	<b>88</b>
<b>5.8</b>	<b>Contexto no SWEETS .....</b>	<b>89</b>
<b>5.9</b>	<b>Considerações Finais .....</b>	<b>92</b>
<b>6</b>	<b><i>Estudo de Caso Konnen</i> .....</b>	<b>94</b>
<b>6.1</b>	<b>Estudo de caso Konnen.....</b>	<b>94</b>
<b>6.2</b>	<b>Implantação do T-SWEETS no Konnen .....</b>	<b>98</b>

6.2.1	Elementos do T-SWEETS no <i>Konnen</i> .....	99
<b>6.3</b>	<b>Implantação do SWEETS no <i>Konnen</i> .....</b>	<b>102</b>
6.3.1	Módulo Fonte de Conhecimento .....	103
6.3.2	Módulo Indexação do Conhecimento no <i>Konnen</i> .....	107
6.3.3	Módulo Geração de Ontologias .....	108
6.3.4	Módulo de Identificação e Recomendação de Especialidades.....	109
<b>6.4</b>	<b>Considerações Finais .....</b>	<b>112</b>
<b>7</b>	<b>Resultados .....</b>	<b>114</b>
<b>7.1</b>	<b>Experimento Preliminar.....</b>	<b>114</b>
<b>7.2</b>	<b>T-SWEETS .....</b>	<b>119</b>
7.2.1	Objetivos do Experimento 1 .....	119
7.2.2	Cenário do Experimento 1 .....	119
7.2.3	Resultados da Primeira Fase do Experimento 1 .....	120
7.2.4	Resultados da Segunda Fase do Experimento 1 .....	123
7.2.5	Objetivos do Experimento 2 .....	128
7.2.6	Cenário do Experimento 2 .....	129
7.2.7	Resultados da Primeira Fase do Experimento 2 .....	130
<b>7.3</b>	<b>SWEETS .....</b>	<b>135</b>
7.3.1	Objetivos do Experimento .....	135
7.3.2	Cenário do Experimento .....	136
7.3.3	Resultados do Experimento – parte 1 .....	137
7.3.4	Análise da qualidade das recomendações .....	141
7.3.5	Objetivos do Experimento – parte 2 .....	142
7.3.6	Cenário do Experimento – parte 2 .....	143
7.3.7	Resultados do Experimento – parte 2 .....	143
<b>7.4</b>	<b>Considerações Finais .....</b>	<b>144</b>
<b>8</b>	<b>Conclusão.....</b>	<b>147</b>
<b>8.1</b>	<b>Contribuições .....</b>	<b>149</b>
<b>8.2</b>	<b>Limitações.....</b>	<b>151</b>
<b>8.3</b>	<b>Trabalhos Futuros .....</b>	<b>152</b>
	<i>Referências Bibliográficas .....</i>	<i>154</i>



## 1 Introdução

A popularização da Web tem contribuído para o aumento no volume de conhecimento produzido pelas pessoas. A direção social que a Web vem ganhando é um dos fatores que pode justificar o excessivo volume de informações disponibilizadas em ambientes virtuais Web. Esta Web social baseia-se na estrutura básica da sociedade, que são as redes sociais (Wellman, 1996; Barnes, 1972), para mapear estas relações sociais em ambientes virtuais (Musiał, 2013). Assim, estes ambientes proporcionam uma facilidade para que as pessoas produzam naturalmente os mais variados conhecimentos.

As plataformas baseadas em redes sociais estabelecem um processo social em que as pessoas participam na criação e compartilhamento de conhecimento e propiciam um ambiente em que estas mesmas pessoas possam aprender umas com as outras. Assim, a possibilidade real das pessoas interagirem umas com as outras atua como um estímulo à colaboração, cooperação e, conseqüentemente, à produção de conhecimento, isto é, há naturalmente a proliferação do conhecimento. De acordo com Lin (2009), em ambientes desta natureza, questionamentos como “quem sabe o que?”, “quem conhece quem?” e “quem sabe o que sobre quem?” podem ser respondidos sem a intervenção explícita das pessoas.

Com o aumento de informações disponibilizadas nestas plataformas virtuais, é imprescindível a existência de mecanismos automatizados que auxiliem na descoberta/classificação de conhecimentos que são especificamente relevantes para um determinado individuo ou situação. Esta classificação inteligente da informação oferece sustentação aos sistemas automatizados de recomendação (Earl, 2001; Fazel-Zarandi, 2001, 2012; Melville, 2010; Silva, 2011). Para a geração de recomendações direcionadas para pessoas específicas (recomendações personalizadas), são considerados os perfis destas pessoas, ou seja, as recomendações de forma não *ad-hoc*. Já as recomendações mais apropriadas para determinadas

situações, isto é, contextos específicos são considerados elementos para personalizar as recomendações de forma mais apropriada.

## 1.1 Motivação

Os sistemas de recomendação são alternativas que estabelecem filtros de informações relevantes às pessoas, isto é, a partir da recomendação de itens (e.g. livros, filmes, artigos, fóruns, entre outros) a estas pessoas. Existem sistemas de recomendação que são projetados para auxiliar a resolução de diferentes problemas (Liu, 2013; Bozzon, 2013, Gollapalli, 2012; Datta, 2011, 2012), e podem, por exemplo, recomendar tópicos de interesse em um fórum de discussão ou até mesmo a própria pessoa criadora do tópico. Estes sistemas que recomendam pessoas especializadas em assuntos específicos são classificados como Sistemas de Recomendação de Especialistas (SRE).

Um SRE pode proporcionar várias vantagens, tais como: (1) a resolução de um problema específico; (2) o auxílio na criação de uma equipe especializada para o desenvolvimento de uma determinada atividade; (3) a realocação de pessoas com habilidades específicas em uma organização; (4) o acesso ao conhecimento dos indivíduos que não estão explicitados, isto é, o conhecimento implícito/tácito dos especialistas; entre outras.

Para que um SRE possa atuar com eficiência, é premissa necessária que haja informações suficientes para que as especialidades dos indivíduos possam ser detectadas. As redes sociais proporcionam um ambiente informal e livre, em que as pessoas colaboram, de forma *ad hoc*, através de interações sem nenhum planejamento – o que, segundo Gutwin (2005), é um facilitador na troca de experiências entre as pessoas - , são ambientes propícios à produção de conhecimento.

Mesmo com os benefícios proporcionados pela facilidade de interação das plataformas sociais, às vezes, o conhecimento produzido pelas pessoas pode não ser suficiente para a inferência das especialidades dos indivíduos. Assim, o ideal é reutilizar o conhecimento produzido pelas pessoas em diferentes ambientes, por exemplo, *Twitter*<sup>1</sup>, *Facebook*<sup>2</sup> ou *Bibso-*

---

<sup>1</sup> <https://twitter.com/>

<sup>2</sup> <https://www.facebook.com/>

*nomy*<sup>3</sup>. Além disso, podem ser usadas também informações públicas de currículos, por exemplo, Página Pessoal, plataforma *Lattes*<sup>4</sup> ou Mendley<sup>5</sup>.

Além disso, ainda podem também ser adotadas estratégias de estímulo à colaboração entre indivíduos em rede social. Assim, uma vez que há uma maior colaboração entre os usuários, há, naturalmente, uma maior produção de conhecimento, que contribui diretamente para uma mais eficaz identificação das capacidades intelectuais delas. Uma estratégia interessante que pode ser adotada é a inferência de relacionamentos ocultos de confiança nestas redes sociais.

Confiança é um fenômeno comum que está presente na sociedade e representa o quanto uma pessoa é confiável para outra (Yamamoto, 1990). Já as redes sociais virtuais são ambientes propícios à descoberta de relacionamentos ocultos de confiança porque elas representam a estrutura básica da sociedade e provêm indicadores para a referida análise. Com a criação destes novos relacionamentos, há uma tendência no aumento do estímulo à colaboração entre os membros da rede social, uma vez que novas relações têm sido estabelecidas.

As características peculiares às redes sociais têm potencial para tornar estas plataformas de recomendação de especialistas mais inteligentes, através da recomendação do especialista mais indicado em uma determinada situação. Por exemplo, o especialista que está a uma distância social menor da pessoa que necessita de ajuda, ou o especialista que possua um nível maior de participação no ambiente possui uma tendência maior à colaboração. Neste sentido, Contexto Computacional é outra área que também pode ser adotada a fim de enriquecer ambientes desta natureza.

Para Dey (2001) os seres humanos, por si só, decorrentes da sua natureza evolutiva, são habilitados a tomarem decisões em situações específicas de acordo com o contexto em que estejam inseridos. Esta natureza evolutiva possui relação direta com a capacidade de aprendizagem contínua; com a capacidade de transmitirem conhecimento uns para os outros, decorrente da linguagem comum que compartilham; com o comum entendimento que possuem a respeito do mundo em que vivem; e com o entendimento implícito em situações cotidianas.

---

<sup>3</sup> <http://www.bibsonomy.org/>

<sup>4</sup> <http://lattes.cnpq.br/>

<sup>5</sup> <http://www.mendeley.com/>

De forma muito diferente atuam as máquinas que não têm as capacidades que os seres humanos possuem. Contudo, há tentativas de tornarem estas máquinas minimamente inteligentes, a partir do uso de técnicas específicas de Inteligência Artificial, por exemplo. Como uma destas tentativas para tornar estas máquinas minimamente “inteligentes” há a área de contexto computacional que objetiva buscar alternativas para representar contextos e, assim, naturalmente, tornar as máquinas capazes de tomarem decisões em situações específicas. Isto é, fazer com que as aplicações façam para os usuários “a coisa certa, no momento certo” (Dey, 2001).

Análise de confiança das pessoas, a partir das relações sociais existentes entre elas, é uma característica essencial que pode ser usada para enriquecer a exploração de contexto computacional. A possibilidade de encontrar relacionamentos ocultos em qualquer ambiente (*e.g.* organizacional) ajuda as pessoas a trabalharem juntas e a compartilharem recursos para alcançarem determinados objetivos. As pessoas tendem a confiar em seus amigos e colegas para buscar informações confiáveis, ajudar a filtrar e interpretar informações, além de buscar referências sobre outras pessoas – e propiciar a criação de novos relacionamentos.

Enfim, um Sistema de Recomendação de Especialistas enriquecido com contexto oferece maior potencialidade na identificação do especialista mais adequado em um determinado instante de tempo. Em muitos casos, identificar a pessoa que detém o conhecimento não é o suficiente, especialmente em casos em que o especialista não tenha disponibilidade, não seja acessível ou não tenha disposição para colaborar com quem necessita de auxílio.

## **1.2 Objetivos**

O presente trabalho tem como propósito usar os conhecimentos produzidos por indivíduos em diferentes plataformas virtuais para propor uma abordagem que identifique as suas especialidades e recomende o especialista mais adequado em um determinado instante de tempo. Como este conhecimento, muitas vezes, pode ser insuficiente, é objetivo também deste trabalho a proposta de uma nova abordagem para identificação de relacionamentos ocultos de confiança entre estes indivíduos a fim de estimular a colaboração entre eles. A identificação destes relacionamentos de confiança é fonte de informações úteis para prover o especialista mais adequado a uma pessoa em um determinado instante de tempo. Alguns objetivos específicos são:

- usar as potencialidades oferecidas pelas redes sociais para melhorar o processo de identificação e recomendação de especialistas aplicados nestes cenários sociais
- utilizar diferentes fontes externas de conhecimento dos indivíduos para tornar mais eficaz a identificação das especialidades destes indivíduos;
- uma nova abordagem para a identificação das especialidades dos indivíduos que pode ser implantada em ambientes virtuais sociais;
- a descoberta de relacionamentos de confiança entre indivíduos em redes sociais como alternativa para estimular a colaboração entre eles e, assim, naturalmente, contribuir para uma maior produção de conhecimento;
- realizar também um experimento com um grupo de pessoas para entender quais elementos podem impactar quando uma pessoa decide colaborar com outra ou não. Esta pesquisa é também essencial para a proposta da nova abordagem para inferência de relacionamentos ocultos de confiança entre indivíduos em um ambiente virtual social;
- implantar as abordagens propostas para identificação e recomendação de relacionamentos ocultos de confiança e especialistas em uma plataforma virtual social;
- realizar experimentos para medir a eficácia ou o impacto das abordagens propostas.

### 1.3 Organização da Tese

O presente trabalho está organizado da seguinte forma:

**Capítulo 2:** apresenta a definição e características inerentes a um Sistema de Recomendação de Especialistas (SREs) e a Contexto Computacional. Além disso, mostra também alguns dos Sistemas de Recomendação de Especialistas disponíveis na literatura, e como cada um destes sistemas explora, mesmo que não explicitamente, os conceitos de contexto computacional;

**Capítulo 3:** apresenta conceitos voltados à área de análise de confiança em ambientes computacionais baseados em redes sociais, já que este é um dos elementos que pode ser explorado em ambientes que utilizam contexto computacional. Além disso, ele apresenta também como os trabalhos disponíveis na literatura adotam as propriedades de confiança do mundo real em ambientes computacionais a fim de inferirem relacionamentos de confiança;

**Capítulo 4:** apresenta a nova abordagem para inferência de relacionamentos ocultos de confiança em ambientes virtuais sociais, intitulada T-SWEETS. São apresentados no

capítulo a motivação para os elementos adotados em T-SWEETS, os pré-requisitos para adoção em diferentes plataforma sociais e o seu formalismo;

**Capítulo 5:** apresenta a abordagem SWEETS para identificação e recomendação de especialistas. No capítulo são apresentadas a arquitetura e os módulos da abordagem e os formalismos adotados para a inferência das especialidades dos indivíduos, além de como a abordagem proposta adota o uso de contexto computacional para prover o especialista de domínio mais adequado;

**Capítulo 6:** apresenta o estudo de caso *Konnen*, uma plataforma virtual social que contém os pré-requisitos necessários para a implantação das abordagens propostas. São apresentados no capítulo o processo de implantação de T-SWEETS e SWEETS na plataforma;

**Capítulo 7:** apresenta os experimentos desenvolvidos e os resultados obtidos no presente trabalho. Inicialmente foram apresentados os resultados de um experimento preliminar que foi aplicado com um grupo de pessoas para entender quais elementos são impactantes quando decidem colaborar/interagir ou não com outras pessoas; em seguida, são apresentados os resultados obtidos a partir dos experimentos realizados com a implantação do T-SWEETS no *Konnen* e, por fim, são apresentados os resultados da eficiência da abordagem apresentada no SWEETS;

**Capítulo 8:** por fim, este capítulo apresenta as considerações finais desta tese; suas contribuições, limitações e possíveis melhorias a partir de trabalhos futuros.

## **2 Sistemas de Recomendação de Especialistas com Contexto**

Este capítulo tem como propósito apresentar a definição e particularidades inerentes aos Sistemas de Recomendação de Especialistas, bem como estes sistemas adotam o conceito de contexto computacional. É apresentado no presente capítulo também a definição para o termo Contexto. O capítulo está organizado da seguinte forma: seção 2.1 apresenta a definição para sistemas de recomendação, enfatizando os Sistemas de Recomendação de Especialistas (SREs), e Contexto Computacional; as seções 2.2, 2.3, 2.4 e 2.5 apresentam SREs disponíveis na literatura, enfatizando o ambiente em que são utilizados (*e.g.* ambiente organizacional), as fontes de informações e o mecanismo utilizado para a inferência de especialidades, além de como estes SREs tem explorado o contexto computacional; por fim, a Seção 2.6 apresenta as considerações finais do capítulo

### **2.1 Definição de Sistemas de Recomendação e Contexto Computacional**

Sistemas de Recomendação (SR), em geral, podem contribuir significativamente para que os usuários consigam localizar informações que sejam especificamente relevantes para eles (Adomavicius, 2005; Anand, 2011; Nie, 2011; Borgatti, 2003). Runte (2000) apud Reichling (2005) classifica estes sistemas de recomendação em dois gêneros: personalizado; em que são consideradas as preferências ou interesses dos usuários para promover recomendações personalizadas (Adomavicius, 2005; Anand, 2011); não-personalizado, que promovem a recomendação não considerando as particularidades de cada usuário. Os SRs podem atuar de duas maneiras distintas: necessitar de uma requisição explícita do usuário para gerar as recomendações; ou gerar as recomendações automaticamente, de acordo com as preferências ou interesses dos usuários.

Da mesma forma que os SRs aproximam os usuários a elementos (e.g. documentos) que sejam de seus interesses, a partir de informações informadas explicitamente ou não, há os Sistemas de Recomendação de Especialistas (SREs), também conhecidos como sistemas localizadores de pessoas (*People-Finder Systems*) (Becerra-Fernandez, 2006), que tem como objetivo principal aproximar as pessoas a determinados especialistas de domínio – um especialista de domínio é um *expert* em um determinado assunto. Para isso, os SREs podem utilizar o conhecimento explícito das pessoas para identificar suas habilidades e, assim, retornar referência ao próprio especialista, o detentor do conhecimento, isto é, tornam possível a exploração do conhecimento tácito das pessoas – o conhecimento que está na mente destas pessoas.

O conhecimento explicitado das pessoas oferece subsídios para criação/representação dos perfis de conhecimentos delas. Esta identificação automática dos conhecimentos dos especialistas não é uma tarefa trivial e pode envolver diferentes áreas da computação, tais como, Web Semântica, a partir do uso de ontologias (Gruber, 1993; Gruber, 1995; Uschold, 1996); Mineração de Dados e Recuperação da Informação, que se baseia em técnicas matemáticas (e.g. modelo vetorial e modelos probabilísticos) para análises de textos planos (Riahi, 2012; Liu, 2013; Kautz, 1997), e Análise de Redes Sociais, que torna possível a identificação das experiências dos indivíduos a partir dos relacionamentos entre eles. Este conhecimento pode estar explicitado em plataformas virtuais de diferentes naturezas, tais como, informações de artigos editados do Wikipédia (Demartini, 2007); informações acadêmicas disponíveis em sites públicos (Datta, 2011), atividades dos usuários em plataformas públicas (e.g. *Facebook* e *Twitter*) (Bozzon, 2013) e ambientes de comunidades virtuais de pergunta-resposta (Riahi, 2012; Liu, 2013).

Em ambientes organizacionais, nos quais há iniciativas de implantação de SRE, como os casos de McDonald (2000), que implantou um SRE em uma companhia de software médica; e de Reichling (2007), que apresenta a implantação de um SRE na Associação das Indústrias Europeias. Alguns benefícios adicionais podem ser proporcionados em detrimento da localização de um especialista. são eles: (1) aumentar o potencial de percepção das atividades na organização; (2) aumentar a colaboração entre as pessoas; e (3) facilitar e agilizar a resolução de problemas (Ackerman, 2003). Outro benefício importante, de acordo com Cohen (2001; Lin (2008), é a construção do capital social: se a interação acontece entre pessoas conhecidas, os laços entre elas são reforçados; porém, se esta interação ocorre entre pessoas desconhecidas, há a possibilidade que as permitem se conhecerem e construírem laços que as aproximem.



Como alternativa para tornar as recomendações personalizadas e, eventualmente, mais convenientes ou eficientes em um determinado instante de tempo, podem ser adotadas estratégias para personalização. Uma das estratégias que pode ser adotada é a exploração da área de Contexto Computacional.

A definição clássica para Contexto Computacional originou-se na computação ubíqua, por Dey (2001), o qual afirma que contexto é qualquer informação que pode ser usada para caracterizar a situação de uma entidade. Uma entidade pode ser uma pessoa, um lugar ou um objeto que pode ser considerado relevante para a interação entre usuário e aplicação. Com base nesta clássica definição, diferentes pesquisadores vêm enriquecendo-a, ou adaptando-a, de acordo com a sua área de pesquisa, a partir da agregação de diferentes elementos e desenvolvendo modelos, frameworks e aplicações baseadas em Contexto. Alguns destes trabalhos são: (Dey, 2001; Greenberg, 2001; Moran, 2001; Halin, 2003; Borges, 2004; Vieira, 2004; Klyne, 2004; Henricksen, 2004; Brézillon, 2005; Henricksen, 2006; Baldauf, 2007; Brézillon, 2007; Castelli, 2007; Freyne (2007; Zimmermann, 2007; Santos, 2008; Vieira, 2008; Vieira, 2008b; Vieira, 2009; Tchuente, 2010; Bettini, 2010; Bettini, 2010; Bettini, 2010; e Fontes, 2011).

Contexto aplicado na área de Sistemas Colaborativos (*e. g. Sistemas de Recomendação*) visa apoiar as interações sociais entre as pessoas, buscando, assim, alternativas para incentivar a colaboração. De acordo com Moran (2001), informações implícitas podem ser utilizadas pelas aplicações para a interpretação de ações explícitas, assim, pode haver personalização dos conteúdos aos usuários (Tchuente, 2010; Halin, 2003) – propiciando, naturalmente, melhorias na comunicação, colaboração e cooperação entre as pessoas. Contexto nestes ambientes colaborativos pode estar restrito a uma única pessoa, a partir do seu perfil; a um grupo ou comunidade; e a uma determinada tarefa que está sendo executada (Borges, 2004).

Existem na literatura vários Sistemas de Recomendação de Especialistas direcionados para os mais variados cenários. As seções 2.2, 2.3, 2.4 e 2.5 apresentam de forma mais detalhada alguns destes SREs. São 4 os SREs apresentados a posteriori: ReferralWeb (Kautz, 1997), ICARE (Petry, 2008), SmallBlue (Lin, 2009) e Conscius (Moraes, 2010). O critério de seleção que reduziu este universo a 4 foi: ambientes que consideram contexto computacional para recomendar um especialista que seja mais adequado. Há casos que existem SREs similares a um dos 4 previamente selecionados, por isso, neste casos há uma seção específica que os apresentam.

## 2.2 *ReferralWeb*

*ReferralWeb* (Kautz, 1997) é um SRE que explora o conceito das redes sociais para promover e filtragem colaborativa com o intuito de proporcionar aos usuários recomendações mais satisfatórias. Assim, baseado na teoria dos “6 graus de separação”, *ReferralWeb* foi projetado para dar prioridade aos especialistas que estão a uma distância social menor do usuário que receberá a recomendação. Desta forma, os desenvolvedores da ferramenta alegam que a comunicação e a colaboração podem ser facilitadas.

*ReferralWeb*, em sua primeira versão, utilizava arquivos dos *e-mails* para inferir os relacionamentos entre as pessoas e, assim, gerar as redes sociais de cada pessoa. Apesar desta fonte de informação ser riquíssima para o propósito do projeto, infelizmente, enfrentaram problemas no que tange à privacidade – o uso destas informações levantou preocupação entre as pessoas e era muito difícil convencê-los do contrário.

Por causa destas dificuldades, foi desenvolvida uma nova versão do *ReferralWeb*, que faz uso de quaisquer documentos publicados na *Web* que possam ser identificados relacionamentos entre as pessoas, a partir da co-ocorrências dos seus nomes. Estas referidas fontes incluem:

- *links* encontrados em *home pages*;
- lista de co-autores em artigos técnicos e citações de artigo;
- interações entre pessoas registradas na *Web*;
- grupos em organizações (por exemplo, departamentos de uma universidade).

O modelo de rede para cada especialista é construído gradativamente. À medida que o usuário se registra no sistema, e usa uma *engine* de busca qualquer, documentos Web são retornados. Depois, é verificada a existência do seu nome nestes documentos e, também, a ocorrência de nomes de outras pessoas – desta forma a rede social é construída.

A rede construída é usada como um guia para encontrar pessoas ou documento em resposta a uma consulta realizada pelos usuários. Assim, os usuários podem fazer pesquisas mais específicas, mais efetivas, a partir da entrada de um tópico (assunto) e a distância social dos especialistas em relação a eles.

### 2.2.1 Inferência de especialidades

Alguns SREs requerem que os usuários entrem manualmente com seu perfil, manifestando suas especialidades ou interesse por assuntos específicos. Contrário a isso, ReferralWeb utiliza técnicas de *data mining* para gerar o perfil do usuário. Assim, os usuários do ReferralWeb não estão limitados a um conjunto de áreas (tópicos, temas) previamente definidos, pois, utiliza um mecanismo de indexação da Web, AltaVista, para estabelecer o relacionamento entre o usuário e um tema qualquer.

No trabalho não é apresentado como são realizadas as inferências das especialidades dos usuários. Apesar disso, acredita-se que seja utilizada a frequência das palavras-chave, informadas no momento em que se busca um especialista, em relação ao perfil de cada usuário para verificar o quanto o usuário está relacionado a estas palavras-chave.

### 2.2.2 Contexto no ReferralWeb

No SRE ReferralWeb o contexto é explorado apenas a partir dos relacionamentos sociais entre os especialistas. Segundo os autores, a maioria dos SRE até aquela época recomendavam especialistas de forma anônima. A partir da exploração destes relacionamentos sociais, podem ser recomendados especialistas que estejam fortemente relacionadas ao usuário que recebe as recomendações, proporcionando, assim, a consulta a especialistas mais confiáveis.

## 2.3 ICARE: *Intelligent Context Awareness for Recommending Experts*

O SRE ICARE (Petry, 2008) explora explicitamente o conceito de contexto para promover as recomendações de especialistas mais apropriados aos usuários, por exemplo, o especialista que tem mais disponibilidade a auxiliá-lo em uma determinada tarefa. Além desta, o ICARE possui outras características fundamentais, utilizar diferentes fontes de experiência para inferir as especialidades dos usuários e é baseado em uma ontologia de domínio.

Duas principais entidades são consideradas para determinar o contexto: o usuário que realizou a requisição de um especialista e os especialistas recomendados. Assim, a relação entre o usuário e cada especialista recomendado é considerada para estabelecer um *ranking* dos especialistas ideais que foram recomendados em um determinado instante de tempo, isto é, para o mesmo usuário, em diferentes momentos, pode haver recomendações de especialistas diferentes.

Os elementos contextuais utilizados pelo ICARE são:

- **disponibilidade:** identifica qual a disponibilidade do especialista. Espera-se que quanto menos ocupado o especialista esteja, mais disposto estará a colaborar com algum usuário que necessite do auxílio na execução de uma determinada tarefa;
- **acessibilidade:** para identificar o quão acessível um especialista está em um determinado instante de tempo. Por exemplo, João, que trabalha em seu escritório com a porta aberta, provavelmente, é mais acessível que seu colega Peter, que trabalha com a porta fechada. Porém, João pode estar demasiadamente ocupado atendendo diferentes pessoas, isto é, apesar de estar acessível em decorrência da porta do seu escritório estar aberta, não tem disponibilidade;
- **nível organizacional:** faz referência a posição que ocupa na hierarquia da organização. Um colaborador pode se sentir mais a vontade para tirar uma eventual dúvida com um especialista que esteja no mesmo nível organizacional que o dele;
- **distância social:** é o número de pessoas que conhecem umas as outras e que separa socialmente o usuário dos especialistas;
- **tarefa atual:** representa a tarefa atual que está sendo executada;
- **reputação:** é o julgamento que os especialistas (uma visão global) recebem pelos usuários que o contataram. É estimado a partir do *feedback* fornecido pelos usuários após o uso do ICARE;
- **interesses:** informação relacionada ao usuário que ajuda a identificar a experiência desejada;
- **grau de especialidade:** significa o quão o usuário é especialista em uma determinada área.

Os elementos contextuais do ICARE foram definidos a partir de um gerenciador genérico de contexto, *CEManTIKA* (*Contextual Elements Management Through Incremental Knowledge Acquisition*) (Vieira, 2007). Este gerenciador define procedimentos e uma infraestrutura para manipular elementos contextuais, promovendo o gerenciamento do contexto do usuário que é compartilhado com o ICARE.

### 2.3.1 Inferência de especialidades

Para que a inferência de especialidades possa ser realizada, o ICARE possui quatro módulos, são eles: Gerador da Base de Especialistas; Definidor de Experiências; Recuperador de Especialistas; e o Apresentador de Especialistas.

#### Módulo Gerador da Base de Especialistas

Este módulo tem as seguintes atribuições: (1) registrar as fontes de especialistas disponíveis; (2) ser configurado para acessar cada fonte de especialistas; e (3) salvar as informações adquiridas das fontes de especialistas fornecidas na Base de Especialistas do ICARE. As recomendações dos especialistas só serão possíveis após estas bases serem povoadas. Ou seja, o conjunto de especialistas armazenados são finitos e definidos.

Quando uma nova fonte de especialistas é inserida, o Gerador da Base de Especialistas recebe uma descrição da informação informada e como acessá-la (ex.: consulta SQL). Além disso, o Gerenciador da Base de Especialistas realiza o mapeamento das informações recuperadas para um conceito da ontologia de domínio.

Algumas experiências podem ser deduzidas a partir de práticas de atividades organizacionais. O Gerador da Base de Especialistas utiliza um conjunto de regras de produção para associar atividades específicas com experiências relacionadas. Estas regras, também, devem ser informadas no momento em que uma nova base de especialistas é inserida.

#### Módulo Definidor de Experiências

Este módulo recebe uma entrada de um usuário e verifica a sua existência na ontologia de domínio, para poder inferir o quanto uma pessoa é especialista em um assunto.

O módulo considera, também, as experiências identificadas em recomendações anteriores. Por isso, são mantidos os históricos das recomendações, as entradas que geraram as recomendações; o *feedback* do usuário; e o contexto similar para o usuário. A similaridade entre dois contextos é calculada a partir de uma fórmula específica para inferência de similaridade. Esta fórmula soma as diferenças absolutas entre as informações contidas em ambos os contextos. Os elementos contextuais considerados são: acessibilidade, disponibilidade, nível organizacional, e tarefa atual.

Ao final, é retornada uma lista de experiências que é ordenada de acordo com o seu peso. O tamanho desta lista é um parâmetro de configuração do sistema.

## **Módulo Recuperador de Especialistas**

Após a definição das experiências, o passo seguinte é realizado pelo módulo Recuperador de Especialistas, que é a recomendação dos especialistas, a partir da identificação dos especialistas correspondentes a cada experiência. As entradas possíveis são oriundas da lista de experiências fornecidas pelo Módulo Definidor de Experiências e a saída é uma lista ordenada de especialistas. São detectados os especialistas para cada experiência retornada do módulo Definidor de Experiências.

Uma vez que estes especialistas são detectados, são ordenados de acordo com o seu contexto. O SRE ICARE classifica (estabelece um *ranking*) os especialistas baseando-se nos seguintes elementos contextuais: (1) grau de especialidade; (2) acessibilidade; (3) disponibilidade; (4) reputação; (5) distância social; e (6) diferença dos níveis organizacionais (hierarquia) que ocupam na organização.

Dentre estes elementos contextuais, os que recebem maior importância são: a distância social entre o usuário e o especialista e a diferença entre os níveis organizacionais que ambos ocupam. Pois, o ICARE considera que quanto menor for a diferença entre os níveis organizacionais de ambos e menor for a distância social aumentam as chances do especialista recomendado colaborar. Apesar destes elementos contextuais receberem maior importância, há a possibilidade de atribuir pesos a estes elementos, podendo, assim, tornar algum elemento mais importante que outro. Enfim, estes pesos são configuráveis pelo sistema.

## **Módulo Apresentador de Especialistas**

Por fim, o último passo no processo é apresentar os especialistas recomendados ao usuário. O principal objetivo deste módulo é a apresentação dos especialistas de acordo com as necessidades ou preferências dos usuários, de acordo com atributos como quantidade e formato de informação. Por exemplo, os usuários podem querer receber todas as informações retornadas no formato HTML.

### **2.3.2 Contexto no ICARE**

O ICARE utiliza contexto para proporcionar recomendações personalizadas, assim, a relação dos especialistas recomendados muda de acordo com o usuário e o instante de tempo. O objetivo destas recomendações personalizadas é dar prioridade aos especialistas mais aptos, em um determinado instante de tempo, a colaborar com as necessidades do usuário.

Para proporcionar estas recomendações personalizadas, alguns elementos contextuais são considerados, são eles: grau de especialidade, disponibilidade, nível organizacional, reputação e distância social. O grau de especialidade representa o quanto o usuário conhece sobre o assunto; a disponibilidade representa o quanto o usuário está disponível em um instante de tempo; o nível organizacional representa qual o cargo o especialista ocupa na organização; a reputação representa o quanto um especialista é confiável dentre um conjunto de pessoas; e a distância social representa qual a distância social que o usuário está do especialista, pois, quanto mais próximo o usuário estiver do especialista, espera-se que ele esteja mais disposto a colaborar. De posse destes elementos contextuais, o ICARE utiliza heurísticas para classificar (estabelecer um ranking) os especialistas mais apropriados a serem recomendados em um instante de tempo  $t$ .

## 2.4 *SmallBlue*

SmallBlue (Lin, 2009) é uma plataforma baseada na teoria de Redes sociais desenvolvidas para a IBM. A ferramenta explora as potencialidades das interações sociais informais para localizar especialistas em um determinado tópico. A IBM pretende comercializá-la, ou seja, oferecê-la como um produto aos seus clientes.

SmallBlue realiza continuamente buscas nos *e-mails* enviados pelos usuários e nas mensagens trocadas de forma síncrona, mensagens instantâneas, para determinar a rede social desses usuários e poder detectar também as suas experiências.

De acordo com Lin (2009), a utilização destas fontes privadas de informações (*e-mail* e mensagens instantâneas) é riquíssima, pois, contém valiosas informações sobre o conhecimento do usuário em relação a um determinado assunto e o conhecimento do conjunto de usuários que se relacionam uns com os outros. Algumas outras questões são levantadas em decorrência da utilização destas fontes, são elas:

- **cobertura:** a utilização de *e-mail* é generalizada. Assim, o conjunto de informações para inferir as especialidades das pessoas não é um problema, que, eventualmente, pode ocorrer em SREs que precisam que os usuários alimentem o sistema para que possa inferir suas especialidades;
- **manutenibilidade:** novos *e-mails* são gerados constantemente;
- **Facilidade de uso:** as pessoas já utilizam *e-mail* no dia-a-dia, assim, não haverá qualquer trabalho adicional para o usuário.

A desvantagem de utilizar dados privados é que capturá-los e usá-los pode violar a privacidade dos usuários. E, por causa desta dificuldade, persuadir as pessoas para deixarem estes dados serem utilizados é um desafio. Por causa desta sensibilidade, os autores desenvolveram uma política que restringe claramente quais dados são coletados, como são utilizados e quais informações são disponibilizadas aos usuários. Assim, SmallBlue não permite que os usuários consigam descobrir o conteúdo das mensagens trocadas entre as pessoas e, tampouco, quem foram as pessoas que se comunicaram por um motivo ou outro.

Smallblue dispõe de uma ferramenta intitulada *Smallblue Egoi*, que torna possível ao usuário visualizar a sua rede social pessoal bem como o capital social de cada contato. Capital social é definido como o número de novos contatos que um usuário pode passar a ter a partir dos seus contatos diretos – incluindo a distribuição organizacional e geográfica destes novos contatos. Conforme mencionado anteriormente, esta é uma informação personalizada, somente o próprio usuário pode visualizá-la.

#### **2.4.1 Inferência de especialidades**

Para identificar os especialistas, SmallBlue dispõe de um módulo intitulado *SmallBlue Find*, que identifica um conjunto de especialistas e os ordena de acordo com a relevância. Para isso, é necessário que o usuário forneça uma *string* de consulta, que é mapeada para um conjunto de palavras-relacionadas. Assim, *SmallBlue Find* agrega os resultados provenientes da pesquisa realizadas por estas palavras-chave e os classifica de acordo com a relevância e estrutura da rede social. É apresentada ao usuário que realizou a busca a relação dos 1.000 primeiros especialistas. Para cada especialista são apresentadas as seguintes informações: fotografia, seu papel na organização e seu *status* (*online* ou não). Os usuários podem personalizar a busca por especialistas da seguinte forma: em uma divisão específica da organização, em um país, em uma comunidade, grupo ou específica distância social.

Para aumentar a probabilidade de uma pessoa que esteja precisando contatar algum especialista receber um retorno, Smallblue aplica o conceito de distância social – mais especificamente, os 6 graus de separação. Assim, os especialistas que estão a uma distância 3 ou inferior recebem prioridade no *ranking* das recomendações, pois, pressupõe-se que os especialistas mais próximos possuem uma maior tendência em oferecer ajuda. Os usuários podem redefinir este critério caso queiram especialistas localizados a uma distância social maior.

De acordo com Lin (2009), SREs devem disponibilizar aos usuários especialistas com informações que os ajudem a entender o quão qualificados os especialistas recomenda-



dos são para responder seus questionamentos. *SmallBlue* trata esta questão da seguinte forma: prover a informação da distância social do usuário em relação ao especialista; e possibilitar que os usuários, a partir de um clique no nome ou fotografia do especialista, tenham informações mais específicas e públicas sobre os especialistas, além de disponibilizar uma lista dos caminhos mais curtos e o limite máximo de 16 caminhos alternativos para que o usuário alcance o especialista. Um caminho curto é 6, por causa dos 6 graus de separação mencionado anteriormente.

*SmallBlue* ainda dispõe de um componente para exibir e analisar as redes sociais dos especialistas em um determinado tópico, *SmallBlue Net*. Normalmente, os melhores especialistas são tratados como os que possuem um maior conhecimento sobre um determinado assunto, mas, há momentos em que é mais conveniente encontrar / contatar um especialista que o usuário interessado conheça – enfim, que possua maior afinidade. Ou, ainda, pode haver a necessidade de encontrar uma comunidade emergente associada a um determinado tópico e identificar qual é o principal especialista naquela comunidade.

#### **2.4.2 Contexto no *SmallBlue***

*SmallBlue* explora o contexto social em que os colaboradores da organização estão inseridos a partir dos relacionamentos existentes entre eles. Assim, têm-se:

- os especialistas que possuem uma distância social 3 ou inferior em relação ao usuário que recebeu as recomendações dos especialistas têm prioridade no *ranking*;
- a possibilidade dos usuários buscarem especialistas inseridos em determinados contextos, tais como, uma comunidade, um grupo, um país ou organização;
- a possibilidade dos usuários visualizarem e analisarem as redes sociais dos especialistas, pois, assim os usuários podem contatar não necessariamente os especialistas que possuem maior conhecimento, mas, qualquer pessoa que conheça o especialista.

#### **4.4.3 Trabalhos similares ao *SmallBlue***

McArthur (2003) propôs um modelo intitulado HALE, que tem como propósito encontrar pessoas que possuam informações, conhecimento ou experiências úteis que possam auxiliar na resoluções de problemas. HALE utiliza as conexões implícitas e explícitas das pessoas a partir

de mensagens trocadas via e-mails. HALe é um modelo proposto para as organizações, apesar de não ter sido explicitamente implantado em uma organização real.

Esse modelo utiliza contexto para identificar redes de pessoas e redes de tópicos. Por exemplo, as redes sensíveis ao contexto das pessoas estão na capacidade que o modelo possui de identificar uma maior participação em determinado documento ou relatório de um conjunto de pessoas que possuem relacionamento entre elas. Já a rede sensível ao contexto de tópicos faz referência a um conjunto de tópicos que uma determinada pessoa tem atuado em relação a um referido documento ou relatório.

Outro trabalho é o apresentado por Tyler (2003), que desenvolveu uma aplicação que aproxima um conjunto de pessoas que possuem conhecimentos similares – o objetivo principal é fazer com que este grupo de pessoas com conhecimentos homogêneos possam compartilhar conhecimento, interagirem, colaborarem e, como consequência, encontrarem soluções para eventuais problemas. Tyler utiliza mensagens de e-mails trocadas entre os colaboradores da organização HP Labs para inferir o conhecimento pertinente a cada colaborador, além de detectar os relacionamentos entre estes colaboradores – a identificação destes relacionamentos permite a extração de redes sociais, que servem com insumo para as montagens destes grupos de colaboradores. Isto é, parte-se da premissa de que, quanto mais próximo um colaborador estiver do outro, maiores são as chances deles colaborarem entre si. Contexto para Tyler está exatamente na possibilidade de reunir um grupo de pessoas em torno de um determinado conhecimento.

## **2.5 *Conscius***

Conscius (Moraes, 2010) é uma ferramenta projetada para localizar especialistas em determinados trechos de códigos fontes de um projeto. Nesta ferramenta, eventuais dúvidas que os desenvolvedores venham a ter em um determinado trecho de código podem ser resolvidas mais rapidamente, já que estes desenvolvedores tem a possibilidade de identificar o especialista de modo mais rápido.

Como fonte de informação, para que possa ser possível gerar as recomendações, Conscius utiliza código fonte, histórico e documentação do projeto e lista de discussões dos desenvolvedores arquivadas para recomendar especialistas em determinados códigos fonte.

Os autores argumentam que a utilização de listas de discussões é uma fonte riquíssima de informações, pois, os desenvolvedores têm o hábito de escrever sobre o que estão tra-

balhando e sobre o que têm conhecimento. Ainda, afirmam que, com a utilização de algoritmos de mineração é possível relacionar as discussões às documentações e aos códigos fonte de um determinado projeto. Os relacionamentos construídos são:

1. **desenvolvedor—código fonte:** a partir do próprio do código fonte;
2. **código fonte—código fonte:** a partir de dependência técnicas entre os códigos;
3. **desenvolvedor—desenvolvedor:** a partir da relação de discussões que aconteceram entre eles;
4. **documentação—código fonte:** a partir da documentação do projeto (javadoc).

Estes elementos indicam quem produziu o conhecimento, além de indicarem o relacionamento entre as atividades do projeto, seus artefatos e a comunicação entre os desenvolvedores. Todo o conhecimento produzido por estes elementos fazem com que o Conscius consiga inferir quem é o especialista em uma determinada parte (pacotes, classes, métodos) do código fonte de um projeto (software).

### 2.5.1 Inferência de especialidades

O SRE Conscius disponibiliza uma área em que o desenvolvedor pode solicitar ajuda sobre um determinado trecho de código e, assim, não é necessário o envio de mensagens em listas de discussões em outro ambiente que não seja a própria ferramenta. A ferramenta identifica as classes referenciadas nas mensagens e, com isso, consegue localizar as demais classes quem possuem dependências com elas.

O Conscius possui um componente chamado Classifier que identifica as palavras-chave de uma Mensagem ou um Javadoc, ignorando as stoptwords e caracteres inválidos. Logo em seguida, outro componente, intitulado Knowledge, recebe a lista de palavras-chave do componente Classifier, as analisa e as associam com os pacotes Javadoc. Feito isso, o componente Knowledge realiza uma classificação automática, a partir da análise de similaridade, baseada em fuzzy (Galho, 2004), para determinar o quão são similares a lista das palavras-chave das mensagens e a lista de palavras-chave de um pacote javadoc (javadoc package).

Por fim, há o componente Recommender que identifica as pessoas mais qualificadas para responder a mensagem original do desenvolvedor (a dúvida). Este componente identifica os desenvolvedores que enviaram mensagens sobre o conhecimento solicitado e, então, calcula o score de comunicação. Para cada mensagem do desenvolvedor, que contém o conhecimento solicitado, o score é incrementado.

O componente Recommender computa, também, o score de desenvolvimento do desenvolvedor em cada classe referenciada na mensagem e em cada classe dependente das classes que são referenciadas na mensagem. Para cada commit efetuado pelo desenvolvedor no sistema de controle de versão seu score é aumentado. Assim, mede-se o nível de contribuição de um desenvolvedor para uma dada classe.

Como resultado de uma solicitação de ajuda realizada por um usuário, são retornados aos usuários os especialistas que possuem um score final maior. O score final é determinado pela soma do score de comunicação e do score de desenvolvimento.

### 2.5.2 Contexto no *Conscius*

Para os desenvolvedores do *Conscius*, contexto no ambiente é definido da seguinte forma:

- **código fonte:** a recomendação de especialistas no *Conscius* é realizada no contexto do código fonte, em que são detectados especialistas em um determinado trecho de código, seja um método, uma classe ou um pacote;
- **troca de mensagens:** a troca de mensagens realizada pelos desenvolvedores referente a um mesmo conhecimento (e.g. classe ou método);
- **melhor relacionamento:** o especialista que possui um melhor relacionamento com o usuário que solicitou ajuda. Este melhor relacionamento é medido pela quantidade de vezes que o especialista respondeu as mensagens enviadas pelo usuário remetente;
- **nível de participação nas classes e trocas de mensagens:** o nível de participação dos desenvolvedores nos arquivos anexos à mensagem também são medidos.

Enfim, o SRE *Conscius* utiliza os códigos fontes de um determinado projeto de software para recomendar especialistas em determinados trechos de código. O *Conscius* identifica os especialistas e os classifica (ranking) de acordo com seus níveis de colaboração. Quanto maior a colaboração de um especialista em um determinado trecho de código, maiores são as chances de este especialista colaborar com o usuário que solicitou ajuda. Para estabelecer este ranking, também é considerada a quantidade de mensagens trocadas entre o usuário que está solicitando a ajuda e o especialista que será recomendado, assim, nos casos em que tenha existido uma colaboração prévia e, quanto maior for este nível de colaboração, há uma tendência que este especialista esteja mais propenso a ajudá-lo.

### 2.5.3 Trabalhos Similares ao *Conscius*

De forma similar ao *Conscius*, *Presley* (Trindade, 2009) é um SRE que recomenda especialistas em determinados trechos de código de um determinado projeto de *software*. *Presley* foi desenvolvido para apoiar o desenvolvimento de *software* distribuído. A percepção dos colaboradores em identificar especialistas em desenvolvimento de *software* desta natureza é enfraquecida naturalmente, já que não há o contato presencial e diário dos desenvolvedores por isso, *Presley* objetiva identificar os especialistas de forma automatizada e, assim, naturalmente, contribuir para que o processo de desenvolvimento de *software* possa ser agilizado.

Contexto para o SRE *Presley* é o mesmo que o SRE *Conscius*, que considera a união entre os elementos de conhecimento (desenvolvedores, documentos e códigos fonte) pertinentes ao desenvolvimento de um determinado projeto para recomendar especialistas que são mais propensos a responderem as eventuais dúvidas que surgirem. De acordo com Trindade (2009), essa peculiaridade que permite a união entre os elementos e, assim, a aplicação de contextos em ambientes desta natureza foi descrita por Ye (2007).

## 2.6 Estudo Comparativo dos SREs Discutidos

O Quadro 2.1 fornece uma visão geral das principais peculiaridades inerentes a cada um dos SREs previamente discutidos. O Quadro 2.1 está organizado da seguinte forma: a primeira coluna identifica o SRE em evidência; a segunda coluna apresenta as fontes de informações que são utilizadas para identificar as especialidades; já a terceira coluna faz menção às Técnicas, Métodos ou Ferramentas que são utilizadas para que o processo de inferência de especialidades possa ser realizado; e, por fim, na quarta coluna são apresentados os elementos que são considerados para determinar o contexto em cada SRE.

**Quadro 2.1:** Visão geral sobre características peculiares aos SREs aprestados

SRE	Evidência de Especialidades	Tecnologia/Técnicas//Metodologia/Ferramentas Utilizadas	Contexto
<b>ReferralWeb (1997)</b>	1) autoria de artigos científicos ou relatórios; 2) interações entre as pessoas na Web	1) análises de redes sociais; 2) mineração de dados na Web	1) Distância Social 2) Grau de Especialidade
<b>ICARE (2008)</b>	informações do currículo Lattes	1) Ontologia de Domínio da ACM; 2) Heurística	1) Disponibilidade; 2) Acessibilidade; 3) Distância social; 4) Nível Organizacional, 5) Tarefa Atual; 6) Reputação; 7) Interesses Habilidades; 8) Grau de especialidade;
<b>SmallBlue (2009)</b>	1) conteúdo dos e-mails corporativos; 2) conteúdos de mensagens instantâneas corporativas	1) <i>data mining</i> ; 2) recuperação de informação; 3) análises de redes sociais.	1) distância social; 2) experiência no contexto ( <i>e.g.</i> comunidade ou um grupo); 3) visualização no contexto social dos especialistas
<b>Conscious (2010)</b>	1) código fonte em java 2) histórico de discussões dos desenvolvedores 3) documentação do projeto	1) <i>data mining</i> 2) cria relacionamentos ( <i>e.g. desenvolvedor—código fonte e desenvolvedor—desenvolvedor</i> )	1) Código fonte; 2) Troca de mensagens; 3) Melhor relacionamento 4) Nível de participação nas classes e trocas de mensagens; 5) Grau de especialidade

Conforme apresentado no decorrer do capítulo, são vários os SREs disponíveis na literatura, e a escolha dos SREs apresentados no Quadro 2.1 foi em decorrência destes explorarem de alguma forma o conceito de contexto computacional. Neste universo de SREs apresentados no Quadro 2.1, com exceção do ICARE, nenhum explora explicitamente o conceito de contexto, por isso, fez-se necessário realizar uma análise minuciosa e crítica para identificar como contexto computacional estava sendo explorando nestes ambientes.

