

UNIVERSIDADE DO EXTREMO SUL CATARINENSE - UNESC

CURSO DE CIÊNCIA DA COMPUTAÇÃO

JOÃO MIGUEL DE OLIVEIRA FRANCISCO

**ANÁLISE DE PONTOS DE PROPAGAÇÃO DE INFORMAÇÃO EM REDES
SOCIAIS COM O USO DO ALGORITMO SOMA-E-PRODUTO**

CRICIÚMA

2012

JOÃO MIGUEL DE OLIVEIRA FRANCISCO

**ANÁLISE DE PONTOS DE PROPAGAÇÃO DE INFORMAÇÃO EM REDES
SOCIAIS COM O USO DO ALGORITMO SOMA-E-PRODUTO**

Trabalho de Conclusão de Curso apresentado para
obtenção do grau de Bacharel no curso de Ciência da
Computação da Universidade do Extremo Sul
Catarinense, UNESC.

Orientadora: Prof.^a MSc . Leila Laís Gonçalves

Co-orientador: Prof. MSc. Kristian Madeira

CRICIÚMA

2012

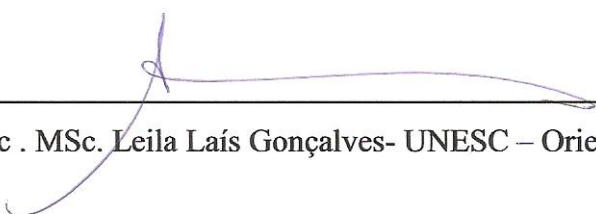
JOÃO MIGUEL DE OLIVEIRA FRANCISCO

**ANÁLISE DE PONTOS DE PROPAGAÇÃO DE INFORMAÇÃO EM REDES
SOCIAIS COM O USO DO ALGORITMO SOMA-E-PRODUTO**

Trabalho de Conclusão de Curso aprovado pela
Banca Examinadora para obtenção do Grau de
Bacharel, no Curso de Ciência da Computação da
Universidade do Extremo Sul Catarinense, UNESC,
com Linha de Pesquisa em Redes Sociais.

Criciúma, 27 de Novembro de 2012.

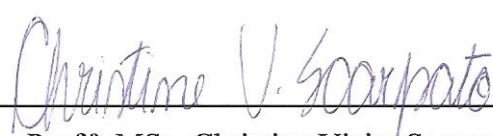
BANCA EXAMINADORA



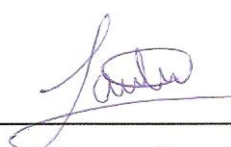
Prof.^a MSc . MSc. Leila Laís Gonçalves- UNESC – Orientador



Prof. MSc. Kristian Madeira - UNESC – Co-orientador



Prof.^a MSc. Christine Vieira Scarpato- UNESC



Prof. MSc. Luciano Antunes - UNESC

Dedico este trabalho aos amigos e familiares.

AGRADECIMENTOS

Primeiramente à minha família por ter me apoiado em todo esse trajeto.

Aos amigos por proporcionarem momentos de descontração e alegria nos momentos difíceis.

A todas as pessoas que conviveram comigo, ao longo desse tempo, tanto no meio acadêmico quanto no profissional.

Às amigas Camila Furlan e Juliana Silva por me nortear para a solução dos problemas que descobri ao longo do desenvolvimento do trabalho.

Aos meus orientadores por me guiarem nos caminhos corretos no desenvolvimento do trabalho e por conseguirem me suportar em toda trajetória.

“A educação é claramente o fator que irá conduzir melhorias na economia em longo prazo. No futuro, software e tecnologia irão permitir que as pessoas aprendam muito com seus colegas.”

Mark Zuckerberg

RESUMO

Com o avanço dos meios de comunicação, em especial da internet, a ambição de informações da sociedade humana tem tido um aumento considerável nos últimos anos. A compreensão do comportamento dos agrupamentos, suas conexões, estruturação, formas de ocorrência da propagação de informação são foco de estudos das Ciências Humanas. Com o avanço dos meios de comunicação, em especial da internet, esse comportamento humano vem sendo praticado em redes sociais virtuais, como o Facebook e o Twitter. O estudo da propagação das informações em redes sociais virtuais envolve questões sociológicas, matemáticas e computacionais como: teorias de Análise Estrutural das Redes Sociais, Teoria dos Grafos e Algoritmos de Inteligência Artificial. Este trabalho teve como objetivo avaliar a aplicação do algoritmo de Soma e Produto na identificação de pontos de propagação de informações em redes sociais virtuais. A partir da pesquisa bibliográfica, abordando as teorias e conceitos das áreas de atuação, foi desenvolvida uma aplicação, formada pela junção do algoritmo Soma e Produto e Grafo Fator, para calcular a propagação de informação em redes sociais virtuais. A validação da proposta foi realizada com a confrontação das características dos atores e seus laços sociais na estruturação da rede, oriundas das teorias da Sociologia, os resultados obtidos com o uso da aplicação na identificação de pontos de propagação de informação. Os resultados da pesquisa foram satisfatórios concluindo que os dados probabilísticos gerados pela aplicação podem dar suporte na identificação de pontos de propagação e servir de apoio à tomada de decisões na escolha de atores da rede que possibilitaram melhor intensidade e fluxo de mensagens na propagação.

Palavras-chave: Propagação de informação. Análise Estrutural das Redes Sociais. Redes Sociais Virtuais. Algoritmo Soma e Produto. Grafo Fator

ABSTRACT

With the advancement of the media, especially the Internet, the ambition of information of human society has experienced a considerable increase in recent years. The comprehension of the behavior in these groups, their connections, their structures, the way that the propagation occurs is focus of study by the human sciences. With the advance of the communications, specially the internet, this human behavior is occurring in virtual social networks, like Facebook or Twitter. The study of the propagation of the information in the virtual social networks, involves sociological, mathematical and computational questions like: structural analysis of the social networks, graph theory and Artificial Intelligence algorithms. This document had the objective of evaluate the application of the algorithm Sum Product for the identification of the propagation points of the information in virtual social networks. Starting from the bibliographic research, addressing theories and concepts of the specific areas, was developed an application, formed by the junction of the Sum Product and Graph Factor algorithms, to calculate the propagation of the information in virtual social networks. The evaluation of the proposition was made by the confrontation of the characteristics of the actors and his social bonds in the structure of the network, coming from the sociological theories, with the obtained results by the use of application in the identification of propagation points of the information. The results were satisfactory, concluding that the probabilistic data generated by the application can support the identification of the propagation points and serve as support for decision making in the choice of actors in the network that can provide better intensity and flow of the messages in the propagation.