Conforme supracitado, o único SRE que usa explicitamente o conceito de contexto computacional é o ICARE (Petry, 2008). Petry (2008) utiliza a plataforma *Lattes* para identificar as especialidades dos usuários e considera os seguintes elementos contextuais: disponibilidade, acessibilidade, nível organizacional, distância social, tarefa atual, reputação, interesses e grau de especialidade.

O ICARE foi apenas um protótipo de um SRE sensível ao contexto - algumas informações de como ele explora estes elementos contextuais foram possíveis de identificar com a leitura dos trabalhos publicados, outras passíveis de serem deduzidas, e outras não. Tanto a disponibilidade, a acessibilidade, quanto o nível organizacional são explicitamente fornecidos pelas pessoas que foram cadastradas no sistema. Da mesma forma, acredita-se que a tarefa atual deve ser informada constantemente pelo usuário. O elemento interesse, que determina as preferências das pessoas não foi possível identificar sua origem, mas acredita-se que estes interesses também devem ser informados pelas pessoas. O grau de especialidade é 0 ou 1, e identifica se uma determinada pessoa conhece ou não determinado assunto, a partir das informações fornecidas no currículo Lattes. A distância social não foi possível identificar como é explorado no ICARE, já que o currículo Lattes não é baseado na teoria de redes sociais e não é realizada nenhuma análise detalhada no currículo para tentar identificar estes relacionamentos sociais. Por exemplo, as relações de autor e co-autores nas publicações científicas disponíveis no currículo Lattes. Já a reputação acredita-se que é formada a partir do *feedback* fornecido pelos usuários do sistema.

Todos os SREs apresentados exploram de alguma forma o relacionamento social entre as pessoas, por exemplo, através de mensagens de e-mails *ReferralWeb* e *SmallBlue*; através de programas de mensagens instantâneas, como *SmallBlue*; a partir dos relacionamentos existentes entre os desenvolvedores de software, como no caso do *Conscious*; ou através da análise da reputação de um especialista como o caso do *ICARE*. Apesar destes SREs explorarem elementos que caracterizam os relacionamentos sociais entre as pessoas, nenhum deles é

explorado explicitamente nas Redes Sociais Web. Um dos objetivos desta tese é o desenvolvimento de um SRE para Redes Sociais Web.

Normalmente, os SREs utilizam técnicas para recuperação e classificação automática da informação para inferir especialidades, por exemplo, *ReferralWeb* e *Conscious*. Há SREs que agregam, também, o conceito de ontologia de domínio com o intuito de adicionar semântica para as recomendações dos especialistas, tornando-as mais eficientes, haja vista que a inferência de especialidades não será realizada simplesmente de forma sintática (*i.e.* a partir da análise de textos planos). Um exemplo de SRE que utiliza ontologia de domínio é o *ICARE*, porém, como o próprio nome diz, permite a inferência de especialidades em um domínio específico.

Essas ontologias, que representam um domínio em particular, são pré-definidas, ou seja, a identificação de especialistas é limitada a um domínio em particular. Contrário a esta restrição, o SRE proposto nesta tese tem como objeto utilizar ontologia para o enriquecimento semântico das recomendações, isto é, a identificação dos especialistas pode ser realizada nos mais variados contextos, pois a ontologia utilizada na referida ferramenta surge à medida que as interações entre os usuários acontecem.

Este estudo foi essencial para que houvesse um entendimento de como os SREs estão lidando com o problema de identificar especialistas de domínio. Preponderante para que seja proposto um SRE que utilize elementos não considerados pelos demais SREs para tornar esta recomendação mais eficaz. Ou ainda, explorar elementos já utilizados por estes sistemas de forma mais eficiente.

## **2.7 Considerações Finais**

Para que houvesse um entendimento sobre a definição e peculiaridades inerentes aos Sistemas de Recomendação de Especialistas, o presente capítulo apresentou discussões que permitem um melhor esclarecimento de quais são as reais importâncias e dificuldades da área.

Também, foram discutidos alguns Sistemas de Recomendação de Especialistas disponíveis na literatura. O estudo destes SREs foi preponderante para que houvesse um entendimento de como os sistemas disponíveis estão lidando com o problema de encontrar pessoas que possuem conhecimentos específicos. Assim, passa-se a ter parâmetros para propor um SRE de especialista diferenciado, isto é, que considere ou agregue elementos distintos para que a identificação destes especialistas possa ser realizada de forma mais eficiente.



Além disso, embora os SREs apresentados explorem de alguma forma os relacionamentos sociais entre as pessoas, nenhum deles possui algum mecanismo automatizado que analise as ações dos especialistas que possam servir de insumo para determinar o quanto um especialista é confiável. Esta análise de confiança de um especialista é importantíssima para recomendar um especialista que possua maior credibilidade na rede, mesmo que este possua menos conhecimento que outro especialista que tenha dificuldades em se relacionar. Neste sentido, o Capítulo 3 apresenta conceitos inerentes à confiança e como esta pode ser medida em ambientes virtuais baseados na teoria das redes sociais.

## 3 Confiança em Redes Sociais

Confiança é um fenômeno comum que está presente na sociedade desde que ela existe (Yamamoto, 1990). Este fenômeno oferece subsídios para apoiar constantemente a sociedade nas tomadas de decisões, que vão desde as mais triviais, quando uma pessoa decide ou não ir ao trabalho, até as mais complexas, como investimento em bolsas de valores.

Com a evolução da sociedade, principalmente, no que concerne o “mundo” computacional, diversas pesquisas nesta área vêm surgindo. Estes estudos buscam, dentre outros fatores, entender como confiança, um dos princípios da sociedade, pode ser mapeada e representada em ambientes computacionais. Neste sentido, capítulo apresenta a definição de confiança (Seção 3.1) e como alguns trabalhos computacionais exploram o referido conceito; as propriedades de confiança, que representam particularidades do conceito de confiança no mundo real e que devem ser capazes de ser mapeadas no mundo computacional são apresentadas na Seção 3.2; as Seções 3.3, 3.4 e 3.5 apresentam alguns dos algoritmos para inferência de confiança disponíveis na literatura; e, por fim, a Seção 3.6 apresenta as considerações finais do capítulo.

### 3.1 Definição

Confiança é um conceito que sustenta a sociedade e diz o quão uma entidade (*e.g.* pessoa ou um agente) é confiável em relação à outra. O estudo de confiança é explorado em diferentes áreas, por exemplo, psicologia (Yamagishi, 1994, 1998), sociologia Murayama (2011), filosofia (Scheffler, 2010) e computação (Golbeck 2008).

A definição para confiança é dada pelos autores das mais variadas formas: Sztompka (1999) & Cook (2001) dizem que confiança se trata do julgamento que uma pessoa faz em

relação à outra; já Hardin (2001) diz que confiança mede expectativas sobre o comportamento das pessoas. Assim, há como identificar as boas ou más intenções das pessoas, o que pode, por exemplo, auxiliar diretamente a identificar se uma pessoa é capaz de respeitar os direitos das outras ou cumprir com suas obrigações. Golbeck (2005, 2008), apresenta outras formas para determinar confiança: experiências passadas, em que a confiança é determinada com base na observação de fatos passados; relacionamentos, em que são identificados o nível do relacionamento (*e.g.* relacionamentos de amizade ou fraternal) para ponderar o processo de inferência de confiança; e opiniões, quando são consideradas as opiniões de terceiros para determinar o nível de confiança.

Complementando as definições previamente apresentadas, Igarashi (2008) diz que confiança pode ser medida seguindo algumas particularidades: (1) *generalizada*, a confiança pode ser determinada em um nível generalizado, isto é, em relação a uma visão de um grupo de entidades; e (2) *particular ou restrita*, a confiança das entidades pode ser medida de acordo com particularidades definidas (*e.g.* os seus relacionamentos) (Yamagishi, 1994 & Yamagishi, 1998).

A confiança generalizada encoraja a exploração dos relacionamentos sociais de forma mais abrangente, pois estimula as pessoas a conhecerem pessoas desconhecidas e, como consequência, cria novas oportunidades de interações (Yamagishi, 1986, 1998). Por exemplo, se uma pessoa  $p$  é confiável perante um grupo  $g$ , e uma pessoa  $p1$  não é confiável perante este mesmo grupo  $g$ , então é mais provável que uma pessoa  $p2$  possa se relacionar com a pessoa  $p$ , *que é confiável perante o grupo  $g$* , do que se relacionar com a pessoa  $p1$ .

Já a confiança *particular ou restrita*, explora os relacionamentos diretos e pode não encorajar as pessoas a procurarem por novas oportunidades sociais, através da formação de novos relacionamentos sociais (Yamagishi, 1986, 1998). As particularidades dos relacionamentos permitem a exploração de características mais específicas das redes sociais como, por exemplo, os níveis de relacionamento (*e.g.* amigo (a), familiar, namorado (a)); a quantidade de amigos comuns entre duas pessoas e que possuam atitudes similares pode ser uma característica que remeta uma pessoa a ter um nível maior de confiança em outra.

Conforme supracitado, confiança pode ser explorada em diferentes áreas, – a área explorada neste trabalho é a computação – mais especificamente, a confiança em ambientes virtuais baseados em redes sociais. Um dos trabalhos pioneiros da computação sobre confiança é o desenvolvido por Marsh (1994), onde foram feitos estudos sobre o conceito em diferen-

tes áreas (*e.g.* sociologia) com o intuito de formalizar o conceito de confiança em ambientes computacionais, baseados em agentes que interagem e mantêm histórico de informações e que observem comportamentos.

Diante destes estudos, Marsh (1994) concebeu um modelo para aplicação em sistemas de inteligência artificial, para auxiliar agentes inteligentes a raciocinarem em situações de incerteza. Para avaliar a precisão deste modelo foi criada uma situação onde alguns agentes inteligentes poderiam interagir ciclicamente. Esta situação foi criada com base no Dilema dos Prisioneiros, um problema clássico da Teoria dos Jogos – um problema para analisar tomadas de decisões em situações onde não se conhecem todas as informações do domínio. Nesta situação, cada agente poderia se mover livremente por um tabuleiro. Caso ocorresse o encontro entre dois agentes no tabuleiro, eles deveriam interagir, nas condições do Dilema dos Prisioneiros, e ajustarem seus valores de confiança de acordo com o resultado desta interação. Este processo se repetia de acordo com o número de movimentos definido para os testes.

Outro trabalho referência, desenvolvido por Golbeck (2005), explora o conceito de confiança aliado às redes sociais e é baseado na definição dada por Deutsch (1962) onde diz que “confiança é a aposta sobre futuras ações de outras pessoas”. Assim, são considerados dois principais componentes para a definição: crença e compromisso. A crença porque representa a visão de uma pessoa sobre determinadas ações de outra pessoa considerada confiável. E o compromisso, que representa o nível de comprometimento de uma pessoa em realizar uma determinada ação. O trabalho desenvolvido por Golbeck (2005) será explorado com mais detalhes na Seção 3.4, já que um dos propósitos do projeto apresentado aqui é a aplicação do conceito de confiança em ambientes computacionais baseados na teoria das redes sociais.

Artz (2007) classificou a confiança no mundo computacional em quatro grandes áreas: políticas de confiança, confiança baseada em reputação, modelos gerais de confiança e confiança em fontes de informação.

- **políticas de confiança:** neste modelo a confiança é baseada em “títulos” que são adquiridos pelos agentes. A quantidade de títulos que os agentes possuem influencia o quanto eles são confiáveis – *e.g.* um título pode ser determinado pelo nível da colaboração que uma pessoa possui em uma rede social Web;
- **confiança baseada em reputação:** este modelo aborda informações dos agentes para inferir o grau de confiança: ações passadas, parcerias, comportamentos, são exemplos destas informações que podem ser utilizados como dados de

entrada para o cálculo do grau de confiança. Este modelo é recomendado para ser aplicado em redes sociais Web, em que as informações de relacionamentos presentes nestes ambientes podem ser exploradas no processo de inferência de confiança;

- **modelos gerais de confiança:** esta área de pesquisa está relacionada às definições e formalizações do conceito de confiança, isto é, obter informações deste fenômeno no mundo real e adaptá-las ao meio computacional. Em trabalhos deste gênero, o principal objetivo é descrever fatores que devem influenciar nos valores da confiança, baseando-se em trabalhos originados principalmente das áreas de psicologia e sociologia;
- **confiança em fontes de informação:** é quando a confiança é aplicada no processo de *Ranking* (classificação) de sistemas de recuperação da informação para auxiliar no processo de determinação de quais conjuntos de informações são mais relevantes que outros – *e.g.* podem ser realizadas análises de quão os autores de um determinados conhecimentos são mais ou menos confiáveis.

Como o propósito do presente trabalho é explorar inferência de confiança em ambientes computacionais baseados em redes sociais, a Seção 3.2 apresenta as propriedades de confiança que são preponderantes para a realização de inferência nestes ambientes.

### 3.2 Propriedades de Confiança

De acordo com Golbeck (2005), os estudos e definições de confianças disponíveis na literatura oferecem subsídios para que algumas propriedades de confiança possam ser definidas. Estas propriedades devem representar explicitamente elementos aos quais os usuários estão acostumados em suas vidas sociais (*e.g.* nível de relacionamento, pai, irmão (a), amigo (a), colega, namorado (a)) e devem ser bem definidas e claras (*e.g.* as pessoas e seus relacionamentos) de forma que possam ser utilizadas em ambientes computacionais.

Segundo Ries (2006), são três as propriedades para confiança discutida na literatura, são elas:

- i. **subjativa:** confiança é subjativa e, portanto, assimétrica, isto é, a confiança de uma pessoa  $p$ , em relação a uma pessoa  $p1$  não é a mesma confiança da pessoa  $p1$  em relação a  $p$ ;

- ii. **contexto:** confiança depende do contexto, ou seja, pode ser medida de uma pessoa em relação à outra, em um grupo ou até mesmo diante de uma determinada situação;
- iii. **dinâmica:** confiança é dinâmica, isto é, muda com o tempo - pode aumentar se as experiências forem positivas, e pode diminuir se as experiências forem negativas.

No âmbito computacional, para a aplicação em redes sociais, Golbeck (2005) define algumas propriedades de confiança: transitividade, que está diretamente relacionada ao item *ii* supracitado, já que considera o contexto social das pessoas; agregabilidade, também relacionado ao item *ii*; e assimetria, relacionado ao item *i*. Uma situação hipotética que representa um conjunto de relacionamentos entre pessoas é mostrado a seguir com o intuito de elucidar estas propriedades (vide Figura 3.2). Seja A um usuário que apenas possui ligações<sup>6</sup> com B e F, também usuários. B possui ligação com um usuário D. D por sua vez se relaciona também ao usuário C. F possui ligação com outro usuário G. C possui ligações com D e G. Assim, as propriedades da confiança podem ser definidas da seguinte forma:

- **transitividade:** a confiança não é perfeitamente transitiva entre usuários. Isto é, conforme a Figura 3.1, se A tem uma confiança alta em B e, além disso, B confia muito em D, isto não quer dizer que A possui alta confiança em D. Porém, há uma teoria de que confiança pode ser passada entre as pessoas. Se A precisa realizar escolhas baseadas nas ações de D, sem conhecê-lo, então A poderia pedir uma opinião a respeito de D para B. A opinião de B seria um dos principais fatores para definir a confiança de A em D nesta circunstância.

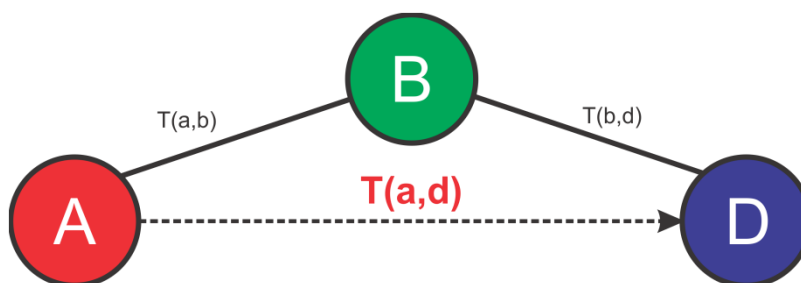


Figura 3.1: Conceito de transitividade

<sup>6</sup> As ligações são bidirecionais (*i. e.* se A está ligado a B então B está ligado a A).

- **agregabilidade:** a transitividade representa a possibilidade de transferir a confiança de pessoa a pessoa. Porém, de acordo com a Figura 3.2, A, ao tentar decidir se confia ou não em C, pode procurar informações a respeito de C em mais de uma fonte (*i. e.* perguntar a várias pessoas o que elas pensam sobre C). Por exemplo, se A pergunta a B sobre C, porém, B não conhece C, então B pergunta a D sobre C. Neste caso A terá uma recomendação a respeito de C proveniente da cadeia iniciada em B. Agora, se além disso, A perguntar a F sobre C, como F não conhece C, F pergunta a G sobre C, então A terá mais uma possível recomendação a respeito de C - desta vez proveniente da cadeia iniciada em F. Para tomar alguma decisão a respeito de C, A considerará as duas recomendações obtidas.

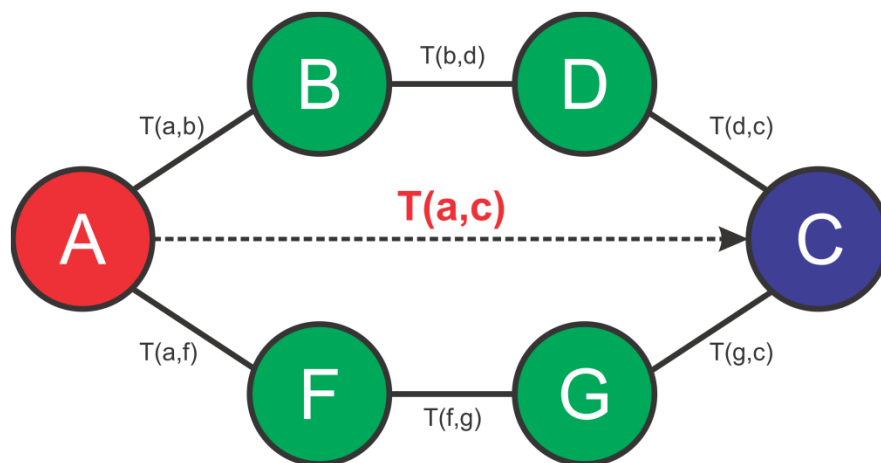


Figura 3.2: Conceito de agregabilidade

- **assimetria:** em geral, a confiança de duas pessoas diferentes a respeito de uma terceira pessoa pode ser diferente. Conforme a Figura 3.2, a opinião fornecida por G sobre C, pode ser completamente diferente da opinião fornecida por D sobre C. Isto ocorre porque G e D podem ter diferentes opiniões sobre o que seria um resultado final positivo, porque estes usuários tiveram diferentes experiências de vida, ou porque seus perfis psicológicos são diferentes. Além disso, a assimetria ocorre quando duas pessoas não mantêm o mesmo nível de confiança entre si, isto é, A confiar muito em B, não quer dizer que B confie muito em A.

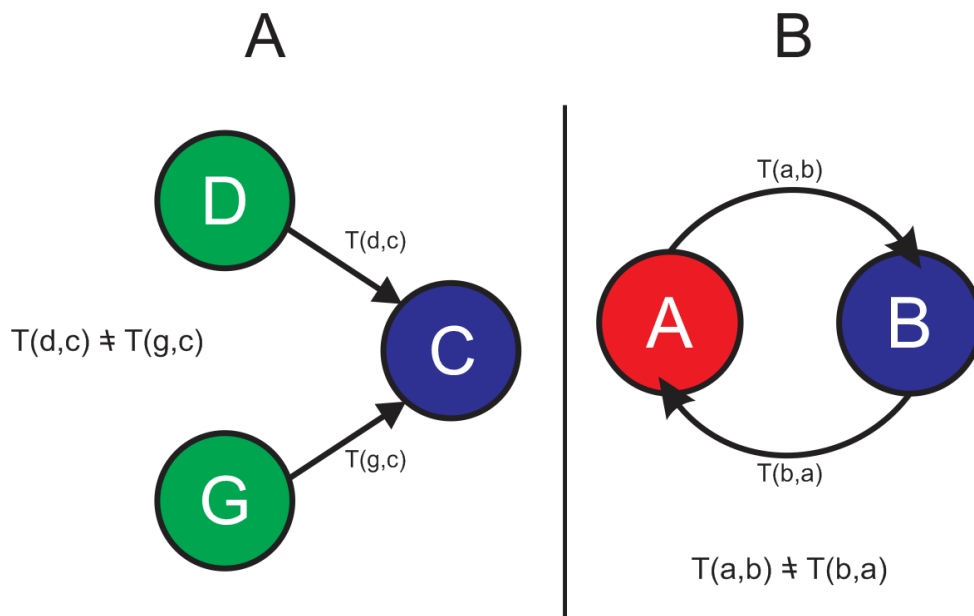


Figura 3.3: Conceito de assimetria

Outra propriedade importante e que deve ser considerada é o valor de confiança. Os relacionamentos diretos provenientes de uma rede social podem possuir um atributo que represente o valor de confiança entre duas pessoas. O valor de confiança pode ser booleano; onde diz se uma pessoa confia ou não em outra; ou pode obedecer a uma faixa de valores; onde diz que uma pessoa tem uma tendência a confiar ou não em outra pessoa (*e.g.* entre 1 e 10, como usado no projeto *FilmTrust*<sup>7</sup>, que personaliza as informações pertinentes a filmes aos usuários de acordo com seus relacionamentos e níveis de confiança entre eles). A escolha da forma como este valor de confiança será representado depende do projetista da aplicação, que, naturalmente, a definirá de acordo com as necessidade da aplicação que utilizará este valor de confiança.

O valor de confiança proveniente de cada relacionamento direto da rede social pode ser explicitamente atribuído pelo usuário ou não. Em casos em que este valor não seja explicitamente atribuído, segundo Golbeck (2005), sistemas automáticos de classificação podem ser utilizados. Por exemplo, os sistemas de classificação podem utilizar mecanismos que detectem as preferências dos usuários para a representação dos perfis de confiança destes usuário, isto é, a confiança dos usuários pode ser atribuída de acordo com suas preferências. Apesar disso, há uma tendência natural para que as aplicações que utilizem atribuição explícita de confiança nos relacionamentos diretos sejam mais eficientes, inclusive, no que

<sup>7</sup> *FilmTrust* - <http://trust.mindswap.org/FilmTrust/>



tange a questão da performance, já que a máquina não terá o trabalho de realizar processamentos complexos para inferir esta informação de confiança.

Da mesma forma que mecanismos podem ser utilizados para inferir a confiança dos relacionamentos diretos, há na literatura um conjunto de algoritmos para inferência de confiança dos usuários em uma rede social em relacionamentos indiretos, por exemplo, a confiança de um usuário A em relação a um usuário C, dado o fato que A possui relacionamento direto com um usuário B, e B por sua vez possui um relacionamento direto com C, pode ser calculada a partir do caminho de confiança  $A \rightarrow B \rightarrow C$ . Dentre estes algoritmos, cada um possui seu conjunto de particularidades: as seções 3.3, 3.4 e 3.5 apresentam e descrevem alguns dos algoritmos disponíveis na literatura voltados para as redes sociais Web, além de alguns dos trabalhos similares.

### 3.3 *FlowTrust*

*FlowTrust* (Wang, 2011) baseia-se na teoria do fluxo em redes para tornar possível a avaliação de confiança em sistemas de gestão de confiança. Assim, a abordagem do *FlowTrust* utiliza a modelagem de um grafo de confiança que pode simular uma rede de confiança em outro domínio. O algoritmo simula grafos da teoria de fluxo em redes elétricas, no qual os valores de confiança representam as capacidades distribuídas na rede. Deste modo, o algoritmo avalia a quantidade máxima de confiança que pode fluir através dos elementos da rede.

Este algoritmo baseia-se também em confiança multidimensional, isto é, são avaliados mais de um elemento na inferência do grau de confiança. Neste algoritmo, são considerados dois elementos para inferir a confiança de um nó  $u$  em relação a outro  $v$ : valor de confiança (*trust*) e nível de confiança (*confidence level*). Ambos os elementos são informações disponíveis para cada aresta do grafo da rede, isto é., sempre que existe uma aresta que liga dois nós, existem valores de *trust* e *confidence level* associados àquela aresta. O elemento *trust* representa a quantidade de confiança que um nó pode atribuir a outro. Já o *confidence level*, é baseado no valor do elemento *trust* e sua interpretação varia de acordo com as métricas de confiança 2, 3 e 4, que serão explicadas na Seção 3.3.1. Mesmo com a característica de multidimensionalidade, o algoritmo *FlowTrust* realiza uma abordagem analítica das informações. Esta abordagem caracteriza-se por considerar todas as informações contidas nas arestas do grafo de confiança, sem excluir os caminhos maiores (determinado pela quantidade de arestas entre um nó  $u$  e  $v$ ), diferente da abordagem simplificada, que apenas considera os menores

caminhos do grafo que ligam dois nós. A abordagem simplificada é utilizada em algoritmos que possuem custo computacional muito alto para sua execução.

A partir dos dois elementos considerados pelo algoritmo foram definidas quatro métricas de confiança: fluxo máximo do valor de confiança; fluxo máximo do nível de confiança; custo mínimo de incerteza com o fluxo máximo do valor de confiança; e mínimo custo de desconfiança com máximo fluxo do nível de confiança. Estas métricas serão explicadas na Seção 3.3.1 durante a explicação do funcionamento do algoritmo.

### 3.3.1 Inferência de confiança

Para inferir a confiança entre dois nós  $u$  e  $v$  pertencentes a um grafo  $G$ , o algoritmo *FlowTrust* cria uma representação do grafo de confiança como um grafo de fluxo. Grafos de fluxo são utilizados para encontrar a melhor maneira de transportar fluido de um ponto a outro por meio de passagens com capacidades limitadas em uma rede. Neste grafo, os dois elementos considerados, *trust* e *confidence level*, são representados de duas maneiras distintas, ora como a capacidade, que diz respeito à quantidade máxima de fluxo que pode passar pela aresta; ora como o próprio fluxo, que diz respeito à quantidade real de fluido que passa pela aresta, isto é, a quantidade de confiança que pode ser passada de um nó a outro.

Para explorar os conceitos da teoria do fluxo, o algoritmo utiliza algumas equivalências da rede de confiança com o grafo de fluxo. Primeiramente, ao se analisar a confiança entre dois usuários  $A$  e  $B$  é necessário que se obtenha o valor total de fluxo que passa entre estes dois nós na rede. Este valor é obtido com o somatório de todos os valores de *trust* (ou, *confidence level*) atribuídos pelo usuário  $A$ , ou ainda, todos os valores de *trust* atribuídos ao usuário  $B$ .

Além disso, para cada unidade de fluxo que passa de um nó a outro é gerado um custo, geralmente representado pelo valor complementar de um dos elementos analisados (ou seja,  $1-trust$  ou  $1-confidence level$ , dado que os valores de confiança pertencem ao conjunto  $[0, 1]$ ). O custo total de um fluxo é dado pelo somatório dos valores de fluxos multiplicados pelo custo de cada aresta (Equação (3.2)). Neste cenário, o problema de inferência de confiança consiste em maximizar a quantidade de fluxo que passa entre os nós e minimizar o custo para a passagem deste fluxo. Para maximizar o fluxo e minimizar o custo, o algoritmo proposto realiza uma busca pelo melhor caminho do grafo, isto é, o caminho que pode transferir o maior fluxo com o menor custo. O valor de fluxo obtido neste caminho representa o valor de confiança.

$$F = \sum_{i:(i,v) \in A} f(i,v) \quad (3.1)$$

$$C = \sum_{i,j \in N} cost(i,j) * f(i,j) \quad (3.2)$$

Nas Equações (3.1) e (3.2),  $A$  e  $N$  representam o conjunto de arestas e de nós que formam a rede, respectivamente. O parâmetro  $v$  representa o nó do qual se deseja informações de confiança, e  $i$  e  $j$  representam os demais nós que conectam o nó  $v$  à origem, isto é, o nó  $u$ . Além disso, as funções  $f(i,j)$  e  $cost(i,j)$  representam respectivamente o fluxo e o seu custo em uma aresta.

Dados os dois elementos a serem ponderados na análise de confiança, e o problema de máximo fluxo e mínimo custo de fluxo, as quatro métricas de confiança utilizadas pelo *FlowTrust* podem ser analisadas em uma rede como representada na Figura 3.4.

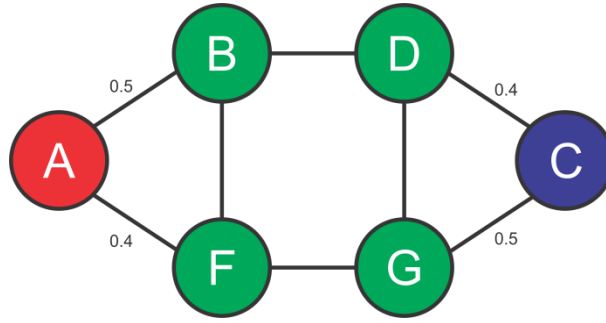


Figura 3.4: Grafo que representa as ligações de confiança de um nó  $a$  em relação a  $c$  (métricas M1 e M2).

Na Figura 3.4 deseja-se encontrar o valor de confiança de  $u$  em relação a  $v$  na rede. A partir desta situação, foram definidas quatro métricas que são fundamentadas na seguinte teoria:

A teoria do fluxo diz que a quantidade máxima de fluído <sup>8</sup>que pode passar de um nó a outro deve ser dada ora pela quantidade máxima de fluído que pode sair do nó origem ( $u$ ), ora pela quantidade máxima que pode chegar ao nó destino ( $v$ ), em uma unidade de tempo. Caso haja diferença nos valores de fluxo de saída de  $u$  e o fluxo de entrada de  $v$ , o valor máximo do fluxo é o menor deles. Além disso, pode ser utilizado - aplicado nas métricas M3 e M4 - o valor do custo deste fluxo. O valor mínimo do custo deve determinar a melhor solução para o problema do fluxo.

<sup>8</sup> Exemplos: informação, água, carga elétrica, etc.

A partir desta teoria são definidas as seguintes métricas:

**M1 - Fluxo Máximo do Valor de Confiança (MaxT)** - esta métrica restringe-se à primeira parte do problema de máximo fluxo e mínimo custo de fluxo. Supondo que os valores exibidos em cada aresta da Figura 3.4 dizem respeito aos valores de *trust*, pode-se observar que o valor total do fluxo, neste exemplo, é dado por  $0.4 + 0.5 = 0.9$ . Além disso, o valor  $\text{MaxP} = 2$  apresentado é a quantidade máxima de caminhos disjuntos (i. e. não possuem nós em comum), assim o valor normalizado de *trust* é  $\frac{\text{MaxT}}{\text{MaxP}} = \frac{0.9}{2} = 0.45$ .

**M2 - Fluxo Máximo do Nível de Confidência (*confidence level*) (MaxC)** - esta métrica é exatamente como a métrica do Fluxo Máximo do Valor de Confiança, porém aplicada aos valores de *confidence level*. No exemplo, para o fluxo de confiança não ocorrem alterações nos valores totais dos fluxos, nem no valor normalizado obtido, uma única alteração ocorre na nomenclatura utilizada na equação de normalização, mudando para  $\frac{\text{MaxC}}{\text{MaxP}}$ .

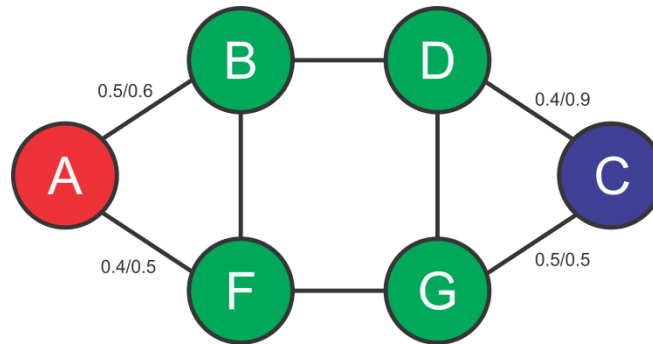


Figura 3.5: Grafo que representa as ligações de confiança de um nó  $u$  em relação a  $v$  (métricas M3 e M4).

**M3 - Custo Mínimo de Incerteza com o Fluxo Máximo do Valor de Confiança ( $\text{Min}\bar{C}\text{MaxT}$ )** - considera o problema de máximo fluxo e mínimo custo de fluxo. Supondo que a Figura 3.5 mostra um grafo em que os dois valores apresentados para cada aresta referem-se ao valor de fluxo de *trust* e a quantidade de fluxo de *confidence level*, respectivamente. Considerando a teoria do fluxo, explicada anteriormente, são explorados os caminhos possíveis na rede de modo a encontrar uma solução para o problema. Uma vez encontrada a solução, é realizado o cálculo do valor normalizado de confiança. Para esta métrica, o fluxo é dado pelo valor de

confiança (*trust*) e o custo deste fluxo pelo valor de incerteza de confiança (1-*confidence level*).

**M4 - Custo Mínimo de Incerteza com o Fluxo Máximo do Valor de Confiança** ( $\text{Min}\bar{T}\text{MaxC}$ ) - este valor é calculado analogamente à métrica M3, porém o fluxo é dado pelo valor de *confidence level* e o valor de incerteza por  $1-\text{trust}$ .

Para que a confiança entre dois usuários  $u$  e  $v$  possa ser inferida pelo *FlowTrust* as seguintes etapas são seguidas:

- fornece-se o grafo direcionado com as especificações dos valores de *trust*, *confidence level* e os nós de origem e destino;
- encontra-se o(s) caminho(s) que soluciona(m) o problema da métrica abordada;
- encontra-se o valor de fluxo nos caminhos encontrados como solução;
- normaliza-se o valor de fluxo encontrado nos caminhos de acordo com o número de caminhos utilizados – o valor normalizado é utilizado como retorno do algoritmo e deve pertencer ao intervalo  $[0, 1]$ .

### 3.3.2 Algoritmos semelhantes

A principal característica do algoritmo *FlowTrust* é sua abordagem analítica de informações sem a perda da multidimensionalidade. A abordagem analítica vem sendo empregada principalmente para resolver o problema de perda de informação da abordagem simplificada. No entanto, a utilização da abordagem analítica traz alguns problemas como, por exemplo, o alto custo computacional dos algoritmos. Para resolver problemas desta natureza, surgem pesquisas que buscam representar grafos de confiança em outros domínios.

Mahoney *et. al.* (2005), intitulado *RelTrust*, apresenta um algoritmo de confiança que representa o grafo de confiança como uma rede em que são analisadas as tendências de falhas entre os nós, ou terminais. Este algoritmo tenta calcular a probabilidade de que o caminho que conecta dois terminais esteja operacional, isto é, possa ser utilizado normalmente. Este cálculo é realizado com base na Teoria de Confiabilidade da Rede. A Teoria da Confiabilidade é uma teoria sobre falhas em sistemas, que permite aos pesquisadores preverem o acontecimento de falhas em sistemas de uma arquitetura com componentes que possuem algum nível de confiabilidade (Gavrilov, 2001). A partir desta teoria é possível calcular a probabilidade de obtenção de um resultado desejado em uma rede, probabilidade esta que é abordada como o valor de confiança inferido.

Outro algoritmo que utiliza a abordagem analítica é o *RN-Trust*, elaborado por Taheerian *et. al.*(2008). Este algoritmo cria uma representação do grafo de confiança no domínio da teoria dos circuitos, no qual o cálculo de equivalência de resistores realiza o trabalho de inferência de confiança. Para o cálculo da equivalência de resistores, cada um dos valores de confiança disponíveis nas arestas do grafo de confiança passa por uma função que resulta em valores de resistência das arestas do circuito. Em seguida é utilizada a teoria dos circuitos para calcular a resistência equivalente do circuito, que representará a confiança inferida.

### 3.4 *SUNNY*

*SUNNY* (Kuter; 2007, 2010) é um algoritmo de análise de confiança fundamentado em modelos probabilísticos. Para realizar a análise de confiança, este algoritmo mapeia uma rede de confiança em um domínio social e cria um grafo de probabilidades. No grafo de probabilidades, cada um dos valores de confiança representa o quanto um nó deve confiar na informação fornecida por outro.

O algoritmo *SUNNY* apresenta uma metodologia para extração das informações de confiança do domínio social. Após sua extração essas informações são submetidas a um algoritmo de inferência de confiança, *TidalTrust* (Golbeck, 2005) – outros algoritmos para inferência de confiança podem ser utilizados. O modelo de extração de informações apresentado no *SUNNY* mapeia algumas informações estatísticas disponíveis no domínio e cria um grafo de probabilidades para que possa ser efetuada a análise de confiança. Para utilizar este mapeamento é necessário que o domínio social disponibilize mecanismos para publicação de conteúdos, além de mecanismos para avaliação dos conteúdos publicados.

Para realizar o mapeamento estatístico o algoritmo considera decisões (e.g. avaliações) realizadas pelos usuários em relação aos conteúdos em comum. O processo consiste em analisar o quanto as decisões de dois usuários são similares em relação a um mesmo conteúdo. Para realizar a análise de similaridade o algoritmo utiliza quatro variáveis, que são:

- **Diferença Total** (*Overall Difference*  $\Theta(u, v)$ ) – esta variável é dada pela média da diferença absoluta de todas as decisões realizadas pelos usuários a conteúdos em comum. Isto é, sejam  $c_i \in C$  os conteúdos aos quais dois nós  $u$  e  $v$  tomaram decisões, onde  $D(x, c_i)$  é o valor da decisão do usuário  $x$  em relação ao documento  $c_i$ , assim a Diferença Total das decisões dos usuários  $u$  e  $v$ , é dada pela Equação (3.3).

$$\Theta(u, v) = \frac{1}{|C|} \sum_{\forall c_i \in C} |D(u, c_i) - D(v, c_i)| \quad (3.3)$$

- **Diferenças Extremas** (*Difference on Extremes*  $\chi(u, v)$ ) – esta variável é dada pela média da diferença absoluta das decisões “extremamente diferentes” realizadas pelos nós  $u$  e  $v$  a conteúdos em comum. Uma decisão é considerada extremamente diferente quando ela é maior que um desvio padrão de  $\theta(u, v)$  – a Equação (3.4) mostra a condição para uma decisão ser considerada extremamente diferente. Assim, para obter-se o valor de  $\chi(u, v)$  pode-se utilizar a Equação (3.3), porém não aplicada ao conjunto  $C$ , mas sim a um  $C'$  de conteúdos que possuem decisões extremamente diferentes entre  $u$  e  $v$ .

$$|D(u, c_i) - D(v, c_i)| > \Theta(u, v) + s(\Theta(u, v))$$

$$s(\Theta(u, v)) = \sqrt{\frac{1}{|C|} \sum_{\forall c_i \in C} [D(u, c_i) - D(v, c_i)]^2} \quad (3.4)$$

- **Diferença Máxima** (*Maximum Difference*  $\nabla(u, v)$ ) – esta variável representa a maior diferença de opiniões entre os nós  $u$  e  $v$ , isto é,  $\nabla(u, v) = \text{Max}(\{|D(u, c_i) - D(v, c_i)| \mid c_i \in C\})$ .
- **Coefficiente de confiabilidade** (*Belief or Disbelief Coefficient*  $\sigma(u, v)$ ) – esta variável indica se o resultado da análise de confiança é um valor que representa crença ou descrença, isto é, se o nó  $u$  deve acreditar ou desconfiar do nó  $v$ . O coeficiente  $\sigma(u, v)$  assume valor 1 quando  $\Theta(u, v)$  (*Diferença Total*) for menor ou inferior que um desvio padrão da média geral das Diferenças Totais  $\Theta$ . Caso o valor de  $\Theta(u, v)$  seja maior ou superior que um desvio padrão da média de  $\Theta$ , tem-se  $\sigma(u, v) = -1$ . Caso  $\Theta(u, v)$  pertença ao intervalo  $[\Theta - s(\Theta), \Theta + s(\Theta)]$  então será necessária mais uma análise para se obter o valor deste coeficiente. Nesta análise  $\sigma(u, v) = \rho$ , onde  $\rho$  é o Coeficiente de Correlação de Pearson aplicado às decisões realizadas pelos nós.

A partir do valor destas quatro variáveis é gerado o mapeamento da rede como um grafo de probabilidades. Neste grafo a probabilidade de um nó confiar na informação fornecida por um nó adjacente é dada pela relação apresentada na Equação (3.5). Caso  $\sigma(u, v) < 0$ ,  $P(u, v)$  deve ser interpretado como um valor que representa o quanto o nó  $u$  deve desconfiar de

uma informação proveniente do nó  $v$ . Caso contrário, o valor de  $P(u,v)$  representa a probabilidade do nó  $u$  confiar em uma informação proveniente de  $v$ .

$$P(u, v) = \begin{cases} \sigma(u, v) \times |1 - 2(w_1\Theta(u, v) + w_2\nabla(u, v) + w_3\chi(u, v))|, & \text{se } \chi(u, v) \text{ existe} \\ \sigma(u, v) \times |1 - 2(w_4\Theta(u, v) + (1 - w_4)\nabla(u, v))|, & \text{caso contrário} \end{cases} \quad (3.5)$$

Após realizar o mapeamento estatístico do domínio social, o algoritmo *SUNNY* utiliza uma técnica de amostragem probabilística para selecionar os nós que serão considerados na análise de confiança. Nesta técnica, com o objetivo de avaliar cada nó da rede, o algoritmo realiza várias simulações sobre o grafo de probabilidades. Durante as simulações, o algoritmo estima limites superiores e inferiores para os valores de  $P(u,v)$ , a fim de determinar a participação, ou não, do nó na análise de confiança. Uma vez estimados os limites para um nó  $n$  são escolhidos  $k$  números aleatórios no intervalo  $[0,1]$  - ( $P(u,v)$  pertence a  $[0,1]$ ). A quantidade de vezes que os números aleatórios forem menores que os valores dos limites determina se um nó será, ou não, considerado na análise de confiança.

### 3.4.1 Inferência de confiança

O algoritmo *SUNNY* utiliza por padrão o algoritmo *TidalTrust* (Golbeck, 2005) para realizar a inferência de confiança – outros algoritmos também podem ser utilizados. O objetivo do algoritmo *SUNNY* é selecionar os melhores nós da rede para utilizá-los na análise de confiança.

O algoritmo *TidalTrust*, como *SUNNY*, realiza uma seleção de nós a serem utilizados na análise de confiança. Nesta seleção são estabelecidos o tamanho máximo dos caminhos e os valores mínimos de confianças entre os nós que poderão ser explorados durante a análise de confiança. Em seguida, com estes valores é possível excluir alguns nós que, segundo Golbeck (2005), reduziriam a precisão da análise de confiança.

Após a seleção dos nós a serem explorados na análise de confiança, o algoritmo *TidalTrust* calcula o valor de confiança entre dois nós como uma média ponderada das informações de confiança extraídas destes nós. A Equação (3.6) mostra como os valores de confiança são calculados, onde  $t_{i,j} = P(i, j)$  quando  $i$  for adjacente a  $j$ .



$$t_{u,v} = \frac{\sum_{j \in \text{adj}(i), t_{i,j} \geq \max} \alpha_j t_{u,j} t_{j,v}}{\sum_{j \in \text{adj}(i), t_{i,j} \geq \max} t_{u,j}} \quad (3.6)$$

A Equação (3.6) representa o cálculo da confiança de um nó  $u$  em relação a  $v$ . Para realizar o cálculo são utilizados os nós intermediários  $j$  tais que o caminho de  $u$  a  $v$  seja menor que o caminho máximo pré-definido, e os valores de confiança entre nós adjacentes sejam maiores que o valor mínimo estabelecido ( $\max$ ). Ainda na Equação (3.6)  $\alpha$  representa o peso da relação de  $u$  a  $v$  por um determinado caminho – o peso pode variar de acordo com o domínio da análise de confiança. Por exemplo, um algoritmo derivado de *TidalTrust* é o *FilmTrust*, utilizado para criar recomendações de filmes, neste caso o peso de cada relação é dado pela valor de avaliação de um nó ao filme em questão.

### 3.4.2 Algoritmos semelhantes

*SUNNY* é um algoritmo que realiza uma seleção de informações que são utilizadas como entrada para efetuar a análise de confiança, realizada pelo algoritmo *TidalTrust*. O algoritmo *TidalTrust* explora o grafo de confiança, gerado pelo *SUNNY*, a fim de inferir os valores de confiança entre usuários, utilizando as informações contidas nos menores caminhos que ligam estes usuários em uma rede social Web. Além do algoritmo *SUNNY*, existem outros algoritmos que exploram características de redes sociais.

O algoritmo apresentado por Katz (2006) infere valores de confiança, derivados de uma rede social Web, e os utiliza para classificar prioritariamente informações. Esta classificação ocorre de maneira a refletir os níveis de confiança dos usuários nos autores das informações. O algoritmo de Katz é aliado a um mecanismo de *ranking* (algoritmo de classificação de informações relevantes) de informações com o objetivo a melhorar o processo de classificação de conteúdos para os usuários. A inferência de confiança efetuada pelo algoritmo de Katz utiliza como entrada as informações de confiança nos caminhos que conectam usuários em uma rede social Web, e retorna valores personalizados (calculados especificamente para cada caso) na mesma escala de valores utilizada pelos usuários para descrever seus relacionamentos.

Outros algoritmos que utilizam modelos de confiança específicos para redes sociais Web são os apresentados por Golbeck (2006). Estes algoritmos utilizam as atribuições explícitas de confiança entre os usuários de uma rede social Web como base para a inferência de confiança. Para o cálculo de confiança são consideradas as informações dos caminhos mais confiáveis da rede social, isto é, caminhos onde os valores de confiança entre os nós são mais altos.

### **3.5 Algoritmo para cálculo de reputação em redes sociais**

O algoritmo de cálculo de reputação em redes sociais proposto por Mamami (2011) cria uma metodologia de análise de confiança baseada nas interações dos usuários de uma rede social Web. Para isso, o algoritmo avalia a quantidade e a qualidade das interações dos usuários na rede. Para realizar esta avaliação o algoritmo utiliza uma modificação do algoritmo *PageRank* (Brin, 1998) utilizado pelo Google.

As informações quantitativas e qualitativas utilizadas pelo algoritmo são obtidas a partir dos mecanismos de interação disponíveis na rede social Web, que podem ser contribuições ou avaliações. A quantidade de contribuições pertinente a cada usuário e a qualidade destas contribuições, medida através das avaliações, serão parâmetros utilizados para o cálculo de reputação. Neste sentido, o domínio social ao qual será aplicado o algoritmo deve possuir mecanismos que permitam aos usuários compartilhar conteúdos, bem como avaliar os conteúdos compartilhados por outros usuários.

Além disso, outra fonte de informação qualitativa é a rede de confiança, na qual os usuários podem atribuir valores de confiança a seus contatos na rede social Web. Os usuários podem atribuir valores de confiança no intervalo  $[-1, 1]$ , que podem significar crença (valores positivos) ou descrença (negativos) entre os usuários. Os valores de confiança devem ser atribuídos de acordo com uma premissa geral “Compartilhar informação de qualidade”. Deste modo, o domínio social também deve possuir um mecanismo que permita a avaliação dos relacionamentos da rede.

#### **3.5.1 Inferência de confiança**

Como supracitado, este algoritmo de cálculo de reputação em redes sociais utiliza uma modificação do algoritmo *PageRank* para realizar a análise de confiança. Esta modificação permite que o algoritmo utilize a informação anteriormente apresentada (interações e rede de confian-

ça). A Equação (3.7) mostra a forma modificada da função  $PR^i(A)$  utilizada no algoritmo *PageRank*.

$$R^i(x) = (1 - d) + \alpha d \left( \text{trust}^{i-1}(x) \right) + \beta d \left( \text{rate}^{i-1}(x) \right) + \gamma d \left( \text{contribution}^{i-1}(x) \right) \quad (3.7)$$

Na Equação (3.7),  $R^i(x)$  é a reputação do usuário  $x$  no tempo  $i$ , além disso, as funções  $\text{trust}^{i-1}(x)$ ,  $\text{rate}^{i-1}(x)$  e  $\text{contribution}^{i-1}(x)$  são, respectivamente, os valores que derivam da rede de confiança, avaliações dos conteúdos compartilhados e quantidade de contribuições do usuário  $x$ . Além destas funções, na Equação (3.7) são apresentados alguns pesos que são fatores moderadores da equação, e devem variar de acordo com o domínio ao qual será aplicado o algoritmo, estes fatores são  $d$ ,  $\alpha$ ,  $\beta$  e  $\gamma$ . Todos estes valores devem estar no intervalo  $[0,1]$ , sendo que  $\alpha + \beta + \gamma = 1$ .

A função  $\text{trust}^{i-1}(x)$  realiza uma análise da rede de confiança – esta rede de confiança considera os valores de confiança recebidos pelo usuário  $x$  de seus relacionamentos diretos (contatos). A função  $\text{trust}$  é uma soma ponderada de valores de confiança recebidos pelo usuário  $x$  em que os pesos são a reputação do usuário que forneceu o valor de confiança. A função  $\text{trust}^{i-1}(x)$  é definida na Equação (3.8), onde  $y_1, \dots, y_n$  são os contatos de  $x$  e  $C(y_j \rightarrow x)$  é o valor proveniente das avaliações realizadas por  $y_j$  aos conteúdos compartilhados por  $x$ .

$$\text{trust}^{i-1}(x) = R^{i-1}(y_1) * C(y_1 \rightarrow x) + \dots + R^{i-1}(y_n) * C(y_n \rightarrow x) \quad (3.8)$$

A função  $\text{rate}^{i-1}(x)$  representa a reputação geral do usuário  $x$  na rede social, isto é, o quanto os usuários da rede, amigos ou não, confiam em  $x$ . Para isso, a função  $\text{rate}$  considera informações das avaliações realizadas por qualquer usuário da rede social, representados por  $z_1, \dots, z_n$  na Equação (3.9), ponderados pela sua reputação  $R(z_j)$ . A Equação (3.9) apresenta a função  $\text{rate}^{i-1}(x)$ .

$$\text{rate}^{i-1}(x) = R^{i-1}(z_1) * C(z_1 \rightarrow x) + \dots + R^{i-1}(z_n) * C(z_n \rightarrow x) \quad (3.9)$$

A função  $\text{contribution}^{i-1}(x)$  analisa a quantidade de contribuições do usuário  $x$ , *i. e.* quantidade de conteúdos compartilhados. A Equação (3.10) mostra a função  $\text{contribution}^{i-1}(x)$ . Nesta equação, é utilizada a função  $C'(x)$  onde é calculado um valor para as contribuições de  $x$ . Uma vez obtido o valor de  $C'(x)$  a função  $\text{contribution}^{i-1}(x)$  é definida como o produto de  $C'(x)$  e  $R^{i-1}(x)$ . A função  $C'(x)$  é a razão da quantidade de contribuições do usuário  $x$  com a quantidade total de contribuições na rede social.

$$contribution^{i-1}(x) = R^{i-1}(x) * C'(x) \quad (3.10)$$

Para realizar a análise de confiança o algoritmo é executado iterativamente até que se obtenha uma situação aceitável dos valores de confiança. Para que os valores de confiança sejam considerados aceitáveis os autores sugerem que o algoritmo seja executado até que  $R^k(x) - R^{k-1}(x)$ , para todo  $x$ , seja menor que um valor máximo estipulado para a rede social. Por exemplo, a diferença entre os valores  $R^i$  em duas iterações diferentes for menor que um valor  $k$  adotado como limite.

### 3.6 Considerações Finais

Para uma compreensão do conceito de confiança, um dos fenômenos que sustenta a sociedade, e como este vem e pode ser explorado em ambientes computacionais, o presente capítulo apresentou a definição de confiança, suas propriedades, e alguns trabalhos que exploram este conceito.

Conforme discutido no decorrer do capítulo, confiança é explorada nas mais variadas áreas, tais como psicologia, sociologia, filosofia e computação. A ênfase do presente trabalho é a computação, por isso, foram apresentados alguns algoritmos para inferência de confiança disponíveis na literatura, são eles: *SUNNY* (Kuter; 2007, 2010); *FlowTrust* (Wang, 2011); e o algoritmo apresentado em Mamami (2011). Além destes algoritmos, nas seções correspondentes, foram apresentados também alguns projetos similares. Existem também outros algoritmos disponíveis na literatura, a escolha destes três foi em decorrência de possuírem características que podem ser exploradas em ambientes sociais na internet, mais especificamente nas Redes Sociais Web.

O algoritmo *SUNNY* é uma evolução do algoritmo *TidalTrust* (Golbeck, 2005). O algoritmo *TidalTrust* infere a probabilidade de um usuário confiar em outro utilizando a estrutura dos relacionamentos proveniente das redes sociais. *TidalTrust* foi precursor de uma série de trabalhos que exploram a estrutura de redes, porém recebe diversas críticas pelos pesquisadores pela perda de informações, pois realiza uma rotina de seleção de fontes de informação (*i. e. nós*) na rede, antes da análise de confiança. Neste sentido, foi apresentado um algoritmo que prevê a utilização completa de todas as informações da rede, o algoritmo *FlowTrust*. Este algoritmo, além de não perder informações, aplica o conceito de multidimensionalidade, isto é, explora mais de um elemento para inferir os valores de confiança. No entanto, as informações exploradas por *FlowTrust* estão todas relacionadas ao mesmo conceito, informações dos

relacionamentos da rede, e ainda são necessárias informações de outras fontes (*e. g.* informações sobre conteúdos e informações contextuais) para concluir se um usuário é realmente confiável. Um algoritmo que explora informações dos documentos e conteúdos publicados pelos usuários em um ambiente colaborativo é o apresentado por Mamami (2011). Este algoritmo considera informações sobre a quantidade e qualidade dos conteúdos postados pelos usuários do domínio social.

O estudo destes algoritmos foi primordial para que houvesse um entendimento claro de como os trabalhos disponíveis na literatura exploram o referido conceito e, assim, como consequência, a obtenção de parâmetros necessários para que seja proposto um novo algoritmo que infira confiança em redes sociais Web. O Quadro 3.1 apresenta de forma resumida as principais características observadas nos algoritmos apresentados.

Quadro 3.1: Principais características dos algoritmos de análise de confiança

Algoritmo	Critérios de Avaliação	Conceitos Utilizados
<b>FlowTrust (2011)</b>	- Grafo de Confiança <sup>9</sup>	- Teoria do Fluxo - Multidimensionalidade
<b>RelTrust (2005)</b>	- Falhas dos nós da rede.	- Teoria da Confiabilidade
<b>RN-Trust (2008)</b>	- Grafo de Confiança <sup>3</sup>	- Resistividade Elétrica
<b>SUNNY (2007, 2010)</b>	- Avaliações dos usuários a conteúdos	- Modelos Probabilísticos
<b>TidalTrust (2005)</b>	- Grafo de Confiança <sup>3</sup> - Atribuições explícitas de confiança entre usuários	- Análises estatísticas e de correlação.
<b>Golbeck (2006)</b>	- Atribuições explícitas de confi-	- Lógica Booleana

<sup>9</sup> Grafo de Confiança refere-se a informações de confiança que podem ser extraídas do domínio de qualquer forma, até mesmo com a utilização de outros algoritmos específicos.

	ança entre usuários	- Análises estatísticas
<b>Mamami (2011)</b>	- Avaliações dos conteúdos publicados pelos usuários - Quantidade de conteúdos publicados	- PageRank(1998) - Multidimensionalidade;

De acordo com o Quadro 3.1, três dos algoritmos (FlowTrust, CircuitTrust e TidalTrust) baseiam-se em Grafo de Confiança, em que são explorados explicitamente os relacionamentos entre nós para inferência de confiança – essencialmente em ambientes baseados em redes sociais Web. A diferença entre eles é que utilizam conceitos diferentes: *FlowTrust* que é baseado na Teoria do Fluxo, que considera todos os caminhos da rede para determinar o melhor caminho para inferência de confiança; *CircuitTrust* que é baseado na Teoria da Resistividade, que desconsidera alguns caminhos da rede de acordo com determinados critérios; e o *TidalTrust*, que busca a confiança nos menores caminhos entre dois nós. Além desses, o *SUNNY* e o *Mamami* consideram as avaliações dos usuários dadas a conteúdos específicos para inferência de confiança.

Há outros critérios e conceitos utilizados por estes algoritmos, conforme supracitado e apresentado no Quadro 3.1, que indica que alguns deles já são mais aceitos na literatura e que provavelmente serão utilizados no algoritmo que será proposto a posteriori. Além desses, a partir do estudo realizado nos mais diferentes trabalhos voltados para a inferência de confiança, alguns outros elementos que podem ser importantes para tornar mais eficaz a inferência de confiança dos usuários em uma rede e que não foram utilizados pelos algoritmos apresentados, podem ser: a similaridade entre as ações dos usuários na rede e o nível de disposição de um usuário em colaborar. Estes elementos ou a combinação de um conjunto de elementos, inclusive, dos elementos já utilizados em outros trabalhos, podem ser utilizados para a proposta de um novo algoritmo que infira confiança em redes sociais Web – um dos propósitos deste trabalho.

## 4 T-SWEETS

Este capítulo apresenta uma nova abordagem para inferência de confiança de relacionamentos em redes sociais, intitulada T-SWEETS. A motivação para o nome surgiu da junção do conceito de *Trust*; da teoria de redes sociais, os 6 graus de separação, em que todas as pessoas estão a uma distância social 6 uma da outra (Smallworld<sup>10</sup>; e da noção de identificação de especialistas (*Experts*), já que a motivação central da presente tese é aplicar o conceito de confiança como apoio aos sistemas de recomendação de especialistas. Assim, tem-se *Trust SmallWorldExpErtS* – T-SWEETS, ou um “açúcar” sobre o processo.

Este capítulo está organizado da seguinte forma: a Seção 4.2 apresenta quais são os pré-requisitos para a implantação do T-SWEETS em um ambiente virtual; a Seção 4.3 apresenta os 4 elementos que foram explorados no T-SWEETS para a inferência de confiança, além de como cada um destes elementos estão relacionados aos trabalhos correlatos apresentados no capítulo 3; e, a Seção 4.5 apresenta como estes 4 elementos são ponderados para a realização da inferência de confiança entre pessoas; e, por fim, a Seção 4.5 apresenta as considerações finais do capítulo.

### 4.1 Introdução

Confiança é um fenômeno comum que está presente na sociedade desde que ela existe (Yamamoto, 1990). Este fenômeno, naturalmente, fornece subsídios para apoiar a tomada de decisões. Ambientes virtuais baseados em redes sociais representam a estrutura básica da socie-

---

<sup>10</sup> Small-world network - [http://en.wikipedia.org/wiki/Small-world\\_network](http://en.wikipedia.org/wiki/Small-world_network)

dade, através das pessoas e seus relacionamentos. Assim, são ambientes propícios para a exploração de análise de confiança, pois possuem indicadores que viabilizam esta análise.

No capítulo 3, foram apresentados os resultados dos estudos em diferentes algoritmos de inferência de confiança disponíveis na literatura. A realização deste estudo teve como objetivo principal entender como inferência de confiança pode ser explorada no contexto de redes sociais. A partir deste estudo, foi possível identificar algumas características específicas e como estas podem ser exploradas no processo de inferência de confiança nestes ambientes virtuais sociais.

A principal motivação para o desenvolvimento de uma abordagem que identifique e recomende relacionamentos de confiança é promover uma maior interação entre os indivíduos de uma rede social. Assim, há naturalmente uma maior produção de conhecimento por estes indivíduos. Um fator que pode contribuir pra que as pessoas interajam mais entre elas é a necessidade que estas podem ter de contatar uma pessoa em particular, por exemplo, para encontrar solução para um determinado problema.

Para entender os fatores que podem contribuir quando uma pessoa decide colaborar com outra ou não, foi aplicado um questionário com um grupo de 53 pessoas. Os resultados desta entrevista forneceram sustentação para a definição dos elementos que seriam adotados na abordagem proposta no T-SWEETS. Algumas constatações foram:

- aproximadamente 56.6% dos entrevistados costumam ajudar as pessoas que possuem afinidade ou alguma relação de confiança;
- aproximadamente 77% das pessoas concordaram que a reputação dos indivíduos é um fator decisivo quando uma pessoa decide interagir ou não com outra;
- o nível de conhecimento das pessoas sobre assuntos específicos foi outro aspecto de forte influência considerado por aproximadamente 90% dos entrevistados;
- e, por fim, aproximadamente 92% dos entrevistados tendem a colaborar com pessoas que possuem conhecimentos e opiniões similares aos deles.

O resultado desta pesquisa é apresentado com mais detalhes no Capítulo 7, Seção 7.1. Estes dados estatísticos e a ausência de um algoritmo na literatura para inferência de relacionamentos ocultos de confiança que adote estes critérios em conjunto forneceram sustentação para a definição do algoritmo T-SWEETS.



## 4.2 Pré-requisitos para utilização do T-SWEETS

T-SWEETS é um algoritmo para inferência de confiança específico para ambientes virtuais baseados na teoria de redes sociais e explora os relacionamentos existentes entre as pessoas e o conhecimento produzido por estas pessoas no ambiente virtual social.

Os relacionamentos de uma rede social são representados no formato de um grafo, e possuem características que podem ser exploradas para realizar o processo de inferência de confiança. Neste sentido, um relacionamento representa a forma como duas pessoas interagem entre si, que pode auxiliar, por exemplo, na identificação do quanto este relacionamento é ou não positivo. Além disso, estes relacionamentos podem receber rótulos específicos, que determinam a natureza dos relacionamentos, por exemplo, se representam a relação entre pai e filho, chefe e empregado, amigo (a), ou namorado(a), ou, ainda mais, podem haver quantificadores que mostrem o quão admirável ou confiável uma pessoa é para outra. Estas características compõem fontes de informações que podem ser potencialmente exploradas para inferência de confiança em ambientes sociais.

O conhecimento produzido pelos usuários é natural e proveniente dos ambientes sociais virtuais, já que a estrutura destes ambientes é propícia à colaboração, isto é, um ambiente propenso à produção de conhecimento, colaboração e cooperação. O conhecimento produzido pelos usuários torna-se outra rica fonte para o auxílio no processo de inferência de confiança.

A partir das características supracitadas, o algoritmo T-SWEETS proposto e desenvolvido neste trabalho explora os relacionamentos diretos e indiretos entre as pessoas; o agrupamento entre pessoas; e pessoas que compartilham (possuem) opiniões e conhecimentos similares. Assim, a Figura 4.1, apresenta a estrutura exigida do ambiente social em que o T-SWEETS pode ser implantado – neste ambiente deve ser possível a produção e/ou compartilhamento de conhecimento entre os usuários.

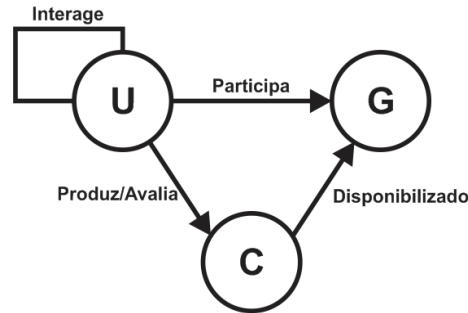


Figura 4.1: Estrutura do ambiente para implantação do algoritmo

Na Figura 4.1, o conjunto de usuários U realiza três funções principais, a interação com outros usuários; a produção de conhecimentos C; e a participação em grupos G. Além disso, há também a possibilidade de compartilhar conteúdos / conhecimentos nos grupos. Para que o T-SWEETS possa realizar a inferência de confiança, o ambiente em que este algoritmo será implantado deve estar adequado à estrutura apresentada na Figura 4.1.

### 4.3 Elementos para Inferência de Confiança

T-SWEETS é um algoritmo de inferência de confiança para redes sociais que explora: os relacionamentos diretos e indiretos entre as pessoas; o agrupamento entre pessoas; pessoas que compartilham (possuem) opiniões e conhecimentos similares, além da qualidade do conteúdo produzido pelas pessoas que compõem a rede.

A definição dos elementos explorados no T-SWEETS foi baseada nas hipóteses apresentadas a seguir:

- H1: pessoas relacionadas ao mesmo contexto tendem a possuir maior nível de confiança uns nos outros;
- H2: pessoas tendem a confiar em usuários que seus amigos de confiança confiam;
- H3: usuários que produzem conteúdos relevantes são mais confiáveis;
- H4: usuários que possuem boas reputações são confiáveis

Assim, os seguintes elementos são explorados no algoritmo T-SWEETS: similaridade entre perfis, relacionamentos de confiança, nível de maturidade e reputação:

- **similaridade entre perfis:** este elemento possui relação direta com H1, uma vez que usuários relacionados ao mesmo contexto (e.g. grupo ou comunidade), isto é, que es-

tão no mesmo ambiente, tendem a ser mais confiáveis. É medida a partir da avaliação de semelhanças entre dois usuários – esta semelhança é determinada por informações comuns relacionadas aos usuários (*e. g. tags* e conteúdos), baseando-se no contexto no qual estão inseridos (*e. g. um grupo de pessoas ou comunidade*).

- **relacionamento de confiança:** na sociedade, é natural que as pessoas procurem por ajuda entre seus amigos (conforme H2). Entretanto, ocorrem situações em que os amigos não podem ajudar. Nestes casos, há a possibilidade de indicação de alguma pessoa de confiança que possa ajudar. Em ambientes virtuais sociais, o relacionamento de confiança pode ser representado, uma vez que há a possibilidade de as pessoas indicarem pessoas de confiança para outros;
- **nível de maturidade:** usuários que produzem conhecimentos relevantes às pessoas que os rodeiam e aos grupos que participam, tendem a ser mais confiáveis que os usuários que não produzem conhecimento de qualidade (de acordo com H3). Para identificar o quão relevantes são os conhecimentos produzidos pelos usuários, podem ser analisadas as avaliações de outros usuários sobre estes conhecimentos. Assim, mede-se a maturidade de conhecimento dos usuários. Quanto maior o nível de maturidade dos conhecimentos produzidos pelos usuários, mais confiáveis são suas produções e, conseqüentemente, mais confiável é o autor;
- **reputação:** as pessoas tendem a possuir maior confiança em pessoas que possuem boa reputação. Da mesma forma, tendem a possuir menor confiança em pessoas que possuem reputação ruim (H4). A reputação no T-SWEETS é inferida a partir das atribuições de confiança que os usuários recebem dos seus relacionamentos.

As seções 4.3.1, 4.3.2, 4.3.3 e 4.3.4 apresentam como os elementos supracitados são calculados no T-SWEETS.

#### 4.3.1 Similaridade entre Perfis

O cálculo da Similaridade entre Perfis das pessoas é importante para a identificação de pessoas que possuem perfis de conhecimento e opinião similares. Nesta seção, são apresentados os quesitos que são considerados para esta análise de semelhança, além da exemplificação de como o referido cálculo é realizado.

Conforme supracitado, para calcular este grau de semelhança, são considerados dois quesitos: (1) perfis de conhecimento do usuário, através de suas produções na rede social (*e.*

g. produções de conteúdos); e (2) perfis de opinião dos usuários, através das avaliações (e.g. gosta ou não gosta) dadas pelos usuários aos conteúdos publicados na rede social. Os usuários com perfis similares possuem maior afinidade, logo, estão mais propensos a interagirem entre eles.

**Definição: Similaridade de Perfis** é a semelhança entre os perfis de opinião e conhecimento dos usuários *a* e *b*.

O processo de análise de semelhança dos conhecimentos e das opiniões dos usuários é apresentado na Figura 4.2.

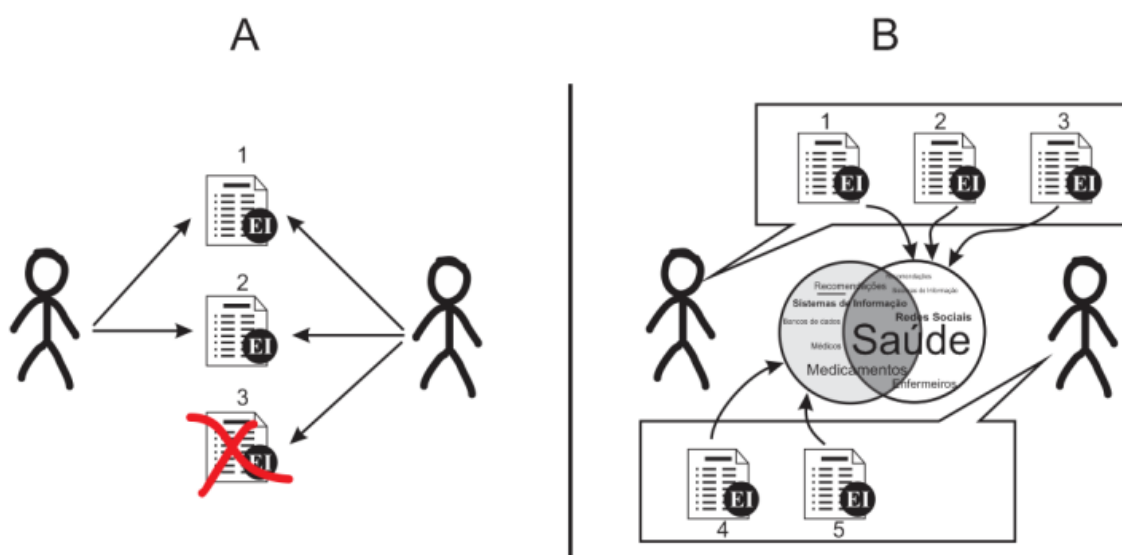


Figura 4.2: Processo de análise da semelhança entre opiniões e conhecimentos dos usuários

Na Figura 4.2A é apresentado o processo de análise de correlação das avaliações dos usuários. Este processo é necessário, pois usuários que avaliam conteúdos de forma semelhante possuem perfis de opiniões semelhantes. Durante este processo são consideradas as notas atribuídas a todos os conteúdos que receberam avaliações dos dois usuários alvos da análise. A Figura 4.2B representa a comparação de conteúdos de autoria dos usuários analisados. Este processo ocorre porque usuários que produzem conteúdos semelhantes, ou relacionados aos mesmos temas, tendem a possuir conhecimentos similares.

A semelhança entre os perfis de opiniões de dois usuários *a* e *b* é dada por:

$$\rho E(a, b) = \frac{\sum_{i \in E(a) \cap E(b)} e(a, i) * e(b, i)}{\sqrt{\sum_{i \in E(a)} e(a, i) * e(a, i)} * \sqrt{\sum_{i \in E(b)} e(b, i) * e(b, i)}} \quad (4.1)$$

Onde:

$a, b$ : usuários de uma rede social

$\rho E(a, b)$ : similaridade entre as avaliações comuns por dois usuários  $a$  e  $b$ ;

$E(a), E(b)$ : conjunto dos conteúdos avaliados pelos usuários  $a$  e  $b$ ;

$i$ : conteúdos publicados na rede social;

$e(a, i), e(b, i)$ : avaliação do usuário  $a$  ou  $b$  a um conteúdo  $i$ .

Para a medida de similaridade entre os perfis de opinião dos usuários  $a$  e  $b$ , é utilizado o modelo algébrico intitulado *Modelo Espaço Vetorial* (Baeza-Yates, 1999). Assim, as avaliações realizadas pelos usuários são representadas em um espaço vetorial. Após esta representação, é necessário utilizar um algoritmo de análise de similaridade para medir a semelhança entre os perfis. No T-SWEETS foi adotada a medida *Coseno*. Outras medidas poderiam ser adotadas, por exemplo, o índice *Jaccard* (Baeza-Yates, 1999).

A Equação (4.1) deriva da equação de produto escalar em espaços Euclidianos (Baeza-Yates, 1999). O valor de  $\rho E(a, b)$  está em uma escala de 0 e 1, sendo que quanto mais próximo de 1, mais semelhantes são as avaliações dos usuários, e quanto mais próximo de 0, menos semelhantes são as avaliações comuns entre  $a$  e  $b$  ( $E(a) \cap E(b)$ ).

Para exemplificar o cálculo da semelhança dos perfis de opinião de dois usuários deve ser considerada a situação hipotética apresentada na Figura 4.3.

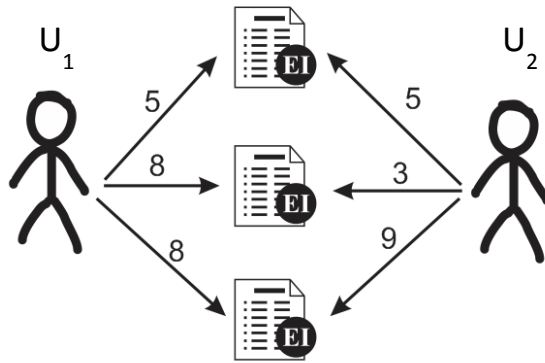


Figura 4.3: Cenário para exemplificar o cálculo de semelhança dos perfis de opinião de dois usuários.

No cenário apresentado na Figura 4.3, dois  $U_1$  e  $U_2$  usuários avaliaram 3 conteúdos em comum. As notas das avaliações dadas pelos usuários aos conteúdos são também apresentadas na Figura 4.3. Aplicando a Equação (4.1) neste cenário, tem-se que  $\rho E(a, b)$ , a similaridade dos perfis de opinião destes dois usuários, é 0,91.

A Equação (4.2) apresenta o formalismo para cálculo da similaridade entre os perfis de conhecimento de um usuário  $a$  em relação a um usuário  $b$ .

$$sK(a, b) = \frac{\sum_{\forall k \in K(a) \cap K(b)} f(k, a) * f(k, b)}{\sqrt{\sum_{\forall k \in K(a)} f(k, a)^2} * \sqrt{\sum_{\forall k \in K(b)} f(k, b)^2}} \quad (4.2)$$

Onde:

$sK(a, b)$ : similaridade entre os perfis de conhecimento dos usuários  $a$  e  $b$ ;

$k$ : um termo;

$K(a)$ : conjunto dos termos extraídos do conhecimento produzido pelo usuário  $a$ ;

$K(b)$ : conjunto dos termos extraídos do conhecimento produzido pelo usuário  $b$ ;

$f(k, a)$ : peso de um termo  $k$  relacionado a um usuário  $a$ ;

$f(k, b)$ : peso de um termo  $k$  relacionado a um usuário  $b$ ;

A Equação (4.2), assim como a Equação (4.1), utiliza a medida de similaridade *Coseno*. O grau de similaridade varia entre 0 e 1 e está diretamente relacionado à similaridade dos perfis dos usuários (conhecimentos similares), isto é, quanto mais próximo de 1 for o resultado, maior a semelhança entre os perfis dos usuário  $a$  e  $b$ , e quanto mais próximo de 0, menor será o grau de semelhança entre os perfis.

Para isso, devem ser extraídos os termos associados aos perfis de conhecimento dos usuários e representado em um *Modelo Espaço Vetorial*. Um termo pode ser, por exemplo, uma palavra de um documento texto, uma palavra-chave (*tag*) que descreve um determinado objeto (*e.g.* artigo, imagem ou planilha) ou uma área de conhecimento informada explicitamente pelo usuário. Assim, cada termo relacionado ao usuário é então contabilizado, a partir da frequência (que determina o peso) com que ocorre e representado em uma dimensão deste espaço vetorial.

O Quadro 4.1 apresenta um cenário de um conjunto de termos que representam os perfis de conhecimento de dois usuários  $U_1$  e  $U_2$ .

Quadro 4.1: Cenário para exemplificar o cálculo de similaridade entre os perfis de conhecimento de dois usuários  $U_1$  e  $U_2$ .

		Termos				
		trust	trustworthiness	social networks	trust modeling	uncertainty
Perfis	$U_1$	3	8	9	5	6
	$U_2$	2	9	7	5	0

A partir desta representação vetorial é possível aplicar a Equação (4.2). O resultado do grau de semelhança entre os perfis de conhecimento dos usuários, determinado por  $sK(a, b)$  é 0,9.

A média aritmética dos dois valores obtidos nas Equações (4.1) e (4.2) representa a semelhança das opiniões e conhecimento de dois usuários, Equação (4.3)

$$S(a, b) = \frac{\rho E(a, b) + sK(a, b)}{2} \quad (4.3)$$

De acordo com os cenários apresentados anteriormente e os respectivos graus obtidos 0,91 e 0,9 para  $\rho E(a, b)$  e  $sK(a, b)$ , é realizada uma média aritmética destes valores, que determina o grau final de similaridade dos perfis de dois usuários  $S(a, b)$ . Nesta exemplificação, tem-se que a similaridade entre os perfis dos usuários  $U_1$  e  $U_2$  é 0,905.

### Relação com Trabalhos Relacionados

O algoritmo SUNNY (2007) (2010) explora o conceito de similaridade entre usuários, a partir das avaliações realizadas pelos usuários na rede social. O T-SWEETS, além do perfil de opiniões (formado pelas avaliações), considera também a similaridade dos perfis de conhecimento, que são as produções dos usuários na rede social. Com esta adição, T-SWEETS tem mais parâmetros para a identificação da semelhança entre os perfis dos usuários.

#### 4.3.2 Relacionamento de Confiança

O elemento Relacionamento de Confiança considera as avaliações de confiança atribuídas explicitamente aos usuários ou o grau de confiança que pode ser inferido entre os membros de

uma rede social. Esta seção apresenta como estes relacionamentos são calculados, além de algumas exemplificações e a relação com os trabalhos relacionados.

**Definição: Relacionamento de Confiança** é a confiança de um usuário  $a$  em relação a um usuário  $b$  atribuída explicitamente por  $a$ ; ou a probabilidade de  $a$  confiar em  $b$  dado que existe um caminho de relacionamento de confiança (e. g.  $a \rightarrow b \rightarrow c$ ) entre eles.

No T-SWEETS, os relacionamentos de confiança exploram os conceitos de *Transitividade*, *Assimetria* e *Agregabilidade* nos casos em que é realizada inferência de confiança entre dois usuários.

A confiança entre as pessoas é um fenômeno que pode ser transitivo, já que as pessoas podem procurar informações a respeito das outras para obter parâmetros que possam ser utilizados no auxílio à identificação do fato de uma pessoa ser ou não confiável.

Contudo, confiança não é um fenômeno totalmente transitivo, isto é, nem toda confiança de uma pessoa pode ser transferida a outra – considerando sua rede de relacionamentos. Isso implica dizer que: se uma pessoa  $a$ , confia muito em uma pessoa  $b$ , e se  $b$ , por sua vez, confia muito em uma terceira pessoa  $c$  - formando assim o caminho de relacionamento  $a \rightarrow b \rightarrow c$  -, não quer dizer que  $a$  confia muito em  $c$ . Entretanto, podem ser utilizadas informações de confiança dos relacionamentos ( $a \rightarrow b \rightarrow c$ ) para inferir o quão a pessoa  $a$  eventualmente poderia confiar em  $c$ . Além disso, dizer que  $a$  confia muito em  $b$  não implica que  $b$  confie muito em  $a$ , neste caso a primeira informação ( $a \rightarrow b$ ) sequer fornece condições para que se obtenha a segunda ( $b \rightarrow a$ ) – por isso, há a *Assimetria* (Golbeck, 2005). A Figura 4.4 apresenta o funcionamento do conceito *Relacionamento de Confiança* em um grafo de confiança.



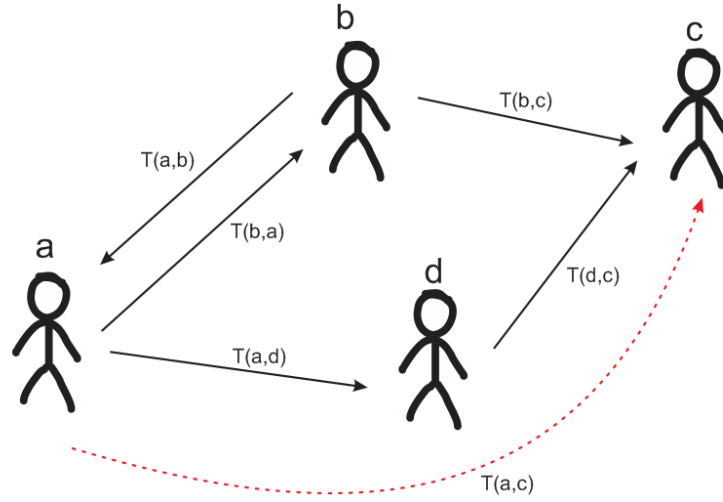


Figura 4.4: Grafo de Confiança

Na Figura 4.4, a linha pontilhada representa o valor de confiança  $T(a, c)$ , que seria o valor de confiança inferido do usuário  $a$  no usuário  $c$ . Além da *Transitividade e Assimetria*, o *Relacionamento de Confiança* também pode explorar o conceito de *Agregabilidade* (Golbeck, 2005), o qual define que se  $a$  confia em mais de uma pessoa que possui informações de confiança sobre  $c$ , então estas informações podem ser consideradas no processo de inferência de confiança. Assim, de acordo com a Figura 4.4, há duas informações de confiança adicionais que podem ser utilizadas no processo de confiança  $T(a, c)$ , que são, respectivamente, os valores de confiança atribuídos pelos usuários  $b$  e  $d$  ao usuário  $c$ .

Diante das condições previamente apresentadas para *Relacionamentos de Confiança*, este pode ser representado pela fórmula a seguir.

$$T(x, y) = \begin{cases} T(x, y), & \text{caso exista um valor atribuído diretamente} \\ \frac{1}{|P(x, y)|} \left\{ \sum_{\forall p \in P(x, y)} \left[ \prod_{\forall \text{adj}(u, v) \in p} T(u, v) \right] \right\}, & \text{em caso contrário} \end{cases} \quad (4.4)$$

Onde:

$x, y$ : usuários;

$P(x, y)$ : é o conjunto dos menores caminhos  $p$  existentes entre os usuário  $a$  e  $b$ ;

$p$ : um caminho no grafo de confiança contido no conjunto  $P(x, y)$ ;

$\text{adj}(u, v)$ : dois usuários,  $u$  e  $v$ , diretamente relacionados (adjacentes) no grafo de confiança;

$|P(x, y)|$ : o número de elementos do conjunto  $P(x, y)$ ;

$T(x, y), T(u, v)$ : grau de confiança de um usuário  $a$  em relação a  $b$  e de  $u$  em relação a  $v$

Na Equação (4.4), a fórmula para o valor de confiança pode variar de acordo com a seguinte situação: caso o usuário  $a$  já tenha atribuído um valor de confiança ao usuário  $b$ , então este será o valor de confiança de  $a$  a  $b$  ( $T(a, b)$ ), caso contrário é utilizada a fórmula apresentada na Equação (4.4), segunda parte. Nesta fórmula,  $P(a, b)$  é o conjunto de todos os caminhos  $p$  que podem ser estabelecidos entre os usuários  $a$  e  $b$  na representação do grafo da rede social, com  $|P(a, b)|$  sendo o número de elementos deste conjunto. Cada elemento deste conjunto é um par  $(u, v)$ , que representa cada aresta, e as informações de confiança relacionadas a ela, que relacionam os usuário  $u$  e  $v$  em um caminho  $p$ .

A priori, foi definida uma distância social 3 para maior caminho entre dois usuários. O número de caminhos foi reduzido para garantir que o algoritmo tenha resultados mais precisos, embora a cobertura seja diminuída. Esta abordagem para seleção de caminhos foi inspirada no algoritmo *TidalTrust* apresentado por Golbeck (2005).

Para exemplificar o cálculo do grau do relacionamento de confiança entre usuários é usado o cenário apresentado na Figura 4.5.

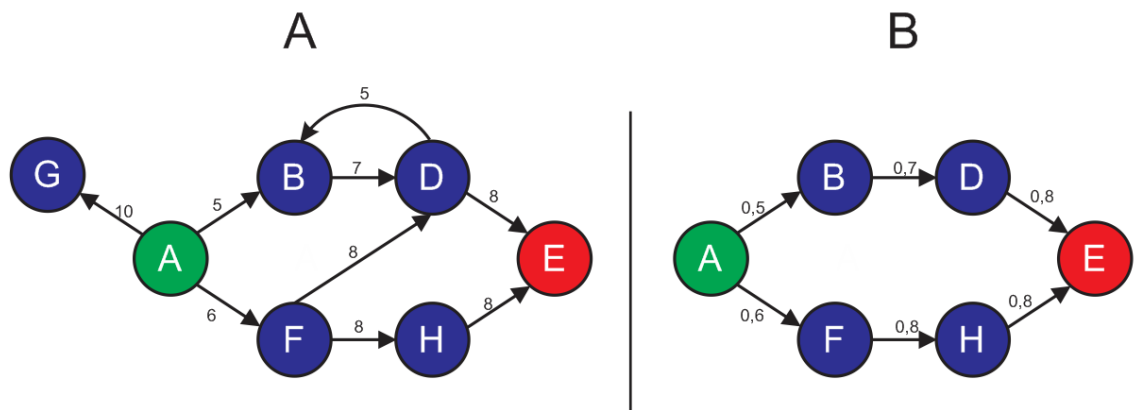


Figura 4.5: Grafo de relacionamentos de confiança: (A) grafo completo e (B) grafo com os menores caminhos.

O objetivo da presente exemplificação é calcular o grau de confiança entre os usuário A e E. A Figura 4.5 – parte A apresenta o grafo completo dos relacionamentos de confiança de uma rede social. Neste grafo, os nós correspondem aos usuários e o valor de cada aresta corresponde ao valor de confiança atribuído de um usuário em relação a outro.

Para que a inferência do grau de confiança entre os usuários A e E seja realizada, é necessário, inicialmente, selecionar os caminhos que ligam os usuários A e E no grafo (Figura 4.5 – parte A). São selecionados os menores caminhos entre A e E. Como resultado a esta seleção de menores caminhos entre A e E, tem-se a representação da Figura 4.5 – parte B.

Nesta exemplificação, os valores de confiança atribuídos explicitamente entre os usuários estão em uma escala [1,10] (Figura 4.5 – parte A). Estes valores devem ser normalizados para uma escala [0,1], conforme representação na Figura 4.5 – parte B. Após a normalização, o passo seguinte é a aplicação da Equação (4.4), o grau  $T(A,E)$  é 0,33.

### Relação com Trabalhos Relacionados

No TidalTrust, além dos menores caminhos, também é estabelecido um limiar de grau de confiança para decidir se um caminho deve ou não ser considerado. No T-SWEETS, este limiar não é adotado, pois acredita-se que a qualidade do algoritmo pode ser comprometida, uma vez que valores baixos de confiança seriam desconsiderados. Se os valores baixos de confiança não fizerem parte da análise, há uma tendência natural para o aumento do grau de confiança inferido entre dois usuários  $a$  e  $b$ .

A definição para inferência dos *Relacionamentos de Confiança* adotada no T-SWEETS é baseada na definição apresentada por (Golbeck, 2005), que confiança não é um fenômeno totalmente transitivo. Apesar desta definição, o algoritmo TidalTrust torna a transitividade de confiança totalmente transitiva em algumas situações. Já o formalismo do T-SWEETS resolve este problema. Para exemplificar esta situação será utilizado o cenário apresentado na Figura 4.5, em que o objetivo é inferir o grau de confiança ( $T(F,E)$ ) entre os usuários F e E. Assim, tem-se o detalhamento do cálculo em ambos os algoritmos no Quadro 4.2.

Quadro 4.2: Demonstração da problemática do algoritmo TidalTrust em relação ao conceito de Transitividade de Confiança.

Algoritmo	Inferência de confiança entre F e E
TidalTrust	$\frac{1}{1,6} (0.8 * 0.8 + 0.8 * 0.8) = 0,8$
T-SWEETS	$\frac{1}{2} (0.8 * 0.8 + 0.8 * 0.8) = 0,64$

Tanto o T-SWEETS quando o TidalTrust explora a transitividade de confiança baseada em um modelo probabilístico. Porém, existem outros algoritmos que exploram a transitivi-

dade de confiança baseando-se em teorias diferentes. Por exemplo, RN-Trust (2008) e Flow-Trust (2011), que baseiam-se teorias que avaliam o máximo de confiança que pode fluir (passar) através dos nós de uma rede. Mais detalhes podem ser encontrados na Seção 3.3.

### 4.3.3 Nível de Maturidade

O nível de maturidade que as pessoas possuem indica o quanto os conteúdos produzidos por estas são aceitos ou não (bem avaliados ou não) por outras. Esta seção apresenta como este grau de maturidade é calculado, uma exemplificação a partir da simulação de um ambiente com conteúdos avaliados, e a relação com os trabalhos relacionados.

**Definição:** *Nível de Maturidade* representa o quanto as produções dos usuários são vistas na rede social. Para isso, são consideradas todas as avaliações dadas aos conteúdos de um usuário  $x$  em relação às produções dos demais usuários  $y$ .

Um dos indicadores de confiabilidade dos usuários é a qualidade dos conteúdos por eles produzidos. Esta qualidade reflete a maturidade dos conhecimentos de seu autor, o que fornece parâmetros para medir o quão confiáveis são seus conhecimentos produzidos.

No T-SWEETS, a qualidade das produções dos autores é calculada a partir do conjunto de avaliações realizadas por outros usuários aos conteúdos dos respectivos autores. Conteúdos que recebem avaliações positivas devem possuir melhor qualidade, além disso, quanto maior o número de avaliações recebidas por um conteúdo maior será a precisão da avaliação da qualidade deste conteúdo. A Figura 4.6 mostra como é o processo que determina o nível de maturidade de um usuário.

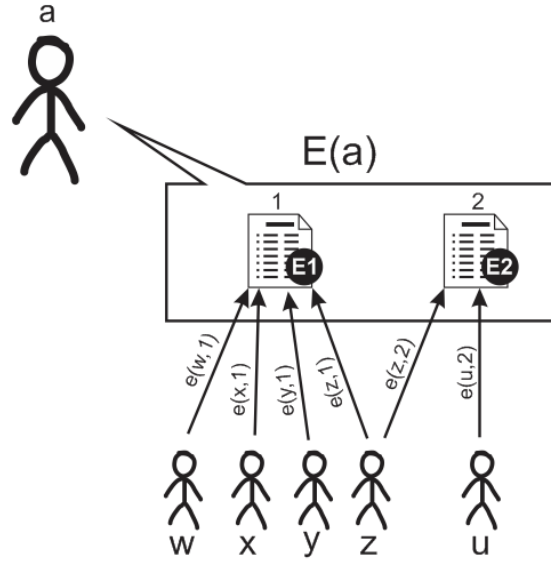


Figura 4.6: Análise do nível de maturidade do usuário  $a$

Na Figura 4.6 o usuário  $a$  produziu dois conteúdos  $\{E1$  e  $E2\}$  que foram avaliados por alguns usuários  $\{w, x, y, z, u\}$ . As avaliações de cada usuário em cada conteúdo, denotadas por  $e(\alpha, \beta)$ , serão utilizadas para calcular a qualidade destes conteúdos e, conseqüentemente, o nível de maturidade do autor destes conteúdos. Assim, o nível de maturidade de um usuário  $a$  pode ser calculado, genericamente, com a Equação (4.5).

$$M(x) = \begin{cases} \frac{\log[\sum_{i \in E(x)} e(i)]}{\max\{\log[\sum_{j \in E(y)} e(j)]: \forall y\}}, & \text{caso } E(x) \neq \emptyset \\ 0, & \text{caso contrário} \end{cases} \quad (4.5)$$

Onde:

$x, y$ : usuários;

$E(x), E(y)$ : conjunto das avaliações atribuídas aos conteúdos de  $x$  e  $y$ , respectivamente;

$i, j$ : cada avaliação do conjunto  $E(x)$  ou  $E(y)$ , respectivamente;

$e(i), e(j)$ : o valor atribuído à avaliação  $i$  ou  $j$ , respectivamente;

$\max\{\log[\sum_{i \in E(y)} e(i)]: \forall y\}$ : maior valor do conjunto formado pelos logaritmos das somas das avaliações atribuídas aos conteúdos dos usuários  $y$ .

De acordo com a representação na Equação (4.5), os usuários que possuem conteúdos avaliados positivamente pelos membros do ambiente social em que T-SWEETS estiver implantado possuem nível maior de maturidade em relação aos usuários que possuem conteúdos avaliados negativamente pelos membros da rede. Assim, é considerado que: usuários que produzem conhecimentos mais relevantes possuem maior maturidade, logo, maior confiança.

Para exemplificar o cálculo do Nível de Maturidade de um usuário em uma Rede Social, é utilizado o cenário hipotético apresentado na Figura 4.7.

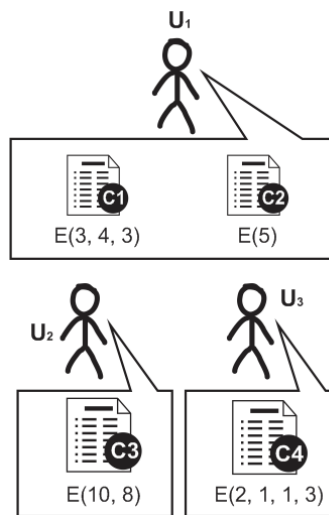


Figura 4.7: Cenário para exemplificar o cálculo do Nível de Maturidade

Neste cenário (Figura 4.7), tem-se:

- 1) o universo de 3 usuários  $U_1$ ,  $U_2$  e  $U_3$ ;
- 2) cada usuário produziu um conjunto  $C_n$  de documentos dispostos da seguinte forma:  
 $U_1 : C_1$  e  $C_2$ ;  $U_2 : C_3$ ; e  $U_3 : C_4$ ;
- 3) alguns destes documentos receberam avaliações por outros usuários, são elas:  $C_1$ : 3, 4, 3;  $C_2$ : 5;  $C_3$ : 10,8 e  $C_4$ : 2,1,1,3.

Assim, as maturidades destes usuários são:  $M(u_1)=0,94$ ,  $M(u_2)=1$  e  $M(u_3)=0,62$ .

### Relação com Trabalhos Relacionados

Um trabalho que inspirou a adoção do nível de maturidade no T-SWEETS foi o (Mamami, 2011), que utiliza uma variação do algoritmo PageRank (Brin, 1998) para ponderar o quanto os itens produzidos por um usuário são bem avaliados. Neste trabalho, a quantidade de avaliações atribuídas por um usuário  $x$  aos conteúdos produzidos por outro usuário  $y$  é utilizada para ponderar inversamente essas mesmas avaliações. Assim, quanto maior o número de

avaliações realizadas por um mesmo usuário  $x$  aos conteúdos de  $y$ , menor é o peso destas avaliações. Esta abordagem é uma tentativa de dar menor importância às avaliações das pessoas que não possuem um comportamento padrão na rede. Porém, existe o problema de que os usuários que estão mais próximos (distância social menor); ou relacionados a um mesmo contexto, possam ser prejudicados, já que, naturalmente, possuem uma tendência maior à colaboração e cooperação.

Este problema não ocorre no T-SWEETS. No T-SWEETS, quanto maior o número de avaliações que um usuário tiver, maiores são as chances de ele ter um nível de maturidade mais elevado na rede, isto é, os usuários que têm maior participação na rede não são prejudicados.

#### 4.3.4 Reputação

A reputação considera o quanto uma pessoa é considerada confiável perante uma visão geral de um universo de pessoas. Assim, possui influência direta na confiança entre as pessoas, pois as pessoas que possuem ações e opiniões mais confiáveis tendem a possuir maior nível de reputação. Neste sentido, a presente seção apresenta como o conceito de reputação foi considerado no presente trabalho, uma exemplificação deste cálculo e a relação com os trabalhos relacionados.

**Definição: Reputação** é a visão que um grupo de pessoas  $g$  tem sobre uma pessoa  $p$ , determinada pelo grau de confiança que  $g$  tem em relação à  $p$ .

No âmbito social virtual, normalmente, é considerada a visão de todos os membros da rede para determinar o grau de reputação de um membro, porém, esta abordagem tem um custo computacional alto. Para evitar este problema, o T-SWEETS não considera todos os membros da rede, permitindo a possibilidade de definir o tamanho da distância social  $d$  de um usuário alvo aos demais usuários da rede. O tamanho  $d$  ideal - que não comprometa o desempenho do algoritmo - ainda não foi definido, pois deve ser realizado um conjunto de experimentos para a referida definição. Por padrão, foi assumido valor 1 para esta distância social.

A Figura 4.8 apresenta um grafo de confiança no qual se objetiva determinar a reputação de um usuário  $a$ .