Keywords: Information propagation, Structural analysis of the social networks, virtual social networks, Sum Product algorithm, Graph Factor.

LISTA DE ILUSTRAÇÕES

Figura 1 - Representação da hierarquia dos tipos de redes.....	17
Figura 2 - Caixa de entrada para atualização do status do Facebook.	19
Figura 3 - Caixa de entrada para atualização de status no Twitter.	19
Figura 4 - Representação simplificada do padrão dos rios e pontes de Königsberg.	22
Figura 5 - Representação de uma rede com TERTIUS GRAUDENS.	31
Figura 6 - Exemplo de um Grafo Fator.....	34
Figura 7 – Igualdade entre três nós.....	35
Figura 8 - Representação da função de igualdade.....	35
Figura 9 - Esquema de passagem de um nó-variável para um nó-função.	36
Figura 10 - Esquema de passagem de um nó-função para um nó-variável.	37
Figura 11 - Exemplo Grafo Fator.....	38
Figura 12 - Grafo Fator representando a fatoração.....	39
Figura 13 - Representação mensagens em um Grafo Fator.	39
Figura 14 – Díade.....	45
Figura 15 – Tríade.....	46
Figura 16 - Grafo 3.	46
Figura 17 - Grafo de definição de uma pseudo rede social.....	47
Figura 18 - Grafo de definição de uma pseudo rede social 2.	48
Figura 19 - Código Exemplo <i>Framework</i> Mallet.	50
Figura 20 - Tabela de registros do banco de dados.....	53
Figura 21 - Exemplo dos dados gerados.....	54
Figura 22 - Classe de representação da tabela do banco de dados para o cálculo na linguagem Java	54
Figura 23 - Gráfico de propagação de uma Díade iniciado no primeiro ator.	55
Figura 24 - Gráfico de propagação de uma Díade iniciado no segundo ator.	56
Figura 25 - Gráfico de propagação de uma Tríade iniciado no primeiro ator.	57
Figura 26 - Gráfico de propagação de uma Tríade iniciado no segundo ator.	58
Figura 27 - Gráfico de propagação de uma Tríade iniciado no terceiro ator.	59
Figura 28 - Gráfico de propagação do Grafo 3 iniciado no primeiro ator.....	60
Figura 29 - Gráfico de propagação do Grafo 3 iniciado no segundo ator.....	61
Figura 30 - Gráfico de propagação do Grafo 3 iniciado no terceiro ator.....	61
Figura 31 - Gráfico de propagação do Grafo 3 iniciado no quarto ator.	62

Figura 32 - Grafo de propagação do Ator 1 e seus vizinhos.	68
---	----

LISTA DE TABELAS

Tabela 1 - Matriz de probabilidade condicional.	49
Tabela 2 - Classes utilizadas do <i>framework</i> Mallet.	51
Tabela 3 - Tabela de valores médios de propagação do grafo representado pela figura 17.	63
Tabela 4 - Tabela de valores médios de propagação do grafo representado pela figura 17 com a ausência do ator 1.	64
Tabela 5- Tabela de valores médios de propagação do grafo representado pela figura 18.	65
Tabela 6 - Tabela de valores médios de propagação do grafo representado pela figura 18 com a ausência do ator 1.	66

LISTA DE ABREVIATURAS E SIGLAS

AERS	Análise d Redes Sociais
API	<i>Application Programing Interface</i>
BP	<i>Belief Propagation</i>
CMS	Comunicação Mediada pelo Computador
CPL	Licença Publica Comum
CRF	Conditional Random Fields
FB	Facebook
GGG	<i>Giant Global Graph</i>
GS	Grafos Sociais
RS	Redes Sociais
RSV	Redes Sociais Virtuais
SP	Soma e Produto
TAP	<i>Topical Affinity Propagation</i>
TG	Teoria dos Grafos
TW	Twitter
UNESC	Universidade do Extremo Sul Catarinense
WWW	<i>Word Wide Web</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 OBJETIVO GERAL	13
1.2 OBJETIVO ESPECÍFICO	14
1.3 JUSTIFICATIVA	14
1.4 ESTRUTURA DO TRABALHO	15
2 REDES SOCIAIS	17
2.1 REDES SOCIAIS VIRTUAIS	18
2.2 ORIGENS DAS REDES SOCIAIS.....	21
2.3 CONSTRUÇÕES DAS RELAÇÕES NAS REDES SOCIAIS.....	24
2.4 TEORIAS EXPLICATIVAS SOBRE PROPAGAÇÃO	28
3 ALGORITMO DE PROPAGAÇÃO	33
3.1 GRAFO FATOR	34
3.2 ALGORITMO SOMA E PRODUTO	35
4 TRABALHOS CORRELATOS	41
4.1 GRAFOS-FATORES E DECODIFICAÇÃO ITERATIVA: NOVAS APLICAÇÕES.....	41
4.2 REDES SOCIAIS VIRTUAIS DE INFORMAÇÃO SOBRE O AMOR	41
4.3 E-CRM E REDES SOCIAIS: UMA INTERAÇÃO POSSÍVEL	42
4.4 OS IMPACTOS DAS REDES SOCIAIS NA COMUNICAÇÃO ORGANIZACIONAL	42
4.5 SOCIAL INFLUENCE ANALYSIS IN LARGE-SCALE NETWORKS	43
4.6 SOCIODRAMA E O TESTE SOCIOMÉTRICO	43
5 PROPAGAÇÃO DE INFORMAÇÃO: AVALIAÇÃO DO ALGORITMO SP	45
5.1 METODOLOGIA.....	45
5.2 GRAFOS DE REPRESENTAÇÃO	45
5.3 ALGORITMO SOMA E PRODUTO	48
5.3.1 Framework Mallet	50
5.3.2 Algoritmo Junction Tree	51
5.4 RESULTADOS OBTIDOS	53
6 CONCLUSÃO	70
6.1 TRABALHOS FUTUROS	70
REFERÊNCIAS.....	72

1 INTRODUÇÃO

O hábito de compartilhar e propagar informações é inerente ao ser humano que desde a Antiguidade busca meios diferentes de realizá-lo. Na Grécia Antiga, grupos de filósofos realizavam fóruns, os homens se reuniam em praças públicas para discutir seus assuntos cotidianos. O comportamento dos agrupamentos, suas conexões, estruturação, formas de ocorrência da propagação de informação são foco de estudo das Ciências Humanas.

Com o avanço dos meios de comunicação, em especial da internet, esse comportamento humano é praticado em redes sociais virtuais, como o Facebook e o Twitter, possibilitando que seus usuários se agrupem, compartilhem e propaguem todo tipo de informações: de triviais a temáticas relevantes.

A propagação da informação em redes sociais virtuais é notória e observada na repercussão de vídeos virais. Apesar de se reconhecer a importância das redes sociais na difusão da informação, o comportamento da propagação a função dos atores da rede e seus agrupamentos na difusão ainda são pouco estudados. Como se formam esses agrupamentos, como se estruturam, como se conectam? Como ocorre a propagação de informação e os pontos de maior probabilidade de ocorrência? Há relação entre o tipo de conexão e a difusão da informação? A qualidade do laço social na conexão influencia no grau de propagação? Esses são alguns dos questionamentos ainda pouco abordados em pesquisas e estudos científicos. Porém, são pontos importantes que começam a ser discutidos por pesquisadores, como Recuero (2005), Cruz (2011). A compreensão desses processos auxilia a entender como as redes sociais podem propagar informações e, até mesmo, boatos, além também da possibilita de planejamento de ações para influenciar esses processos.

O estudo da propagação das informações envolve questões sociológicas, matemáticas e computacionais como: teorias de Análise Estrutural das Redes Sociais, Teoria dos Grafos e Algoritmos de Inteligência Artificial.

Este trabalho tem como tema principal o estudo da propagação das informações em redes sociais, levantando questões sociológicas e matemáticas, assim formando uma base de conhecimento apoio para o trabalho. A construção e a estruturação da rede social tem subsídio na Teoria dos Grafos (TG) utilizada para representação da rede e suas conexões. O entendimento do comportamento as características de atores, a conectividade, os laços sociais e a difusão da informação são realizados a partir das teorias da área de Análise Estrutural das Redes Sociais (AERS) abordando questões como comportamento, transmissão de

informações e relacionamentos na sociedade. A Teoria dos Grafos, na sociologia, foi um ponto forte para o avanço do estudo de Análise Estrutural aplicado em redes sociais. Segundo Recuero (2005, p.02) “um grafo é uma representação de um conjunto de nós conectados por arestas que, em conjunto, formam uma rede”.

O foco principal da investigação em AERS é dar subsídio para validação da influência da conectividade da rede na difusão de informações. Partindo do princípio que quando mais conectada é a rede, quanto mais visíveis são as possibilidades de interação entre seus atores, maiores as chances de que uma informação seja propagada.

Para o estudo e a identificação de pontos de propagação da informação em uma rede social foi proposta a aplicação do algoritmo de Soma e Produto (SP) e Grafo Fatores (GF). Os algoritmos de Soma e Produto, como outros algoritmos de passagem de mensagens, funcionam basicamente passando mensagens de um nó para outro calculando a probabilidade de propagação até as marginais dos grafos. Algoritmos aplicados em grafos normalmente são de grande complexidade computacional, aplicando certos algoritmos de estudo em grafos usando a gigante massa de dados adquirida em redes sociais, em alguns casos é totalmente inviável pelo alto custo computacional. O algoritmo SP é um, entre vários, que tem como apoio de funcionamento a transferência de mensagens entre nós em grafos. Destacando o algoritmo denominado “*Belief Propagation*” conhecido na comunidade de inteligência artificial, o qual tem o funcionamento parecido com o algoritmo SP.

Com a aplicação do algoritmo SP em grafos busca-se uma possível solução na identificação de pontos de propagação de informação em RS. Os grafos que representam redes sociais fictícias são formados por nós (pessoas e seus perfis nas RS), ligações entre os nós (relações entre os participantes da rede e o grau de ligação). A avaliação do uso do algoritmo para o objetivo proposto dar-se-á partir da confrontação dos resultados obtidos na sua aplicação e das proposições com relação à propagação das informações defendidas pelas teorias de Análise Estrutura de Redes Sociais.

1.1 OBJETIVO GERAL

Avaliar a aplicação do algoritmo de Soma e Produto para a identificação de pontos de propagação de informações em redes sociais.

1.2 OBJETIVO ESPECÍFICO

Para alcançar o objetivo geral deste trabalho destacam-se os seguintes objetivos específicos:

- a) descrever técnicas de análise de redes sociais;
- b) identificar estruturas e elementos compositivos das redes sociais;
- c) compreender o algoritmo de SP na identificação de pontos de propagação;
- d) aplicar o algoritmo de SP na identificação de pontos de propagação de informação em RS;
- e) confrontar os pontos de propagação de informação resultantes do algoritmo SP e das teorias das AERS.

1.3 JUSTIFICATIVA

A compreensão do comportamento da formação de grupos, estrutura e conexões, como é feita a transmissão e a probabilidade de mensagens entre atores e ainda, a associação entre os tipos de relacionamentos com a propagação auxiliam a entender as redes sociais do ponto de vista da propagação de informação.

Com a aplicação do algoritmo Soma e Produto junto à definição das redes por meio do Grafo Fator obtém-se uma possível correlação dos elementos compositivos de uma rede social com o levantamento dos processos para o estudo da passagem de mensagem.

Baseando-se em estudos na área da sociologia são levantadas teorias que explicam o comportamento em redes sociais reais, assim há referencial teórico que confrontado a junção do algoritmo com a representação específica de grafos, têm-se uma ferramenta cujo dados auxiliam a tomar ações ou medidas para alterar o nível de propagação da informação.