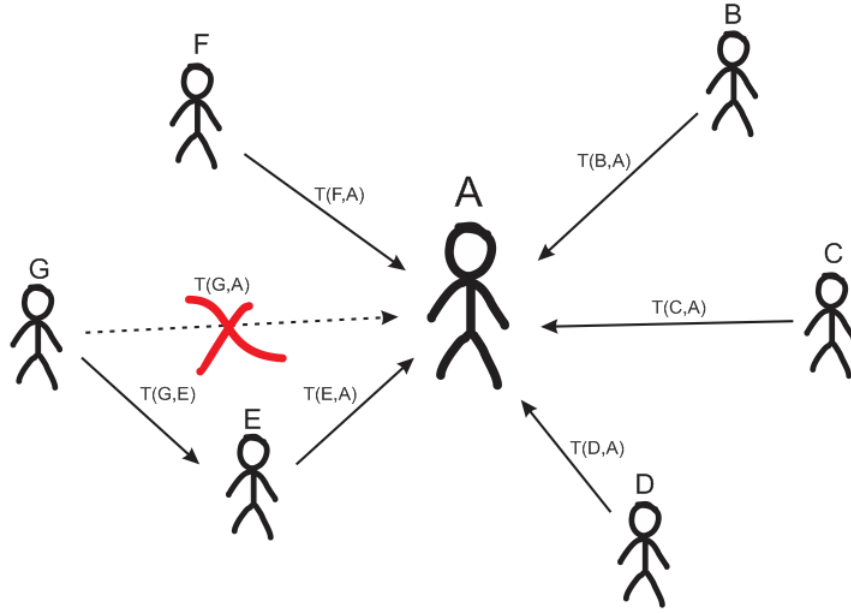


Figura 4.8: Grafo de confiança para cálculo de reputação do usuário A

Na Figura 4.8 as setas não pontilhadas representam os valores de confiança entre os usuários no grafo de confiança e a seta pontilhada representa um valor de confiança inferido para um relacionamento indireto. Conforme supracitado, T-SWEETS considera todas as avaliações diretas de confiança dadas a um usuário alvo para o cálculo da reputação. Assim, os valores de  $T(G, A)$  e  $T(G, E)$ , não irão influenciar na reputação do usuário A. Diante destas condições apresentadas, o cálculo de reputação pode ser realizado pela Equação (4.6).

$$R(a) = \frac{\sum_{u \in adj_1(a)} T(u, a)}{|adj_1(a)|} \quad (4.6)$$

Onde:

$R(a)$ : grau de reputação de um usuário  $a$ ;

$adj_1(a)$ : conjunto de usuários adjacentes que atribuíram confiança ao usuário  $a$ ;

$|adj_1(a)|$ : quantidade de usuários do conjunto  $adj_1(a)$ ;

$T(u, a)$ : grau de confiança do usuário  $u$  em relação a  $a$ .

A Equação (4.6) consiste em uma média aritmética dos valores de confiança atribuídos diretamente ao usuário  $a$ . O conjunto  $adj_1(a)$  contém os usuários que realizaram as atribuições de confiança a  $a$ , sendo que  $|adj_1(a)|$  é a quantidade de usuários neste conjunto. Enquanto  $T(u, a)$  é o valor de confiança do usuário  $u$  em relação a  $a$ , como apresentado nas seções anteriores.



Para exemplificar o cálculo da reputação de um usuário  $U_1$ , o Quadro 4.3 apresenta os dados do grafo de confiança de um conjunto de usuários em relação ao usuário  $U_1$ .

Quadro 4.3: Graus de confiança dos relacionamentos de um grupo de usuários  $U$  em relação ao usuário  $U_1$ .

	$U_2$	$U_3$	$U_4$	$U_5$	$U_6$	$U_7$
$U_1$	8	7	5	7	6	2

Os graus dos relacionamentos de confiança do conjunto de usuários em relação ao usuário  $U_1$  está em uma escala de 0 a 10. Este valor é, ao final, normalizado, para uma escala de 0 a 1. Com a aplicação da Equação (4.6) nos dados do Quadro 4.3, tem-se que a reputação de  $U_1$  ( $R(u_1)$ ) é 0,58.

#### Relação com Trabalhos Relacionados

De maneira similar ao T-SWEETS, o trabalho de (Mamami, 2011) também utiliza as mesmas fontes de informação para medir a reputação de um usuário na rede – que são as relações de confiança entre os usuários. A diferença em relação ao T-SWEETS é que, assim como no cálculo de *Maturidade do Usuário*, os usuários que contribuem mais na rede social têm menor importância. Esta abordagem, conforme já mencionado, prejudica os usuários que estão mais próximos (distância social menor); ou relacionados a um mesmo contexto, uma vez que estes usuários possuem uma tendência maior à colaboração e cooperação.

#### 4.4 Inferência de Confiança

As medidas apresentadas anteriormente, nas seções 4.3.1, 4.3.2, 4.3.3 e 4.3.4, devem ser ponderadas para se obter o valor de confiança entre dois usuários. Neste sentido, a Equação (4.7) apresenta a maneira como as características do T-SWEETS estão relacionadas no algoritmo.

$$C(a, b) = \frac{\alpha(S(a, b)) + \beta(T(a, b)) + \gamma(M(b)) + \delta(R(b))}{\alpha + \beta + \gamma + \delta} \quad (4.7)$$

Na Equação (4.7) os valores obtidos em cada uma das características do T-SWEETS (*Compatibilidade entre Usuários* ( $S(a, b)$ ), *Relacionamento de Confiança*, ( $T(a, b)$ ) *Nível de Maturidade* ( $M(b)$ ) e *Reputação* ( $R(b)$ )) são utilizados em uma média ponderada para inferir o quão um usuário  $a$  confia em um usuário  $b$ . Os valores  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$  determinam a importância

cia de cada um dos elementos considerados na inferência de confiança. A importância de cada elemento possui dependência direta com o ambiente em que o algoritmo T-SWEETS for implantado. Neste trabalho, assumiram-se pesos iguais para cada um dos elementos, uma vez que ainda não foi realizado um estudo mais aprofundado para mensurar a importância de cada um destes elementos.

## 4.5 Considerações Finais

O presente capítulo apresentou a definição e o formalismo dos elementos de um novo algoritmo para inferência de confiança em ambientes virtuais sociais, intitulado T-SWEETS, além dos pré-requisitos necessários para a sua implantação nestes ambientes.

T-SWEETS utiliza as produções e avaliações dos usuários na rede social como insumo para a inferência de confiança dos usuários. Este algoritmo é composto por 4 elementos: *Similaridade entre Perfis*; *Relacionamento de Confiança*; *Nível de Maturidade* e *Reputação*. A partir destes quatro elementos abordados, tem-se parâmetros para inferir o grau de confiança de um usuário em relação a outro.

Existem diversos algoritmos disponíveis na literatura que abordam inferência de confiança, conforme apresentados no Capítulo 3. O estudo destes algoritmos foi essencial para entender suas peculiaridades e assim, naturalmente, obter parâmetros para auxiliar na definição dos elementos que seriam utilizados no T-SWEETS. O Quadro 4.4 mostra uma visão geral de como estes 4 elementos são explorados nos trabalhos relacionados (apresentados no Capítulo 3).

Quadro 4.4: Resumo dos algoritmos que influenciaram na definição do T-SWEETS

	TidalTrust	SUNNY	Mamami & Gerosa	FlowTrust	T-SWEETS
Similaridade de Perfis		Utiliza as avaliações dos usuários a conteúdos em comum para avaliar a similaridade entre os perfis.			Considera os conteúdos produzidos no ambiente e as avaliações dos usuários aos conteúdos em comum para medir a similaridade entre os perfis.
Relacionamentos de Confiança	Baseado em probabilidade condicional.	Utiliza o algoritmo <i>TidalTrust</i> para realizar a inferência de relacionamentos de confiança ocultos, com base nas informações de similaridade de perfis dos usuários.		Utiliza a teoria de fluxos para inferir relacionamentos de confiança a partir do grafo de confiança.	Baseado em probabilidade condicional.
Nível de Maturidade			Baseia-se nas avaliações recebidas pelos conteúdos produzidos pelos usuários para inferir os valores de maturidade.		Baseia-se nas avaliações recebidas pelos conteúdos produzidos pelos usuários para inferir os valores de maturidade.
Reputação			Utiliza uma variação do algoritmo <i>PageRank</i> para avaliar o grafo de confiança rede social para inferir a reputação dos membros desta rede.		Inferir a reputação dos usuários da RSW a partir de informações do grafo de confiança.

Todos os algoritmos apresentados no Quadro 4.4 exploram o conceito de relacionamento de confiança, porém de maneiras diferentes. O TidalTrust (Golbeck, 2005), o algoritmo referência no campo de pesquisa sobre confiança em redes sociais, e o SUNNY (Kuter 2007, 2010), uma extensão do TidalTrust, são os que mais se aproximam da abordagem adotada no T-SWEETS. Estes algoritmos utilizam três características básicas de confiança, *Transitividade*, *Agregabilidade* e *Assimetria* para realizar o processo de inferência. Estas mesmas características são exploradas no T-SWEETS, porém com uma diferente perspectiva que, naturalmente, origina alterações no processo de inferência de confiança. No T-SWEETS a confiança de um usuário em relação a outro (*Relacionamento de Confiança*), baseia-se nos conceitos de *Assimetria* e *Agregabilidade* para tornar a confiança parcialmente transitiva.

A motivação para a utilização desta abordagem é em decorrência da definição dada para transitividade de confiança por Golbeck (2005). Golbeck diz que confiança não é um fenômeno totalmente transitivo, porém o processo de inferência de confiança estabelecido em seu trabalho torna a confiança completamente transitiva em determinadas regiões do grafo de confiança. Neste sentido, a inferência de confiança do T-SWEETS foi alterada para que essa definição de transitividade seja preservada em todas as regiões do grafo.

O algoritmo SUNNY, além de utilizar o mesmo processo de inferência de confiança do TidalTrust, explora mais um conceito para análise de confiança. Este conceito está relacionado à *Similaridade de Perfis* no T-SWEETS. O SUNNY analisa a semelhança de perfis de usuários a partir de uma análise estatística comparativa entre as avaliações realizadas por dois usuários a conteúdos em comum, que são compartilhados em um ambiente colaborativo. No T-SWEETS as avaliações realizadas pelos usuários também são analisadas, no entanto é utilizado o algoritmo de similaridade Coseno para realizar esta análise. Além disso, o T-SWEETS ainda realiza uma comparação dos perfis de conteúdos dos usuários para identificar os usuários que possuem conhecimentos similares.

Outro algoritmo que considera as avaliações a conteúdos é o apresentado por Mamami (2011), que analisa as avaliações recebidas pelos conteúdos de um usuário. Esta característica determina o que foi denominado como *Nível de Maturidade* dos usuários no T-SWEETS, que deve fornecer informações sobre o quão confiáveis são as informações produzidas por uma pessoa. Ambos os trabalhos, Mamami e T-SWEETS, consideram as avaliações dos usuários para determinar o *Nível de Maturidade* dos usuários, porém exploram diferentes abordagens. Conforme apresentado na Seção 4.3.3, a abordagem do trabalho de Mamami (2011) penaliza

as pessoas que possuem maior participação no ambiente colaborativo social. Já o formalismo deste elemento no T-SWEETS não apresenta esta limitação.

Ainda no algoritmo apresentado por Mamami é avaliado o nível de reputação dos usuários. De maneira similar ao T-SWEETS, o trabalho de Mamami (2001) também utiliza as mesmas fontes de informação para medir a reputação de um usuário na rede social – que são as relações de confiança entre os usuários. A diferença em relação ao T-SWEETS é que, assim como no cálculo de *Maturidade do Usuário*, os usuários que contribuem mais na rede social tem menor importância. Esta abordagem, conforme já mencionado, prejudica os usuários que estão mais próximos (distância social menor); ou relacionados a um mesmo contexto, uma vez que estes usuários possuem uma tendência maior à colaboração e cooperação.

## 5 SWEETS

Este capítulo tem o objetivo de apresentar a abordagem para identificação e recomendação de especialistas adotada no presente trabalho, intitulado SWEETS. A motivação para o nome SWEETS (*SmallWorldExpErtS*) origina-se da teoria de redes sociais dos 6 graus de separação, que diz que todas as pessoas estão a uma distância social 6 uma das outras.

O capítulo está organizado da seguinte forma: a Seção 5.1 apresenta uma breve contextualização para fundamentar a abordagem apresentada no presente trabalho; a Seção 5.2 apresenta a arquitetura e os módulos da abordagem para identificação e recomendação de especialistas, além dos formalismos adotados para inferência de especialidades; a Seção 5.8 apresenta como Contexto Computacional se enquadra na abordagem proposta e, por fim, a Seção 5.9 apresenta as considerações finais.

### 5.1 Introdução

Foi realizado um levantamento detalhado na literatura sobre os Sistemas de Recomendação de Especialistas (SRE) existentes, conforme apresentado no Capítulo 2. A partir desta análise, foi dada ênfase aos SREs que exploraram de alguma forma o conceito de contexto computacional, isto é, sistemas que forneçam os especialistas mais adequados para apoiar o usuário em suas necessidades. Assim, forneceu sustentação para a proposta de uma abordagem diferenciada para a identificação e recomendação de especialistas, intitulada SWEETS. O SWEETS possui algumas particularidades, são elas?

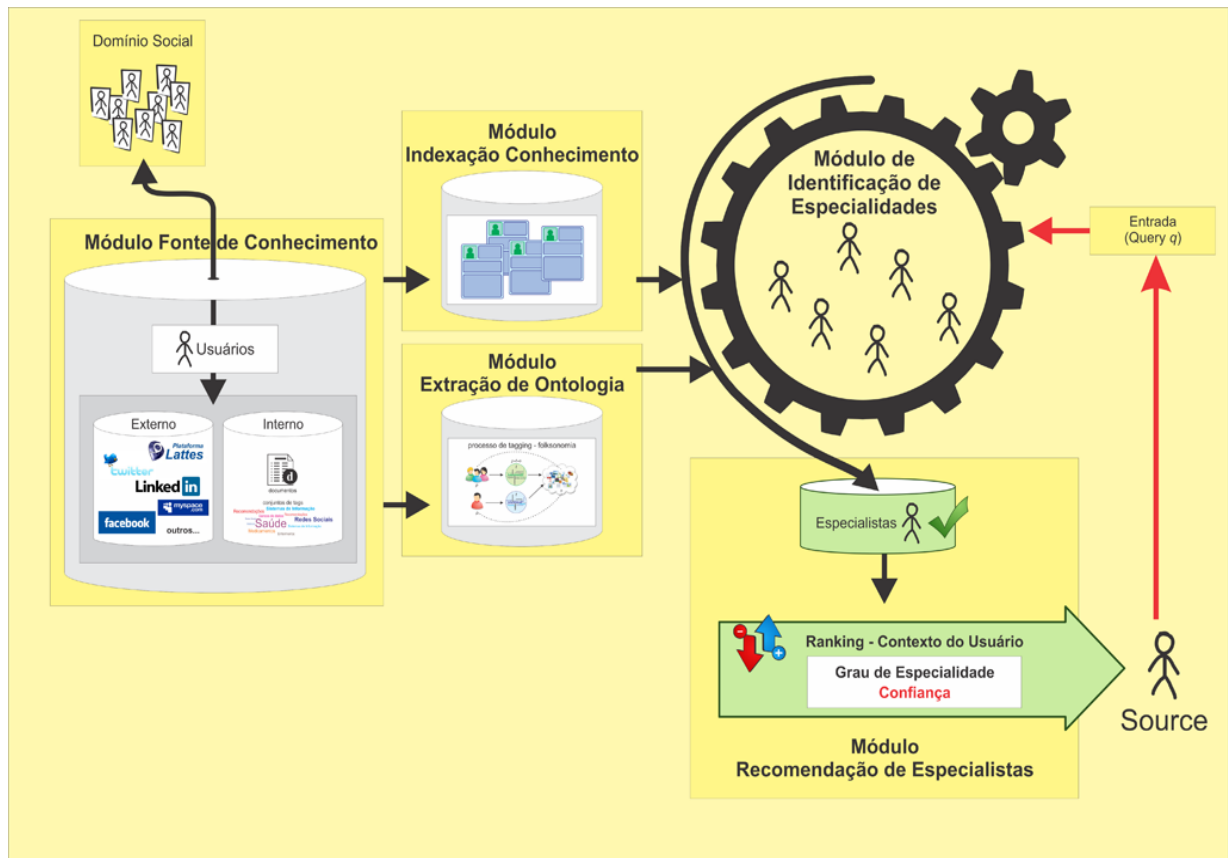
- considera as opiniões dos usuários para poder determinar o nível de especialidade dos candidatos a especialistas;

- considera a credibilidade da opinião dos usuários, a partir do nível de reputação e maturidade que eles têm no ambiente, para calcular o grau de especialidade de um candidato a especialista em um assunto qualquer. A maturidade de um indivíduo é determinada pela qualidade do conteúdo por ele produzido, enquanto a reputação de um indivíduo representa a visão geral que as pessoas têm em relação a ele. Ambos são apresentados no Capítulo 4, nas seções 4.3.3 e 4.3.4, respectivamente.
- utiliza fontes externas de conhecimento para enriquecer o processo de inferência de especialidades ou evitar o problema de insuficiência de conhecimento para a inferência das especialidades. O volume de conhecimento disponível das pessoas tem relação direta com a eficiência da abordagem para inferência das especialidades dos candidatos a especialistas;
- utiliza as produções de conhecimento (e.g. artigos, livros ou produções de qualquer natureza) dos indivíduos para inferência das especialidades;
- utiliza o conhecimento (e.g. posts em um fórum de discussão) produzido pelos usuários no ambiente em que o SWEETS estiver implantado para enriquecer a inferência das especialidades;
- utiliza características semânticas para inferência das especialidades, a partir do surgimento de ontologias provenientes de folksonomia. Folksonomia (Wal, 2007) é um mecanismo de *tagging* social em que as pessoas colaboram livremente para a sua criação e surge a partir de associações entre pessoas, *tags* (palavras-chave ou conceitos) e objetos (instâncias), em que uma pessoa pode categorizar um objeto com determinadas *tags*, criando assim associações ternárias entre estes elementos. Assim, uma folksonomia ( $F$ ) é definida por um conjunto de anotações  $F \subseteq A \times C \times I$ , onde  $A$  são as pessoas,  $C$  as *tags* (conceitos) e  $I$  os objetos categorizados. Como a folksonomia permite a classificação livre de objetos pelos usuários de uma plataforma social, esta ontologia não fica restrita a um domínio específico;
- utiliza os relacionamentos de confiança entre as pessoas para prover o especialista mais adequado em um determinado instante de tempo  $t$ .

A Seção 5.2 apresenta o SRE SWEETS, sua arquitetura, e a especificação de seus módulos necessários à identificação e recomendação de especialistas de domínio.

## 5.2 Arquitetura do SWEETS

*SWEETS* é um sistema de recomendação de especialistas não pró-ativo que pode ser implantado em qualquer ambiente computacional: (i) que seja baseado na teoria de redes sociais, isto é, ambientes em que as pessoas mantêm relacionamentos sociais; e (ii) que tenha conhecimentos associados aos usuários, ou seja, conhecimentos produzidos por cada usuário no ambiente em que o *SWEETS* estiver implantado, ou conhecimentos provenientes de ambiente externos.



**Figura 5.1:** Arquitetura do SWEETS

A Figura 5.1 apresenta a arquitetura do sistema de recomendação de especialistas *SWEETS*, que está dividido em módulos: Fonte de Conhecimento; Indexação de Conhecimento; Extração de Ontologia; Inferência de Especialidades; e, por fim, o Módulo de Recomendação de Especialistas. As especificações de cada um destes módulos são apresentados nas seções seguintes.

### 5.3 Módulo Fonte de Conhecimento

No Módulo Fonte de Conhecimento serão identificados os conhecimentos produzidos ou de interesse dos usuários, além das relações existentes entre as pessoas. A extração destes relaci-

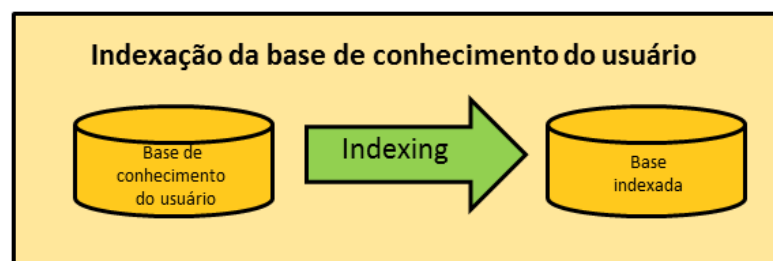


onamentos sociais será preponderante para que o contexto dos relacionamentos possa ser explorado a posteriori, quando as recomendações personalizadas serão realizadas.

De acordo com Ehrlich (2008), um dos maiores desafios em projetar um sistema de recomendação de especialista é permitir que as pessoas desenvolvam rapidamente e com segurança seu perfil de experiência com o mínimo possível de esforço e, ainda, que a privacidade seja mantida. Estas questões conflitantes conduzem à projeção e desenvolvimento de sistemas que facilitem e agilizem a construção dos perfis das pessoas. Para isso, podem ser desenvolvidos mecanismos que busquem o conhecimento produzido por estas pessoas em diferentes ambientes (*e. g.*: aplicações disponibilizadas na Web), para o mapeamento dos seus perfis com maior agilidade. Por isso, no Módulo Fonte de Conhecimento há dois tipos de classificações de fontes de conhecimentos dos usuários: Interna e Externa. A base interna representa os conhecimentos produzidos no ambiente em que o SWEETS estiver implantado, já a base externa representa os conhecimentos inerentes de fontes externas, por exemplo, *LinkedIn*, *Plataforma Lattes* e *Facebook*.

#### 5.4 Módulo Indexação de Conhecimento

Uma vez que a fonte de conhecimento foi definida, é realizada a indexação do conhecimento dos usuários do sistema, por meio do Módulo Indexação de Conhecimento. Neste momento, o conhecimento do usuário produzido tanto no ambiente interno quanto nos ambientes externos serão indexados de forma que os processos de inferência de especialidades e recomendação de especialistas possam ser realizados a posteriori.



**Figura 5.2:** Indexação da base do conhecimento

Esta indexação (conforme Figura 5.2) pode ser realizada periodicamente com um prazo previamente estabelecido, por exemplo, a cada 5 ou 7 dias, e tem como propósito prover um rápido e eficiente meio de acesso ao conhecimento produzido pelos usuários. Neste processo de indexação, deverão ser definidos critérios para estabelecer a relevância (maior ou menor peso) de cada termo associado ao usuário – termos associados aos usuários são termos que re-

presentam o conhecimento produzido por eles. Por exemplo, os termos de um texto plano presentes no corpo de um documento podem possuir menor relevância que os termos (palavras-chave) associados a um determinado documento. Já que as palavras-chave são termos chave que descrevem um documento e os textos planos possuem um maior volume de termos, o que, normalmente, torna mais difícil de identificar a relevância de cada termo presente no documento.

## 5.5 Módulo Geração de Ontologia

O Módulo Geração da Ontologia Simples (*Lightweight Ontology*) permite a adição de uma camada semântica para a realização da inferência de um tipo de especialidade (*especialidade inferida*, que será apresentada na Seção 5.6). A ontologia surge a partir da folksonomia, que oferece parâmetros para a identificação do conhecimento de um ou um grupo de usuários.

A ontologia que pode surgir a partir de *folksonomias* é classificada como Ontologia simples (*Lightweight Ontology*), já que explora o conceito bipartido (conceitos e instâncias) de ontologias. Assim, tem-se uma ontologia  $O_{c,i}$  - uma ontologia de conceitos (tags) em relação às instâncias (objetos) descritas por eles. Esta ontologia surge a partir da co-ocorrência dos conceitos em relação aos objetos. Assim, quanto mais os conceitos co-ocorrerem em relação aos objetos há um indicador de que estão fortemente relacionados, isto é, quanto maior a co-ocorrência, maior o peso do relacionamento semântico entre os conceitos.

As palavras-chave associadas aos trabalhos de caráter científico, por exemplo, monografias, dissertações e artigos, também remetem ao conceito de folksonomia, uma vez que os autores destas publicações utilizam um conjunto  $n$  de palavras-chave para categorizar suas publicações. Por isso, as palavras-chave que classificam estes trabalhos científicos também podem ser utilizadas como alternativa para o enriquecimento da Ontologia  $O_c$ . A Figura 5.3 apresenta uma situação hipotética de parte de uma simples ontologia.

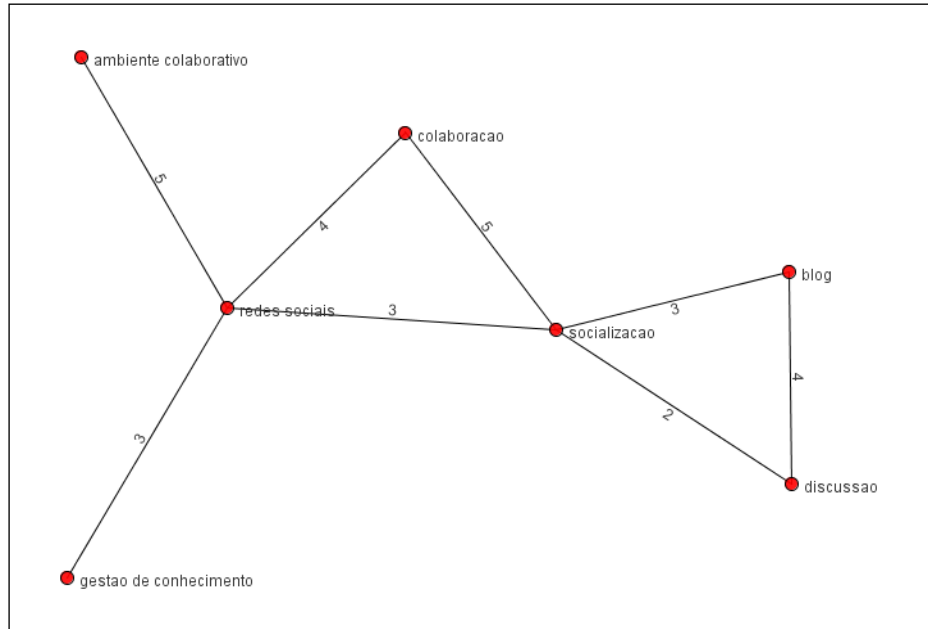


Figura 5.3: Parte de uma ontologia  $O_{ci}$ , hipotética.

Os conceitos representados na Figura 5.3, através dos nós, e as relações entre eles, representados através das arestas, indicam que os respectivos conceitos co-ocorreram em um determinado objeto  $o$ , o que determina a relação entre conceito e instância mencionada anteriormente. O peso de cada aresta, representado através de rótulos de números (Figura 5.3), representa a quantidade de vezes que estes mesmos conceitos co-ocorreram em diferentes instâncias de objetos. Assim, quanto maior for este peso, maior a evidência de que estes termos estejam fortemente relacionados, isto é, há uma maior densidade semântica.

## 5.6 Módulo Identificação de Especialidades

O Módulo Identificação de Especialidades utiliza o conhecimento pertinente a cada usuário do ambiente em que SWEETS estiver implantado para inferir as possíveis especialidades dos candidatos a especialistas.

O processo de inferência de especialidades no SWEETS explora os 4 tipos de habilidades que o trabalho de Fazel-Zarandi (2011, 2012) sugere que um especialista possa ter: (1) *habilidade demonstrada*; (2) *habilidade sugerida*; (3) *habilidade declarada*; e (4) *habilidade refutada*. Fazel-Zarandi (2011, 2012) utiliza o termo “*habilidade desconhecida*” para o que foi chamado no presente trabalho de “*habilidade informada*”. Fazel-Zarandi classifica esta categoria como desconhecida, pois são habilidades que as pessoas ainda não concordaram que o referido especialista a tenha.

- **habilidade demonstrada:** quando o usuário prova, por meio da prática, que ele possui uma determinada habilidade, por exemplo, através de certificações;
- **habilidade sugerida:** quando uma pessoa indica que outra possui conhecimento em determinada área;
- **habilidade informada:** quando o usuário autodeclara uma habilidade como, por exemplo, através do preenchimento do seu perfil;
- **habilidade refutada:** quando o usuário não demonstra, na prática, o conhecimento atribuído a ele.

A exploração dos tipos de habilidades supracitados, com exceção da *habilidade demonstrada e informada*, exige a participação coletiva dos usuários para a construção do perfil dos usuários candidatos a especialistas. Em decorrência disso, por causa destas diferentes opiniões, há uma tendência que os perfis dos usuários mapeiem com maior eficácia as suas especialidades, já que os perfis dos candidatos a especialistas serão construídos de acordo com a sabedoria coletiva.

Neste trabalho, como alternativa para enriquecer os perfis de especialidades dos usuários, é adotada uma categoria de habilidade intitulada *habilidade calculada*. Para a inferência desta categoria de habilidades, são utilizados os conhecimentos inerentes aos usuários para a inferência de possíveis habilidades que estes eventualmente venham a ter. Uma vez que um conjunto de habilidades  $H$  é inferido para um usuário  $u$ , estas podem ser, a posteriori, enquadradas em duas das 4 categorias listadas acima: *habilidade informada* e *habilidade sugerida*.

### 5.6.1 Processo de identificação de especialidades

Para que as especialidades dos usuários possam ser inferidas, inicialmente, faz-se necessário ponderar os 4 tipos de especialidades apresentados na Seção 5.6. No trabalho de Fazel-Zarandi (2011, 2012), uma habilidade  $h$  pode assumir apenas um estado em um instante de tempo  $t$ . Já no SWEETS, uma habilidade em uma área  $h$  pode estar classificada em mais de um tipo de habilidade em um tempo  $t$  – desta forma, espera-se explorar ao máximo as produções realizadas pelos usuários no ambiente no qual o SWEETS estiver implantado. Por exemplo, *João* é especialista (possui habilidades) em uma área  $h$  por possuir essa habilidade demonstrada através de uma *certificação*. À medida que *João* colabora com outras pessoas, estas percebem que *João* possui conhecimento sobre a área  $h$ , então estas pessoas sugerem que *João* possua conhecimento sobre  $h$  – assim, naturalmente, esta habilidade  $h$  passa a pertencer também à outra categoria, *habilidade sugerida*.

Essas categorias devem ser ponderadas de acordo com o cenário em que a abordagem SWEETS for implantada. A partir da definição dos pesos para cada categoria de habilidade, pode ser realizada a inferência das especialidades dos usuários e, a posteriori, a recomendação dos respectivos especialistas. A seguir é apresentada a formalização de todo este processo.

### ***Habilidade Demonstrada***

Conforme supracitado, a determinação da *habilidade demonstrada* considera as publicações dos indivíduos. Para que possa ser calculada, é necessário que exista apenas autores associados às publicações, além das palavras-chave, que representam as habilidades analisadas. Uma vez definidos estes pesos, o passo inicial para o cálculo da especialidade de um usuário em uma área  $h$  é realizado pela Equação (5.1).

$$h_{dem}(u, h) = \frac{1}{w_{t_{accl}} + w_{t_{tc}} + w_{t_{re}}} \times \sum_{\forall t_{u,h} \in T_{u,h}} \frac{w_t \times \sum_{\forall p_t \in t_{u,h}} |a_{p_t}|^{-1}}{|t_u|} \quad (5.1)$$

Onde:

$h_{dem_{u,h}}$ : a habilidade demonstrada de um usuário  $u$  em uma área  $h$ ;

$T_{u,h}$ : conjunto de publicações de um usuário  $u$  em uma área  $h$ , relacionadas a um tipo (conforme Quadro 6.3);

$t_{u,h}$ : subconjunto de  $T_{u,h}$ , que agrupa os trabalhos por tipo

$w_t$ : peso de uma publicação científica do tipo  $t$ ;

$p_t$ : uma publicação do conjunto  $t$ ;

$|a_{p_t}|$ : quantidade de autores da publicação  $p$ ;

$|t_u|$ : quantidade de publicações de um usuário em um conjunto  $t$ ;

$w_{t_{accl}}$ ,  $w_{t_{tc}}$  e  $w_{t_{re}}$ : peso das publicações apresentados no Quadro 6.3.

Inicialmente, na Equação (5.1) são identificadas todas as publicações de um usuário  $u$ , nas diferentes categorias  $T$  (e. g. *artigos em revistas*, *artigos completos* ou *resumos expandidos*) em uma área específica  $h$ .

Em seguida, para cada publicação é importante identificar a quantidade de autores existentes para mensurar o grau de *expertise* do usuário. Se uma produção científica possuir apenas um autor, há como afirmar que o autor é o detentor do conhecimento que originou o trabalho. Já nos casos em que há vários autores para uma publicação, não há como afirmar

que todos os autores possuem o mesmo conhecimento nas áreas abordadas no trabalho. Da mesma forma, não há como afirmar também qual dos autores possui um nível maior de conhecimento de uma área em relação à outra, já que não há um padrão que possibilite esta identificação. Inclusive, não há qualquer simples padrão de apresentação dos autores de uma publicação (*e. g.* autor principal é o primeiro da lista de autores). Por isso, a solução adotada no SWEETS para mensurar o nível de *expertise* do usuário em uma área  $h$  em relação a uma produção científica foi a distribuição homogênea do conhecimento entre o conjunto de autores da publicação.

Para considerar o nível de *expertise* do usuário  $u$  em uma área  $h$  em relação ao contexto que o SWEETS estiver implantado, foi considerada na Equação (5.1) a relação do grau de *expertise* de  $u$  em um tipo  $t$  pelo total de publicações de um tipo  $t$  em uma área  $h$ . Este resultado é, finalmente, multiplicado pelo  $w_t$ , considerando assim as ponderações para cada categoria de publicação apresentadas no Quadro 6.3. Este processo se repete para cada categoria de publicações ( $t$ ), e o somatório dos valores obtidos para cada categoria representa o grau de *expertise* do usuário em uma área  $h$  para a categoria *habilidade demonstrada* ( $h_{dem_{u,h}}$ ).

Após isso, é realizada a normalização do grau de *expertise* do usuário através da divisão do somatório dos pesos para cada categoria. Com isso, tem-se o grau de especialidade do usuário em uma escala de 0 a 1, sendo que quanto mais próximo de 1, maior o nível de especialidade do usuário, e quanto mais próximo de 0 menor o nível de especialidade.

Conforme supracitado, para o cálculo da *habilidade demonstrada* de um usuário  $u$  em uma área  $h$  ( $h_{dem_{u,h}}$ ) são consideradas todas as publicações de  $u$  em um tipo  $t$ , em relação a todas as publicações da área  $h$  no tipo  $t$ . Com isso, há um problema de não considerar a escala do grau de especialidade adequadamente. Para minimizar este problema, é necessário um passo adicional, apresentado na Equação (5.2).

$$h_{dem}(u, h) = \left( \frac{h_{dem}(u, h)}{\max(h_{dem}(h))} \right) \times w_{h_{dem}} \quad (5.2)$$

Onde:

$h_{dem_{u,h}}$ : a *habilidade demonstrada* de um usuário  $u$  em uma área  $h$ ;

$\max(h_{dem}(h))$ : maior habilidade demonstrada entre o conjunto de usuários em uma área  $h$ ;

$w_{h_{dem}}$ : peso da *habilidade demonstrada*.

Para normalizar o grau de especialidade do usuário em uma área “h”, foi adotada uma abordagem da área de Recuperação da Informação, intitulada *Term Frequency (tf)*. Esta abordagem considera a razão entre a frequência de um termo “t1” em um documento pela frequência do termo que tem maior número de ocorrências neste documento. Assim, a Equação (5.2) considera a divisão do grau de *expertise* em uma habilidade *h* de um usuário *u* ( $h_{dem_{u,h}}$ ) pelo maior grau de especialidade de um usuário *u* ( $\max(h_{dem}(h))$ ) dentre o conjunto de usuários do ambiente em que SWEETS estiver implantado, multiplicado pelo respectivo peso do tipo de habilidade ( $w_{h_{dem}}$ ). Como resultado, tem-se o valor final da *habilidade demonstrada* ( $h_{dem_{u,h}}$ ).

Para a exemplificação do cálculo do grau de habilidade de um usuário *u* (João) em uma área de especialidade *h* (Ontologia) na categoria *Habilidade Demonstrada*, considerar o cenário apresentado no Quadro 5.1.

Quadro 5.1: Cenário para exemplificação da *Habilidade Demonstrada*

Tipo das publicações de João	Publicações	Qtd. de autores
Artigos completos publicados em periódicos	1	2
Capítulos de livros	1	2
Trabalhos completos publicados em anais de congressos	1	3
Resumos expandidos publicados em anais de congressos	1	1

Diante do cenário apresentado no Quadro 5.1 e com o formalismo da Equação (5.1), tem-se:

$$h_{dem}(u,h) = ((5 \times (1/2) + 1/2)/2 + (2 \times 1/3)/1 + (1 \times 1/1)/1)/5+2+1$$

Assim, o grau não normalizado da habilidade demonstrada de João ( $h_{dem}(u,h)$ ) é 0,52. Há, também, no ambiente outro usuário, Pedro, que tem grau demonstrado de especialidade 0,7 nesta mesma área *h*. Diante destes valores, o passo final para determinar o grau final da especialidade demonstrada destes especialistas na área *h* é realizar a normalização, a partir do

formalismo apresentado na Equação (5.2). Com isso, tem-se no ambiente Pedro com grau 5 e João com 3,72.

### ***Habilidade Informada***

A *habilidade informada* é aquela em que o próprio especialista afirma ter (e. g. através do cadastro do seu perfil). No trabalho de Fazel-Zarandi (2011, 2012), esta categoria de habilidade é definida como *desconhecida* porque não há uma comprovação real que o especialista tenha, por exemplo, através de uma aclamação social (opinião dos usuários) ou certificações na área. A definição desta categoria de habilidade no presente trabalho é dada pela Equação (5.3).

$$hinf(u, h) = \begin{cases} w_{h_{des}}, & \text{se } h \in H_u \\ 0, & \text{caso contrário} \end{cases} \quad (5.3)$$

Onde:

$hinf(u, h)$ : grau da *habilidade informada* de um usuário  $u$  em uma área  $h$ ;

$w_{h_{des}}$ : peso da *habilidade*

$H_u$ : conjunto de habilidades informadas do usuário  $u$ .

De acordo com a Equação (5.3), se o usuário informar que possui uma habilidade  $h$ , então o próprio peso pré-definido para esta categoria de habilidade determina o grau de especialidade do usuário ( $hinf(u, h)$ ).

Dando continuidade ao exemplo apresentado no cálculo da *Habilidade Demonstrada*, João informou explicitamente que possui conhecimento na área  $h$  (*Ontologia*), por isso, tem  $hinf(u, h)$  igual a 3 (de acordo com a definição dos pesos apresentadas no Quadro 6.2), enquanto Pedro não informou que tem experiência em  $h$  e, por isso, tem o valor 0 para  $hinf(u, h)$ .

### ***Habilidade Sugerida***

As habilidades sugeridas são aquelas em que há uma intervenção explícita de outros usuários. Assim, o grupo de usuários pertinentes ao ambiente é responsável por sugerir que uma pessoa possui uma determinada especialidade. O grau da *habilidade sugerida* é determinado pela Equação (5.4).



$$hsug(u, h) = \sum_{\forall u \in U} v_r + v_m \quad (5.4)$$

Onde:

$hsug(u, h)$ : grau da *habilidade sugerida* de um usuário  $u$  em uma área  $h$ ;

$v_r$ : grau de reputação de um usuário  $v$ , que informa que o usuário  $u$  é especialista em  $h$ ;

$v_m$ : grau de maturidade de um usuário  $v$ , que informa que o usuário  $u$  é especialista em  $h$ .

Como no cálculo do grau da *habilidade sugerida* ( $hsug(u, h)$ ) são consideradas as opiniões dos usuários, é importante considerar tanto a maturidade dos usuários no ambiente ( $v_m$ ), qualidade do conhecimento produzido, quanto a reputação destes usuários ( $v_r$ ) perante os demais usuários do ambiente. Assim, é considerada a credibilidade de cada usuário para a determinação da *habilidade sugerida*. O Capítulo 4 mostra como são determinados o nível de maturidade e a reputação dos usuários no presente trabalho.

A normalização do valor obtido na Equação (5.4) é realizada pela Equação (5.5).

$$hsug(u, h) = \left( \frac{hsug(u, h)}{\max(hsug(h))} \right) \times w_{hsug} \quad (5.5)$$

Onde:

$hsug(u, h)$ : grau da *habilidade sugerida* de um usuário  $u$  em uma área  $h$ ;

$\max(hsug(h))$ : maior grau de *habilidade sugerida* em uma área  $h$ ;

$w_{hsug}$ : peso da *habilidade sugerida*.

Nesta normalização, apresentada na Equação (5.5), é realizada a divisão do grau da *habilidade sugerida* de um usuário  $u$  em uma área  $h$ , pelo maior grau de *habilidade sugerida* nesta área  $h$  no ambiente em que o SWEETS estiver implantado. Assim, tem-se um valor em uma escala de 0 a 1. Após isso, é realizada a multiplicação pelo respectivo peso da *habilidade sugerida*,  $w_{hsug}$ , que representa o valor final da *habilidade sugerida*  $hsug(u, h)$ .

*João* agora foi sugerido por *Francisco* como especialista em  $h$  (*Ontologia*). *Francisco* possui *Nível de Maturidade* no ambiente de 0,73 e *Reputação* de 0,64. Com isso, *João* tem grau de especialidade para *habilidade sugerida* ( $hsug(u, h)$ ) é igual a 1,0. Como *Pedro* não recebeu uma sugestão realizada por outros usuários, seu grau de  $hsug(u, h)$  é 0.

***Habilidade Refutada***

Assim como a *habilidade sugerida*, a *habilidade refutada* também depende da intervenção explícita dos usuários do ambiente em que o SWEETS estiver implantado. O grau da *habilidade refutada* é determinado pela Equação (5.6)

$$href(u, h) = \sum_{\forall u \in U} u_r + u_m \quad (5.6)$$

Onde:

$href(u, h)$ : grau da *habilidade refutada* de um usuário  $u$  em uma área  $h$ ;

$u_r$ : grau de reputação de um usuário  $u$ ;

$u_m$ : grau de maturidade de um usuário  $u$ .

Quando uma pessoa solicita ajuda de um determinado especialista em uma área  $h$ , e este especialista não comprova que realmente possui conhecimento na área  $h$ , o grau geral de especialidade do usuário nesta área é então diminuído, já que o peso desta categoria de habilidade é negativo, conforme apresentado no Quadro 6.2. Conforme definido na Equação (5.6), o cálculo da *habilidade refutada* considera tanto a reputação ( $u_r$ ) quanto o nível de maturidade ( $u_m$ ) do usuário no ambiente, para considerar a importância da opinião do usuário no ambiente.

A Equação (5.7) apresenta a normalização do grau obtido na Equação (5.6):

$$href(u, h) = \left( \frac{href(u, h)}{\max(href(h))} \right) \times w_{href} \quad (5.7)$$

Onde:

$href(u, h)$ : grau da *habilidade refutada* de um usuário  $u$  em uma área  $h$ ;

$\max(href(h))$ : maior grau de *habilidade refutada* em uma área  $h$ ;

$w_{href}$ : peso da habilidade refutada.

A etapa de normalização, apresentada na Equação (5.7), é primordial para manter o grau  $href(u, h)$  na escala. O resultado proveniente da normalização realizada na Equação (5.7) será em uma escala entre 0 e -1, sendo que quanto mais próximo de -1, maior será o índice de refutação do especialista na área  $h$ , e quanto mais próximo de 0, menor é este índice de refutação.

Dando continuidade na exemplificação do cálculo do grau de habilidade dos candidatos a especialistas *João* e *Pedro*, tem-se agora a seguinte situação: *Francisco* solicitou uma vez ajuda sobre *h* (*Ontologia*) a *João* e duas vezes a *Pedro*. Após isso, *Francisco* avaliou as solicitações de ajuda como úteis ou não úteis, fornecendo assim parâmetros para refutar ou não o quanto estas pessoas (*João* e *Pedro*) realmente conhecem o assunto *Ontologia*. *Francisco* avaliou todas estas solicitações de ajuda como não úteis, refutando assim o nível de conhecimento destes usuários em *h*. A reputação de *Francisco* é 0,64 e a maturidade é 0,73. Com isso, tem-se:

$$H_{refJoão,h} = 0,73 + 0,64 = 1,37$$

$$H_{refPedro,h} = (0,73 + 0,64) + (0,73 + 0,64) = 2,74$$

Após a normalização, aplicando a Equação (5.7), tem-se como valor final da habilidade refutada ( $href(u, h)$ ) de ambos, *João* e *Pedro*, -0,5 e -1, respectivamente.

### ***Habilidade calculada***

Esta categoria de habilidade é calculada automaticamente pelo ambiente. Para isso, são utilizados os conhecimentos produzidos pelos usuários para o cálculo desta habilidade. Uma vez identificado este conjunto de habilidades, são realizadas recomendações para o usuário candidato a especialista. Se o usuário admitir que possui uma ou várias das habilidades recomendadas, então, esta(s) habilidade(s) passa(m) para outra categoria, de *habilidade informada*. E se os usuários do ambiente sugerirem que um determinado usuário possui conhecimento em alguma das habilidades calculadas, então esta habilidade passa para a categoria de *habilidade sugerida*.

Para que esta categoria de habilidade possa ser explorada é necessário possuir no ambiente de implantação do SWEETS: conhecimentos associados aos usuários e exploração de conceitos inerentes à folksonomia.

Nesta categoria, o universo de especialidades possíveis para cada usuário é limitado às marcações realizadas nos objetos em que o SWEETS estiver implantado (que originam a *folksonomia*) e às palavras-chave oriundas das produções científicas. As produções científicas podem ser originadas de diferentes fontes, por exemplo, produções dos usuários disponibilizadas internamente no ambiente; produções disponíveis na Plataforma *Lattes* (base de dados de currículos de pesquisadores brasileiros); no Mendeley (é uma rede social acadêmica e um gerenciador de referências bibliográficas online); na LinkedIn, uma rede social profissional; e

no BibSonomy (um sistema para organização e compartilhamento de *bookmarks* e publicações).

Para conceber modelos de redes de folksonomias em um nível abstrato, pode ser utilizado um sistema de grafo tripartido (West, 2001) (Cranston, 2009) em que o conjunto de vértices é definido em 3 (três) conjuntos disjuntos:  $A = \{a_k, \dots, a_{k-1}, a_k\}$ ,  $C = \{c_1, \dots, c_{l-1}, c_l\}$ ,  $I = \{i_m, \dots, i_{m-1}, i_m\}$  em que, cada um desses conjuntos correspondem, respectivamente, aos atores (usuários), conceitos (*tags*) e as instâncias anotadas (*e. g.* documentos, sites, imagens, entre outros.). Dessa forma, um sistema de *tagging* social permite que usuários etiquetem objetos com conceito, criando assim associações ternárias entre o *usuário*, o *conceito* e o *objeto*. Assim, uma folksonomia ( $F$ ) é definida por um conjunto de anotações  $F \subseteq A \times C \times I$ . A partir disso, é possível utilizar o tradicional modelo bipartido de ontologia (conceitos e instâncias) –  $O_{ci}$  –, fundamental para a inferência da *habilidade calculada* do SWEETS.

Para a criação da ontologia  $O_{ci}$ , são utilizadas as co-ocorrências (Seção 5.5) das *tags* (conceitos) em relação aos *itens* (instâncias). Os relacionamentos entre os termos / conceitos da ontologia  $O_{ci}$ , são ponderados pela quantidade de instâncias ( $I$ ) que são etiquetadas (*tagged*) com ambos os termos, ou seja, o número de vezes que os termos co-ocorrem em instâncias diferentes. Este é um método básico de mineração de texto, em que os termos são geralmente associados pela sua co-ocorrência em documentos (Cutting, 1992) (Feldman, 1998), (Baeza-Yates, 1999), (Mika, 2007) e (Pound, 2010).

Uma vez que a ontologia  $O_{ci}$  tenha sido criada, é realizada a extração de cada conceito que possui relacionamentos maior que 1, seus relacionamentos e respectivos pesos. Sabe-se que essa quantidade mínima deve ser maior que 1, porém, não há uma definição exata do valor ideal. A quantidade mínima de relacionamentos é configurável e, por isso, pode ser modificada a qualquer instante.

Após isso, estas informações são representadas em um espaço vetorial  $\vec{Wo}_{t_i} = [Wo_{t_1}, Wo_{t_2}, \dots, Wo_{t_{n-1}}, Wo_{t_n}]$ , em que  $\vec{Wo}_t$  representa o vetor de pesos do conceito chave e seus relacionamentos. O peso do conceito chave (conceito em evidência) possui maior relevância que qualquer outro peso do relacionamento, por isso, este peso é determinado por  $\max(\vec{Wo}) + 1$ . Não se sabe realmente o quanto o peso deste conceito-chave deve ser maior que os demais. Talvez, afirmar que seu peso deve ser maior em 1 que qualquer outro peso dos relacionamentos não seja o suficiente. O ideal é implantar esta abordagem em algum ambiente

e, durante um período de tempo, realizar sucessivas análises da eficácia dos resultados a partir do incremento do peso deste conceito-chave, até que se chegue a um “bom” valor. Com isso, para cada conceito e seus relacionamentos haverá uma representação vetorial, formando assim um conjunto de vetores.

Diante destas representações vetoriais, é possível realizar a indexação da base de conhecimento de cada usuário. Essa indexação é executada utilizando todo o conhecimento produzido pelos usuários no ambiente no qual o *SWEETS* estiver sendo implantado. Com isso, para cada vetor  $\vec{Wo}_t$  haverá um vetor  $\vec{Fu}_t = [Fu_{t_1}, Fu_{t_2}, \dots, Fu_{t_{m-1}}, Fu_{t_m}]$  equivalente, com a diferença que  $\vec{Fu}_t$  representa a frequência (peso) de cada termo em relação ao usuário e  $\vec{Wo}_t$  representa os pesos dos termos na  $O_{ci}$ . É importante ressaltar que  $n$ , o tamanho do vetor  $\vec{Wo}_t$ , e  $m$ , o tamanho do vetor  $\vec{Fu}_t$  são iguais, sendo  $Fu_{t_m} \geq 0$  e  $Wo_{t_m} \geq 0$ .