A utilização de grafos para representar o mapeamento das redes sociais não é aproveitada apenas na área da sociologia, na matemática, juntamente com a computação, a representação de grafos é usada como apoio para a resolução de vários problemas. A correlação do resultado obtido no estudo da propagação da informação nas redes sociais pode ser feita em outras áreas cujo representação do problema é visualizado pela passagem de mensagens entre pontos.

A escolha das redes sociais Facebook e Twitter é comprovada por dados retirados do site Alexa (2012). Segundo o site, as redes apresentam o privilégio de ocupar a segunda e oitava posição no *ranking* de sites mais acessados respectivamente, tornando-se os associados que mais acessados na internet. A correlação dos sites é observada nas seguintes perguntas propostas pelo site: “*Quais os sites que os usuários visitam imediatamente antes do facebook.com*”¹ e “*Onde é que os visitantes vão após deixar facebook.com?*”² onde o Twitter obteve o quarto lugar em ambas as questões com 1,68 e 1,87 respectivamente, perdendo somente para os gigantes Google e Yahoo. Já no Twitter, o Facebook obteve o primeiro lugar em ambos os questionamentos com 12,89 e 12,58 por cento de acessos anteriores e posteriores à rede.

Trabalhos de autores como Recuero e Zago (2010), defendem que o processamento de informações do Twitter é carente de estudos, e levantam a seguinte indagação: como estudar a difusão da informação nas redes sociais? Outros estudos como os feitos por Kleinberg (2007) sobre cascatas de informação, e sobre a estrutura das redes e sua influência na propagação das informações por Adar e Adamic, 2005, são exemplos de como esta área tem muitas questões a serem respondidas a fim de que os acadêmicos tenham esclarecimento sobre o assunto.

1.4 ESTRUTURA DO TRABALHO

O presente trabalho de conclusão de curso está estruturado em seis capítulos descritos a seguir. O capítulo 1 apresenta a problemática do trabalho, seus objetivos, o porquê da escolha do tema e a estrutura dos capítulos.

No capítulo 2 são discutidos conceito, origem e evolução de Rede Social e sua virtualização. São levantadas também as principais características e teorias na área de sociologia que embasam a constituição de redes sociais e explicam o comportamento das pessoas nas mesmas.

O algoritmo Soma e Produto e a representação de grafos denominada Grafo Fator são abordados no capítulo 3. E ainda o funcionamento da representação de um Grafo Fator é com as definições e para uma aplicação correta. Em seguida é abordado o algoritmo por meio de definições matemáticas, área de sua aplicação junto com suas regras de passagens de

¹ Tradução de “*Which sites did users visit immediately preceding facebook.com?*”

² Tradução de “*Where do visitors go after leaving facebook.com?*”

mensagens.

No capítulo 4 são apresentados os trabalhos correlatos que abordam, em linhas gerais ou específicas, os objetos de estudo deste trabalho: redes sociais, algoritmo Soma e Produto, propagação de informação. O objetivo é obter maior entendimento sobre o assunto e como e propagado no ambiente acadêmico, em âmbito regional, nacional e mundial.

A aplicação prática da fundamentação teórica encontra-se no capítulo 5 dividido em quatro etapas: metodologia, construção do grafo representando redes sociais, aplicação do algoritmo SP e resultados obtidos.

Por último, no capítulo 6, encontram-se as conclusões do trabalho e as recomendações para trabalhos futuros.

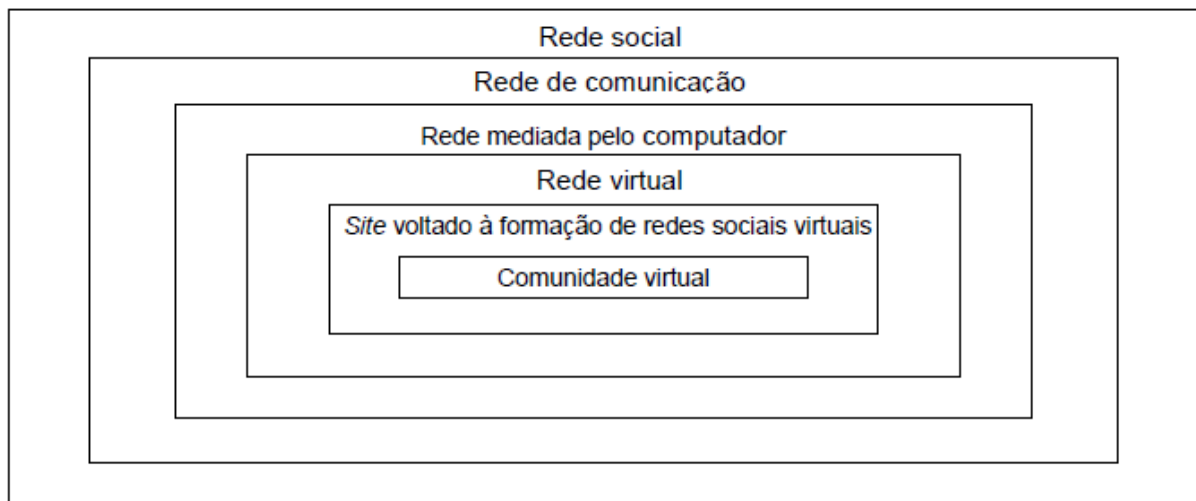
2 REDES SOCIAIS

A tecnologia está presente e tem um papel fundamental nas mudanças da sociedade, percebendo assim que as redes sociais virtuais provocaram uma profunda alteração na forma de relacionamento. Recuero (2009) define Redes Sociais Virtuais (RSV) “como agrupamentos complexos instituídos por interações sociais apoiadas em tecnologias digitais de comunicação”, que disponibilizam um vasto meio de fornecimento de informações do ciberespaço. Esta definição ajuda a entender “a criação das estruturas sociais; suas dinâmicas, tais como a criação de capital social e sua manutenção, a emergência da cooperação e da competição; as funções das estruturas e, mesmo, as diferenças entre os variados grupos e seu impacto nos indivíduos.”.

O estudo de RS apoia-se na necessidade da construção empírica, qualitativa e quantitativa a partir de observações sistemáticas procurando assim verificar padrões expressos tanto nas sociedades virtuais como reais, explorando elementos individuais como em grupos (RECUERO, 2009).