Por fim, é possível realizar o cálculo do grau da *habilidade calculada* dos usuários em relação aos conceitos. Para isso, são utilizadas as representações vetoriais de  $\vec{Fu}_t$  e  $\vec{Wo}_t$ . O cálculo é realizado a partir do valor do *Coseno*, apresentado na Equação (5.8).

$$\text{hinfer}(u, h) = \frac{\sum_{i=1}^n (Wo_{t_i} \cdot Fu_{t_i})}{\sqrt{\sum_{i=1}^n (Wo_{t_i})^2} \cdot \sqrt{\sum_{i=1}^n (Fu_{t_i})^2}} \quad (5.8)$$

Onde:

$\text{hinfer}(u, h)$ : grau da *habilidade inferida* de um usuário  $u$  em uma área  $h$ ;

$t$ : cada elemento dos conjuntos  $\vec{Wo}_t$  e  $\vec{Fu}_t$ .

$\vec{Wo}_t$ : vetor de pesos do conceito em evidência (área  $h$ ) e seus relacionamentos;

$\vec{Fu}_t$ : vetor de pesos do conceito em evidência e seus relacionamentos relacionados a um usuário  $u$ .

O grau  $\text{hinfer}(u, h)$ , que representa a *habilidade inferida* de um usuário  $u$  em uma área  $h$ , está em uma escala entre 0 e 1, sendo que quanto mais próximo de 1, maior é o nível de conhecimento que o usuário possui em relação a esta área.

Uma das vantagens desta abordagem para identificar as especialidades dos indivíduos é que a pessoa que recebe as recomendações pode ser informada de porque recebeu o

referido especialista como sugestão. Por exemplo, *João* foi recomendado como especialista em *Redes Sociais*, pois *Redes Sociais* tem forte relação com *colaboração*, *socialização*, *gestão de conhecimento e ambiente colaborativo*, e *João* está fortemente relacionado a estes conceitos, pois tem conhecimento produzido sobre estes assuntos.

### **Exemplo de cálculo da *Habilidade Calculada***

Para exemplificar o cálculo da habilidade na categoria *Calculada*, é apresentado o seguinte cenário hipotético: calcular o grau da habilidade do usuário *João* no termo “*Processos de Software*”.

O primeiro passo é selecionar o termo chave (*Processos de Software*), seus termos (conceitos ou palavras-chave) e pesos relacionados a ele na ontologia  $O_{ci}$ . Com isso, tem-se, “*Modelo em cascata*” com peso 3, “*Prototipação*” com peso 3; “*Modelos Incrementais*” com peso 4 e “*Espiral*” também com peso 4. Após identificados, esses pesos são representados em um espaço vetorial para uma posterior análise. Essa representação vetorial é dada por  $\overrightarrow{Wo}_{t_i} = [3, 3, 4, 4]$ . O termo chave é o principal, ou seja, deve ter maior relevância e tem que ser considerado neste espaço vetorial. Por isso, seu peso foi definido como  $\max(\overrightarrow{Wo}) + 1$ , conforme apresentado anteriormente. Neste exemplo em particular, o termo chave é *Processos de Software* e tem peso 5. Assim, a nova representação vetorial é  $\overrightarrow{Wo}_{t_i} = [5, 3, 3, 4, 4]$ .

O passo seguinte é identificar o quanto *João* está associado com cada um destes conceitos. Para isso, são verificadas as ocorrências destes termos nos conhecimentos produzidos por *João*, por exemplo, em uma postagem em um fórum. Assim, tem-se, nas produções de *João* os termos/conceitos e respectivos pesos: “*Processos de Software*” com peso 15; “*Modelo em cascata*” com peso 2, “*Prototipação*” com peso 2; “*Modelos Incrementais*” com peso 2 e “*Espiral*” também com peso 2. Denotando na seguinte representação vetorial:  $\overrightarrow{Fu}_{t_i} = [15, 2, 2, 2, 2]$ .

Com a obtenção destas duas representações vetoriais  $\overrightarrow{Wo}_{t_i}$  e  $\overrightarrow{Fu}_{t_i}$ , é possível aplicar o algoritmo do *Coseno* e assim obter o grau de especialidade de *João* em *Processos de Software* na categoria *Habilidade Calculada*, que é 0,76.

## Habilidade Geral Inferida

Uma vez realizados os cálculos das habilidades: *demonstrada*, *informada*, *sugerida* e *refutada*, é possível determinar o grau final da habilidade de um usuário  $u$  em uma área  $h$ .

$$GE(u, h) = \frac{h_{dem}(u, h) + h_{des}(u, h) + h_{sug}(u, h) + h_{ref}(u, h)}{w_{h_{dem}} + w_{h_{des}} + w_{h_{sug}} + w_{h_{ref}}} \quad (5.9)$$

Onde:

$GE(u, h)$ : grau geral da habilidade do usuário  $u$  em uma área  $h$ ;

$h_{dem}(u, h)$ : grau da *habilidade demonstrada* do usuário  $u$  em uma área  $h$ ;

$h_{des}(u, h)$ : grau da *habilidade informada* do usuário  $u$  em uma área  $h$ ;

$h_{sug}(u, h)$ : grau da *habilidade sugerida* do usuário  $u$  em uma área  $h$ ;

$h_{ref}(u, h)$ : grau da *habilidade refutada* do usuário  $u$  em uma área  $h$ ;

$w_{h_{dem}}$ : peso da *habilidade demonstrada*;

$w_{h_{des}}$ : peso da *habilidade informada*;

$w_{h_{sug}}$ : peso da *habilidade sugerida*;

$w_{h_{ref}}$ : peso da *habilidade refutada*;

Na Equação (5.9) são considerados os 4 níveis de habilidades definidas anteriormente para inferir o *grau final de especialidade* de  $u$  em relação a  $h$ . O resultado final está em uma escala de 0 a 1. Quanto mais próximo de 1, maior será o nível de conhecimento do especialista em  $h$ .

Uma visão sumarizada dos resultados das exemplificações do cálculo de especialidades dos candidatos a especialistas *João* e *Pedro* é apresentada no Quadro 5.2.

Quadro 5.2: Visão sumarizada das habilidades de *João* e *Pedro* no assunto *Ontologia*

Habilidade ( $h$ )	Grau <i>João</i>	Grau <i>Pedro</i>	Peso ( $w_h$ )
Demonstrada ( $(h_{dem}(u, h))$ )	3,72	5	5
Informada $h_{inf}(u, h)$	3	0	3
Sugerida ( $(h_{sug}(u, h))$ )	1	0	1
Refutada ( $(h_{ref}(u, h))$ )	-0,5	-1	-1

A partir do cenário apresentado no Quadro 5.2 e a aplicação da Equação (5.9), tem-se:

$$GE(Jo\tilde{a}o, Ontologia) = ((3,72 + 3 + 1 + (-0,5))/5+3+1+(-1)) = 0,9$$

$$GE(Pedro, Ontologia) = (5 + 0 + 0 + (-1))/(5+3+1+(-1)) = 0,5.$$

Isto é, o grau final ( $GE(u, h)$ ) de *João* em *Ontologia* de 0,9, enquanto o grau final de especialidade *Pedro* em *Ontologia* é 0,5.

### **5.7 Módulo Recomendação de Especialistas**

Uma vez que as especialidades dos usuários foram identificadas, é possível realizar a recomendação de pessoas classificadas como especialistas em uma determinada área. O grau de especialidade em uma determinada área, inferido no Módulo Identificação de Especialidades, está em uma escala entre 0 e 1 [0...1]. Neste trabalho, foi determinado que usuários que possuem grau de especialidade superior ou igual a 0,7 são classificados como especialistas e, portanto, podem ser recomendados.

A definição deste limiar é flexível e pode ser definida de acordo com o ambiente em que o SWEETS for implantado. Quanto mais próximo de 1 for esse limiar, maior será a precisão na recomendação de especialidade, porém, é diminuída, naturalmente, a cobertura (*recall*) das recomendações. Assim, potenciais especialistas podem ficar fora do conjunto de especialistas recomendados.

Além do grau de especialidade, é também considerado neste módulo o grau de confiança do usuário que solicitou ajuda em relação aos especialistas recomendados. Assim, o usuário tem a liberdade de buscar por ajuda de uma pessoa que esteja mais próxima a ele, ao invés de necessariamente buscar por ajuda de uma pessoa que seja classificada como especialista em uma determinada área.

Neste sentido, as recomendações dos especialistas podem ser classificadas de duas maneiras: pelo grau de especialidade ou pelo nível de confiança. Se for considerado o nível de confiança, há uma relação direta com o contexto do usuário que solicitou ajuda, logo, o conjunto de recomendações pode ser diferente em instantes de tempo diferentes, já que a confiança de uma pessoa em relação à outra é dinâmica, isto é, muda com o tempo – conforme apresentado no Capítulo 3.



## 5.8 Contexto no SWEETS

A exploração de contexto no SWEETS é proveniente das relações sociais entre os indivíduos que oferecem sustentação para a descoberta de relacionamentos ocultos de confiança. Desta forma, é possível recomendar os possíveis especialistas mais confiáveis para um determinado indivíduo. Muitas vezes, o usuário que possui o maior nível de conhecimento em um assunto  $h$  pode não ser o mais disposto a colaborar com um usuário  $u$ , por diferentes motivos, por exemplo, o especialista pode estar ocupado, pode não ser propenso à colaboração, ou pode não haver um relacionamento prévio entre eles (usuário e especialista), seja este relacionamento de confiança ou não..

A Figura 5.4 apresenta um exemplo em que um conjunto de especialistas em uma área  $h$  é recomendado para um usuário ( $U_{source}$ ). Além do grau de especialidade de cada especialista na área  $h$ , denotado por  $(GE(u, h))$ , é apresentado também o grau de confiança do usuário que recebe a recomendação ( $U_{source}$ ) em relação ao usuário recomendado ( $U_{sink_n}$ ), que é denotado por  $T(U_{source}, U_{sink_n})$ . Ambos os graus,  $GE(u, h)$  e  $T(U_{source}, U_{sink_n})$  estão em uma escala de 0 a 1. No cenário apresentado (Figura 5.4), os especialistas  $U_{sink_1}$  e  $U_{sink_2}$  são aqueles que possuem maior nível de conhecimento na área  $h$ , porém, o nível de confiança entre o usuário que recebeu a recomendação ( $U_{source}$ ) e eles é menor, o que pode ser um empecilho à colaboração. Já os especialistas  $U_{sink_3}$ ,  $U_{sink_4}$  e  $U_{sink_5}$  possuem um nível de especialidade menor, porém, o usuário  $U_{source}$  tem maior nível de confiança neles, logo, há uma tendência natural que haja uma facilidade na colaboração entre eles. Destes, provavelmente, os mais indicados para serem recomendados seriam os especialistas  $U_{sink_4}$  e  $U_{sink_3}$ , nesta ordem de prioridade, pois são eles que possuem o maior nível de confiança do usuário  $U_{source}$  e maior nível de conhecimento em  $h$ .

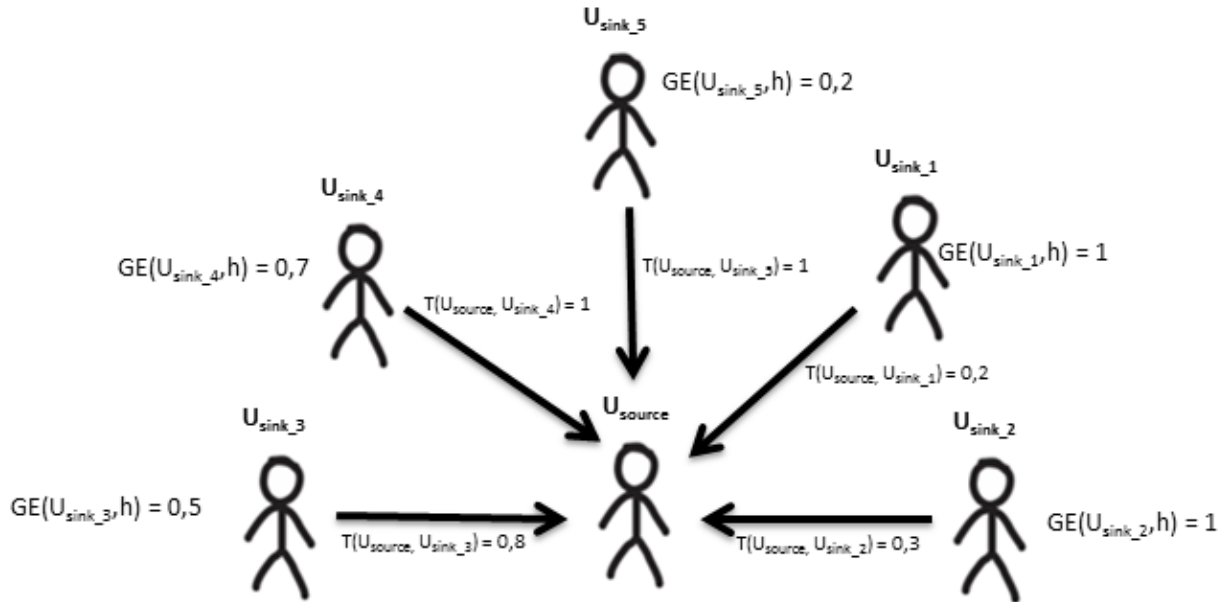


Figura 5.4: Exemplo de recomendação de especialistas em uma área  $h$  para um usuário  $U_{source}$

Outros elementos contextuais, adotados na abordagem do SWEETS, foram usados para mensurar as habilidades nas categorias *Sugerida e Refutada*. Ambas as categorias baseiam-se no *Nível de Maturidade e Reputação* dos indivíduos em relação aos demais usuários no ambiente. O *Nível de Maturidade* de um indivíduo é determinado de acordo com a relevância dos conteúdos produzidos por ele no ambiente. Esta relevância é determinada pelas avaliações dos demais usuários aos conteúdos produzidos por ele. E a *Reputação* determina a visão geral que os indivíduos da rede têm em relação a um determinado usuário. Para isso, são utilizadas as avaliações diretas de confiança atribuídas às relações sociais dos indivíduos. Com a exploração destes dois elementos, o cálculo das habilidades nestas duas categorias (*Sugerida e Refutada*) baseia-se no contexto das relações sociais das pessoas que fazem parte do ambiente.

Além disso, a identificação destes relacionamentos de confiança também é baseada em contexto computacional, pois é fundamentada em características como: pessoas que compartilham (possuem) opiniões e conhecimentos similares, relacionamentos diretos e indiretos entre as pessoas, e a visão que um grupo de pessoas tem em relação à outra. Estas características forneceram sustentação para a definição dos 4 elementos definidos: Similaridade entre Perfis, *Nível de Maturidade*, Relacionamentos de Confiança e Reputação.

Para determinar a semelhança entre os perfis das pessoas, são consideradas as opiniões e os conhecimentos semelhantes entre as pessoas. Assim, tanto as opiniões quanto

os conhecimentos similares são os elementos contextuais que os unem, conforme pode ser visualizado na Figura 5.5.

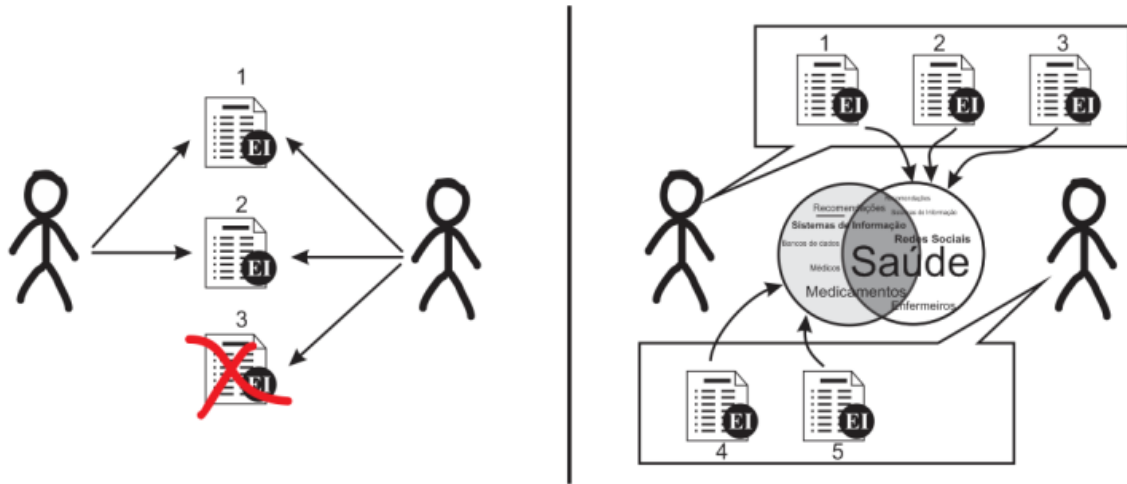


Figura 5.5: Análise de Opiniões similares entre pessoas (A)

Análise dos conhecimentos similares entre pessoas (B)

O Nível de Maturidade de uma pessoa é determinado pela visão que um grupo de pessoas tem em relação ao conhecimento que é produzido por ela. Neste cenário, o contexto é considerado como sendo a visão comum e compartilhada que um grupo de pessoas tem sobre as produções de conhecimento dos membros no ambiente (Figura 5.6) - a partir das avaliações dadas aos conteúdos,  $E(a)$ .

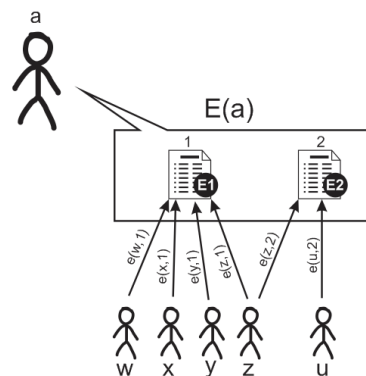


Figura 5.6: Cenário para análise do nível de maturidade de um usuário.

O contexto das relações sociais é utilizado para determinar o grau do elemento Relacionamento de Confiança e Reputação (Figura 5.7). Enquanto o grau dos Relacionamentos de Confiança entre pessoas é atribuído explicitamente por eles ou inferido a partir de um caminho de confiança que os une, a Reputação é determinada por uma visão que um grupo de pes-

soas tem em relação a outra. Assim, ambos os elementos são estabelecidos pelo contexto das relações de confiança que unem os membros da rede.

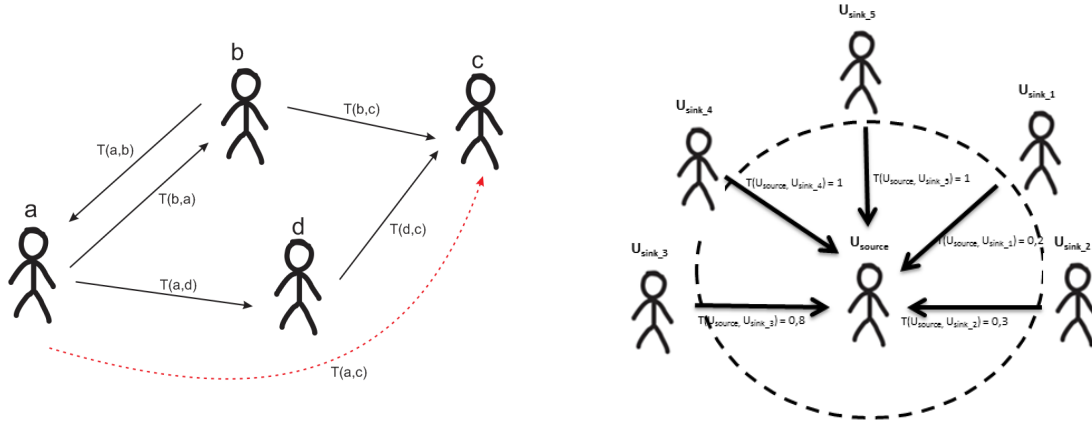


Figura 5.7: (A) Grafo de Confiança entre os membros de uma rede social. (B) Visão que um grupo de pessoas tem em relação a outra.

Ainda há outros elementos contextuais que podem ser explorados no SWEETS, tais como, distância social entre os indivíduos, mensurável em ambientes virtuais sociais; disponibilidade, que determina se o especialista está ou não disponível em um determinado instante de tempo  $t$ ; e pré-disposição à colaboração, um indicador que qualifica o nível de pré-disposição à colaboração de um especialista, *e.g.*, através de ações passadas. A adoção destes elementos é condicionada às características do ambiente em que o SWEETS for implantado.

## 5.9 Considerações Finais

SWEETS é uma abordagem que permite a identificação e recomendação de especialistas. Esta abordagem pode ser implantada em qualquer ambiente computacional que possua conhecimentos associados aos usuários e que seja baseado na teoria de redes sociais. Apesar disso, há parte da abordagem que pode ser explorada em ambientes não sociais, que é a inferência de habilidades nas categorias *Demonstrada* e *Informada*, pois os usuários não precisam de interação entre eles para fornecer subsídios necessários para os seus cálculos.

As demais categorias de habilidades adotadas no SWEETS foram: *Sugerida*, *Refutada* e *Calculada*. As habilidades *Sugeridas* e *Refutadas* precisam explicitamente das interações dos usuários no ambiente. As *Sugeridas* são aquelas em que os próprios usuários sugerem que o possível especialista possui. Da mesma forma, as *Refutadas* são as que os próprios usuários sugerem que um indivíduo não tem. Como ambas dependem da intervenção explíci-

ta dos usuários, é importante considerar a reputação e a maturidade dos usuários que sugerem ou refutam habilidades para um determinado indivíduo. Embora esta estratégia tenha sido adotada na abordagem do SWEETS, ainda há algumas falhas. O grau final de especialidade de um usuário  $u$  em um assunto  $h$  pode ser prejudicado se este possuir alto nível de colaboração no ambiente.

O grau final de especialidade de um indivíduo é determinado pelas 4 categorias de habilidades supracitadas, porém, ainda foi adotado no SWEETS uma quinta categoria de habilidade, a *Calculada*. Esta categoria de habilidade é baseada em folksonomia e ontologia (*lightweight*) para o seu cálculo. Habilidades inferidas nesta categoria podem também ser usadas no cálculo final de especialidade de um candidato a especialista em um assunto  $h$ . É necessário apenas que, uma vez que esta habilidade tenha sido calculada, os usuários sugiram ou refutem que um possível especialista a tenha ou não.

Outra importante estratégia adotada no SWEETS é a possibilidade de utilização do conhecimento dos usuários produzidos em ambientes externos ao que o SWEETS estiver implantado. O conhecimento produzido no ambiente em que SWEETS estiver implantado pode ser limitado e, com isso, a inferência das especialidades pode ser ineficiente. Quanto maior for o volume de conhecimento dos usuários disponível, maior a eficiência na inferência das especialidades destes indivíduos.

Em decorrência do SWEETS ser projetado para ser implantado em plataformas virtuais sociais, as relações sociais destes indivíduos podem ser exploradas para recomendar o especialista mais adequado para um determinado usuário. Vários elementos podem ser considerados, tais como, distância social dos usuários ou disponibilidades destes. Inicialmente, foi considerado no presente trabalho o nível das relações de confiança entre os indivíduos para prover o especialista mais adequado a um determinado usuário. O nível das relações de confiança entre os indivíduos é calculado a partir da abordagem proposta do T-SWEETS apresentada no Capítulo 4.

## 6 Estudo de Caso *Konnen*

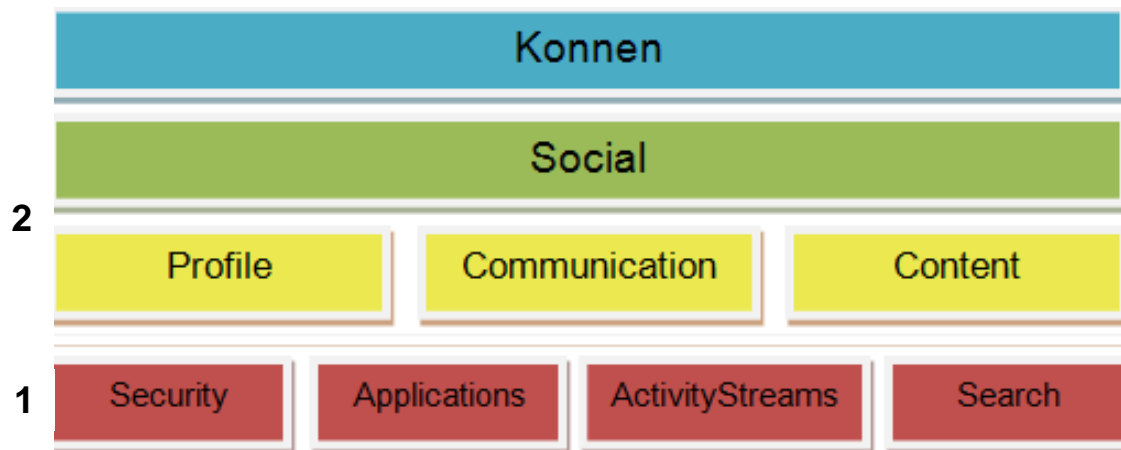
O presente capítulo apresenta o estudo de caso *Konnen*, que é uma plataforma de redes sociais que possui as características exigidas para a implantação das abordagens propostas neste trabalho: T-SWEETS, para recomendar relacionamentos ocultos de confiança entre os usuários da rede; e SWEETS, para recomendar especialistas em determinados assuntos. O capítulo está organizado da seguinte forma: Seção 6.1 apresenta a estrutura básica e necessária do *Konnen* para o trabalho; Seção 6.2 apresenta o processo de implantação do T-SWEETS na plataforma e a Seção 6.3 apresenta o processo de implantação do SWEETS no *Konneni*.

### 6.1 Estudo de caso *Konnen*

*Konnen* (*Knowledge Organization in a Native Network ENvironment*) é uma plataforma de redes sociais projetada para gestão organizacional. Esta plataforma está sendo desenvolvida em caráter experimental e faz parte do grupo de pesquisa Engenharia Inteligente de Dados do CELP/ULBRA<sup>11</sup>. A Figura 6.1 apresenta a arquitetura desta plataforma.

---

<sup>11</sup> <http://www.ulbra-to.br>

Figura 6.1: Arquitetura do *Konnen*

Conforme pode ser observada na **Figura 6.1**, a arquitetura da plataforma *Konnen* é organizada em camadas, assim permite ao sistema uma maior flexibilidade no que tange à projeção, ao desenvolvimento e à implantação de novos módulos ou subsistemas. A camada 1 da arquitetura não possui relação direta com os usuários do sistema, isto é, são aplicações importantes para aspectos internos do sistema, mas que não possuem uma interfaces para que os usuários possam interagir diretamente com elas. Alguns exemplos de aplicações destas camadas são: *Security*, responsável por definir as políticas de segurança de acesso à plataforma; *Search*, sistema de indexação e busca; e *ActivityStreams*, que faz o gerenciamento do *log* das ações dos usuários na plataforma, assim há como notificar aos demais membros da plataforma as ações dos usuários em um instante de tempo *T*. Já a camada 2 da arquitetura contém as aplicações de comunicação (*Communication*) e Conteúdo (*Content*), que provêem aos usuários funcionalidades específicas para a produção e compartilhamento de conhecimento. Além dessas, há também o módulo de *Profile*, que permite a definição dos perfis de opinião e conhecimento dos usuários. A Figura 6.2 apresenta uma visão da tela pós-login da plataforma *Konnen*.

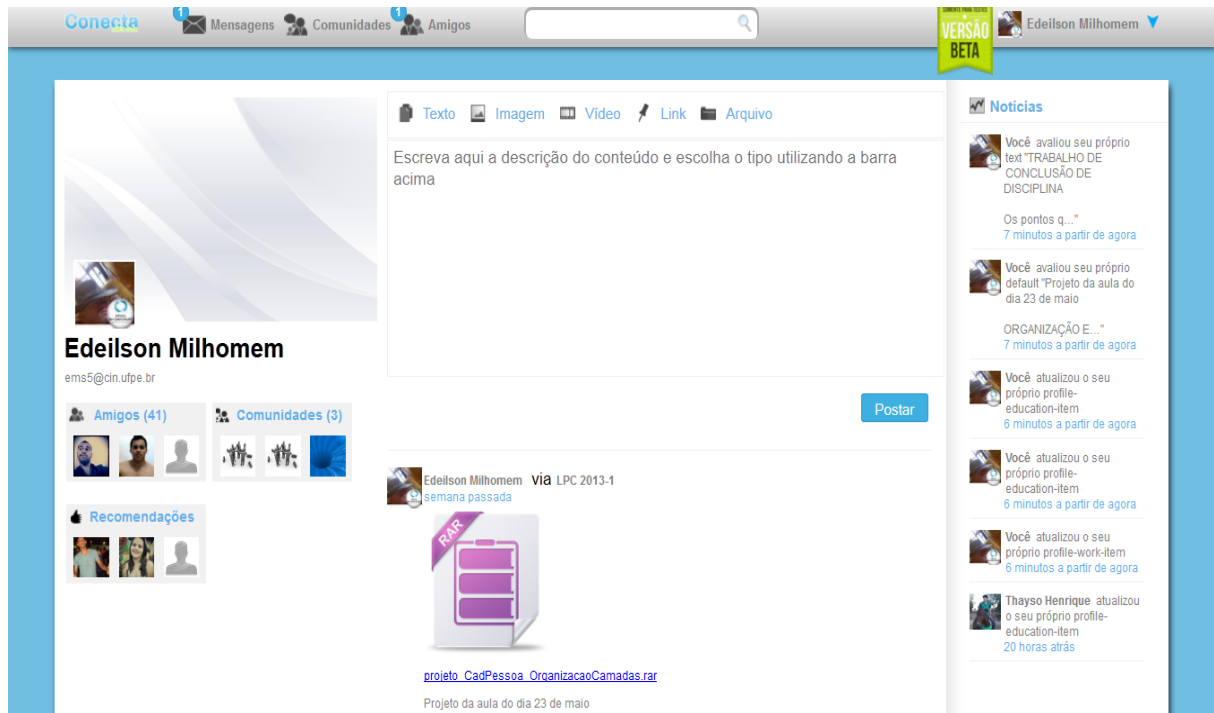


Figura 6.2: Tela principal pós-login do *Konnen*

As funcionalidades do *Konnen* que são primordiais para o presente trabalho são:

- Perfis:** cada usuário da plataforma possui seu perfil. Parte do perfil é preenchida explicitamente por estes usuários com informações simples, tais como, nome completo, apelido, endereço de e-mail e conta do Skype. Há a segunda parte do perfil que é determinada automaticamente pelo sistema de maneira transparente para os usuários, ou seja, não é explicitamente informada pelos usuários. Este perfil determinado implicitamente é composto por três partes: (1) a gerência do índice de atividades dos usuários no ambiente, que é determinado à medida que eles participam das atividades de produção e compartilhamento dos conhecimentos publicados na rede social; (2) a determinação do conjunto de assuntos mais relevantes para cada usuário é de acordo com o conhecimento que estes produzem no ambiente, isto é, todo o conteúdo textual produzido pelos usuários é representado em centróides, que contêm um conjunto de termos (palavras) e respectivos pesos, que representam a frequência que estes termos aparecem. Termos como preposições, artigos e interjeições são desconsiderados, uma vez que são classificados como termos irrelevantes; e (3) a identificação dos assuntos de interesses dos usuários, a partir dos termos mais relevantes presentes nos itens de conhecimento qualificados pelos usuários. Estes itens são



aqueles em que os usuários atribuem uma nota que determina o nível de interesse dos usuários em relação ao item qualificado.

- **Objetos:** O sistema permite a adição de conhecimento através de vários tipos de objetos, tais como documentos, artigos, planilhas, vídeo ou URLs que referenciam recursos externos, por exemplo, uma notícia. Com o objetivo de promover uma maior colaboração entre os indivíduos da rede, qualquer objeto pode receber comentários. Assim, aumenta-se a probabilidade de surgirem novos diálogos sobre o conhecimento representado no objeto. Os objetos podem ser postados na linha do tempo dos usuários ou em uma comunidade específica. Além disso, é possível também que qualquer usuário compartilhe com qualquer indivíduo da sua rede de relacionamentos, ou com uma comunidade, um objeto postado na rede.
- **Relacionamentos:** O *Konnen* permite a criação explícita de relacionamentos a partir de convites de amizade enviados de uma pessoa A para uma pessoa B. Assim, um usuário pode adicionar qualquer membro da rede a sua rede de contatos. Os relacionamentos estabelecidos não são necessariamente bidirecionais. No momento em que o usuário A adiciona o usuário B como seu contato, o usuário B é notificado, e poderá também adicionar o usuário A como seu contato, ou então ignorar o aviso e manter o relacionamento unidirecional de A para B.
- **Comunidades:** as comunidades representam um grupo que reúne um conjunto de usuários que possuem algum interesse em comum, isto é, é um local em que os membros da rede social podem promover discussões sobre assuntos específicos. Uma comunidade oferece um fórum com o principal objetivo de dar suporte à criação e ao compartilhamento de conhecimento entre os membros da rede. Qualquer membro pode iniciar novos tópicos sobre qualquer assunto de interesse. O sistema permite a criação de comunidades por qualquer membro da rede.
- **Mecanismo de Avaliação:** Existe também na plataforma a possibilidade dos usuários realizarem avaliações sobre determinados itens. Estes itens podem ser os conteúdos produzidos pelos usuários ou os próprios usuários provenientes da rede de contatos de um usuário. A avaliação sobre os conteúdos pode ser realizada de duas maneiras distintas: (1) a partir da adição de comentários aos objetos ou posts, enriquecendo assim, naturalmente, estes itens com novos conhecimentos e diálogos; (2) a partir de uma avaliação positiva ou negativa destes itens realizada pelos usuários, gosta (*like*) e não gosta (*unlike*), isto é, uma avaliação qualitativa. Assim, cada

membro da rede pode atuar como um revisor dos conteúdos postados pelos demais membros da rede.

- **Folksonomia:** Na plataforma existe a possibilidade dos usuários classificarem o conhecimento armazenado, através do uso de folksonomia. Assim, existe uma facilidade na descoberta de informações publicadas na rede pelos usuários. Isto é, os usuários podem classificar os conteúdos publicados na plataforma de uma forma social e colaborativa (*e.g.* comunidades e comentários). Para isso, é necessário apenas que os usuários associem tags aos itens que são classificados. Uma alternativa importante que poderia ser adotada pelo *Konnen* para descoberta de conhecimento em plataformas sociais, já que explora o conceito de folksonomia é a adoção de nuvem de tags (*tagcloud*). A nuvem de tags permite a visualização de todos os marcadores criados pelos usuários, destacando aqueles que mais ocorrem, *i.e.*, os marcadores que mais ocorrem (maior frequência) têm maior tamanho em relação àqueles que menos ocorrem. Assim, os usuários poderiam rapidamente acessar qualquer conteúdo associado com uma tag específica.

## 6.2 Implantação do T-SWEETS no *Konnen*

T-SWEETS, conforme apresentado no Capítulo 4, foi projetado para ser implantado em ambientes virtuais sociais. Estes ambientes virtuais sociais devem possuir algumas características básicas, são elas:

1. a representação de relacionamentos diretos e indiretos entre pessoas;
2. a existência de colaboração entre as pessoas, a partir da produção e compartilhamento de conhecimento;
3. a avaliação dos conhecimentos produzidos pelos usuários.
4. a atribuição de um grau de confiança de um usuário  $u$  em relação a um usuário  $a$ ;

O *Konnen* atende todas as exigências necessárias para a implantação do T-SWEETS. O quesito Relacionamentos (Seção 6.1) contempla o item 1 da lista de pré-requisitos supracitados, a partir da possibilidade de criação de relacionamentos diretos entre as pessoas, que subsidia a descoberta de relacionamentos indiretos. Os Objetos, Comunidades e Relacionamentos (Seção 6.1) oferecem suporte ao item 2. Já o Mecanismo de Avaliação oferece suporte aos itens 3 e 4.

O T-SWEETS implantado no *Konnen* permite a descoberta de relacionamentos ocultos de confiança entre os membros da rede social. A descoberta destes relacionamentos de confiança é um elemento de apoio ao Modelo 3C: comunicação, cooperação e coordenação (Ellis, 1991). Os relacionamentos de confiança entre as pessoas em ambientes colaborativos sociais servem como estímulo à comunicação e cooperação e pode facilitar a coordenação, pois a pessoa que coordena atividades em grupo pode ter informações de confiança sobre as pessoas do grupo. Como consequência, as pessoas têm uma tendência a produzirem um volume maior de conhecimento, que pode servir de insumo para a inferência de novos relacionamentos de confiança – de acordo com a Figura 6.3.

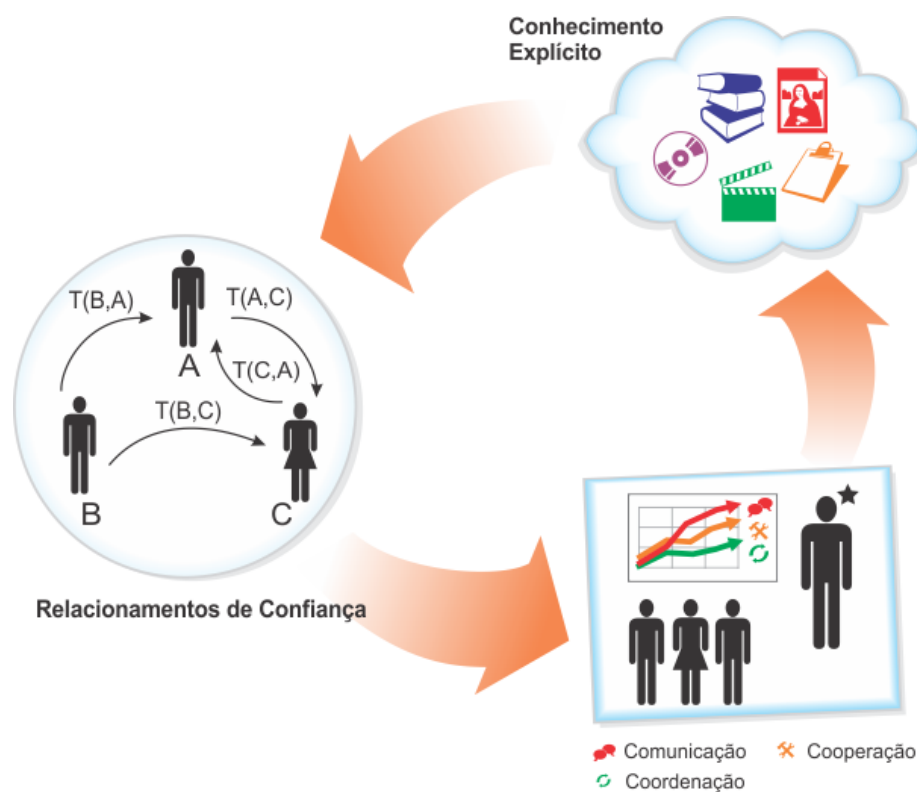


Figura 6.3: Relações de confiança como apoio ao Modelo 3C

A análise do impacto de um sistema que recomende relacionamentos ocultos de confiança em uma plataforma virtual social foi um dos objetivos que motivou a implantação do T-SWEETS na plataforma de aprendizagem organizacional *Konnen*.

### 6.2.1 Elementos do T-SWEETS no *Konnen*

Os elementos do T-SWEETS são: Similaridade entre Perfis dos Usuários, Relacionamentos de Confiança, Nível de Maturidade e Reputação.

- **Similaridade entre Perfis:** a similaridade entre os perfis considera duas categorias dos perfis dos usuários: conhecimento e opinião. O perfil de conhecimento considera todas as produções dos usuários no ambiente. Assim, são extraídos os termos (palavras) dos documentos textuais produzidos no ambiente e representado em um espaço vetorial determinado por um centroide. Um centroide contém um conjunto de termos e respectivos pesos. O peso de cada termo é determinado pela frequência deste em relação às produções do usuário. Os termos que possuem menor relevância, tais como artigos, preposições, conjunções são desconsiderados. Esta é uma estratégia da área de Recuperação da Informação intitulada *stopwords* (Baeza-Yates, 1999). Os perfis de opinião dos usuários consideram todas as avaliações que estes usuários fazem no ambiente. No *Konnen* é possível avaliar objetos, posts, comentários aos posts de forma positiva (*like*) ou negativa (*unlike*). Com estas avaliações é possível obter os usuários que possuem opiniões similares, a partir do conjunto de avaliações aos itens em comum. As equações de similaridade (4.1) e (4.2) (Capítulo 4) entre os perfis de conhecimento e de opinião dos usuários são baseadas na equação do *Coseno*. E o grau final da similaridade entre os perfis é dado pela equação (4.3).
- **Relacionamentos de Confiança:** Os relacionamentos de confiança consideram as avaliações explícitas de confiança que os usuários atribuem a outros. Assim, é possível inferir a probabilidade de um usuário qualquer confiar em outro, dado que estes não estejam conectados explicitamente (relacionamentos diretos). No *Konnen*, é possível os usuários avaliarem explicitamente o nível de confiança que possuem em outros, a partir da atribuição de uma nota, que varia no intervalo de 0 a 10.
- **Nível de Maturidade:** objetiva mensurar o quanto as produções dos usuários são vistas pelos demais usuários no ambiente. No *Konnen* há o mecanismo de avaliação dos conteúdos (posts, objetos e comentários) que oferece suporte para determinar a maturidade dos usuários no ambiente. Este mecanismo de avaliação é binário: gosta (*like*) e não gosta (*unlike*). A expectativa é que se a escala de avaliação fosse maior, por exemplo, 1-5, a avaliação da maturidade do usuário poderia ser mais eficiente. Em contrapartida, acredita-se que uma ava-

liação neste nível poderia prejudicar o interesse dos usuários em realizar avaliações no ambiente.

- **Reputação:** a reputação de um indivíduo é determinada pela visão geral que um grupo de pessoa tem em relação a outra. No *Konnen* é possível mensurar a reputação de um indivíduo a partir das atribuições de confiança que são dadas explicitamente entre os usuários.

Conforme apresentado no Capítulo 4, uma vez obtidos o grau de cada um dos elementos supracitados, é possível determinar o grau final de confiança que um usuário possa ter em relação a outro. As recomendações dos possíveis relacionamentos de confiança a um usuário é ordenada de forma decrescente de acordo com este grau final. Porém, para que o usuário possa ter parâmetros para decidir se aceita ou não a recomendação gerada, foi desenvolvido um *widget* que apresenta ao usuário a justificativa do que foi considerado para recomendar aquele usuário – conforme apresentado na Figura 6.4 (fora de foco propositalmente para proteger a imagem do usuário).



Figura 6.4: *Widget* para justificação das recomendações

Quando o usuário passa o mouse sobre as imagens dos perfis das pessoas recomendadas o *widget* com as justificativas é apresentado. Nesta justificativa, para cada elemento do T-SWEETS é apresentado o percentual equivalente que foi inferido, por exemplo, o usuário João foi recomendado para Pedro porque tem 75% de reputação e 68% de maturidade no ambiente. Somente usuários que tiveram um grau final de confiança maior ou igual a 70% foram recomendados.

Com a implantação do T-SWEETS na plataforma *Konnen*, há uma expectativa que este mecanismo de recomendação de relacionamentos de confiança possa subsidiar uma maior interação e cooperação entre os membros da rede. Esta maior interação entre os membros contribui diretamente para uma maior produção/explicação de conhecimento destes indivíduos.

O conhecimento explicitado destes indivíduos é premissa fundamental para a identificação automática das especialidades/habilidades destas pessoas e assim, naturalmente, pode tornar a recomendação de especialistas de domínio mais eficiente. Neste sentido, a Seção 6.3 apresenta a implantação do sistema de recomendação SWEETS no *Konnen*.

### **6.3 Implantação do SWEETS no *Konnen***

SWEETS, conforme apresentado no Capítulo 5, foi projetado para ser implantado em diferentes plataformas, desde que estas atendam a alguns requisitos, são eles:

- Ambientes virtuais sociais, ou seja, ambientes em que as pessoas mantêm relacionamentos sociais, que subsidia a interação entre os indivíduos;
- E conhecimentos associados aos usuários, que sustentam a descoberta das especialidades dos usuários.

Uma das vantagens do formalismo proposto pelo SWEETS é que este permite a exploração de conhecimento dos usuários produzidos em ambientes externos ao que estiver implantado. Assim, há como ter maiores parâmetros para identificar com mais eficiência as especialidades de cada usuário.

A plataforma *Konnen* atende a todas as exigências supracitadas e a arquitetura proveniente da implantação do SWEETS no *Konnen* é apresentada na Figura 6.5.

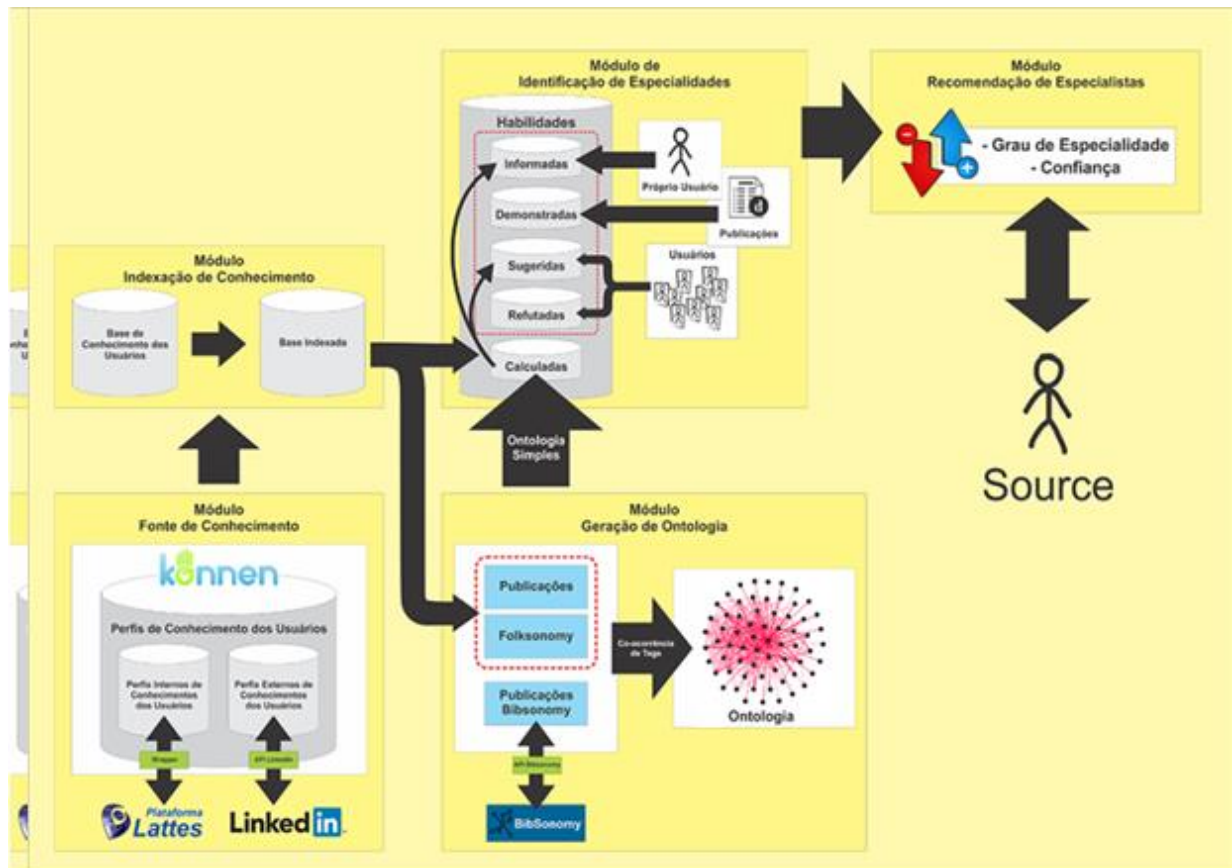


Figura 6.5: Arquitetura do SWEETS implantado na plataforma *Konnen*

A arquitetura (Figura 6.5) está de acordo com a arquitetura básica do SWEETS (Figura 5.1), mas com as suas particularidades. O processo de integração e o conjunto de particularidades adotadas são apresentadas nas seções seguintes.

### 6.3.1 Módulo Fonte de Conhecimento

O Módulo Fonte de Conhecimento representa a fonte de conhecimento dos usuários. Estes conhecimentos podem ser apenas os conhecimentos produzidos internamente no ambiente em que o SWEETS estiver implantado, ou podem ser os conhecimentos produzidos em diferentes plataformas externas. A implantação do SWEETS no *Konnen*, além de explorar este conhecimento interno ao *Konnen*, também explorou os conhecimentos produzidos na plataforma *Lattes*<sup>12</sup> e na Rede Social *LinkedIn*<sup>13</sup>.

<sup>12</sup> É uma plataforma nacional que disponibiliza uma base de currículos dos pesquisadores brasileiros. Disponível em: <http://lattes.cnpq.br/>

<sup>13</sup> Maior rede social profissional do mundo com mais de 225 milhões de profissionais na plataforma. Disponível

Para a extração de conhecimento destas plataformas, foi desenvolvido um mecanismo automático de coleta de conhecimento destes usuários para realizar a integração dos perfis. Cada uma das plataformas possui peculiaridades diferentes em sua arquitetura, logo, o mecanismo de coleta desenvolvido foi independente. A extração dos dados da plataforma *Lattes* foi realizada a partir de um *wrapper*, em que extrai as informações públicas disponíveis no currículo *Lattes* e as representam em um modelo relacional. Já para a extração dos dados da *LinkedIn* foi usado o protocolo de comunicação *OAuth*. O *OAuth* é um protocolo aberto que permite o acesso seguro às ferramentas de autenticação por aplicações Convencionais e Web (PBWorks, 2010). A Figura 6.6 apresenta uma arquitetura do processo de integração dos perfis dos usuários do *Konnen* com as plataformas externas.

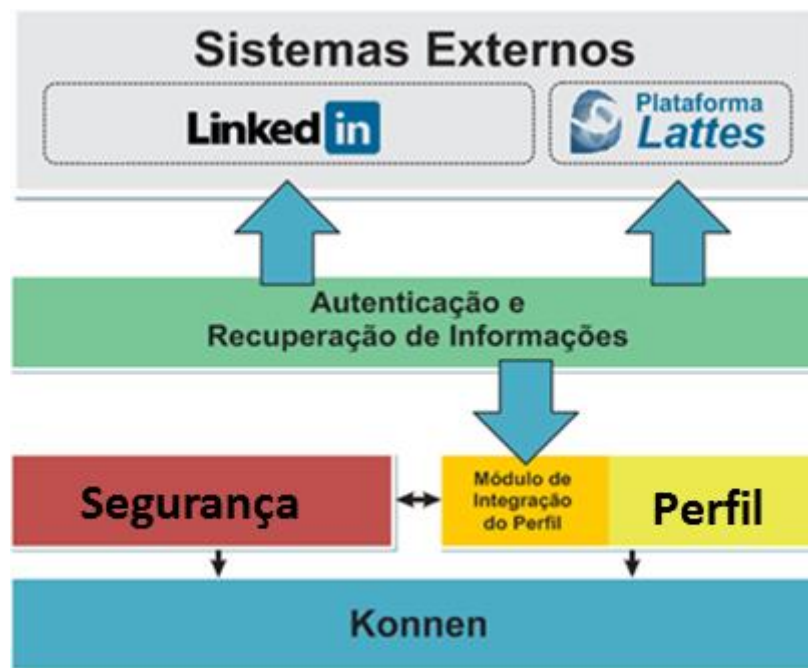


Figura 6.6: Integração dos perfis externos do *LinkeDin* e *Lattes* ao do *Konnen*

A integração dos perfis dos usuários no *Konnen* está relacionada ao aplicativo *Profile* e *Security*. Há uma relação com o *Profile* porque é este que gerencia as informações relacionadas aos usuários e as compartilha com os demais aplicativos do *Konnen*. Já a relação com o aplicativo *Security* é porque há um controle de acesso dos usuários, isto é, o sistema executa funções de autenticação na plataforma. As operações de autenticação e recuperação de informações são realizadas a partir de alguns protocolos descritos a seguir.

### ***Integração com o LinkedIn***

em: <https://www.linkedin.com/>



Conforme já mencionado, a integração com o perfil da rede social *LinkedIn* é baseada no protocolo de comunicação *OAuth*. O protocolo é baseado em dois tipos de sistema: cliente (que solicita e recupera informações) e fornecedor (que disponibiliza as informações). O cliente é a plataforma *Konnen* e o fornecedor é a *LinkedIn*. Ao utilizar o protocolo *OAuth* é necessário que o sistema cliente possua um conjunto de credenciais (*client identifier* e *client secret*) que o permitirão autenticar-se no sistema fornecedor. O *client identifier* é o identificador público do sistema cliente do processo de autenticação, este identificador é enviado para o sistema fornecedor, visivelmente na *url* de redirecionamento. Por outro lado, o *client secret* é uma chave secreta que é utilizada na comunicação direta entre o sistema cliente e o fornecedor, sem interação com o usuário, assim não é possível visualizá-la. A partir destas credenciais, é possível que o sistema cliente consiga fazer a autenticação e recuperação de informações do usuário, uma vez que a autenticação seja realizada com sucesso e a autorização seja concedida para o acesso de informações. O processo de recuperação de informações da rede social *LinkedIn* é apresentado na Figura 6.7.

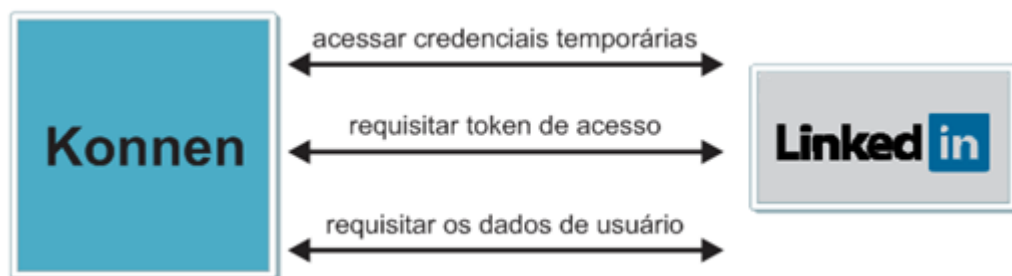


Figura 6.7: Procolo *OAuth* para conexão com a *LinkedIn*.

Para que haja esta comunicação, o primeiro passo é a plataforma cliente (*Konnen*) requisitar as credenciais de acesso ao fornecedor (*LinkedIn*). Uma vez que o cliente tenha obtido as credenciais temporárias, o cliente pode redirecionar um usuário para o sistema fornecedor. Neste momento, o usuário terá que fornecer os dados de autenticação da plataforma fornecedora e permitir que o sistema cliente possa acessar suas informações disponibilizadas no fornecedor. Se a autenticação for realizada, então a importação dos dados dos usuários disponíveis no *LinkedIn* é realizada.

As informações do perfil do *LinkedIn* que podem ser integradas ao perfil da plataforma *Konnen* foram:

1. básicas do usuário, *e.g.* e-mail;
2. imagem do perfil do usuário;
3. profissões do usuário;

4. informações educacionais do usuário;
5. línguas que os usuários têm conhecimento;
6. habilidades dos usuários
7. informações sobre cursos que o usuário cursou.

Deste universo de informações importadas, aquelas que foram utilizadas no presente trabalho foram as “habilidades dos usuários”, que fornecem parâmetros para identificação das especialidades dos indivíduos. As informações sobre línguas, cursos e profissões dos usuários também poderiam ser exploradas, pois fornecem parâmetros que podem ser utilizados para a identificação das especialidades dos indivíduos. Todavia, ainda não foi considerada uma forma de adotá-las no formalismo do SWEETS.

### ***Integração com a Plataforma Lattes***

Conforme já mencionado, para integração do perfil do usuário do *Konnen* com o perfil do da Plataforma *Lattes* foi desenvolvido um *Wrapper*, que coleta automaticamente as informações públicas da plataforma.

O perfil público de uma pessoa na plataforma *Lattes* não disponibiliza todas as informações previamente cadastradas por ele na plataforma. Porém, há alternativas que podem ser usadas para acessar estas informações: totalmente completas ou praticamente completas. As totalmente completas estão em um documento XML, que está disponível para *download* na plataforma; e as informações parcialmente completas estão na opção detalhada, também disponíveis na plataforma. A Figura 6.8 mostra o protocolo de comunicação com a plataforma *Lattes*.

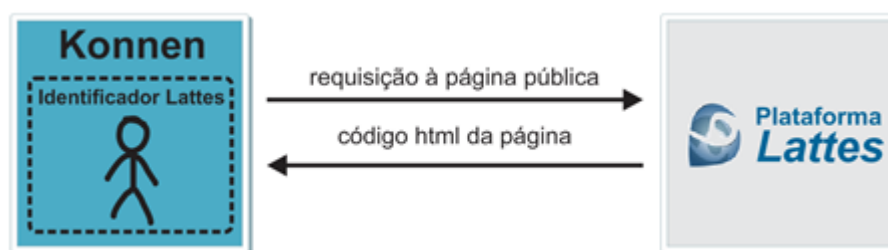


Figura 6.8: Procolo de comunicação com a plataforma *Lattes*

Com estas possibilidades de acesso aos dados dos usuários, foi desenvolvido um mecanismo em que o usuário informa o identificador público do seu currículo na plataforma *Lattes*, e, a partir disso, o sistema faz a extração de dados desta página pública do perfil do usuário. Neste momento, é importado também o documento XML, que contém o currículo com-

pleto do usuário. É possível, também, o usuário fazer apenas *upload* do seu currículo *lattes*, após tê-lo importado diretamente da plataforma *Lattes*.

As informações dos currículos disponíveis na plataforma *Lattes* que podem ser integradas aos perfis dos usuários no *Konnen*, foram:

1. gerais, tais como, nome e sobrenome e sexo;
2. educacionais: formação acadêmica e titulação;
3. atuação profissional: locais, profissões e datas;
4. projetos de pesquisa onde tem atuação;
5. áreas de atuação;
6. produções e publicações: científicas, técnicas e artísticas;
7. eventos que o usuário participou.

Dentre este conjunto de informações, as exploradas para a identificação das especialidades dos candidatos a especialistas foram as “produções e publicações”. As informações de publicações na plataforma *Lattes* estão organizadas em categorias, são elas: Livros ou capítulos de livros, periódicos, completos em anais de eventos e resumos expandidos em anais de eventos. Em decorrência desta categorização e a disponibilização das palavras-chave destas publicações no perfil público do *Lattes*, estas foram as informações consideradas para a identificação das especialidades dos usuários no presente trabalho.

Com a realização de adaptações no formalismo do SWEETS, ainda poderiam ser usadas outras informações que fornecem parâmetros para a identificação das especialidades dos indivíduos, são elas: informações de projetos de pesquisa onde tem atuação; de atuação profissional; educacionais; ou de áreas de atuação;

A partir destas fontes externas de conhecimento dos usuários: *LinkedIn* e *Lattes*, acredita-se que haja conhecimento suficiente para a inferência das habilidades dos usuários. A seção 6.3.2 apresenta o processo de indexação do conhecimento dos usuários no *Konnen*.

### **6.3.2 Módulo Indexação do Conhecimento no *Konnen***

A indexação do conhecimento realizada pelo presente módulo é importante para otimizar o processo de inferência de especialidades dos candidatos a especialistas. Todo o conhecimento proveniente do Módulo Fonte de Conhecimento é então armazenado em uma base de índices.

A base de índices representa uma visão lógica dos perfis dos usuários. A visão lógica é representada por um conjunto de palavras-chave (termos) associadas aos seus respectivos pesos, ou seja, índices-termos. O peso determina o grau de relevância para um determinado termo. Esta relevância do termo é baseada na técnica TF (*Term Frequency*) (Baeza-Yates, 1999). O TF é determinado pela razão da frequência de um termo pelo termo que mais ocorre. Assim, é possível definir um conjunto de termos que melhor representa o perfil de um usuário. O Quadro 6.1 apresenta a exemplificação de uma visão lógica do perfil de um usuário.

Quadro 6.1: Visão lógica do perfil indexado de um usuário

	$k_1$	$k_2$	...	$K_{n-1}$	$K_n$
$p_1$	$w_{1,1}$	$w_{1,2}$	...	$w_{1,n-1}$	$w_{1,n}$

Conforme representação no Quadro 6.1,  $k_i$  representa os  $i$  termos,  $p_j$  representa os  $j$  perfis do usuário e  $w_{i,j}$  representa os pesos dos termos no perfil. ( $k_i$ ,  $p_j$ ), sendo que  $w_{i,j} \geq 0$ . Uma vez que a base de índices esteja criada, é possível realizar um rápido acesso ao perfil de conhecimento do usuário.

Na visão lógica de termos-índices do perfil de um usuário no *Konnen*, apresentada no Quadro 6.1, são consideradas as seguintes fontes de informação: posts e comentários dos usuários em sua linha do tempo ou comunidades que participam, postagens de objetos (.doc e rtf), links e os conteúdos são importados da Plataforma *Lattes e LinkedIn*. As palavras chave provenientes das publicações científicas disponíveis na Plataforma *Lattes* podem ser compostas e também são consideradas na indexação.

Este processo de indexação foi programado na plataforma *Konnen* para ser realizado em um período de a cada 7 dias e durante os horários que tiverem menos uso na plataforma (nas madrugadas). Para isso, foi utilizado um Framework para agendamento de tarefas Quartz.NET<sup>14</sup>. Assim, o perfil de conhecimento do usuário é atualizado automaticamente e de maneira periódica.

### 6.3.3 Módulo Geração de Ontologias

A ontologia (*Lightweight Ontology* -  $O_{c,i}$ ) originada no Módulo Geração de Ontologia é baseada na categorização dos objetos de uma folksonomia. Esta ontologia é importante para a in-

<sup>14</sup> Quartz.NET. Disponível em: <http://quartznet.sourceforge.net/index.html>

ferência de uma das categorias de habilidades proposta na abordagem do presente trabalho, *Habilidade Calculada* (Capítulo 5).

Assim, o surgimento da ontologia é baseado na co-ocorrência das tags que classificam um determinado objeto no *Konnen*. Além disso, as palavras-chave associadas aos trabalhos de natureza científicas, e.g. monografias, dissertações e artigos, também referem-se ao conceito de folksonomia, uma vez que os autores destas publicações usam um conjunto de palavras-chave para categorizarem estas publicações.

Como alternativa para o enriquecimento da ontologia  $O_{c,i}$ , foram usadas também as palavras-chave que categorizam as publicações científicas. Mesmo com esta estratégia, o número de termos na ontologia  $O_{c,i}$  ainda estaria reduzido. Por isso, foram usadas também as publicações e suas palavras-chaves disponíveis no BibSonomy<sup>15</sup>. Foi desenvolvido um mecanismo que importa automaticamente as informações do BibSonomy, que utiliza a API (*Application Programming Interface*) pública disponibilizada pela própria plataforma. A partir disso, as publicações e referidas palavras chave são utilizadas para o enriquecimento da ontologia ( $O_{ci}$ ).

Neste processo de importação, é utilizada uma palavra chave para dar início ao processo. A partir desta palavra chave, busca-se por trabalhos disponíveis na plataforma *Bibsonomy* que a citam, assim, naturalmente, tem-se todos os termos que costumam co-ocorrer em relação a esta palavra chave. O ideal é que esta palavra chave seja escolhida de acordo com a relevância que esta tem para o ambiente em que o mecanismo para a identificação de especialista estiver sendo implantado. No *Konnen*, dentre o grupo de usuários que participou dos experimentos, a palavra mais relevante era *Ontologia*. Mais detalhes sobre são apresentados no capítulo correspondente.

#### 6.3.4 Módulo de Identificação e Recomendação de Especialidades

Em relação ao Módulo Identificação de Especialidades, são identificadas as habilidades dos indivíduos em 4 categorias: *Demonstrada*, *Sugerida*, *Informada* e *Refutada* – conforme apresentadas no capítulo 5. Para o presente cenário foram considerados os pesos para cada categoria, conforme o Quadro 6.2.

---

<sup>15</sup> É um sistema social de bookmarks e compartilhamento de publicações. Disponível em: <http://www.bibsonomy.org/>

**Quadro 6.2:** Pesos para cada habilidade adotados no SWEETS

Habilidade ( $h$ )	Peso ( $w_h$ )
Demonstrada ( $w_{h_{dem}}$ )	5
Informada ( $w_{h_{des}}$ )	3
Sugerida ( $w_{h_{sug}}$ )	1
Refutada ( $w_{h_{ref}}$ )	-1

De acordo com a representação no Quadro 6.2, a *habilidade demonstrada* possui maior relevância, já que há uma comprovação para a respectiva habilidade. No SWEETS, foram exploradas as publicações científicas dos usuários disponibilizadas na Plataforma *Lattes* para a representação desta categoria de habilidade. O segundo nível de importância das habilidades são as *informadas*, que consideram as habilidades fornecidas explicitamente pelos candidatos a especialistas. Tanto a *habilidade sugerida* quanto a *refutada* possuem intervenção explícita dos demais usuários que compõem o ambiente em que o SWEETS estiver implantado. Por isso, elas possuem pesos equivalentes, com a diferença que a *habilidade sugerida* é ponderada positivamente, e a *refutada* ponderada negativamente.

Não se sabe se as proporções destes pesos são ideais, porém, de acordo com os critérios apresentados anteriormente, acredita-se que o nível de importância esteja adequado. O ideal é desenvolver um trabalho de análise junto aos dados do ambiente em que o SWEETS estiver implantado para uma definição mais acertada destes pesos.

Para a categoria *Habilidades Demonstradas* foram usadas as publicações científicas das pessoas como fonte de conhecimento validada. A Plataforma *Lattes*, como fonte de informações sobre publicações científica explorada no SWEETS, organiza estas publicações em diferentes categorias, conforme apresentadas no Quadro 6.3.

**Quadro 6.3:** Categorias para organização das publicações científicas na Plataforma Lattes e respectivos pesos adotados no SWEETS.

Categorias ( $t$ )	Pesos ( $w_t$ )
Artigos completos publicados em periódicos e capítulos de livros publicados ( $w_{t_{accl}}$ )	5

Trabalhos completos publicados em anais de congressos ( $w_{t_{tc}}$ ).	2
Resumos expandidos publicados em anais de congressos ( $w_{t_{re}}$ )	1

Os pesos para as publicações apresentados no Quadro 6.3, que foram os adotados no SWEETS, estão em uma escala de 1 a 5, em que quanto mais próximo de 5, maior a relevância da publicação científica, da mesma forma, quanto mais próximo de 1, menor é a relevância da publicação. Estes pesos foram definidos de acordo com o grau de dificuldade de publicação para cada categoria (*e. g.* resumo expandido e trabalho completo). Porém, não se sabe se as proporções entre os pesos das categorias são as mais adequadas. Além disso, há situações que o peso de uma conferência poderia ser maior que de um periódico, dependendo dos critérios de aceitação que são utilizados. Por isso, o ideal é que seja utilizada uma fonte de conhecimento que mapeie a relevância dos eventos ou periódicos, por exemplo, o Qualis da CAPES.

As *Habilidades Informadas* são informadas explicitamente pelo usuário em seu perfil pessoal no *Konnen* ou *LinkedIn*. Já as habilidades nas categorias *Informada* e *Refutada* não têm sido aplicadas no presente estudo, embora façam parte da abordagem apresentada no SWEETS. Ambas as categorias dependem de interações dos usuários no ambiente. Para mensurar a eficiência destas habilidades, o ideal é que sejam implantadas e testadas em um ambiente organizacional real em que as pessoas o usam para o desenvolvimento de suas atividades diárias, como os trabalhos de Costa (2009). Além destas categorias de habilidades, há também a categoria *Habilidade Calculada*. Esta categoria explora as produções (*e.g. posts* em sua linha do tempo ou comunidades que participam) realizadas pelo usuário no *Konnen* como alternativa para enriquecer o perfil de especialidades do usuário. Os termos e seus relacionamentos definidos na ontologia  $O_{ci}$  que é originada no Módulo Geração de Ontologia é a base para inferência desta categoria de habilidade. Assim, o universo de especialidades possíveis para esta categoria está limitado às marcações (tags) usadas para categorizar os objetos que subsidiaram o surgimento da ontologia  $O_{ci}$  – informações mais detalhadas sobre o assunto são apresentadas no Capítulo 5.

Uma vez que o conjunto de especialistas em assuntos específicos tenha sido identificado pelo sistema, é possível realizar as recomendações para os usuários do *Konnen*. Como não há um conjunto de usuários usando assiduamente a ferramenta, não foi desenvolvida uma interface no *Konnen* que possibilite a recomendação de especialistas para os usuários. Apesar

disso, por causa da abordagem adotada no presente trabalho, é possível uma pessoa ordenar os especialistas recomendados para ele de acordo com o grau de especialidade ou de acordo com o que acredita que esteja mais propenso a colaborar, a partir do grau de confiança que o usuário que recebeu a recomendação tem em relação ao especialista recomendado.

## 6.4 Considerações Finais

Este capítulo apresentou a ferramenta de redes sociais *Konnen*, dando ênfase às principais funcionalidades que são essenciais ao presente trabalho, que são: Perfis, que representam as preferências dos usuários; Objetos, que permitem a adição de conhecimento no ambiente e, assim, fornecem mais parâmetros para o enriquecimento dos perfis dos usuários; Relacionamentos, que permitem às pessoas estabelecerem relações sociais e, assim, propiciar uma maior interação na plataforma; Comunidade, que reúne um grupo de pessoas que possuem interesses em comum; e o Mecanismo de Avaliação, que permite às pessoas avaliarem outras pessoas e os conteúdos produzidos por elas.

Depois disso, são apresentadas as implantações do T-SWEETS, que identifica e recomenda relacionamentos de confiança, e do SWEETS, que identifica e recomenda especialistas em assuntos específicos, no *Konnen*.

No processo de implantação do T-SWEETS no *Konnen* foram apresentados seus elementos e como cada um destes elementos é usado na plataforma. A Similaridade entre Perfis, que considera as preferências dos usuários para aproximar usuários que possuem conhecimentos em comum. Os Relacionamentos de Confiança, que consideram as avaliações explícitas de confiança que os usuários atribuem uns aos outros para calcular a probabilidade de um usuário confiar em outro, considerando que estes usuários não estejam diretamente conectados; O Nível de Maturidade, que faz uso do Mecanismo de Avaliação para determinar o quanto as produções dos usuários são positivas ou negativas perante os membros do ambiente. E, por fim, a Reputação, que considera uma visão geral que os indivíduos da rede têm em relação a um determinado usuário, que também é baseada nas atribuições de confiança que os usuários atribuem uns para os outros.

Quanto à implantação do SWEETS no *Konnen* também foram apresentados como cada um dos módulos é integrado ao ambiente. A integração do Módulo Fonte de Conhecimento com as plataformas *Lattes* e *LinkedIn*, que são essenciais para que haja parâmetros de inferência das especialidades dos usuários. A importação dos dados do sistema BibSonomy, como



fonte de informação para a geração da ontologia, que é premissa necessária para o cálculo de uma das categorias de habilidades, a Calculada. A Indexação do Conhecimento do usuário, como alternativa para tornar o mecanismo de cálculo de especialidades mais eficiente. E, por fim, os módulos de identificação e recomendação de especialidades, que utilizam as saídas dos módulos anteriores para determinar o grau final de especialidade de um usuário em um assunto e, com isso, recomendá-lo.

.

## 7 Resultados

Uma vez que foram definidas as abordagens do SWEETS, para identificação e recomendação de especialistas em ambientes virtuais, e do T-SWEETS, para a descoberta de relacionamentos ocultos de confiança em plataformas sociais, o presente capítulo apresenta os experimentos que foram realizados e os resultados obtidos. O experimento preliminar, apresentado na Seção 7.1, tem como objetivo a identificação de elementos que podem influenciar na decisão de colaboração entre pessoas; a Seção 7.2 apresenta os experimentos desenvolvidos com o T-SWEETS, que objetivam mensurar o impacto que a implantação da referida abordagem teria em um ambiente virtual baseado em redes sociais; na Seção 7.3 são apresentados os experimentos com o SWEETS, que têm como propósito apresentar as potencialidades oferecidas por sua abordagem, além de mostrar como o grau de confiança entre as pessoas podem influenciar na qualidade das recomendações. E, por fim, a Seção 7.4 apresenta as considerações finais.

### 7.1 Experimento Preliminar

Antes da formalização do algoritmo T-SWEETS e da inferência de habilidades (especialidades) do SWEETS, foi desenvolvido um trabalho de pesquisa de campo junto a um conjunto de pessoas para definir e medir alguns elementos que poderiam ser explorados no processo de inferência de confiança e entender o comportamento destas pessoas ao decidir por colaborar ou não com outra pessoa. Dentre os elementos analisados, buscou-se entender como as pessoas consideram o conceito de “confiança” (por exemplo, antes de buscar ajuda de outra pessoa a respeito de uma questão ou de um problema) e a importância deste conceito em ambientes colaborativos, isto é, se a confiança pode ser um dos pré-requisitos ao incentivo à colaboração.

O questionário da pesquisa, cujas características são demonstradas a seguir, foi submetido, via e-mail e pela plataforma CONECTA, a um universo de 103 pessoas, das quais um conjunto de 53 pessoas a responderam. As pessoas selecionadas para a pesquisa foram de diferentes áreas do conhecimento (*e. g.* computação, jornalismo, publicidade e psicologia) e que trabalham cotidianamente com computador.

Conforme já mencionado, alguns dos questionamentos foram direcionados para entender que critérios as pessoas costumam considerar ao buscar ajuda de outra pessoa. O primeiro deles foi:

**Q1** “*Nas últimas 5 vezes em que você solicitou ajuda, se você obteve um pronto retorno, responda a seguinte pergunta: Em quais das categorias as pessoas que te ajudaram se enquadram? Em uma escala de 1 a 5, assinale 1 para categoria que mais ocorre e 5 para a categoria que ocorre com menor frequência. Obs.: Os valores escolhidos não podem se repetir.*” – o resultado é mostrado no Gráfico 7.1.

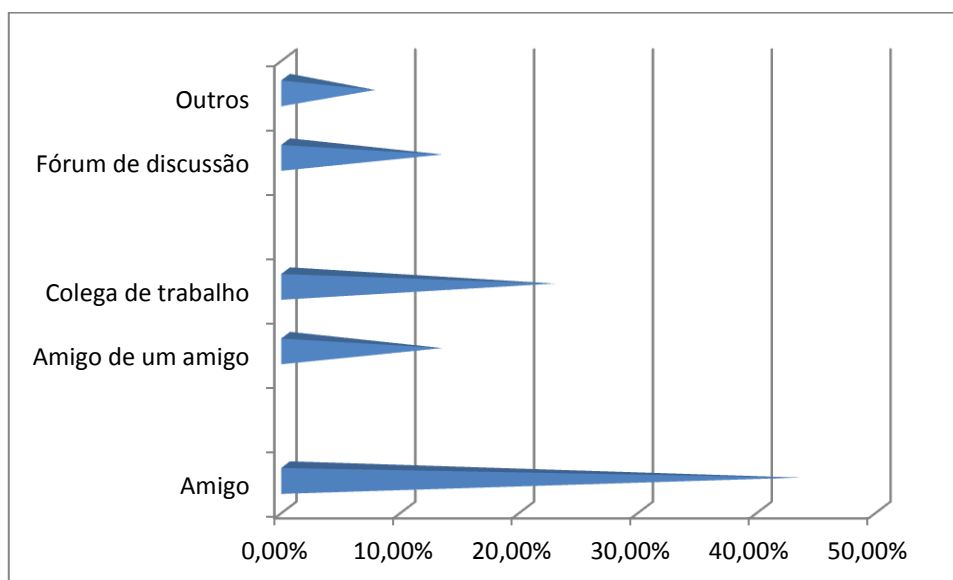


Gráfico 7.1: Conjunto de respostas à questão Q1.

Como resultados, obteve-se que 43,4% das pessoas costumam prioritariamente buscar ajuda com seus amigos (conforme o Gráfico 7.1), destes, embora não esteja explicitado no gráfico, como segunda e terceira opções, 47,83% costumam buscar ajuda com um amigo de um amigo e 43,3% com um colega de trabalho. Já 22,64% costumam prioritariamente buscar ajuda com um colega de trabalho, e os amigos ficam para um segundo plano. Os resultados obtidos são compreensíveis, uma vez que é natural as pessoas procurarem ajuda em seu contexto de convívio social.

Com isso, 66,04% das pessoas costumam prioritariamente buscar ajuda com pessoas relacionadas aos seus contextos sociais. Assim, prover um ambiente colaborativo social em que as pessoas tenham a possibilidade de aumentar seus laços sociais (quantidade de relacionamentos) é uma alternativa interessante para auxiliar no processo de resolução de problemas.

Uma das perguntas objetivou entender critérios que as pessoas consideram ao decidir colaborar com outra ou não.

**Q2:** “Das últimas 5 vezes que você ajudou alguém, houve alguma prioridade em ajudar um grupo específico de pessoas? Se sim, quais? Marque as opções em uma escala de 1 a 5, onde 1 indica prioridade maior e 5 prioridade menor”.

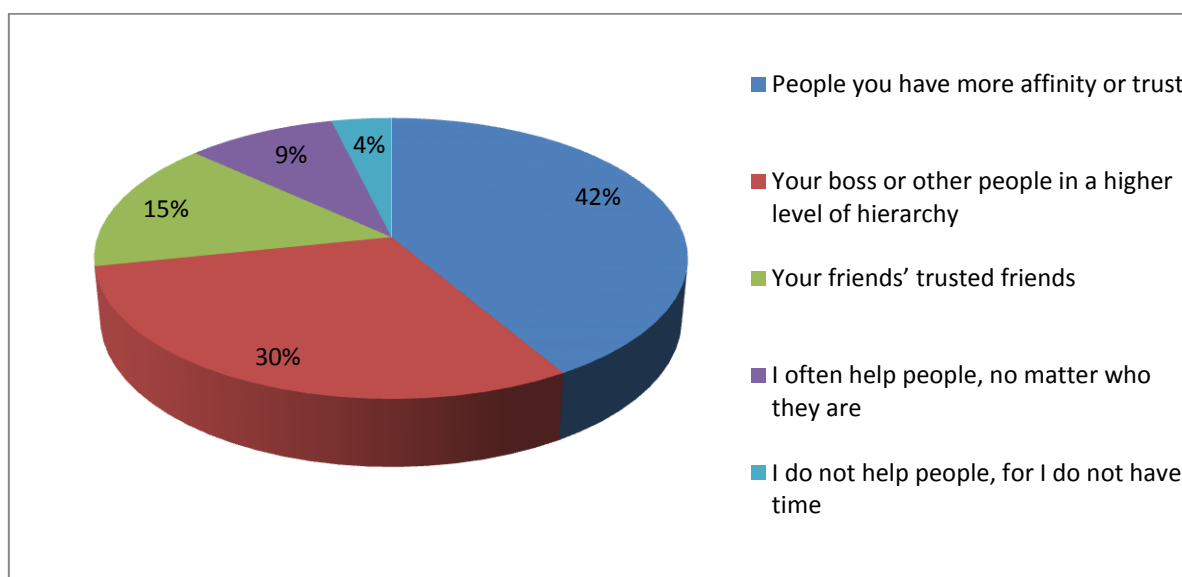


Gráfico 7.2: Conjunto de respostas à questão Q2.

Como respostas a este questionamento, constatou-se que aproximadamente 57% (42% mais 15% - Gráfico 7.2) das pessoas ajudam pessoas que possuem afinidade ou confiança direta ou indiretamente com elas.

Além disso, a reputação foi outro elemento importante considerado na pesquisa, o que auxilia a determinar o quanto uma pessoa é confiável perante um grupo de pessoas. Assim, foi aplicado o seguinte questionamento:

**Q3:** “Antes de solicitar ajuda a alguém, você costuma avaliar a reputação desta pessoa entre o grupo de pessoas do seu convívio?”.

Como respostas à Q3, 47,17% dos entrevistados a consideram diretamente ao decidirem buscar pela ajuda de uma determinada pessoa. Já 22,64% dos respondentes, disseram que

a consideram apenas às vezes. A maioria destes alegaram que a consideram apenas às vezes porque esta análise de reputação é feita implicitamente, isto é, fazem naturalmente, sem perceber.

Outro elemento que auxilia as pessoas decidirem se confiam ou não em outras é o nível de conhecimento que estas possuem sobre um assunto específico, este também foi um quesito abordado na pesquisa. Uma das questões voltadas para este contexto foi:

**Q4:** “Em uma escala de 1 a 5, considerando 1 para maior prioridade e 5 para menor prioridade: quanto você acredita que o nível de conhecimento de uma pessoa sobre assuntos específicos é considerado quando você decide confiar ou não nesta pessoa.?”.

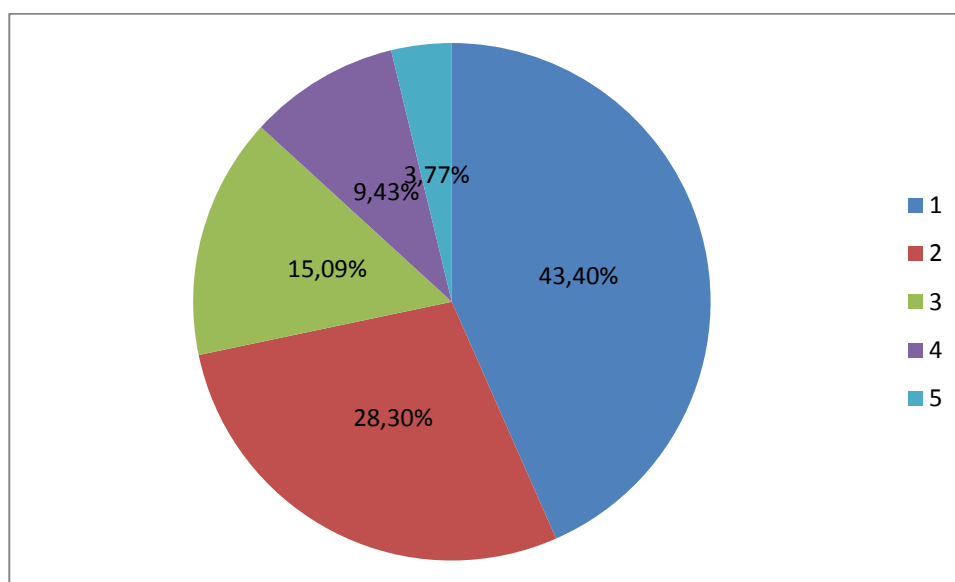


Gráfico 7.3: Conjunto de respostas à questão Q4.

De acordo com os dados apresentados no Gráfico 7.3, 71,70% das pessoas entrevistadas consideram que o nível de conhecimento que as pessoas possuem sobre assuntos específicos é primordial ao decidirem se confiam ou não nestas pessoas.

Nesta direção, foi questionado aos entrevistados o quanto eles acreditam que o nível de conhecimento de uma pessoa sobre um problema específico pode influenciar na decisão desta pessoa prestar ou não auxílio, isto é, decidir ou não colaborar. Como resposta, constatou-se que 79,25% das pessoas consideram que o nível de conhecimento sobre assuntos específicos é primordial na decisão de colaborar ou não com outra pessoa.

Ainda, sobre o nível de conhecimento que as pessoas possuem, foi realizado o seguinte questionamento:

**Q5:** “O quanto você acredita que as pessoas que possuem perfis de conhecimento similares ao seu estão propensas a colaborar com você, mediante a solicitação de ajuda? Responda de acordo com a escala de probabilidades “Muito alta”, “Alta”, “Baixa”, e “Muito baixa”. Obs: Se achar necessário, justifique sua resposta.”.

Como respostas à Q5, 92,45% dos entrevistados acreditam que as pessoas que possuem perfis de conhecimento e interesses similares estão mais propensas a colaborar entre si. Alguns respondentes justificaram que é natural as pessoas que possuem conhecimentos similares colaborarem e interagirem mais entre si, já que os interesses são comuns e, assim, as discussões tem uma tendência maior e natural a fluírem com maior facilidade.

Além desses, outro quesito importante levantado no questionário foi a importância da receptividade das pessoas na colaboração. Para isso, a seguinte questão foi aplicada:

**Q6:** “Antes de solicitar ajuda a alguém, você costuma avaliar a receptividade da pessoa com relação a você ou com relação às outras pessoas. A receptividade é o quanto esta pessoa mostra-se disposta a colaborar. .Obs.:Se achar necessário, justifique sua resposta. ”

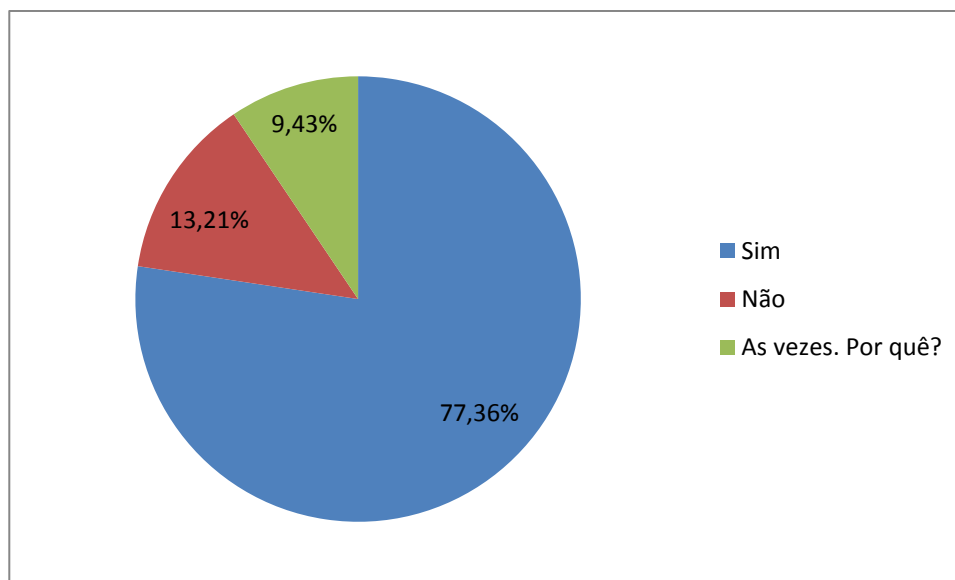


Gráfico 7.4: Conjunto de respostas à questão Q6.

Conforme o Gráfico 7.4, 77,36% dos entrevistados costumam avaliar o quanto as pessoas estão pré-dispostas a colaborar. De acordo com os entrevistados, a maioria das vezes, esta análise é realizada de forma implícita através de interações passadas. 9,43% das pessoas alegaram que realizam esta análise de receptividade somente às vezes. Dentre estes, alguns

alegaram que realizam esta análise se não estiverem em uma situação de emergência, por exemplo, necessitando de ajuda para resolver um problema urgente.

## 7.2 T-SWEETS

Foram realizados duas classificações de experimentos com o T-SWEETS: a primeira (seção 7.2.1) foi voltada para identificar o impacto que uma abordagem para inferência de relacionamentos ocultos confiança poderia ter após ser implantada em um ambiente colaborativo baseado em redes sociais; a segunda classificação dos experimentos foi para medir a eficácia do T-SWEETS em relação a alguns resultados aprestados no trabalho de Leskovec (2010), que apresenta um estudo de análise de confiança em relacionamentos implícitos que são extraídos de algumas fontes de dados disponíveis na web. Além disso, apresenta também uma análise em relação ao algoritmo clássico da literatura para inferência de confiança em ambientes sociais, *TidalTrust* (Golbeck, 2005).

### 7.2.1 Objetivos do Experimento 1

Conforme mencionado no Capítulo 6, o algoritmo T-SWEETS foi desenvolvido e integrado à instância *Conecta* da plataforma *Konnen*, um ambiente colaborativo educacional baseado em Redes Sociais, que está em fase experimental. Esta integração junto à rede *Conecta* tinha os seguintes objetivos:

1. Utilizar as particularidades de redes sociais para sugerir novos relacionamentos de confiança;
2. Explorar o conhecimento e interações das pessoas para auxiliar no processo de identificação de confiança;
3. Promover maior colaboração entre os usuários da plataforma.

Dentre estes objetivos, o 3 não há como garantir que este possa ser alcançado. Apesar de que, com a possibilidade de criação de novos relacionamentos oriundos das recomendações do T-SWEETS, há uma forte tendência de que os usuários colaborem com maior frequência.

### 7.2.2 Cenário do Experimento 1

O período de análise deste experimento foi de 16 semanas, de agosto a dezembro de 2012, em 3 turmas de graduação, contendo respectivamente, 22, 14 e 21 alunos. As escolhas

destas turmas foram de acordo com o período de cada uma no curso: uma que estava iniciando o curso; outra em um período intermediário (meio do curso); e a última turma no fim do curso. O objetivo era selecionar estudantes que não tinham contatos prévios. Mesmo assim, havia dois alunos que estavam em duas das turmas do experimento simultaneamente.

O experimento foi dividido em duas fases de 8 semanas cada: a primeira sem a integração do T-SWEETS à plataforma e a segunda fase com a integração do respectivo algoritmo. Assim, esperava-se medir o impacto que o referido algoritmo viria a ter no ambiente. Os resultados destas duas fases do experimento são apresentados nas seções 7.2.3 e 7.2.4.

### 7.2.3 Resultados da Primeira Fase do Experimento 1

Durante a primeira fase do experimento foram criados 249 relacionamentos. O Gráfico 7.5 apresenta a criação destes relacionamentos em relação ao tempo, no qual é possível notar que o momento inicial foi quando surgiram mais relacionamentos. Isso é natural e esperado, uma vez que até este momento não existiam usuários na plataforma. Após este primeiro momento, o surgimento de novos relacionamentos na rede diminuiu gradativamente, tendendo a zero.



Gráfico 7.5: Primeira Fase do Experimento: Relacionamentos vs Tempo.

A partir destes relacionamentos, surgiu um conjunto de interações na plataforma – alguns dos resultados provenientes destas interações são apresentados no Quadro 7.1.



Quadro 7.1: Resultados das postagens da primeira fase do experimento

Primeira fase do Experimento	
quantidade de mensagens privadas trocadas entre usuários	24
média de troca de mensagens privadas entre os usuários	1,56
quantidade total de postagens	278
quantidade total dos usuários que realizaram pelo menos uma postagem	41
quantidade de postagens que receberam comentários	34
quantidade total de comentários das postagens	81
média de comentários das postagens	2,38

Das 278 postagens apresentadas no Quadro 7.1, 131 são do tipo arquivos (*e.g.* pdf, doc e rtf), 24 são imagens (*e. g.* jpeg e png), 16 são links, 134 são textos (tradicionais posts em fóruns de discussão) e 4 vídeos. Estas postagens em relação ao tempo podem ser visualizadas no Quadro 7.1.



Gráfico 7.6: Primeira Fase do Experimento: Conteúdos Publicados vs Tempo.

Da mesma forma como os relacionamentos, o instante inicial foi o momento que mais foram realizadas postagens – também natural, já que neste momento houve um contágio geral dos usuários em explorar a ferramenta. Já na semana 2, a queda no número de publicações foi grande, mas, na semana 3, voltou a subir até pouco mais de 50% em relação às publicações do momento inicial. Após isso, a quantidade de publicações somente diminuiu, o que deixa clara a necessidade de haver mecanismos para incentivar a colaboração entre os usuários do ambiente.

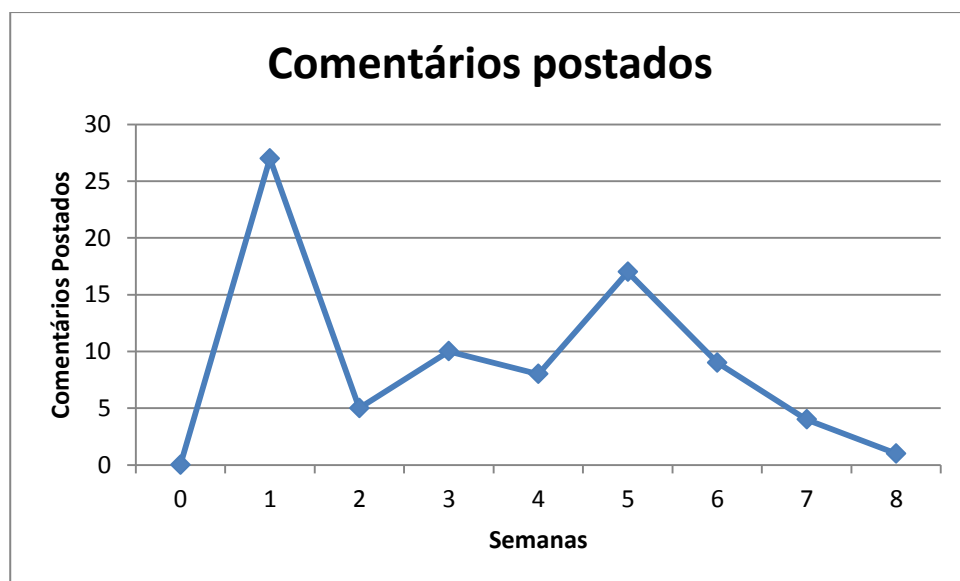


Gráfico 7.7: Primeira Fase do Experimento: Comentários às Postagens vs. Tempo.

O Gráfico 7.7 apresenta a quantidade de comentários realizados às postagens em relação ao tempo – similar aos relacionamentos e às postagens, houve uma queda grande nas últimas semanas da primeira fase do experimento, o que enfatiza a necessidade de existir na plataforma alternativas para manter os usuários motivados a produzir conhecimento.

#### 7.2.4 Resultados da Segunda Fase do Experimento 1

A segunda fase do experimento (a partir da semana 8) foi quando o algoritmo T-SWEETS foi implantado na plataforma CONECTA. Com a quantidade de usuários e relacionamentos existentes na plataforma CONECTA, ainda seria possível a criação de novos 5256 relacionamentos. Do total de relacionamentos possíveis, 49,79% foram recomendados pelo T-SWEETS, porém, foi estabelecido o limiar para filtrar as recomendações de confiança mais relevantes. O limiar definido foi de 0,7, em uma escala entre 0 e 1. Não há conhecimento se este é o limiar ideal – para a definição de um limiar ideal é necessário realizar uma série de experimentos e analisar cuidadosamente os resultados para esta definição.

Com a definição deste limiar, foram recomendados 11,04% do total de relacionamentos de confiança (49,79%) que poderiam ser recomendados. O Gráfico 7.8 mostra a criação destes relacionamentos em relação ao tempo.

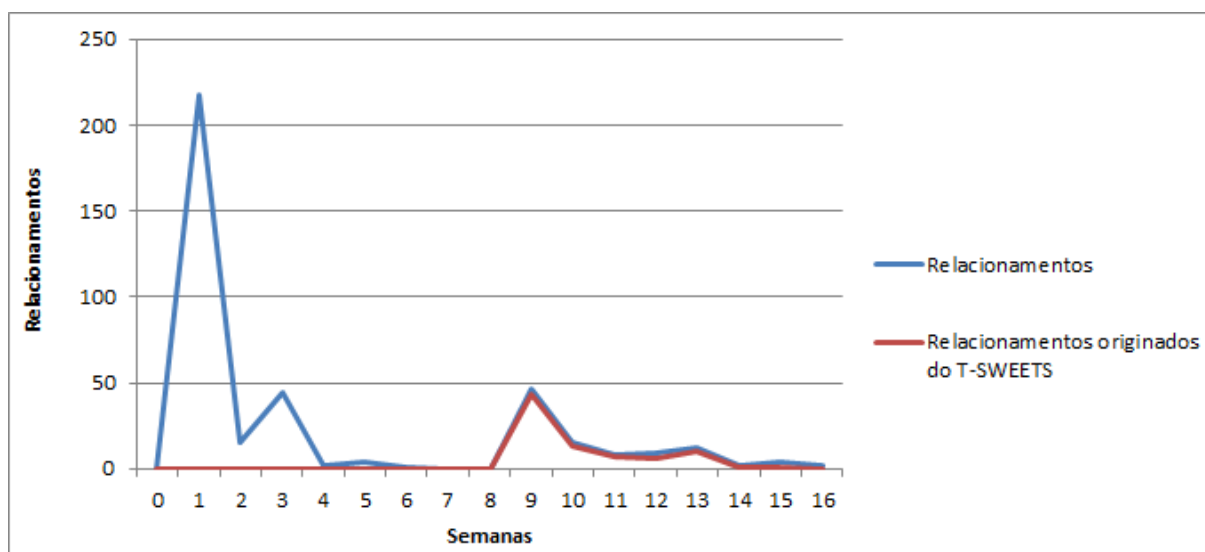


Gráfico 7.8: : Primeira e Segunda Fase do Experimento: Relacionamentos vs. Tempo.

O Gráfico 7.8 mostra que houve um pico no aumento de relacionamentos na rede social. Este pico é natural e esperado, uma vez que há uma tendência natural para as pessoas se

contagiarem diante de novidades. O mais importante é que, conforme o Gráfico 7.8, os relacionamentos continuaram acontecendo à medida que o tempo foi passando. Embora, novamente, da mesma forma como no final da primeira fase do experimento, os novos relacionamentos chegaram a zero na fase final, mas isso também é compreensível, dado que este era o momento em que as disciplinas estavam sendo encerradas e os alunos praticamente não acessavam a plataforma.