Park (2003) apud Cruz (2011) define os tipos de redes em: rede social, rede de comunicação, rede mediada pelo computador e rede virtual, localizando uma dentro da outra como demonstra a Figura 1.

Figura 1 - Representação da hierarquia dos tipos de redes.



Fonte: CRUZ (2011, p. 44).

Visto que a definição de redes sociais nas diversas áreas de estudo abordadas neste trabalho apresenta conceitos diferentes, bordar-se como Redes Sociais Virtuais (RSV), aquelas que se referenciam a redes sociais na internet, e como Rede Sociais (RS) as redes de

pessoas aplicadas ao estudo na área das ciências humanas.

Neste trabalho abordar-se-á as RSV Facebook e Twitter, por conterem diferentes características, as quais geram um conjunto que forma um sistema interligado completo; proporcionando assim um amplo estudo sobre a propagação da informação na internet, percebendo a relação mutua entre as RSV.

2.1 REDES SOCIAIS VIRTUAIS

O advento da Internet trouxe diversas mudanças para a sociedade. A possibilidade de expressão e socialização através das ferramentas de comunicação mediada por dispositivos eletrônicos proporcionaram aos atores construir, interagir e se comunicarem com outros atores, deixando rastros que permitem o reconhecimento dos padrões de suas conexões e a visualização de seus dados. É neste âmbito que a rede como metáfora estrutural para a compreensão dos grupos expressos na Internet é utilizada através da perspectiva de rede social (RECUERO, 2009).

As RSVs passaram de novidade a fenômeno global nos últimos anos. No primeiro bimestre de 2010, o Facebook ultrapassou a Google em número de visitas às suas páginas. De uma maneira indefensável, este dado indica que a internet tornou-se um meio onde a convivência social, com a ajuda de ferramentas, é utilizada comparada ao uso de ferramentas de pesquisa, indicando em uma escala considerável que *sites* de RSV estão satisfazendo, alguns anseios humanos muito básicos, mudando bruscamente a forma como as pessoas veem a vida social dentro e fora da internet (RUSSEL, 2011).

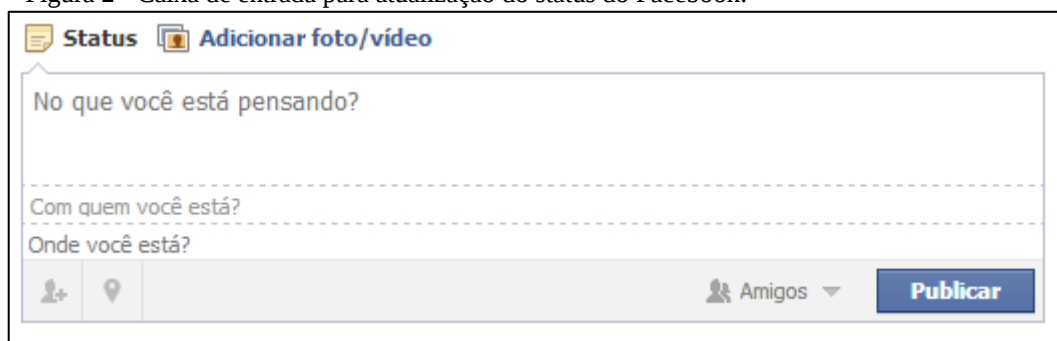
O termo Rede Social Virtual pode ser tratado igualmente como *sites* de redes sociais, termo esse entendido por Boyad e Elisson (2007) apud Recuero (2009) como: “[...] sistemas que permitem i) a construção de uma persona através de um perfil ou página pessoal; ii) a interação através de comentários; e iii) a exposição pública da rede social de cada ator”.

O Facebook, lançado em 2004, foi um sistema criado pelo americano Mark Zuckerberg enquanto este era aluno de Harvard. A ideia era focar em alunos que estavam saindo do secundário (*High School*, nos Estados Unidos) e aqueles que estavam entrando na universidade. O foco inicial do FB era criar uma rede de contatos em um momento crucial da vida de um jovem universitário: o momento em que este sai da escola e vai para a universidade. Isso nos Estados Unidos o que geralmente representa uma mudança de cidade e um aspecto novo das suas relações sociais.

Já o Twitter, foi lançado após dois anos do lançamento do Facebook como um projeto da empresa Odeo, fundado por Jack Dorsey, Biz Stone e Evan Williams e é definido por muitas pessoas como “*microblogging*” pelo fato de permitir atualizações de apenas 140 caracteres. (RECUERO, 2009).

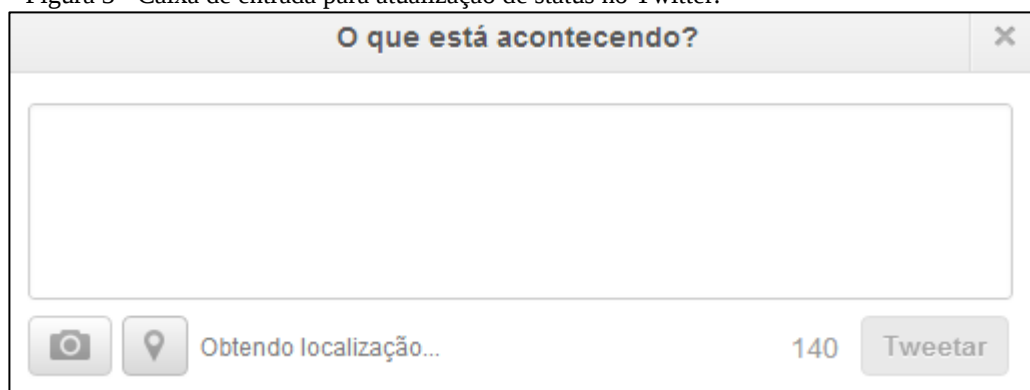
A natureza da interação do usuário com os sistemas apoia-se em perguntas, guiando assim os usuários a inserção de informações sobre seu cotidiano. O Facebook proporciona uma maior interação, por meio da inserção de comentários baseados, por exemplo, nas perguntas “*No que você está pensando?*”, “*Com quem você estava?*” e “*Onde você estava?*” demonstrada na Figura 2. Enquanto no Twitter, por sua forma de interação simplificada, a interação baseia-se na pergunta “*O que está acontecendo?*” e na localização, onde sendo que os comentários postados são de no máximo 140 caracteres como pode ser observado na Figura 3.

Figura 2 - Caixa de entrada para atualização do status do Facebook.

A screenshot of the Facebook status update interface. At the top, there are two tabs: 'Status' (selected) and 'Adicionar foto/vídeo'. Below the tabs is a large text input area with the placeholder text 'No que você está pensando?'. Below this, there are two smaller input areas with dashed lines and placeholder text: 'Com quem você está?' and 'Onde você está?'. At the bottom, there are icons for adding friends and location, a dropdown menu for 'Amigos', and a blue 'Publicar' button.

Fonte: Do autor.

Figura 3 - Caixa de entrada para atualização de status no Twitter.

A screenshot of the Twitter status update interface. At the top, there is a header bar with the text 'O que está acontecendo?' and a close button (X). Below the header is a large text input area. At the bottom, there are icons for adding photos and location, a text label 'Obtendo localização...', a character count '140', and a 'Tweeter' button.

Fonte: Do autor.

As duas redes foram projetadas inicialmente para proporcionar uma grande interação tanto com os usuários quanto com outros sites; as vastas APIs disponibilizadas por elas, permitem extrair inúmeros dados sobre os usuários e também aos seus dados e a

interação de outros, facilitando a inserção de conteúdo na rede social.

Leblanc (2011) define que atualmente as RSV contêm três modelos de relacionamentos (seguidor, conexão e grupo), com propriedades individuais de alcance e complexidade diferentes, as quais estão presentes nas RSV em estudo.

No modelo de seguidor (*Follower model*), o utilizador sempre interage com a maioria ou todos os seus seguidores. Quando você postar alguma coisa sobre uma rede desenvolvida em torno de um modelo de seguidor, seu alcance potencial é o número de seguidores que você tem. As relações sociais formadas dentro desse tipo de modelo tendem a ser limitadas e superficiais. O foco aqui é mais no "o quê" (mensagens e atualizações que são postadas) e menos no "quem" (pessoa a publicá-las). Implementadores do modelo tendem a limitar a quantidade de informação que pode ser postada em uma única mensagem ou atualização e pode fornecer apenas uma quantidade muito limitada de informação social sobre o indivíduo. Isso significa que você não vai ter a mesma profundidade das relações sociais que você poderia a partir de um modelo de relacionamento mais centrado.

O modelo de conexão (*Connection model*) centra-se na conexão das pessoas, interações com outros indivíduos singulares ou objetos. Para este fim, as RSV geralmente permitem ao usuário o perfil altamente detalhado e rico e permite amarrá-los a outros grupos, organizações e movimentos. Este modelo concentra-se na partilha de experiências sociais e não na rápida propagação da informação para grandes grupos. Os usuários desse modelo tendem a compartilhar muitos dos eventos mais importantes de suas vidas, sendo boas ou más, com os seus amigos através de texto, fotos e vídeos.

O modelo de grupo (*Group model*) é baseado nas interações e experiências compartilhadas de um grupo de indivíduos com interesses em comum, ou pessoas com interesses semelhantes, fundos, ou situações. As pessoas podem pensar nisso como sair com um grupo de amigos, que possuem coisas em comum que querem passar algum tempo juntos, compartilhando experiências e aprendendo com o outro. Pode haver um ou muitos grupos em sua rede e alguns podem sobrepor-se quando há pessoas que ligam as lacunas entre as várias áreas de interesse.

Se o seu serviço inteiro é construído em um modelo desse tipo, o agrupamento vem naturalmente. O problema reside na segmentação de um Grafo Social (GS) relevante, quando a pessoa deseja oferecer um serviço que não está apenas centrado nas relações de grupo, mas também sobre o indivíduo, seu perfil, e as interações estreitas com outros usuários.

O Twitter segue o modelo de seguidor, os usuários postam mensagens ("*tweets*") que chegam a seus seguidores, que por sua vez podem "*retuitar*" os *posts* (na maioria dos casos, dando a atribuição ao usuário original propriamente dito) para seus seguidores, aumentando assim o alcance global. Embora o Twitter forneça mecanismos para envio direto de mensagens entre os usuários ou menciona entre os grupos de seguidores, ele é construído em torno da ação de um usuário postando atualizações a todos os seus seguidores. Na sua essência, é um mecanismo para a rápida disseminação de informações entre muitos usuários.

Facebook é um exemplo ideal para o modelo de conexão. O corpo do *site* é construído através do perfil social dos usuários, que permite aos indivíduos alcançar e compartilhar postagens com o outro e interagir em um nível muito pessoal por meio de construções, como mensagens diretas, comentários do status, grupos, páginas, aplicativos e mais, levando a uma interação mais pessoal. Mesmo que o principal mecanismo por trás do modelo de relacionamento que o Facebook oferece como plataforma, seja o de conexão, a empresa integrou várias ferramentas, tais como grupos e páginas, que permitem aos usuários se organizarem em grupos.

O estudo das duas RSV proporciona assim, um estudo completo, por abranger todos os tipos de relacionamentos definidos por Leblanc (2011) como modelo de seguidor pelo Twitter e modelo de conexão e grupo com o Facebook.

2.2 ORIGENS DAS REDES SOCIAIS

A Teoria dos Grafos (TG) proporciona um entendimento geral das várias áreas dentro da computação. Destacam-se as áreas das redes de computadores, a qual foi o berço da atual rede mundial de computadores, conhecida como *Word Wide Web* (WWW), criada por Tim Berners-Lee qual mais tarde sugeriu a mudança do termo WWW para Grafo Global Gigante (GGG), tradução de "*Giant Global Graph*", explicando que os termos "web" e "grafo" podem ser livremente intercambiados nos contextos da definição de uma topologia para a internet.