Um fato importante é que 81,82% dos novos relacionamentos criados, na segunda fase do experimento, originaram-se das recomendações de relacionamentos de confiança sugeridas aos usuários, que são originadas do T-SWEETS. A Figura 7.1 apresenta os grafos dos relacionamentos criados na plataforma antes e após a implantação do T-SWEETS.

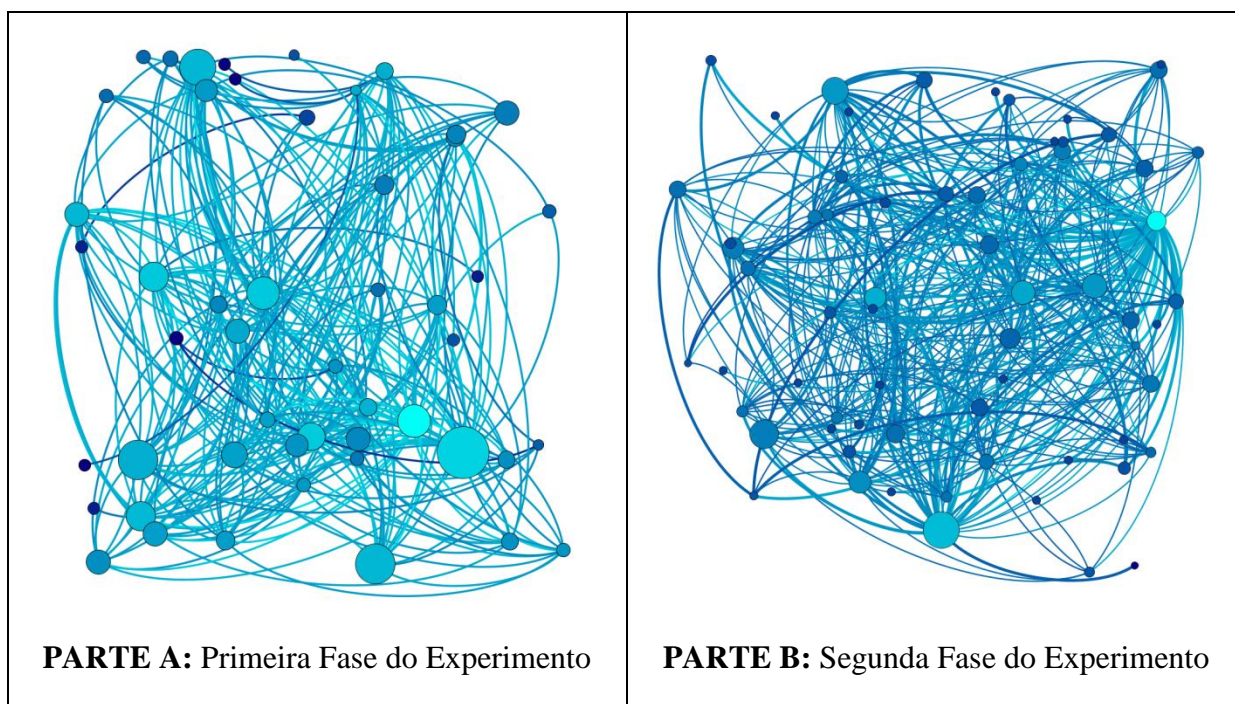


Figura 7.1: Grafo dos Relacionamentos do Conecta durante as 16 Semanas do Experimento

Na parte B da Figura 7.1, é apresentado o grafo com todos os relacionamentos, enquanto na parte A da Figura 7.1 é apresentado apenas o grafo com os relacionamentos oriundos da primeira fase do experimento. As arestas mais escuras da parte B da Figura 7.1 representam os relacionamentos originados da implantação do T-SWEETS. A dimensão dos nós representa a quantidade de relacionamentos que um usuário possui, isto é, quanto maior for a quantidade de relacionamentos de um nó, maior é a sua dimensão. Já a intensidade da cor de um nó representa a conectividade de um usuário na rede. Esta conectividade representa a distância social média entre os usuários e objetiva mensurar a facilidade de um usuário estabelecer comunicação com outro na rede.

De acordo com a Figura 7.1, há nós com uma grande quantidade de relacionamentos, porém com uma conectividade baixa ou média, em contrapartida existem nós com quantidade de relacionamentos reduzida, porém, com uma conectividade alta. Isso é justificável, uma vez que há nós que são classificados como “mais centrais”, ou mais influentes. Assim, se o nó possui uma maior quantidade de relacionamentos não há garantia de que este nó tenha maior conectividade. Porém, o estímulo ao acréscimo dos relacionamentos pode contribuir fortemente para o aumento das conectividades destes nós. Este acréscimo de relacionamentos pode contribuir também para uma maior interação e colaboração na rede social – uns dos princípios que sustentam estes ambientes virtuais sociais.

A Figura 7.2 apresenta os mesmos grafos da Figura 7.1 com a diferença que há a identificação explícita dos nós, a partir dos identificadores dos usuários.

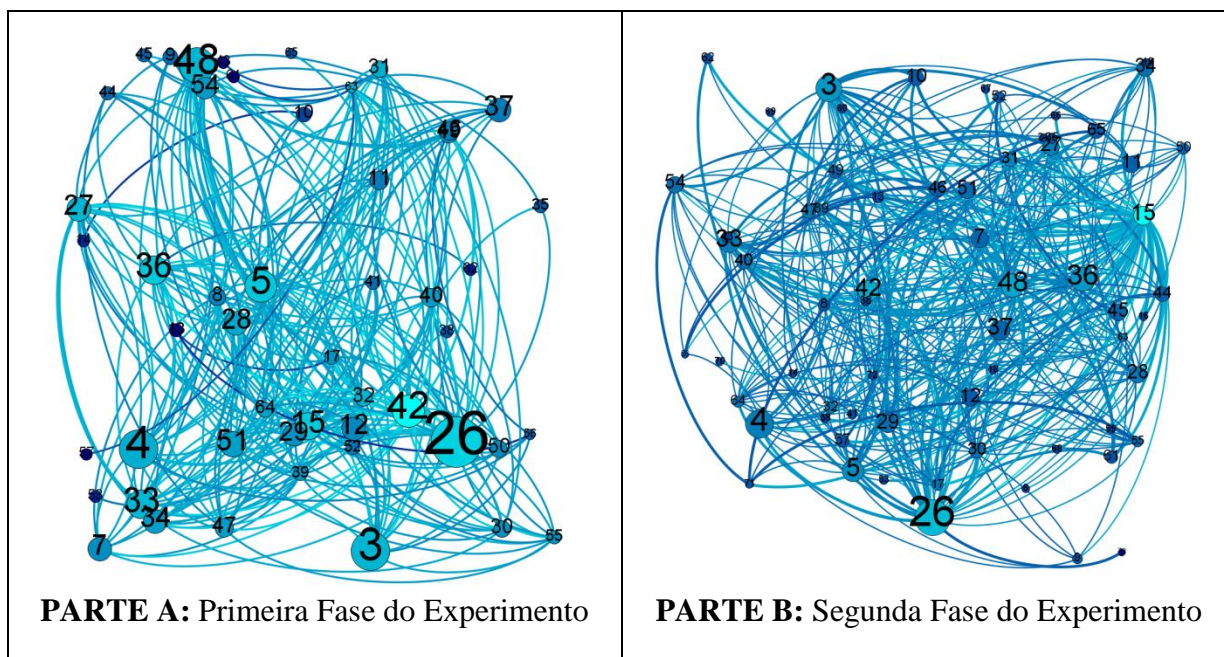


Figura 7.2: Grafo dos Relacionamentos do Conecta com os nós identificados durante as 16 Semanas do Experimento

A identificação explícita dos usuários nos grafos apresentados na Figura 7.2 é importante para a visualização pontual de que a situação de um determinado indivíduo na rede pode ser modificada à medida que novos relacionamentos surgem com o tempo. Por exemplo, os usuários 3, 4, 5 e 42 aumentaram poucos relacionamentos, só que o suficiente para que suas conectividades fossem aumentadas. Já com o usuário 15 ocorreu o processo inverso, também aumentou pouco a quantidade de relacionamentos, porém diminuiu a sua conectividade; o usuário 38 tinha alta conectividade e poucos relacionamentos, com o tempo, aumentou a



quantidade de relacionamentos e continuou com a conectividade alta. A Figura 7.3 mostra um cenário organizado em camadas que enfatiza a distância social entre os nós.

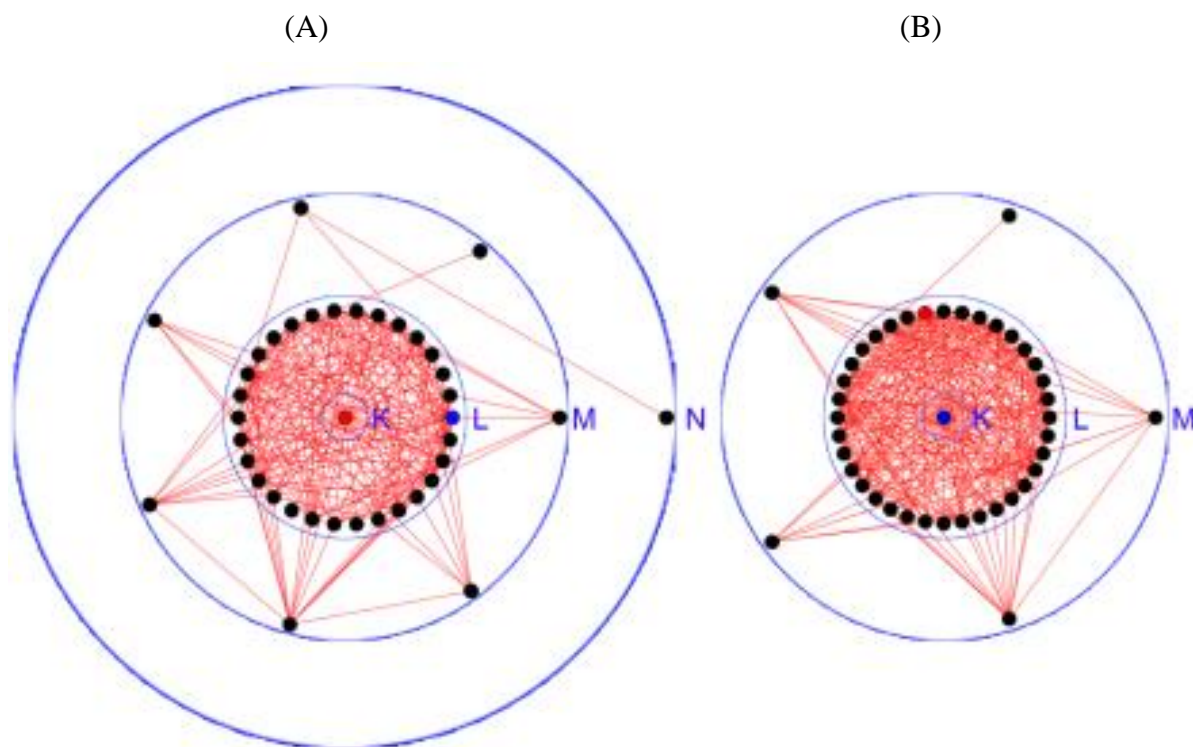


Figura 7.3: Grafo organizado em camadas que apresenta a distância social entre os nós durante o experimento.

A Figura 7.3 mostra uma visão dos relacionamentos da rede organizada em camadas. Para isso, foi utilizado o algoritmo de segmentação *LayeredLayout* (Kuchar, 2011). Nesta organização, o nó mais central (em vermelho, Figura 7.3 – cenário A) é a base para a definição das camadas. A quantidade de camadas representa a densidades das conexões dos usuários na rede, assim, o maior o número de camadas significa que mais dispersos estão os usuários na rede. Os nós que estão em uma mesma camada significa que estão a uma mesma distância social média do nó mais central. Na Figura 7.3 – cenário B, a quantidade de camadas é inferior ao cenário A (antes das recomendações), assim é possível afirmar que os usuários estão a uma distância social do nó mais central. Além disso, a quantidade de nós na camada M do cenário B também diminuiu – outro indicador que os usuários estão mais próximos após a implantação do mecanismo de recomendação. No cenário B, o nó mais central também foi modificado em relação ao cenário A. O nó que tem maior conectividade no cenário B é o de cor azul.

A Figura 7.4 apresenta os nós agrupados de acordo com os grupos que pertencem e depois de acordo com os seus relacionamentos.

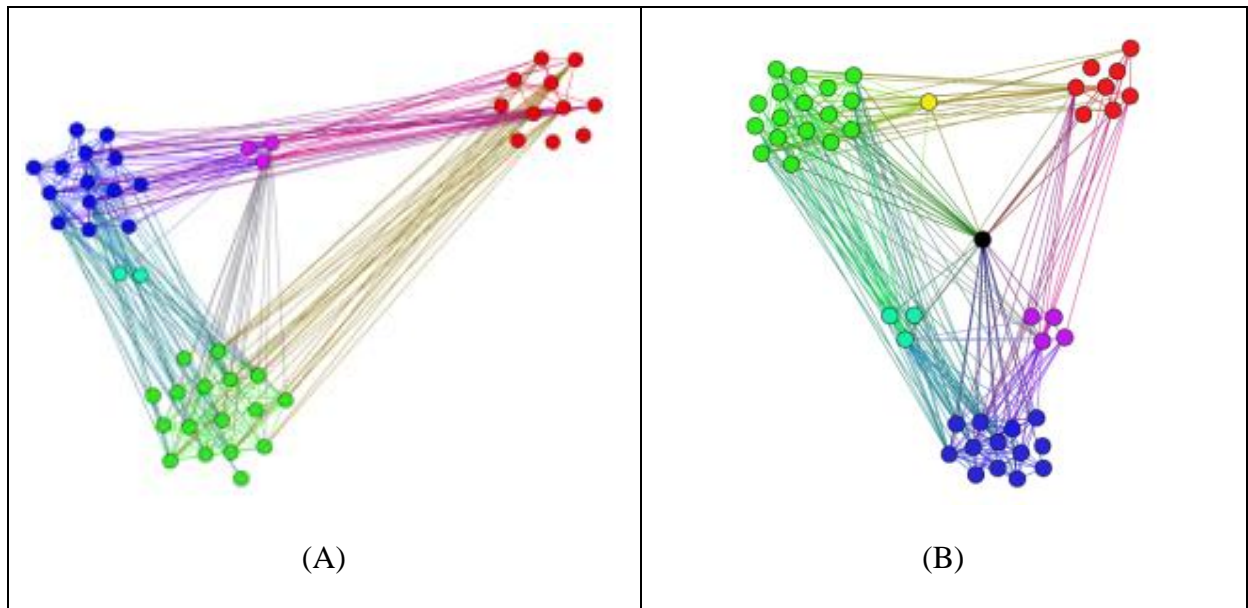


Figura 7.4: Usuários do experimento agrupados em grupos (turmas) que pertencem e seus relacionamentos

Na Figura 7.4 – cenário B é possível observar que há uma maior quantidade de relacionamentos entre os usuários que estão em grupos diferentes, se comparado ao cenário A. Os usuários que antes estavam isolados aos relacionamentos internos dos grupos em que pertenciam, passam a ter uma maior conectividade na rede através da criação de novos relacionamentos com usuários que pertencem aos outros grupos.

Com o acréscimo no número de relacionamentos na rede, é natural que haja uma maior interação entre os usuários. O Gráfico 7.9 apresenta os resultados das postagens, comentários às postagens e mensagens enviadas aos usuários na segunda fase do experimento.

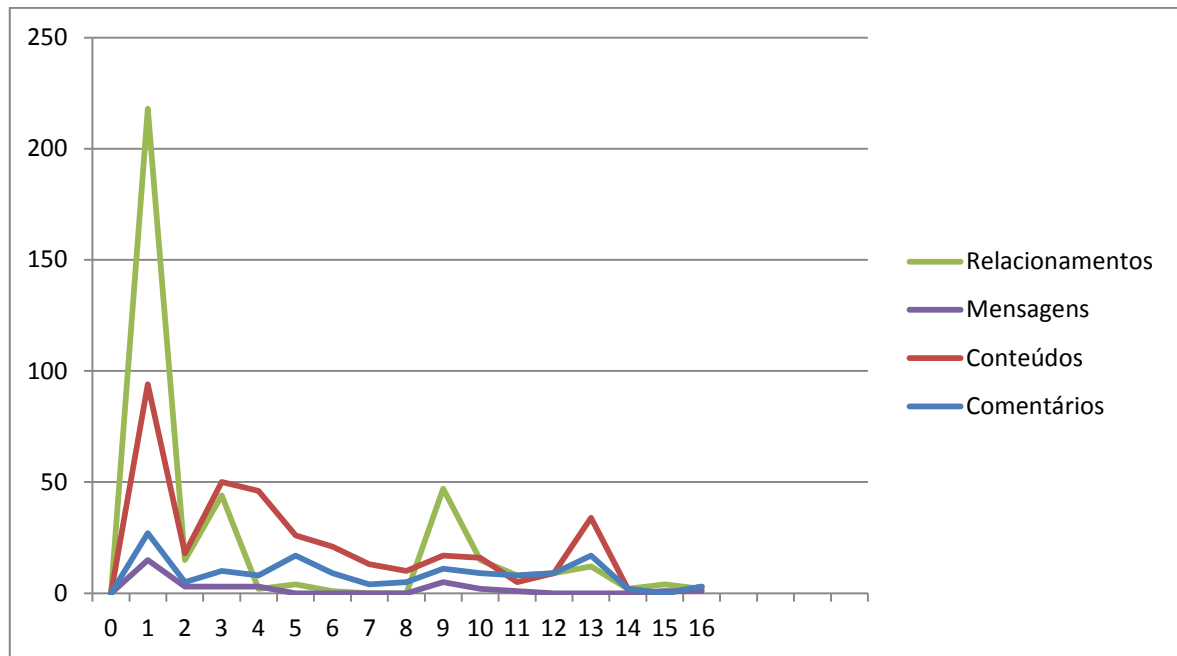


Gráfico 7.9: Relacionamentos Criados, Mensagens Enviadas, Conteúdos Postados e Comentários em Relação ao Tempo.

Com o aumento no número de relacionamentos na rede, há uma tendência natural para que as mensagens enviadas entre os usuários e os comentários às postagens realizados pelos usuários também aumentem, pois quanto maior for o número de relacionamentos, maior será o número de pessoas que acompanham as produções e avaliações realizadas uns pelos outros. Nas últimas semanas houve uma queda acentuada nas interações realizadas pelos usuários, conforme já mencionado anteriormente, em decorrência do período escolar estar em fase de encerramento.

Em decorrência do aumento deste número de relacionamentos, era esperado que o número de mensagens privadas trocadas entre os usuários também aumentasse, porém, isso não ocorreu (Gráfico 7.9) – isso foi uma surpresa. Para tentar entender este resultado, foi realizada uma pesquisa informal com alguns usuários, quando se constatou que haviam alguns problemas técnicos no módulo de mensagens e, por isso, dificultava o uso dos usuários e estes não se sentiam motivados a utilizá-lo.

### 7.2.5 Objetivos do Experimento 2

Este experimento tem como objetivo medir a eficácia do T-SWEETS, primeiramente, em relação aos resultados de uma análise de relacionamentos de confiança oriundos do trabalho



desenvolvido por Leskovec (2010); e depois apresentar uma análise em relação ao algoritmo TidalTrust (Golbeck, 2005).

### 7.2.6 Cenário do Experimento 2

Foram utilizadas como fonte de relacionamentos de confiança, dados oriundos do *Epinions*<sup>16</sup>, *Slashdot*<sup>17</sup> e *Wikipedia*<sup>18</sup>. *Epinions* é um ambiente em que os usuários podem realizar avaliações a determinados produtos (e.g. notebook e smartphone) ou avaliar pessoas (os próprios avaliadores). A partir das avaliações realizadas entre estes avaliadores forma-se uma rede social de relacionamentos de confiança. *Slashdot*, um site voltado para publicações de notícias de tecnologia. Neste ambiente, os usuários podem associar tags a outros usuários classificando-os como amigos ou inimigos. *Wikipedia*, uma enciclopédia online em que as pessoas podem colaborar livremente para o seu desenvolvimento. Os usuários podem passar para a categoria de administradores da enciclopédia, a partir das avaliações que podem receber de outros usuários, que são positivas, negativas e neutras.

Estas fontes de dados são oriundas do trabalho de Leskovec (2010), e estão disponíveis no projeto SNAP<sup>19</sup>, desenvolvido por pesquisadores de *Stanford*. O trabalho de Leskovec (2010) analisa estes relacionamentos a partir de associações ternárias (tríades), que são relacionamentos formados pelo conjunto de três nós, conforme apresentado na Figura 7.5.

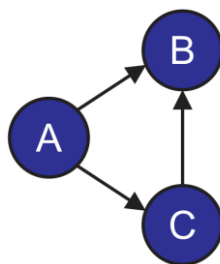


Figura 7.5: Estrutura de uma tríade

Cada relacionamento forma uma relação de confiança de valores  $\{-1, +1\}$ , em que o valor positivo indica que há a confiança entre os nós, enquanto o valor negativo indica a au-

---

<sup>16</sup> *Epinions.com: Product Reviews and Consumer Reports*: <http://www.epinions.com/>

<sup>17</sup> *Slashdot: News for nerds, stuff that matters*: <http://slashdot.org/>

<sup>18</sup> Wikipédia, a enciclopédia livre: <https://pt.wikipedia.org/>

<sup>19</sup> Stanford Large Network Dataset Collection: <http://snap.stanford.edu/data/#signnets>

sência de confiança. De acordo com este cenário, o Quadro 7.2 apresenta a proporção dos dados originados de cada uma das fontes supracitadas.

Quadro 7.2: Classificações e quantidades de tríades nas fontes de dados do Epinions, Slashdot e Wikipédia (Leskovec, 2010)

Tríade $T_i$		$ T_i $	$p(T_i)$
$T_3$	+++	11.640.257	87%
$T_2$	+- -	947.855	7,1%
$T_1$	++ -	698.023	5,2%
$T_0$	- - -	89.272	0,7%

As *tríades* podem variar conforme o valor atribuído aos relacionamentos na rede social, conforme apresentado no Quadro 7.2. Por exemplo,  $T_3$  é uma *tríade* na qual os três relacionamentos possuem pesos positivos, já em  $T_2$  apenas um relacionamento é positivo, enquanto que os outros dois são negativos. Para executar o algoritmo T-SWEETS nas bases supracitadas, o tamanho do caminho entre dois nós também foi limitado a 3 – tamanho esse limitado pela tríade.

### 7.2.7 Resultados da Primeira Fase do Experimento 2

De acordo com as associações ternárias mencionadas anteriormente e os valores positivos e negativos, alguns cenários para a inferência de relacionamentos de confiança podem ser representados. A Figura 7.6 apresenta estes cenários.

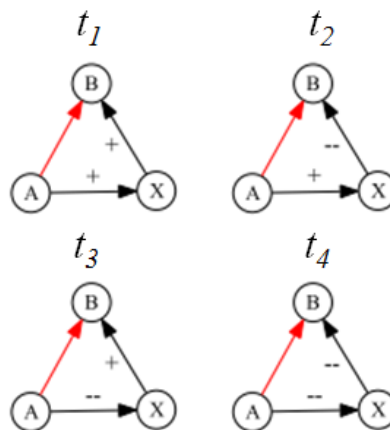


Figura 7.6: Possibilidades de tríades que podem ser usadas nas análises

Leskovec (2010) apresentou um estudo de algumas possibilidades para a construção de tríades, de acordo com os valores  $\{-1, +1\}$ . Neste estudo, foi apresetado que, dado dois relacionamentos com valores de confiança quaisquer (+1 ou -1), qual a probabilidade do terceiro ser positivo ou negativo. O relacionamento, representado pela seta vermelha na Figura 7.6, representa este terceiro relacionamento. Este quadro de probabilidades pode ser visualizado no Quadro 7.3.

Quadro 7.3: Probabilidades de inferência de relacionamentos positivos e negativos para cada tríade. (Leskovec, 2010, adaptada)

$t_i$	$P(+)$	$P(-)$
1	0,97	0,03
2	0,54	0,46
3	0,30	0,70
4	0,23	0,77

De acordo com as probabilidades apresentadas no Quadro 7.3, é possível executar o algoritmos T-SWEETS e verificar a sua eficácia. Além do T-SWEETS, também, foi realizada uma análise em relação ao algoritmo clássico *TidalTrust* (Golbeck, 2005). Assim, para cada uma das configurações apresentadas (Figura 7.6 e Quadro 7.3), o Quadro 7.4 apresenta os resultados obtidos.

Quadro 7.4: Porcentagem da quantidade de relacionamentos positivos em cada tríade

$t_i$	$P(+)$		
	Leskovec (2010)	TidalTrust	T-SWEETS
1	97%	100%	100%
2	54%	0%	0%
3	30%	0%	0%
4	23%	100%	100%

De acordo com os resultados que podem ser visualizados no Quadro 7.4 pode-se observar que tanto os resultados obtidos do T-SWEETS, quanto do *TidalTrust*, são valores muito determinísticos nos cenários apresentados. Isto é, ou o grau de confiança foi nulo ou foi integral (100%). Contudo, esta deficiência pode ser justificável, uma vez que foram utilizados valores inteiros positivos e negativos. Para evitar esta limitação, o ideal é utilizar valores fracionados em uma escala de  $[-1...+1]$ .

### **Problema**

Leskovec (2010) faz uso dessas tríades para analisar os seguintes princípios: “o amigo do meu amigo é meu amigo”; “o amigo do meu inimigo é meu inimigo”; “o inimigo do meu amigo é meu inimigo”; e “o inimigo do meu inimigo é meu amigo”. Além disso, considera apenas valores inteiros de confiança,  $\{-1,+1\}$ . Essas características inviabilizam a utilização de um caminho no grafo social maior que 3, já que tornaria ainda mais ineficaz a avaliação de confiança entre dois indivíduos. Essa ineficácia deve-se, principalmente, por tornar totalmente transitiva a relação de confiança entre dois nós A e B, independente do tamanho do caminho que os une. Esta ineficácia também ocorre em um caminho 3, porém, nestes casos, há uma representação dos princípios apresentados no início desta seção (por exemplo, o amigo do meu inimigo é meu inimigo) – o propósito de Leskovec.

Tanto o *T-SWEETS* quanto o *TidalTrust*, realizam inferência com maior eficácia com entradas positivas fracionárias no intervalo  $[0,1]$ . Assim, é possível a obtenção de valores de confiança não discretos entre indivíduos da rede. Ou seja, a inferência de confiança não totalmente transitiva, um dos princípios da análise de confiança.

### **Análise com valores fracionados**

Para realizar uma análise sobre valores fracionados e positivos, foi usado o conjunto de dados apresentados no trabalho de Golbeck (2005), que são usados para analisar o algoritmo *Tidaltrust*.

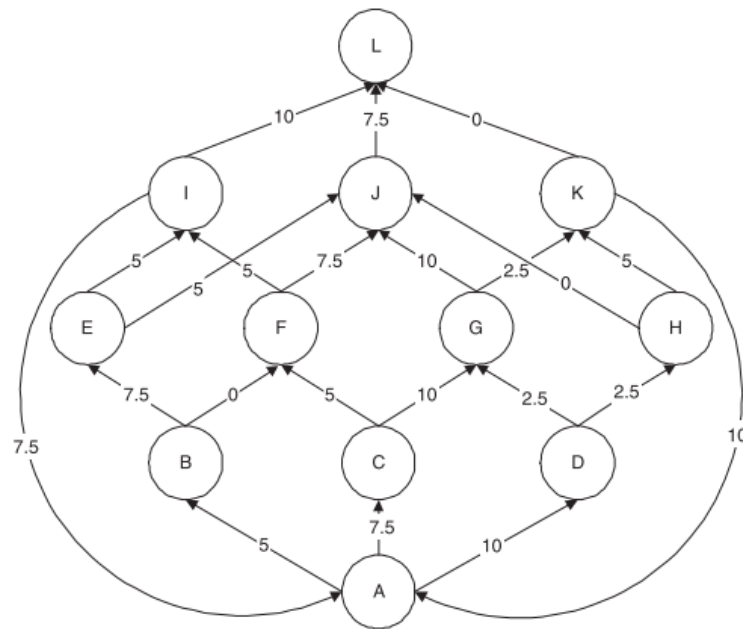


Figura 7.7: Grafo de relacionamentos utilizado para analisar os algoritmos (Taherian, 2008)

Após a aplicação de ambos os algoritmos, TidalTrust e T-SWEETS, foram inferidos valores de confiança entre os nós apresentados no grafo da Figura 7.7. Os valores inferidos são apresentados nos Quadros 7.5 e 7.6. Os valores destacados são os graus de confiança que foram atribuídos diretamente pelos usuários a outros usuários.

Quadro 7.5: Resultados após a execução do algoritmo T-SWEETS aplicado no cenário da Figura 7.7.

	A	B	C	D	E	F	G	H	I	J	K	L
A	-	5	7.5	10	3.7	2	5	2.5	1.3	3.2	0.2	1.2
B	0.7	-	0.5	0.7	7.5	0	0.3	0.9	0.9	5	2.5	0.9
C	2.2	1.1	-	2.2	0.8	5	10	0.5	2.5	6.9	2.5	3.4
D	0.9	0.5	0.7	-	0.3	0.2	2.5	2.5	0.1	2.5	0.9	0.6
E	3.7	1.9	2.8	3.7	-	0.7	1.8	0.9	5	5	0	4.4
F	3.7	1.9	2.8	3.7	1.4	-	1.8	0.9	5	7.5	0	5.3
G	2.5	1.3	1.9	2.5	0.9	0.5	-	0.6	0.3	10	2.5	3.7
H	5	2.5	3.7	5	1.9	1	2.5	-	0.7	0	5	0
I	7.5	3.7	5.6	7.5	2.8	1.4	3.7	1.9	-	2.4	0.1	10
J	-	-	-	-	-	-	-	-	-	-	-	7.5
K	10	5	7.5	10	3.7	1.9	5	2.5	1.3	3.2	-	0
L	-	-	-	-	-	-	-	-	-	-	-	-

Quadro 7.6: Resultados após a execução do algoritmo TidalTrust aplicado no cenário da Figura 7.7. (Taherian, 2008)

	A	B	C	D	E	F	G	H	I	J	K	L
A	-	5	7.5	10	7.5	5	2.5	2.5	10	5	2.5	7.5
B	7.5	-	7.5	10	7.5	0	5.7	2.5	5	5	2.5	8.8
C	7.5	5	-	10	7.5	5	10	2.5	5	10	2.5	7.5
D	10	5	7.5	-	7.5	3	2.5	2.5	5	5	3.8	3
E	7.5	5	7.5	10	-	3	5.7	2.5	5	5	2.5	8.8
F	7.5	5	7.5	10	7.5	-	5.7	2.5	5	7.5	2.5	7.5
G	10	5	7.5	10	7.5	3	-	2.5	5	10	2.5	7.5
H	10	5	7.5	10	7.5	3	5.7	-	5	0	5	0
I	7.5	5	7.5	10	7.5	5	5.7	2.5	-	10	2.5	10
J	-	-	-	-	-	-	-	-	-	-	-	7.5
K	10	5	7.5	10	7.5	5	2.5	2.5	5	10	-	0
L	-	-	-	-	-	-	-	-	-	-	-	-

Quadro 7.7: Média e desvio padrão dos valores dos Quadros 7.5 e 7.6.

	Média	Desvio Padrão
T-SWEETS	4,73	3,59
TidalTrust	6,27	3,13

De acordo com o que pode ser observado no Quadro 7.7, a média dos valores de confiança entre os relacionamentos é maior no algoritmo *TidalTrust*. O valor menor obtido no T-SWEETS é em decorrência de que este se baseia no princípio que quanto maior os caminhos de confiança que unem dois nós, menor é a relação de confiança entre eles. Assim, T-SWEETS considera que pessoas que estejam mais próximas, tem maior tendência que haja uma maior relação de confiança entre elas, já que esta relação de confiança é baseada em informações (e.g. trocas de experiências) que as pessoas possuem umas sobre as outras.

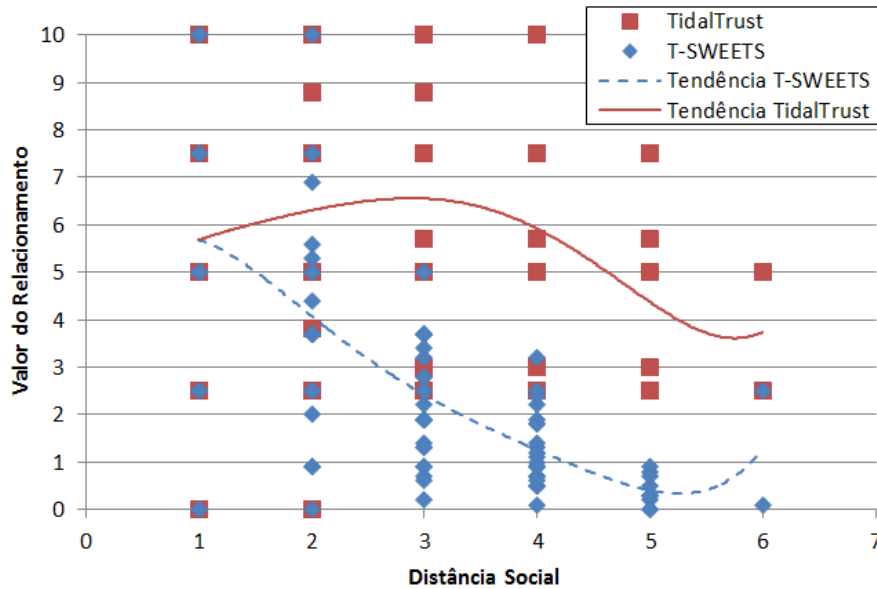


Figura 7.8: Dispersão dos valores obtidos para o grafo da Figura 7.7 com os algoritmos *TidalTrust* e *T-SWEETS*.

Além da dispersão (existem pontos sobrepostos) dos valores obtidos com a execução dos algoritmos *TidalTrust* e *T-SWEETS*, no cenário apresentado anteriormente, são aprestadas também as tendências dessas dispersão, considerando o aumento na distância do caminho social que liga dois nós no grafo. Essas linhas de tendências foram calculadas a partir da interpolação dos pontos na dispersão em um polinômio de grau 5. De acordo com o gráfico apresentado na Figura 7.8 (linha pontilhada) é possível observar que no *T-SWEETS* a tendência é decrescente, na maior parte de sua extensão, à medida que a distância social aumenta. Ao mesmo tempo, o algoritmo *TidalTrust* apresenta um comportamento sem padrão em relação à distância social. Diante da premissa que quanto maior a distância social entre dois nós há uma menor relação de confiança entre eles, o *T-SWEETS* apresenta um comportamento mais adequado.

## 7.3 SWEETS

### 7.3.1 Objetivos do Experimento

Conforme mencionado no Capítulo 5, o *SWEETS*, um sistema que identifica e recomenda especialistas de domínio, foi desenvolvido e implantado à plataforma *Konnen*, um ambiente colaborativo educacional baseado em Redes Sociais, que está sendo desenvolvido em caráter experimental. Esta integração à plataforma tinha os seguintes objetivos:

1. utilizar características inerentes às plataformas baseadas em redes sociais em um sistema que identifique especialistas de domínio;
2. identificar especialistas de domínio;
3. recomendar o especialista mais adequado para uma determinada pessoa;
4. promover melhorias na colaboração entre os usuários da plataforma;
5. incentivar uma maior produção de conhecimento dos usuários na plataforma;
6. agilizar o processo de resolução de um determinado problema.

Dentre estes objetivos, os objetivos 4, 5 e 6 só poderão ser alcançados com uma análise dos resultados gerados pelo SWEETS durante um período de tempo mais longo, por exemplo, 1 a 2 anos. Além disso, é necessário ter um maior número de usuários usando o sistema como um pré-requisito para o desenvolvimento de suas atividades cotidianas. Não é o caso, já que o SWEETS foi implantado na plataforma *Konnen* em caráter experimental e com um número reduzido de pessoas utilizando-o. Para uma análise mais detalhada, sobre o impacto real desta plataforma, o ideal seria a implantação deste sistema de recomendação de especialistas em uma organização que utilize uma plataforma baseada na teoria de redes sociais que promova a interação entre os colaboradores e faça a gestão de conhecimento de toda a organização.

### 7.3.2 Cenário do Experimento

O experimento foi realizado com um grupo de 23 pessoas, que, conforme mencionado anteriormente, teve como objetivo mensurar a eficiência da abordagem proposta para identificar e recomendar especialistas de domínio. Alguns critérios foram adotados para a seleção deste conjunto de pessoas, são eles: (1) possuir currículo na plataforma *Lattes*; (2) ter um perfil na rede social *LinkedIn*; e (3) possuir de alguma forma um contato prévio entre eles, isto é, algum relacionamento social, por exemplo, um amigo, conhecido, colega de trabalho ou de aula. Os critérios 1 e 2 são importantes para que haja uma garantia da exploração de fonte externa de conhecimento e, assim, evitar a ausência de conhecimento para identificação das especialidades. Já o critério 3 é importante para que se tenha parâmetros provenientes das relações sociais, que são importantes para identificar o especialista de domínio mais propenso a colaborar.

Depois, foi criado um perfil para cada uma destas pessoas na plataforma *Konnen*. Estas pessoas foram direcionadas a editarem seus perfis seguindo as seguintes instruções: (1) informar o endereço do currículo na Plataforma *Lattes*; e (2) vincular seu perfil na plataforma



*Konnen* ao perfil da plataforma *LinkedIn*, isto é, utilizar a funcionalidade disponível na plataforma *Konnen* e informar os dados de autenticação da plataforma *LinkedIn* e, com isso, importar automaticamente os seus dados da *LinkedIn*.

Como este grupo de pessoas não utiliza a plataforma *Konnen* para o desenvolvimento de suas atividades diárias, então, é natural que na plataforma não possua produções (e.g. postagens em grupos de discussão) destas pessoas. Estas produções são fontes de conhecimento necessárias para a inferência das experiências dos candidatos a especialistas na categoria *Habilidade Calculada* (Capítulo 5). Por isso, estas pessoas foram convidadas a realizarem um conjunto de publicações na plataforma. Para que houvesse uma fonte mais rica de conhecimento, estas pessoas foram orientadas a postarem parte dos conteúdos das suas publicações científicas. Assim, evita-se um esforço adicional no que tange à produção/criação de conhecimento e tem-se um conjunto de postagens direcionadas (não aleatórias).

### 7.3.3 Resultados do Experimento – parte 1

São quatro as categorias de habilidades (Capítulo 5) utilizadas na abordagem proposta neste trabalho para a inferência do grau de especialidade do usuário em um assunto *a*, são elas: *Habilidade informada*; *Demonstrada*, *Sugerida* e *Refutada*.

As habilidades na categoria *Informada* são aquelas que são informadas pelos usuários na Plataforma *LinkedIn* e *Lattes*. Como são as habilidades que os próprios especialistas assumem que tem, então não há necessidade de realizar nenhum cálculo adicional. É necessário apenas considerá-las no processo de inferência do grau final de especialidade do candidato a especialista em um assunto *a*.

Para a categoria *Habilidade Demonstrada*, foram utilizadas os dados das publicações científicas das pessoas disponíveis na Plataforma *Lattes* – conforme mencionado no Capítulo 6. As especialidades dos indivíduos são determinadas a partir das palavras chave associadas às publicações científicas dos usuários disponíveis na Plataforma *Lattes*, em que cada uma possui seu grau de importância de acordo com a categoria que pertence (e.g. publicações em periódicos ou conferência). Este grau de importância atribuído a cada categoria é configurável de acordo com o ambiente em que estiver implantado. Mais informações sobre o assunto são apresentadas no Capítulo 5. Uma vez que o nível de experiências das pessoas é mensurado na categoria de *Habilidades Demonstradas*, foi aplicado um questionário para estes indivíduos, a fim de identificar a eficiência da abordagem.

Não foi possível analisar nos experimentos a eficiência dos cálculos das *Habilidades Sugeridas* e *Refutadas*, apesar de que ambas são apresentadas na abordagem proposta no presente trabalho. Estas categorias de habilidades dependem fortemente de frequentes interações dos usuários no ambiente para que possam ser mensuradas. Um cenário ideal para que estas categorias de habilidades possam ser avaliadas adequadamente é um ambiente organizacional em que as pessoas o utilizam para o desenvolvimento de suas atividades diárias. Estas atividades devem envolver interações constantes com outras pessoas, por exemplo, através de fóruns de discussão. Assim, passa-se a ter parâmetros ideais para a refutação ou a sugestão de que uma determinada pessoa possua ou não uma habilidade específica.

Para determinar o grau final de especialidade de uma pessoa em um assunto  $a$ , é aplicada a Equação (5.9) apresentada no Capítulo 5. Nesta equação, são considerados os 4 tipos de habilidades apresentados na abordagem proposta no presente trabalho. Porém, para este experimento, a equação foi ajustada de forma a desconsiderar as categorias de habilidades *Sugerida* e *Refutada*, já que não foi possível aplicá-las no presente experimento. A Equação (5.9) representa uma média ponderada, por isso o ajuste realizado não foi complexo – foi necessário apenas desconsiderar os valores destas categorias e seus respectivos pesos no cálculo. O grau final de especialidade de uma pessoa em um assunto  $a$  está em uma escala de 0 a 1. Neste experimento, foram considerados que as pessoas que possuem grau final de especialidade maior ou igual a 0,8 em um assunto  $a$  são classificadas como especialistas.

Para mensurar a eficiência da abordagem adotada para o cálculo da habilidade na categoria *Calculada*, devem ser utilizadas as produções dos usuários e a ontologia  $O_{ci}$ . A ontologia  $O_{ci}$  é composta por relacionamentos entre conceitos (*tags*) associados a um determinado peso. Ambos são determinados pela co-ocorrência dos conceitos (*tags*) nos elementos utilizados para a criação da *folksonomia*. Quanto maior o peso entre os conceitos, maior a frequência que eles co-ocorrem, e maior a relação semântica entre eles (Capítulo 5).

Na plataforma *Konnen* o número de objetos *tagged* era bastante limitado, aproximadamente 9, com um total de 16 tags diferentes. Esta limitação seria um problema para mensurar a eficiência da abordagem proposta para a inferência da habilidade na categoria *Calculada*. Por isso, foi utilizada uma fonte externa de objetos *tagged*, *BibSonomy*, um sistema social de *bookmark* e compartilhamento de publicações. Para selecionar tags relevantes ao contexto dos usuários e as relações semântica entre elas, foi identificado o termo que mais ocorre em relações aos usuários no experimento – foram consideradas também as palavras-chave associada

às publicações dos usuários nesta análise. Este termo serviu de entrada para a importação das informações do *BibSonomy*. O termo mais frequente entre os usuários foi “*Social Networks*”.

A ontologia  $O_{ci}$  possui um total de 955 tags. Para criá-la ou atualizá-la, exige-se um alto custo computacional. Por isso, este processo é realizado periodicamente e pode ser configurado. Inicialmente, foi definido um intervalo de 15 dias. A Figura 7.9 mostra uma representação  $O_{ci}$ .

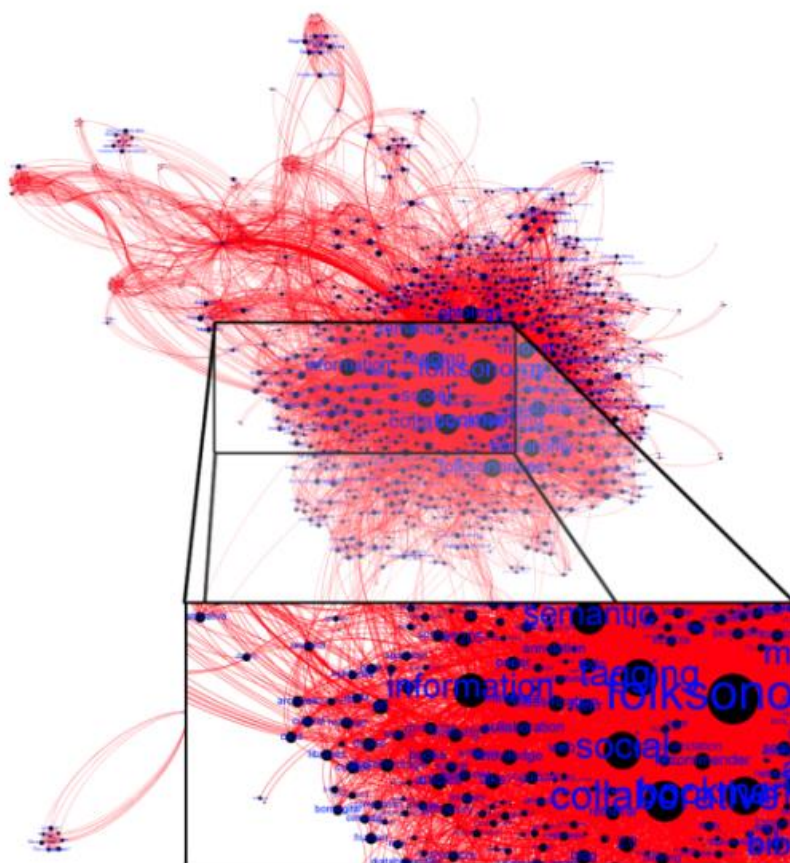


Figura 7.9: Lightweight Ontology  $O_{ci}$ .

A ontologia  $O_{ci}$  possui 9291 relacionamentos, por isso, o emaranhado de informações apresentadas na Figura 7.9 não está claro. Para que fosse possível visualizar a evidência semântica entre os termos, foi utilizada uma escala logarítmica com base 1,01 para normalizar os pesos dos primeiros 15 termos que mais co-ocorrem, e uma escala de 1,02 nos termos restantes. Assim, os termos que possuem maior semântica entre eles são evidenciados, conforme parte destacada na Figura 7.9. Dentre este conjunto de termos que mais co-ocorrem (destacam-se) estão “*folksonomy*”, “*tagging*”, “*collaborative*” e “*social*”.

Surgiram na ontologia tags que têm significados idênticos representados em relacionamentos distintos como, por exemplo, “*social networks*” e “*social network*”. Este problema era esperado, já que os usuários têm a liberdade de associar livremente os conceitos (*tags*) aos itens (publicações), e na abordagem adotada não é realizado um tratamento para evitar esta situação. Além disso, podem existir erros de digitação. A livre descrição é uma característica de *folksonomia*. Esses problemas comprometem o enriquecimento da ontologia, logo, podem tornar o cálculo da *Habilidade Calculada* menos eficiente.

O passo seguinte para o cálculo da habilidade na categoria *Calculada* é indexar a base de conhecimento dos usuários. Para isso, são utilizadas as produções textuais (textos planos de qualquer natureza, e.g. post em uma comunidade ou o conteúdo de um documento *rtf*) dos usuários que compõem os perfis pré-processados destes usuários. Estes perfis possuem os termos mais relevantes e a frequência (peso) com que eles ocorrem em relação ao usuário.

Para realizar a indexação, são considerados os conceitos/termos da ontologia  $O_{ci}$  e seus respectivos relacionamentos. Não são considerados todos os conceitos, pois a decisão de utilizá-los está limitada à quantidade de relacionamentos que possuem. A quantidade mínima de relacionamentos deve ser maior que 1. Neste experimento, foram considerados os termos que possuem no mínimo 4 relacionamentos.

O mínimo de 4 relacionamentos foi adotado em razão de: quando o experimento foi executado com no mínimo 3 relacionamentos foi identificado um conjunto grande de especialidades para os usuários; quando adotou-se um valor mínimo de 5 relacionamentos as especialidades foram reduzidas drasticamente. Por isso, chegou-se a um valor intermediário, 4. É provável que cada experimento diferente necessite de uma análise similar para determinar este valor mínimo ideal de relacionamentos. Existem no experimento 408 relacionamentos com esta configuração.

Além disso, estes relacionamentos devem possuir um peso maior que 1. Os relacionamentos que possuem peso mínimo 1 possuem evidência de pouca densidade semântica entre eles, logo, não são considerados relevantes para o experimento. Para definir o peso ideal a ser considerado, é importante que sejam conduzidos experimentos para realizar esta análise.

O grau inferido para a habilidade na categoria *Calculada* também está em uma escala de 0 a 1. Para que uma pessoa possa ser classificada como possível especialista em um assunto *a*, é necessário que este grau esteja acima de um limiar. O limiar adotado neste trabalho foi 0,8.

Não foram realizados testes exaustivos para determinar qual o limiar ideal, pois isso implicaria em sucessivas modificações e análises no experimento.