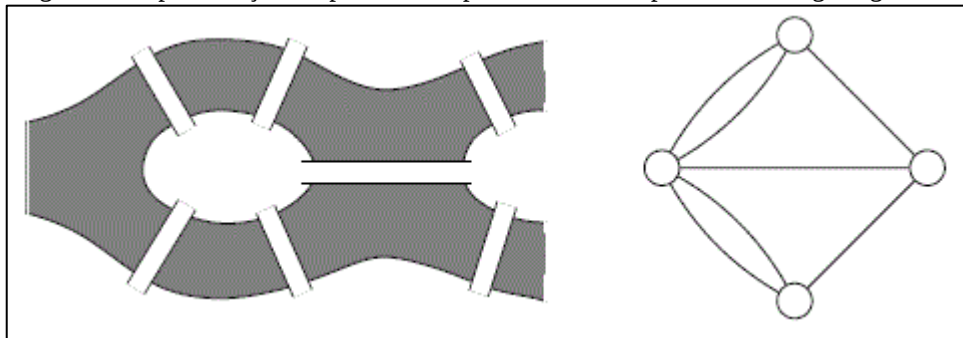
Segundo Recuero (2009), um grafo é a representação de uma rede constituída de nós e arestas que conectam os nós. A teoria dos grafos é uma parte da matemática aplicada que se propõem estudar os diferentes tipos de grafos. Essa representação de rede pode ser utilizada como modelo metafórico para diversos tipos de situação, amplamente usadas para mapear uma rede social, definida como:

“um conjunto de dois elementos: atores (pessoas, instituições, ou grupos; os nós da rede) e suas conexões (interações ou laços sociais) [...] Uma rede, assim é uma metáfora para observar os padrões de conexão de um grupo social, a partir das conexões estabelecidas entre os diversos atores. A abordagem de rede tem, assim, seu foco na estrutura social, onde não é possível isolar os atores sociais e nem suas conexões” (RECUERO, 2009).

O estudo de RS teve seu início graças aos primeiros teoremas de redes no estudo da matemática não-linear em que surgiram as primeiras ideias sobre a teoria dos grafos. Posteriormente, foi aplicada em estudos na área de física que ajudaram no surgimento de importantes assuntos como a física quântica e a teoria do caos; contribuindo também com vários estudos para entendimento comportamental e estrutural entre pessoas, aplicados nas chamadas ciências sociais. Visto que o estudo sobre esse assunto se tornou fundamental em diversas áreas, a visão passou a ser amplamente multidisciplinar, conhecido atualmente como Análise Estrutural das Redes Sociais.

Ëuler foi um matemático que começou o estudo geral de redes. Por meio de sua pesquisa a respeito de um problema chamado Enigma das Pontes de Königsberg, publicado em um artigo no ano de 1736, que tinha como fundamento um conto folclórico que desafiava a travessia da cidade, cruzando uma única vez todas as pontes, onde a cidade era cortada pelo rio Pregolva que tinha sete pontes os quais faziam a ligação entre diversos pontos da cidade. Ele provou que a travessia era impossível, criando assim a teoria dos grafos, sendo representadas por áreas e nós, cujo pontes eram as arestas e os nós eram as partes terrestres (BARABASI; BUCHANAM; WATTS, 2006; RECUERO, 2008).

Figura 4 - Representação simplificada do padrão dos rios e pontes de Königsberg.



Fonte: Barabási, Buchanan e Watts (2006, p. 3).

Segundo Recuero (2008), a teoria de redes foi usada profundamente no entendimento da interação nos níveis subatômicos o que ajudou Einstein a comprovar que a energia e a matéria se comportam do mesmo modo, proporcionando um grande avanço na

física quântica. Além de ajudar outros físicos e matemáticos como Heisenberg, Ashby e Winer.

O termo redes complexas refere-se a um grafo que apresenta uma estrutura topográfica não trivial, composto por um conjunto de vértices (nós) que são interligados por meio de arestas (NEWMAN; BARABÁSI; WATTS, 2006). Esse termo foi usado para o entendimento da interação das massas de ar pelo meteorologista Edward Lorenz que originou entre outras teorias, a teoria do caos, com grande impacto na matemática não-linear nos anos 60, mas também em outras áreas, como na biologia, educação e comunicação com a abordagem da complexidade relacional (RECUERO, 2009).

Em ciências humanas, a explicação dos fenômenos sociais se ocorre por meio do estudo das relações entre os atores, sendo individual ou coletivamente em redes públicas. A análise casual tenta estabelecer as determinantes pontuando a importância de cada uma, já a análise descritiva tenta mostrar a divisão entre as etapas da evolução. Sendo que a análise feita em ações costuma ser mais racional por ter um foco em políticas públicas, levando diretamente nos interesses dos atores (LEMIEUX; OUIMET, 2004).

A chamada Análise Estrutural das Redes Sociais é normalmente usada para explicar a forma das relações entre os atores sociais, que tentam apresentar as ligações entre os atores no âmbito do estruturalismo, linguístico e antropológico. Esses estudos podem ser divididos basicamente em dois níveis o primeiro é aplicado na análise sobre uma visão de regularidade e organização dando uma explicação descritiva e informativa sobre o processo, já a de segundo nível trata de uma análise estrutural, com um resultado baseado em uma análise totalmente empírica (LEMIEUX; OUIMET, 2004).

Na aplicação de técnicas para a compreensão dos relacionamentos de um grupo de pessoas interconectadas surge o desafio da limitação, um dos grandes problemas para a abordagem de estudo, que muitas vezes influencia nas propriedades dos relacionamentos, e leva a um impacto grande na forma de se observar a rede. A limitação é feita por meio de um corte entre os relacionamentos, definido em TG como Sub-Grafo cuja definição é um grafo obtido pela supressão de vértices de um grafo original ou, o que a isso equivale, separando-se de um acompanhado de todas as ligações entre seus vértices, excluindo-se as demais, que deixam de existir em ausência de ou menos um dos vértices que a compõem (BOAVENTURA, 2003).

2.3 CONSTRUÇÕES DAS RELAÇÕES NAS REDES SOCIAIS

Em ciências humanas, as análises correlacionais são feitas através do comportamento e propriedades dos atores em um grafo definido por Aguiar (2007, apud Cruz (2011)) como:

- a) dados relacionais: são os que fazem parte dos laços sociais;
- b) dados de atributos: são os atributos, qualidade, características, definidos como capital social, com uma visão individual de cada ator.