### **7.3.4 Análise da qualidade das recomendações**

Para mensurar a eficiência da abordagem proposta neste trabalho, foram desenvolvidos e aplicados questionários personalizados para as pessoas do experimento. A adoção destes questionários foi a melhor alternativa encontrada para identificar se uma pessoa possui ou não uma determinada especialidade em um assunto  $a$ . A premissa que fundamenta esta estratégia é: para determinar se uma pessoa possui ou não conhecimento em um assunto  $a$ , a melhor alternativa é perguntar para a própria pessoa candidata a especialista.

Uma vez que foram identificadas as experiências dos usuários, o questionário personalizado para os usuários foi aplicado. Cada questionário continha as especialidades que o candidato foi identificado como especialista e a possibilidade da própria pessoa indicar o quanto conhece (em uma escala de 1-5) daquele respectivo assunto. Foram utilizadas as 10 primeiras especialidades inferidas de cada pessoa.

De posse destes dados, é possível fazer uma análise em relação aos resultados obtidos a partir da aplicação do método proposto neste trabalho e, então, medir a eficiência desta abordagem. Para evitar uma avaliação tendenciosa, os usuários candidatos a especialistas não foram informados que tinham sido identificados como possíveis especialistas nos assuntos apresentados nos questionários.

Com os resultados da aplicação destes questionários, foi aplicada a métrica de precisão para analisar a qualidade destas recomendações geradas. *Precisão* é a razão entre a quantidade de itens relevantes de um conjunto-resposta de uma requisição e a quantidade total de itens desse conjunto-resposta (ou seja, todos os itens, sejam eles relevantes ou não) (Baeza-Yates, 1999). Neste caso, a quantidade de especialidades detectadas corretamente pelo total de especialidades detectadas. Assim, obteve-se uma precisão de aproximadamente 91%.

Além deste, foi aplicado um segundo questionário para analisar se a inferência do grau de importância para cada habilidade foi determinada corretamente, isto é, se um especialista possui maior nível de conhecimento em um assunto  $X$  em relação a um assunto  $Y$ . Uma média aproximada de 74% das especialidades inferidas estavam corretas em relação ao número total de usuários do ambiente.

Estes resultados são considerados positivos e ressaltam que os pesos para cada categoria de habilidade considerados na presente abordagem (Capítulo 5) estavam corretos. Entretanto, futuros experimentos podem ser executados para analisar se estes pesos são ideais e se podem ser considerados nos mais variados ambientes.

Para mensurar a qualidade das recomendações geradas pela abordagem adotada na categoria *Habilidade Calculada*, foi desenvolvido também um questionário personalizado e aplicado a este conjunto de indivíduos. Neste questionário, os candidatos a especialistas foram perguntados qual o nível de conhecimento e interesse que eles possuem nos assuntos em que foram detectados como especialistas. Da mesma forma que no questionário anterior, para evitar uma avaliação tendenciosa, estes indivíduos não foram informados que haviam sido detectados como possíveis especialistas nestes respectivos assuntos presentes no questionário. O resultado apresentado na Figura 7.10. representa uma média geral entre o grupo de usuários.

		Interest's Degree				
		1	2	3	4	5
knowledge's degree	1	60%				
	2					
	3	22%				
	4					
	5					

Figura 7.10: Qualidade das recomendações na categoria *Habilidade Calculada*

Novamente, foi adotada a métrica de precisão. Conforme apresentado na Figura 7.10 foi obtida uma precisão aproximada de 60%. Esta precisão é determinada pela razão entre a quantidade de especialidades corretamente detectadas e a quantidade total de especialidades detectadas para cada usuário. O percentual aproximado de especialidades erradas que foram inferidas foi de 22%, enquanto o restante, 18%, é um valor possivelmente justificável, uma vez que se refere aos assuntos em que os indivíduos possuem interesses, (ou, por exemplo, são co-autores de publicações que contemplam esses assuntos), mas ainda não são considerados especialistas.

### 7.3.5 Objetivos do Experimento – parte 2

O objetivo da segunda parte do experimento foi identificar o quanto os relacionamentos de confiança entre os indivíduos podem ser impactantes quando uma pessoa decide buscar por

ajuda de um especialista. Para isso, foram consideradas as especialidades de indivíduos inferidas pela abordagem proposta no SWEETS.

### 7.3.6 Cenário do Experimento – parte 2

Conforme mencionado na Seção 7.3.2, foram selecionadas 22 pessoas para atingir os objetivos do presente experimento. Os critérios para seleção destas pessoas foram:

1. possuírem currículos na plataforma *Lattes*;
2. possuírem publicações científicas;
3. possuírem conta no *LinkedIn*;
4. já possuírem algum contato prévio entre eles, isto é, se conhecerem.

Estas pessoas selecionadas teriam que ter informações suficientes para a inferência de suas especialidades através da abordagem do SWEETS. Por isso, a definição dos critérios 1, 2 e 3 são fundamentais. Além disso, é importante que estas pessoas se conheçam previamente, pois, assim, elas naturalmente possuem um conhecimento prévio das relações de confiança entre elas. Isto é, o grau de confiança que possuem umas nas outras e os parâmetros (e.g. reputação ou nível de conhecimento) para a definição deste grau.

### 7.3.7 Resultados do Experimento – parte 2

Foram importadas as informações destes usuários na Plataforma *Lattes* e *LinkedIn* e, logo em seguida, executada a abordagem do SWEETS para a inferência de especialidades. As especialidades possíveis de serem inferidas foram nas categorias *Informada* e *Demonstrada*: as *Informadas*, que são provenientes da plataforma *LinkedIn* e as especialidades *Demonstradas*, que são oriundas das publicações científicas disponíveis na plataforma *Lattes*. Se uma especialidade estiver nas duas categorias, então foi calculada uma média aritmética entre elas para que fosse obtido o grau final de especialidade (conforme apresentado no Capítulo 5, Equação (5.9)).

Foi obtida uma média de 11 especialidades por pessoa. Foram agrupados todos os especialistas de acordo com a área. Assim, havia uma lista dos especialistas áreas/assuntos específicos. Esta relação foi essencial para o passo seguinte, que foi a aplicação de um questionário personalizado para cada indivíduo. Nestes questionários personalizados, foram apresentados os especialistas que possuíam especialidades em uma mesma área *a*. Depois disso, foi feito a cada uma das pessoas o seguinte questionamento:

**Q1:** *“Se você precisasse de ajuda para sanar um problema específico, qual das pessoas relacionadas a seguir você procuraria prioritariamente? E quais delas ficariam como segunda ou terceira opção? Enumere sua resposta em uma escala de 1 a 3, sendo que 1 é para primeira opção e 3 para última. Considerando esta mesma escala, 1-3, informe qual o nível de confiança que você possui em cada especialista indicado e em que circunstância o conheceu. Se achar necessário, justifique suas escolhas.”*

Foram selecionadas cinco especialidades comuns entre os indivíduos para a aplicação deste questionário. Assim, os entrevistados responderam este questionamento cinco vezes, uma para cada especialidade. Os entrevistados não foram informados do nível de especialidade no assunto *a*, que cada especialista havia sido detectado.

Após a realização destas entrevistas, foram realizadas análises das respostas dos indivíduos em relação aos graus de especialidades dos especialistas em assuntos específicos inferidos pelo SWEETS. Uma das constatações desta análise foi que em aproximadamente 75% dos casos as pessoas procurariam prioritariamente por ajuda de pessoas que acreditam que têm maior relação de confiança, proporcionada por afinidades em comum ou um relacionamento prévio existente (*e.g.* colegas de pós-graduação ou que já trabalharam juntos). Em aproximadamente 56% dos casos, a terceira opção dos indivíduos era o especialista que possuía um maior nível de conhecimento no assunto especificado. Porém, ainda assim, nas justificativas das pessoas, elas ainda alegavam que tinham algum grau de relacionamento com estes especialistas.

## **7.4 Considerações Finais**

Os relacionamentos de confiança entre as pessoas é naturalmente uma premissa necessária na sociedade para colaboração entre indivíduos. Assim, a descoberta de novos relacionamentos em ambientes sociais pode ser um fator de estímulo à colaboração. Para que pudesse haver um entendimento de quais fatores podem influenciar na descoberta destas novas relações de confiança, foi realizado um experimento com um grupo de pessoas. Algumas constatações gerais desta pesquisa foram: em aproximadamente 90% dos casos as pessoas consideram o nível de conhecimento que uma pessoa tem em assuntos específicos ao decidirem confiar ou não nela; ou a semelhança entre os perfis de conhecimento e opinião delas em relação à outra pessoa; a reputação dos indivíduos foi também considerada como um fator de confiança entre indivíduos, de acordo com aproximadamente 77% das pessoas.



Este experimento foi fundamental para auxiliar na definição dos elementos a serem considerados no formalismo do T-SWEETS (Capítulo 4). T-SWEETS foi implantado em uma plataforma virtual social intitulada *Konnen* (Capítulo 6), que está sendo adotada em fase experimental em algumas turmas de graduação de uma instituição de ensino superior. Como resultado desta implantação, foi possível realizar um experimento com algumas turmas de graduação para mensurar o impacto que a abordagem para inferência e recomendação de relacionamentos ocultos de confiança poderia proporcionar ao ambiente. A expectativa era que T-SWEETS pudesse atuar como um fator de estímulo à colaboração entre indivíduos, isto é, um fator que pudesse contribuir para que as pessoas produzissem conhecimento com maior frequência no ambiente. Os resultados do experimento apresentados no presente capítulo comprovam esta tendência, uma vez que aproximadamente 82% dos novos relacionamentos estabelecidos no *Konnen* foram oriundos das recomendações dos relacionamentos de confiança provenientes do T-SWEETS.

Uma vez que confiança é um elemento decisivo quando uma pessoa decide colaborar ou não com outra, a adoção deste elemento nos sistemas de recomendação de especialistas para recomendar um especialista mais confiável pode ser um diferencial positivo. Às vezes, o especialista com um maior nível de conhecimento sobre um determinado assunto pode não ser a pessoa mais indicada para atender a uma solicitação. A adição da inferência destes relacionamentos de confiança é um das principais diferenças do SRE SWEETS proposto no presente trabalho.

Um ambiente ideal para mensurar o impacto que a adoção desta estratégia poderia trazer seria um ambiente em que as pessoas utilizassem um ambiente virtual social como uma plataforma de uso cotidiano em seu ambiente de trabalho. Por causa desta limitação, foram definidos alguns critérios para a seleção de algumas pessoas para a realização de um experimento e, assim, mensurar os impactos que uma abordagem para a inferência de relacionamentos de confiança poderia trazer para um sistema de recomendação de especialista.

O primeiro critério estabelecido foi estas pessoas já possuírem algum contato prévio, ou seja, se conhecerem a partir de experiências passadas (*e.g.* colegas de pós-graduação). Esta noção de confiança que eles têm entre si, a partir destas experiências passadas, auxiliaria nas escolhas futuras que eles poderiam fazer no momento de buscar por ajuda. Depois disso, foi aplicada a abordagem proposta no SWEETS para a inferência das especialidades destas pessoas. De posse destas especialidades, os especialistas que possuíam especialidades em comum foram agrupados de forma que as pessoas pudessem ser questionadas sobre quem elas procu-

rariam (dentre aqueles especialistas) caso precisassem de alguma ajuda sobre aquele assunto em específico. Uma das constatações desta análise foi que em aproximadamente 75% dos casos as pessoas procurariam por ajuda de pessoas que acreditam que têm maior afinidade ou um relacionamento prévio. O especialista que possuía um nível maior de conhecimento foi apenas a terceira opção em aproximadamente 56% dos casos – e, ainda assim, as pessoas alegavam que tinham algum relacionamento prévio com estes especialistas.

Outra importante constatação com os experimentos aplicados no SWEETS foi que a precisão na detecção das especialidades das pessoas foi de aproximadamente 91%. Ainda era importante mensurar o quanto a ordem do grau de especialidades destes indivíduos estava sendo detectada corretamente. Isto é, se um especialista possui um nível de conhecimento maior em um assunto *X* em relação a outro assunto *Y* ou vice-versa. A média da precisão desta ordem foi de aproximadamente 74%.

## 8 Conclusão

Uma comum e conhecida prática entre os seres humanos é a procura por informações entre pessoas de grupos que se relacionam. Isso porque as pessoas tendem a dar uma maior credibilidade às informações oriundas das suas redes sociais de contatos – colegas e amigos (Nardi, 2000; Borgatti, 2003; Plickert, 2007). Segundo a análise que consta em Gartner (2002) *apud* Lin (2008), as pessoas costumam obter 50 a 75% de suas informações diretamente de outras pessoas.

Embora as redes pessoais de relacionamentos possam atuar como uma alternativa para se obter respostas rápidas, às vezes elas não são suficientes para alcançar diretamente quem tem informações necessárias sobre um determinado assunto. Assim, essas redes têm como característica o alcance limitado. As pessoas em uma rede social podem atuar como intermediadores, facilitando assim o contato com pessoas ainda desconhecidas (Ehrlich, 2008) e, com isso, provendo uma maior interatividade, comunicação e colaboração entre as pessoas.

Em uma ferramenta com contexto social, se alguém recebe uma determinada requisição de informação, é mais provável esta pessoa responder à requisição se esta vier de uma pessoa amiga ao invés de uma pessoa desconhecida. Assim, segundo Ehrlich, (2008) um sistema que identifique especialistas poderia refletir o contexto social em que as pessoas estão inseridas. Segundo Ehrlich (2008), há algum tempo, pesquisadores da área argumentam que todo sistema de rede social deveria ser adicionado de tecnologias capazes de localizar especialistas em assuntos específicos. Assim, a procura por pessoas certas para sanar um determinado problema poderia ser bem mais eficiente.

Neste sentido, a presente tese teve como objetivo principal apresentar uma abordagem que identifique especialistas de domínios em ambientes virtuais sociais, intitulada SWEETS e como os relacionamentos ocultos de confiança entre os membros destas redes sociais podem

enriquecer o processo de identificação e recomendação destes especialistas de domínio. Além disso, apresentou também um novo método que se baseou no contexto das relações sociais destes indivíduos para a inferência de relacionamentos ocultos de confiança, intitulado T-SWEETS.

Um dos objetivos do T-SWEETS é promover uma maior interação entre os membros de uma rede social e, assim, naturalmente, proporcionar uma maior produção de conhecimento por partes destes membros. Desta forma, fornecendo um maior insumo para a inferência das especialidades destes membros. Neste sentido, foi realizado um experimento junto a um grupo de pessoas para identificar quais elementos poderiam ser impactantes como um fator decisivo na colaboração.

A partir desta entrevista, chegou-se a algumas constatações, são elas: aproximadamente 57% das pessoas costumam ajudar pessoas que possuem afinidade ou confiança, seja direta ou indireta; aproximadamente 92% das pessoas têm uma tendência a colaborar com pessoas que possuem opiniões ou conhecimentos similares aos delas; aproximadamente 77% das pessoas acreditam que a reputação de um indivíduo é um fator decisivo de colaboração; e aproximadamente 90% dos entrevistados acreditam que o nível de conhecimento que as pessoas têm sobre assuntos específicos também é um fator determinante de colaboração, pois pessoas que possuem maior nível de conhecimento em assuntos específicos têm uma tendência natural a produzirem conhecimento de maior nível de qualidade.

Estes resultados ofereceram sustentação para a definição dos 4 elementos adotados no T-SWEETS são eles: *Similaridade entre Perfis*, *Relacionamentos de Confiança*, *Nível de Maturidade dos Usuários*; e *Reputação*. T-SWEETS foi então desenvolvido e implantado em uma rede social para a realização de um experimento que durou 4 meses: 2 meses sem a implantação do T-SWEETS e 2 meses após a sua implantação. O objetivo principal era verificar o impacto que uma abordagem para inferência de relacionamentos ocultos de confiança poderia trazer em um ambiente virtual social. Como resultado deste experimento, constatou-se que aproximadamente 82% dos novos relacionamentos que surgiram depois dos 2 primeiros meses iniciais foram oriundos das recomendações do T-SWEETS. Assim, a expectativa é que a adoção de estratégias desta natureza possa contribuir fortemente para uma maior colaboração entre os membros de um ambiente virtual social e, assim, naturalmente, uma maior e mais frequente explicitação de conhecimento.

A descoberta destes relacionamentos ocultos de confiança é importante também para a recomendação do especialista mais indicado para uma pessoa em um determinado instante de tempo  $t$ . Às vezes, a pessoa que possui um maior nível de conhecimento em um assunto  $a$  pode não ser a mais indicada para prestar auxílio à outra. Neste sentido, foi realizado um experimento com um grupo de pessoas especialistas em alguns assuntos em comum para identificar quem (tipo de especialista) elas procurariam quando eventualmente necessitassem de ajuda em um assunto  $a$ . Como resultado desta análise, chegou-se à conclusão que em aproximadamente 75% dos casos as pessoas buscam prioritariamente por ajuda de pessoas que acreditassem que têm maior relação de confiança, proporcionada por afinidades em comum ou um relacionamento prévio existente (*e.g.* colegas de pós-graduação ou que já trabalharam juntos). A procura pelo especialista que possui um maior nível de conhecimento ficou apenas na terceira posição na lista de prioridade, mas, ainda assim, houve justificativas em que as pessoas alegavam possuírem algum grau de relacionamento com estes especialistas.

## 8.1 Contribuições

As principais contribuições desta tese são:

- C1: a descoberta de um conjunto de fatores a partir uma entrevista com um grupo de pessoas para entender quais elementos são considerados como um fator decisivo na decisão de colaboração entre indivíduos. Foi constatado que as pessoas costumam considerar os seguintes fatores: a afinidade ou confiança; a reputação; capacidade cognitiva, a partir da explicitação de conhecimento de qualidade dos indivíduos; e a semelhança entre os perfis de conhecimento e opiniões similares aos seus.
- C2: uma nova abordagem para inferência de relacionamentos ocultos de confiança em ambientes virtuais sociais;
- C3: a implantação deste método para inferência de relacionamentos de confiança em um ambiente virtual social para mensurar o impacto que este eventualmente, poderia ter proporcionar;
- C4: uma abordagem para inferência das especialidades dos indivíduos em plataformas virtuais, sejam elas sociais ou não. Esta abordagem é baseada em 4 tipos de habilidades: Informada, Demonstrada, Sugerida e Refutada;
- C5: a proposta de outra categoria de especialidade: Habilidade Calculada. Esta categoria baseia-se nas produções dos usuários e em folksonomia, que sustenta o surgimento de ontologias independentes de domínio, o que torna a estratégia flexível;

C6: A proposta de utilização de fontes externas de conhecimentos, tais como, Facebook, LinkedIn, Twitter e Plataforma Lattes, para evitar o problema de ausência de conhecimento dos indivíduos no processo de inferência de suas especialidades;

C7: E, por fim, a utilização dos relacionamentos de confiança entre os indivíduos, que são oriundos do contexto das relações sociais em que estão inseridos, para prover o especialista mais propenso a colaborar em uma determinada situação.

### Publicações

Contribuições	Publicações
C2	<ol style="list-style-type: none"> <li>1. SILVA, E. M. et al. Recommendation of Trust Relationships in Social Networks. In: IADIS International Conference - Collaborative Technologies, 2013, Prague. IADIS International Conference - Collaborative Technologies, 2013.</li> <li>2. SILVA, E. M. et al. T-SWEETS: An Alternative to the Stimulus Collaboration from Trust Inference in Social Networks. In: Brazilian Symposium on Collaborative Systems, 2012, São Paulo. 2012 Brazilian Symposium on Collaborative Systems, 2012. p. 168-173.</li> </ol>
C1, C2 e C3	<ol style="list-style-type: none"> <li>3. Silva, E. M. et al. Inferring Hidden Trust Relationships in Social Networks for Encouraging Collaboration and Cooperation among Individuals, 2013, CRIWG - 19th International Conference on Collaboration and Technology, Wellington, New Zeland.</li> </ol>
C5	<ol style="list-style-type: none"> <li>4. SILVA, E. M. et al. Recommending Knowledge in a knowledge based Social Network. Journal of Applied Computing Research, 2011, v. 1, p. 20-32.</li> </ol>

### Outras Publicações

Publicações
<ol style="list-style-type: none"> <li>5. SILVA, E. M. ; COSTA, R. A. ; SCHMITZ, L. R. B. ; MEIRA, S. R. L. . Recommending Knowledge in a knowledge based Social Network. Journal of Applied Computing Research, v. 1, p. 20-32, 2011.</li> <li>6. SILVA, E. M. ; COSTA, R. A. ; SCHMITZ, L. R. B. ; MEIRA, S. R. L. . SWEETS: um Sistema de Recomendação de Especialistas aplicado a uma plataforma de Gestão de Conhecimento. Revista de Informática Teórica e Aplicada (Impresso), v. 18, p. 83-111, 2011.</li> <li>7. Souza, J. G. ; SILVA, E. M. ; BRITO, P. F. ; COSTA, J. A. F. ; SALGADO, A. C. ; MEIRA, S. R. L. . Using Graph Clustering for Community Discovery in Web-Based</li> </ol>

Social Networks. In: International Conference on Swarm Intelligence, 2013, Harbin. ICSI 2013, 2013. v. 7929.

8. BRITO, D. M. ; BRITO, P. F. ; SILVA, E. M. . Ontologia baseada em Objetos de Aprendizagem na definição de Conteúdos de uma Rede Social Acadêmica. In: 23º - Simpósio Brasileiro de Informática na Educação, 2012, Rio de Janeiro. 23º - Simpósio Brasileiro de Informática na Educação, 2012.

## 8.2 Limitações

A abordagem adotada no SWEETS para identificação das especialidades das pessoas é baseada em 4 categorias de habilidades: *Informada*, *Demonstrada*, *Sugerida* e *Refutada*. O formalismo adotado na categoria *Demonstrada* baseia-se nas publicações científicas dos autores e respectivas categorias a que pertencem para mensurar o nível de conhecimento que estes têm em assuntos abordados nestas publicações. O grau de conhecimento nos assuntos abordados em cada publicação é distribuído de maneira homogênea entre os autores, isto é, se uma publicação  $p$  tem 2 autores e engloba o assunto  $a$ , o nível de conhecimento sobre  $a$  é distribuído na mesma proporcional aos 2 autores. Esta é uma limitação na abordagem, pois há autores que conhecem mais um assunto específico em relação a outro.

Embora as habilidades nas categorias *Sugerida* e *Refutada* façam parte da abordagem adotada no SWEETS, infelizmente elas não puderam fazer parte dos experimentos. Para que houvesse uma validação adequada das habilidades nestas categorias seria necessário um ambiente real em que as pessoas o usassem para o desenvolvimento de suas atividades no dia-a-dia, por exemplo, um ambiente organizacional. Este cenário ideal seria importante, pois ambas as habilidades necessitam de interação e intervenção explícita dos usuários do ambiente, já que são os próprios usuários que sugerem ou refutam a habilidade de um candidato a especialista.

As 4 categorias de habilidades supracitadas determinam o grau final de especialidade em um assunto  $a$  de uma especialista  $e$ , porém há outra categoria de habilidade, a *Calculada*, que infere possíveis especialidades das pessoas a partir de suas produções (e.g. documentos) e ontologia ( $O_{ci}$ ) baseada em folksonomia. Estas especialidades podem, em um passo seguinte, mudar de categoria: para *Informada* ou *Sugerida*, dependendo das ações dos usuários. A adoção de folksonomia para o surgimento de ontologia traz consigo alguns problemas, já que uma das principais características de folksonomia é a liberdade que os usuários têm em categorizar objetos. Alguns destes problemas são: as entradas dos usuários podem conter erros de digita-

ção ou podem ser inconsistentes. Isto é, ora o usuário associa a *tag* “*Rede Social*” a um item, ora associa uma *tag* diferente a outro item, mas que expressam o mesmo significado, por exemplo, “*Redes Sociais*”, o que compromete diretamente a qualidade da  $O_{ci}$ . Além disso, quanto mais “rica” for a folksonomia, ou seja, possuir um número maior de classificações, maior é a tendência na eficiência para identificação das especialidades dos indivíduos.

### 8.3 Trabalhos Futuros

São vários os projetos que ainda podem ser desenvolvidos para que o presente trabalho possa obter um maior nível de profundidade, dentre eles:

- considerando a limitação da distribuição homogênea de conhecimento para a inferência das especialidades na categoria *Demonstrada*, pode ser desenvolvido um mecanismo que identifique dentre o conjunto de autores de uma publicação quais são aqueles que possuem um nível maior ou menor em cada assunto abordado no trabalho. E, a partir desta análise, a distribuição do nível de conhecimento nas áreas específicas de cada publicação pode ser realizada de maneira proporcional;
- a implantação da abordagem para a identificação e recomendação de especialistas em um ambiente real organizacional. Com isso, seria possível obter parâmetros que poderiam auxiliar na identificação de quão precisa é a abordagem completa apresentada no SWEETS;
- a partir da agregação/unificação do conhecimento dos usuários produzidos em diferentes plataformas (*e.g.* *LinkedIn* e *Lattes*) em um único repositório, oferecendo subsídios para a identificação das especialidades dos usuários, pode ser proposto e desenvolvido um modelo geral para representação deste conhecimento. Por exemplo, o vocabulário *FOAF* (*Friend of a Friend*)<sup>20</sup> foi desenvolvido com o propósito de uma representação única para cada pessoa na Web, suas atividades e respectiva rede de relacionamentos. Desta maneira, poderia ser desenvolvida uma forma de representar o conhecimento pertinente ao usuário e agregá-lo a uma representação do FOAF. Assim, aonde quer que esteja uma pessoa que compõe a rede, seu conhecimento e relacionamentos estarão agregados com ela.
- para diminuir as redundâncias dos conceitos representados na ontologia  $O_{ci}$  e aumentar o peso semântico entre esses conceitos, que é imprescindível para o cálculo da ha-

---

<sup>20</sup> The Friend of a Friend (FOAF) project - <http://www.foaf-project.org/>



bilidade na categoria *Calculada*, podem ser aplicados algoritmos de *stemming* (redução da palavra ao seu radical);

- enriquecer a ontologia  $O_{ci}$  utilizando um *Thesaurus* como, por exemplo, o *thesauro WordNet* (Fellbaum, 1998). Assim, para cada conceito da ontologia  $O_{ci}$  seria verificada sua existência no *thesaurus* e os respectivos sinônimos, hierarquias e relacionamentos entre os demais conceitos e, caso exista, todas essas informações seriam agregadas à ontologia  $O_{ci}$ ;
- podem ser utilizadas técnicas de aprendizagem de máquina, por exemplo, *clustering*, para o desenvolvimento de mecanismos de recomendação de *tags* aos usuários, assim, sempre que estes usuários forem categorizar um objeto da folksnomia, *tags* mais relevantes para aquele objeto podem ser recomendadas;
- para tornar o processo de inferência dos relacionamentos de confiança mais eficaz, podem ser agregados novos elementos aos algoritmos T-SWEETS para identificar quais são os usuários que não possuem um comportamento padrão no ambiente, ou seja, os usuários que realizam avaliações tanto do conteúdo quanto dos usuários de forma aleatória, sem nenhum padrão. Comportamentos desta natureza podem prejudicar o desempenho do T-SWEETS.

## Referências Bibliográficas

- Ackerman, M.; Pipek, V.; Wulf, V.; (2003): *Sharing Expertise: Beyond Knowledge Management*, The Mit Press.
- Adomavicius, G.; Tuzhilin, A; (2005): Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, Issue. 6, pp. 734-749.
- Anand, D.; Bharadwaj, K.; (2011) : Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities, *Expert Systems with Applications*, Vol. 38, Issue 5, pp. 5101-5109.
- Artz, D.; Gil, Y. (2007): A Survey of Trust in Computer Science and the Semantic Web. *Journal of Web Semantics: Science, Services and Agents on the World* 40.
- Baeza-Yates, R.; Ribeiro-Neto, B.; (1999): *Modern Information Retrieval*, ACM Press, New York, USA.
- Baldauf, M.; Dustdar, S.; e Rosenberg, F.; (2007): A survey on context-aware systems, In: *Int. J. Ad Hoc Ubiquitous Comput.*, Geneva, Switzerland, Inderscience Publishers, Vol. 2, Issue. 4, p. 263-277.
- Barnes, J. A., (1972): *Social networks*, Addison-Wesley Module in Anthropology, vol. 26, pp.1-29,
- Becerra-Fernandez, Irma; (2006): Searching for experts on the Web: A review of contemporary expertise locator systems. *ACM Trans. Internet Technol.*, v. 6, pp. 333-355.

- Bettini, C.; Brdiczka, O.; Henricksen, K.; Indulska, J.; Nicklas, D.; Anand, Ranganathan, Riboni, D. (2010): A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, Vol. 6, Issue: 2, Publisher: Elsevier B.V., pp. 161-180
- Borgatti, S.P.; Cross, R. (2003): A Relational View of information seeking and learning in social networks. *Management Science*, Vol. 49, Issue. 4, pp. 432-445.
- Borges, M. R. S.; Brézillon, P.; Pino, J. A.; Pomerol, J.-C; (2004): Bringing context to CSCW, In: *Proc. of the 8th International Conference on Computer Supported Cooperative Work in Design (CSCWD'04)*, Xiamen, China, pp. 161-166.
- Bozzon, A.; Brambilla, M.; Ceri, M; Silvestri, M. Vesci, G.; (2013): Choosing the right crowd: expert finding in social networks. *Proceedings of the 16th International Conference on Extending Database Technology*. New York, NY, USA: ACM Press, pp. 637-648.
- Brézillon, P. (2007): Context modeling: Task model and model of practices. In: *Proc. of the 6th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT'07)*, LNAI 4635, Roskilde, Denmark, pp. 122-135.
- Brézillon, P.; Araujo, R. M.; (2005): Reinforcing Shared Context to Improve Collaboration. *Revue d'Intelligence Artificielle*, Vol. 19 Issue. 3, pp. 537-556.
- Brin, S.; Page L (1998): The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, Vol. 30, Issue. 1-7 pp. 107-117.
- Castelli, G.; Rossi, A.; Mamei, M.; Zambonelli, F.; (2007): A simple model and infrastructure for context-aware browsing of the world, in: *Proceedings of the 5th IEEE Conference on Pervasive Computing and Communications*, IEEE Computer Society, White Plains, NY, pp. 229 - 238.
- Cohen, D.; Prusak, L. (2001): *In Good Company: How Social Capital makes Organizations Work*, Knowledge Management, Harvard Business School Press, Boston.
- Cook, J. A.; Gouda, M. G. (1997): Flow Theory. *Networking*, IEEE/ACM Transactions, Vol. 5, Issue. 5, pp. 661-674.
- Cook, K. S. & Cooper, R. M. (2001): Experimental studies of cooperation, trust, and social exchange. In: E. Ostrom & J. Walker, eds. *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, New York: Russell Sage Foundation, vol. VI, pp. 209-244.

- Costa, R. A.; Silva, E. M.; Neto, M. G.; Delgado, D. B.; Ribeiro, R. A; Meira, S. R. L; (2009): Social knowledge management in practice: a case study, International conference on Groupware: Design, Implementation, and Use, Springer-Verlag Berlin, Heidelberg, pp. 94-109.
- Cranston, D. W. (2009): Regular bipartite graphs are antimagic, Journal of Graph Theory, Vol. 60, Issue 3, pp. 173-182.
- Cutting, D. R.; Karger, D. R.; Pedersen, J. O.; Tukey, J. W. (1992): Scatter/Gather: a Cluster-based Approach to Browsing Large Document Collections, In Proceedings of the ACM SIGIR, New York, NY, USA , pp. 318-329.
- Datta, A.; Yong, J. T. T.; Ventresque, A.; (2011): T-RecS: Team Recommendation System through Expertise and Cohesiveness. Proceedings of International World Wide Web Conference Committee. Hyderabad, India: ACM Press. pp. 201-204.
- Demartini, G. (2007): Finding experts using wikipedia. Proceedings of the Workshop on Finding Experts on the Web with Semantics. (FEWS2007) at ISWC/ASWC2007, Busan, South Korea, pp. 33-41.
- Deutsch, M. (1962): Cooperation and Trust. Some Theoretical Notes. in Jones, M.R. (ed) Nebraska Symposium on Motivation. Univer. Nebraska Press, xiii, 330 pp.
- Earl, M. (2001): Knowledge management strategies: Toward a taxonomy. Journal of Management Information Systems, Vol. 18, Issue 1, pp. 215-233
- Ehrlich, K.; Shami, N.S.; (2008): Searching for expertise, in: ACM Conference on Human Factors, in Computing Systems, New York: ACM Press, pp. 1093-1096.
- Fazel-Zarandi, M.; Fox, M. S. (2011): Constructing expert profiles over time for skills management and expert finding. Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, Graz, AUT, New York, NY, USA, pp. 5:1-5:6
- Fazel-Zarandi, M.; Fox, M. S. (2012): An Ontology for Skill and Competency Management. Proceedings of the 7th International Conference on Formal Ontology in Information Systems.
- Feldman., R.; Dagan, I.; Hirsh, H. (1998): Mining Text Using Keyword Distributions, Journal of Intelligent Information Systems. Vol. 10, Issue. 3, p. 281 – 300.

- Fellbaum, C. (1998): WordNet - An Electronic Lexical Database. Cambridge, MIT Press, pp. 423.
- Fontes, L. M. O.; Neto, F. M. M.; Pontes, A. Á. A.; Campos, G. A. L.; (2011): An agent-based architecture for supporting the workgroups creation and the detection of out-of-context conversation on problem-based learning in virtual learning environments, pp. 1175-1180. SAC - ACM Symposium on Applied Computing.
- Freyne, J.; Varga, E.; Byrne, D.; Smeaton, A. F; Smyth, B.; Jones, G.J.F.; (2007): Realising context-sensitive mobile messaging. In: MONET, 25-30 November, Vilamoura, Portugal.
- Galho, T. S.; Moraes, S. M. W. (2004): Categorização Automática de Documentos de Texto utilizando Lógica Difusa. Logos, Vol. 15, Issue. 1, pp. 91-104.
- Gartner; (2002): The Knowledge Worker Investment Paradox.
- Gavrilov, L. A; E Gavrilova, Natalia S.; (2001): The Reliability Theory of Aging and Longevity. Journal of Theoretical Biology, Vol. 213, Issue 4, pp. 527–545.
- Golbeck, J. (2008): Trust on the World Wide Web: A Survey. Foundations and Trends in Web Science, Vol. 1: Issue. 2, pp 131-197.
- Golbeck, J. (2005): Computing and Applying Trust in Webbased Social Networks. Ph.D. Dissertation, University of Maryland, College Park, MD, USA
- Golbeck, J.; Hendler, J. (2006): Inferring binary trust relationships in Web-based social networks, Journal ACM Transactions on Internet Technology, Vol. 6, Issue. 4.
- Gollapalli, S. D.; Mitra, P.; Giles, C. L. (2011): Ranking authors in digital libraries. Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, ACM Press, New York, NY, USA, pp. 251-254.
- Gollapalli, S. D.; Mitra, P.; Giles, C. L.; ( 2012): Similar researcher search in academic environments. Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries, ACM Press, New York, NY, USA, pp. 167-170.
- Greenberg, S. (2001): Context as a Dynamic Construct, Human Computer Interaction, Vol. 16, Issue. 2-4, pp. 257-268.
- Gruber, T. R. (1993): Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino and R. Poli, editors, Formal Ontology in Conceptual Analysis and Knowledge Representation, Deventer, The Netherlands, Kluwer Academic Publishers.

- Gruber, T.; (1995): Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *International Journal Human-Computer Studies*, vol. 43, ed 5-6, p. 907-928.
- Gutwin, C.; Greenberg, S.; Blum, R.; Dyck, J.; (2005): Supporting Informal Collaboration in Shared-Workspace Groupware. Computer Science Department, University of Saskatchewan Technical Report HCI-TR-05-01.
- Halin, G.; Hanser, D. (2003): Towards contextual visualization of collaborative design. Proceedings of the 15th French-speaking conference on human-computer interaction on Conference Francophone sur l'Interaction Homme-Machine, New York, NY, USA, pp. 208-211.,
- Hardin, R. (2001): Gaming trust. In: E. Ostrom & J. Walker, eds. *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, New York: Russell Sage Foundation, Vol. VI, pp. 80–102.
- Henricksen, K.; Indulska, J.; (2006): Developing Context-Aware Pervasive Computing Applications: Models and Approach, *Pervasive and Mobile Computing Journal*, Vol. 2, Issue. 1, pp. 37-64.
- Igarashi, T.; Kashima, Y.; Kashima, E. S; Farsides, T.; Kim, U.; Strack, Fritz, Werth, Lioba, Yuki, M.; (2008): Culture, trust, and social networks, *Asian Journal of Social Psychology*, Vol. 11, Issue: 1, pp. 88-101.
- Katz Y.; Golbeck J. (2006): Social network-based trust in prioritized default logic, Proceedings of the 21st National Conference on Artificial Intelligence, pp.1345-1350..
- Kautz, R.; Selman, B.; Shah, M. (1997): ReferralWeb: Combining Social Networks and Collaborative Filtering, *Communications of the ACM*, Vol. 40, Issue. 3, p. 63-65.
- Klyne, G.; Reynolds, F.; Woodrow, C.; Ohto, H.; Hjelm, J.; Butler, M. H.; Tran, L. (2004): Composite capability/preference profiles (CC/PP): Structure and vocabularies 1.0, W3C Recommendation, Tech. Rep., W3C, January.
- Kuchar, Layered Layout. GitHub, (2011): Disponível em: <<https://github.com/jaroslav-kuchar/Layered-Layout>>. Acesso em 29 de julho de 2013.
- Kuter, U.; Golbeck, J.; (2007): SUNNY: a new algorithm for trust inference in social networks using probabilistic confidence models, Proceedings of the 22nd national conference on Artificial intelligence, Vancouver, British Columbia, Canada, pp.1377-1382.

- Kuter, U.; Golbeck, J.; (2010): Using probabilistic confidence models for trust inference in Webbased social networks. *ACM Trans. Intern. Tech.* Vol. 10, Issue. 2.
- Leskovec, J.; Huttenlocher, D.; Kleinberg, J.: (2010) Signed networks in social media, *CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1361-1370.
- Lin, C.-Y.; Ehrlich, K.; Griffiths-Fisher, V.; (2008): Desforges, C., *SmallBlue: People Mining for Expertise Search, Multimedia, IEEE*, Vol. 15, Issue. 1, pp. 78-84.
- Lin, C-Y.; Cao, N.; Liu, S. X.; Papadimitriou, S.; Sun, J.; Yan, X. (2009): *SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence, ICDE '09. IEEE 25th International Conference on*, Shanghai, China, pp. 1483–1486.
- Liu, D.-R. Chen; Y.; Kao; W.; Wang, H.; (2013): Integrating expert profile, reputation and link analysis for expert finding in question-answering Websites. *Information Processing and Management*, Vol. 49, Issue. 1, pp. 312-329.
- Mamami, E. Z. S.; Gerosa, M. A., (2011): *Cálculo de Reputação em Redes Sociais. VIII Simpósio Brasileiro de Sistemas Colaborativos*, Paraty, Rio de Janeiro, Brazil, pp. 202–207.
- Marsh, S. (1994): *Formalising Trust as a Computational Concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling.
- McArthur, R.; P. Bruza; (2003): Discovery of Implicit And Explicit Connections Between People Using Email Utterance. In: *ECSCW 2003 – Proceedings of the Eighth Conference on Computer Supported Cooperative Work*, Helsinki, Finland, pp. 21–40.
- McDonald, D. W. (2000): *Supporting Nuance in Groupware Design: Moving from Naturalistic Expertise Location to Expertise Recommendation*, Tese de Doutorado, University of California, Irvine.
- Melville, P.; Sindhvani, V.; (2010): *Encyclopedia of Machine Learning*. Springer, , ch. Recommender Systems, pp. 829–838.
- Mika, P. (2007): Ontologies are us - A unified model of social networks and semantics, *Journal of Web Semantics*, Vol. 5, Issue. 1, pp. 5-15.
- Moraes, A. K. O.; Trindade, C. C; Barbosa, Y. A. M.; Silva, E. V. F. ; Meira, S. R. L. (2010): Recommending experts using communication history. In: *2nd International Workshop on Recommendation Systems for Software Engineering*, Cape Town, South Africa. Proceed-

- ings of the 2nd International Workshop on Recommendation Systems for Software Engineering (RSSE '10), ACM Press, New York, NY, USA, pp. 41-45.
- Moran, T. P., Dourish, P. (2001): Introduction to the Special Issue on Context-Aware Computing, *Human Computer Interaction*, Vol. 16, Issue. 2, 3 e 4, pp. 87-95.
- Murayama, H.; Taguchi, A.; Ryu, S.; Nagata, S.; Murashima, S. (2011): Institutional trust in the national social security and municipal healthcare systems for the elderly and anxiety with respect to receiving elderly care in Japan: a cross-sectional study, *Oxford Journals* (Oxford University Press), *Health Promot.*
- Musiał, K.; Kazienko, P. (2013): Social networks on the Internet, *WWW*, Vol. 16, pp. 31–72
- Nardi, B. A.; Whittaker, S.; Schwarz, H.; (2000): It's not what you know, it's who you know: Work in the information age. *Peer-Reviewed Jornal On The Internet*, 2000
- Nie, D.; Zhang, Z.; Zhou, J.; Fu, Y. (2011): Evaluating the Effects of Trustworthiness and Preference in Recommender Systems. 2nd International Workshop on Social Recommender Systems, Hangzhou, China.
- Pbworks, Online Team Collaboration.; (2010):Williams, A.; Messina C. OAuth: FrontPage. Janeir. Disponível em: < <http://wiki.oauth.net> >.
- Pessoa, R. M.; Calvi1, C. Z.; Filho, J. G. P.; Andreão, R. V. (2006): Aplicação de um middleware sensível ao contexto em um sistema de telemonitoramento de pacientes cardíacos. XXXIII Seminário Integrado de Software e Hardware (SEMISH'06), Campo Grande-MS, Brasil. SBC Press, pp 1-15.
- Petry, H.; Tedesco, P.; Vieira, V., Salgado, A. C.: (2008): ICARE: A Context-Sensitive Expert Recommendation System, *ECAI 2008 - Workshop on Recommender Systems*, pp. 53-58.
- Pimentel, M.; Fuks, H. (2011): *Sistemas Colaborativos*, 1ª edição, Editora: Campus / Elsevier.
- Plickert, G.; Côtéa, R. R.; Wellman, B.; (2007): It's not who you know, it's how you know them: Who exchanges what with whom?. *Social Networks*, Vol. 29, Issue. 3, pp. 405-429.
- Pound, J.; Mika, P.; Zaragoza, H. (2010): Ad-hoc object retrieval in the Web of data, *Proceedings of the 19th international conference on World wide Web*, New York, NY, USA, pp. 771-780.



- Reichling, T.; Schubert, K.; Wulf, V. (2005): Matching Human Actors Based on their Texts: Design and Evaluation of an Instance of the Expert Finding Framework. In: International ACM Siggroup Conference on Supporting Group Work, Proceedings, New York, ACM Press, pp. 61-70.
- Reichling, T.; Veith, M.; Wulf, V., (2007): Expert Recommender: Designing for a Network Organization, Computer Supported Cooperative Work (CSCW), Vol. 16, No. 4-5, pp. 431-465.
- Riahi, F.; Zolaktaf, Z.; Shafiei, M.; Milios, E. (2012): Finding expert users in community question answering . Proceedings of the 21st international conference companion on World Wide Web, New York, NY, USA, pp. 791-798.
- Ries, S.; Kangasharju, J.; & Muhlhauser, M.; (2006): A Classification of trust systems, 782 LNCS 4277/2006. Berlin: Springer, pp. 894–903
- Runte, M.; (2000): Personalisierung im Internet – Individualisierte Angebote mit Collaborative Filtering., Deutscher Universitäts-Verlag, Wiesbaden.
- Santos, V. V.; (2008): CEManTIKA: A Domain-Independent Framework for Designing Context-Sensitive Systems, Tese de Doutorado, Universidade Federal de Pernambuco, PE.
- Scheffler, I. (2010): Worlds of Truth: A Philosophy of Knowledge By Israel iScheffler, Worlds of Truth: A Philosophy of Knowledge, Oxford University Press, vol. 70, Issue 2, pp. 383-385,
- Silva, E.M. et al. (2011): Recommending knowledge in a knowledge based social network, Journal of Applied Computing Research, Vol. 1 Issue 1, pp. 20-32.
- Sztompka, P. (1999): Trust: A Sociological Theory, Cambridge: Cambridge University Press.
- Taherian, M.; Amini, M., and Jalili, R. (2008): Trust Inference in Web-Based Social Networks using Resistive Networks, Proc. of ICIW, Washington, DC, USA, pp. 233-238.
- Tchunte, D.; Canut, M.-F.; Jessel, N.B.; Péninou, A.; El Haddadi, A.; (2010): Visualizing the Evolution of Users' Profiles from Online Social Networks. International Conference on Advances in Social Networks Analysis and Mining, Odense, Denmark, pp. 370 – 374.
- Trindade, C. C.; Moraes, A. K. O.; Barbosa, Y. A. M.; Albuquerque, J. O.; Meira, S. R. L. (2009): Um Sistema de Recomendação de Especialistas em Desenvolvimento Distribuído de Software: Requisitos, Projeto e Resultados Preliminares. In: VI Simpósio Brasileiro de

- Sistemas Colaborativos (SBSC), Fortaleza-CE. VI Simpósio Brasileiro de Sistemas Colaborativos.
- Tyler, J. R., D. M. Wilkinson and B.A. Huberman (2003): Email as Spectroscopy: Automated Discovery of Community Structure Within Organizations. In M. Huysman, E. Wenger and V. Wulf (eds): *Communities and Technologies 2003*, Dordrecht: Kluwer, pp. 81–96.
- Uschold, M. (1996): Gruninger, M., *Ontologies: principles, methods and applications*, The Knowledge Engineering Review, Vol. 11, Issue. 2 pp. 93-136.
- Viana, W., Bringel Filho, J., Gensel, J., Villanova-Oliver, M., Martin, H. (2008): PhotoMap: from location and time to context-aware photo annotations. *Journal of Location Based Services*, Vol. 2 Issue 3, pp. 211-235.
- Vieira, V.; Brézillon, P.; Salgado, A. C.; Tedesco, P.; (2008b): A Context-Oriented Model for Domain-Independent Context Management, *Revue d'Intelligence Artificiel*, Vol. 22, Issue. 5, pp. 609-627.
- Vieira, V.; Mangan, M. A. S.; Werner, C. M. L.; Mattoso, M. L. Q.; (2004): Ariane: An Awareness Mechanism for Shared Databases, In: *Proc of the X International Workshop on Groupware (CRIWG'04)*, San Carlos, Costa Rica, pp. 92-104.
- Vieira, V.; Tedesco, P.; e Salgado, A. C.; (2009): A Process for the Design of Context-Sensitive Systems, In: *Proc. of the Int. Conference on Computer Supported Cooperative Work in Design*, IEEE, Washington, DC, USA, pp. 143-148.
- Vieira, V.; Tedesco, P.; Salgado, A. C.; Brézillon, P.; (2007): Investigating the Specifics of Contextual Elements Management: The CEManTIKA Approach, *Modeling and Using Context (CONTEXT'07)*, Berlin, Heidelberg, pp. 493-506.
- Wal, T. V.; (2007): Folksonomy coinage and definition. Disponível online em: <<http://www.vanderwal.net/folksonomy.html>>.
- Wang, G.; Wu J. (2011): FlowTrust: Trust Inference with Network Flows. In *Frontiers of Computer Science. China*, Vol. 5, Issue. 2, pp. 181-194.
- Wellman, B., et al. (1996): Computer networks as social networks: collaborative work, telework, and virtual community. *Annu. Rev. Sociol.* Vol. 22, Issue. 1, pp. 213–238.
- Yamagishi, T.; Yamagishi, M. (1994): Trust and commitment in the United States and Japan. *Motivation and Emotion*, Vol. 18, Issue 2, pp. 129–166.

- Yamagishi, T. (1986): The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, Vol. 51, Issue 1, pp. 110–116.
- Yamagishi, T.; Cook, K. S.; Watabe, M. (1998): Uncertainty, trust and commitment formation in the United States and Japan. *American Journal of Sociology*, Vol. 104, Issue 1, pp. 165–194.
- Yamamoto, Y. (1990): A Morality Based on Trust: Some Reflections on Japanese Morality, *Journal Philosophy East & West*. Vol: 40, Issue. 4, pp. 451-469.
- Ye, Y.; Nakakoji, K. and Yamamoto, Y. (2007): Reducing the cost of communication and coordination in distributed software development, *Lecture Notes in Computer Science*, 2007, LNCS 4716, pp. 152–169.
- West, D. B. (2001): *Introduction to Graph Theory*, Second Edition, Prentice Hall, Upper Saddle River, N. J..
- Zimmermann, A.; Lorenz, A.; Oppermann, R.; (2007): An Operational Definition of Context, In: *Proc. of the 6th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT'07)*, LNAI 4635, Roskilde, Denmark, pp. 558-571.