Para a compreensão de uma rede tem é preciso definir os elementos compositivos para ter início a caracterização de suas propriedades. Estes conceitos fundamentais de uma RS são definidos por Wasserman e Faust (1994, apud CRUZ, 2011) por meio de sete (7) elementos compositivos:

- a) ator: é um distinto indivíduo, corporação ou unidade social coletiva;
- b) laço social: é o que estabelece a ligação entre um par de atores;
- c) díade: consiste em um par de atores e no possível laço entre eles, assim, analisa se os laços são recíprocos ou se ocorrem laços de múltiplas relações;
- d) tríade: é o grupo de três atores e os possíveis laços entre eles;
- e) sub-grupo: é um sub-conjunto de atores dentro da rede e todos os laços entre eles;
- f) grupo: é um conjunto finito de atores que conceitualmente, teoricamente ou por razões empíricas são definidos como finitos em determinada avaliação da rede;
- g) relação: é a coleção de laços de um específico tipo entre membros de um grupo.

Recuero (2009) afirma também que os atores em uma RS são os elementos representados pelos nós, ou seja, as pessoas envolvidas na rede em estudo e que fazem parte do sistema. Eles formam e moldam as estruturas sociais por meio das interações obtidas pelos laços sociais. Em uma RSV a autora defende que os atores não necessariamente são representações reais de uma pessoa humana. Devido a distância entre os envolvidos nas interações mediadas por computador, a identificação torna-se discernível. Assim trabalha-se o

ator em um RSV como uma identificação no ciberespaço, definidas como conta e *profile*³ nas diversas RSV. Recuero (2009) também defende que o laço social não faz parte da rede, mas sim da conexão formada por “interação social entre os atores”.

Assim pode-se compreender definições que ajudam a levantar a importância de um ator em uma rede, proporcionando o entendimento inicial das relações em uma RSV que foram citadas nesse capítulo como: orientação, densidade, conexidade e localização.

Os laços sociais definem que as conexões entre os atores são formadas por meio das interações na sociedade. As conexões formam o principal foco de estudo em RS, por meio delas explica-se todo o estudo da estrutura de uma RS.

Uma RS pode ter uma representação em forma de grafos. Alguns conceitos já aplicados a grafos foram amplamente abordados no estudo comportamental e estrutural das RS, tais como: orientação, densidade, conexidade e posição que serão discutidos a seguir.

A orientação ou não das relações entre os nós tem um papel muito importante para a compreensão das relações, formando um entendimento paralelo entre a sociologia e a TG. Nas relações orientadas, denominadas na TG como arcos ou arestas, permite ter mais de um canal de comunicação entre os atores, podendo ser bidirecional ou unidirecional. Já o comportamento da transmissão das mensagens em relações não orientadas definidas como arcos ou arestas em TG, leva à irrelevância a direção da comunicação no relacionamento em comum entre os atores. A orientação ou não das mensagens em grafos sociais está ligada diretamente ao estudo da importância de um ator na rede em estudo (LEMIEUX; OUIMET 2004, BOAVENTURA 2003).

Segundo Lemieux e Ouimet (2004, p. 20) densidade em um grafo é “a razão entre as relações existentes e as relações possíveis”, assim representando a proporção entre o número de arestas e vértices. O número de relações varia em função da orientação ou não das ligações. Em grafos orientados o número de possíveis ligações se multiplica por dois (2). Em um grafo representando por uma função f , em que o número máximo de autores é denotada por $f(f - 1)$, a densidade dar-se-á pela notação D , definida pela quantidade de ligações L da rede, dividida pelo número máximo de ligações. A equação (1) representa a fórmula da densidade em grafos orientados e pela equação (2) em grafos não orientados.

³ Consistem de informações pessoais, como nome, aniversário, sites, interesses, fotos, localização, e uma série de outros detalhes que é fornecido pelo usuário como identificação para outros atores na rede (Leblanc, 2011).

$$D = \frac{2L}{f(f-1)} \quad (1)$$

$$D = \frac{L}{f(f-1)} \quad (2)$$

A análise de passagens de um ponto a outro em redes sociais é discutido em TG como caminho, designando uma sequência de relações orientadas. Já em sociologia, isto é chamada de cadeia. Tendo em vista um nó ou ator para o estudo em sociologia, o termo utilizado em grafos com raiz é dito como fonte, e o final do caminho dito em grafos como nó folha, em sociologia entendesse como alvo e todos os atores que estão fazendo a ligação são chamados de intermediários. O ator dominante é que iniciou o processo nos casos de grafos que tenham orientação; ao contrário de redes que não possuem orientação, em que todos os atores serão considerados dominantes (LEMIEUX; OUIMET, 2004).

A noção de conexidade está relacionada à possibilidade da passagem de um vértice a outro em um grafo através das ligações existentes, traduzindo o “estado de ligação” de um grafo, que se comporta diferentemente quando determinado grafo é orientado ou não. A ideia de passagem de mensagens está ligada diretamente à conexidade, a qual é usada em grafos orientados, já em grafos não orientados é definida pela simples existência de uma conexão entre dois nós. Um grafo conexo é definido pela existência de pelo menos um par de vértices unidos através de uma cadeia; em um grafo não conexo pode se obter pelo menos dois sub-grafos conexos (BOAVENTURA, 2003; LEMIEUX, OUIMET, 2004). O nível de conexidade em uma rede têm definições tanto no âmbito social quanto matemático, tais como:

- a) não-conexo: quando um ou vários nós (atores) são isolados, não tendo a existência de nenhum ator dominante, formando uma estrutura desintegrada das relações;
- b) quase-fortemente-conexo: a existência de somente um ator dominante, não existindo uma uniconexão em ambos os sentidos em um par de atores, representando uma estrutura hierárquica, total ou parcial, tendo pelo menos dois atores sem conexão diretas;
- c) semi-fortemente conexo: a existência de somente um ator dominante que tenha uma conexão em pelo menos um sentido, possibilitando uma

estruturação estratificada tornando possível a distinção de no mínimo dois sub-grafos quase-fortemente-conexo ;

- d) fortemente conexo: quando todos os atores são dominantes .

A posição de um ator em um grafo têm duas visões distintas, sendo divididas pela orientação ou não das relações. Em grafos orientados a distinção das posições dos atores é feita pelo domínio no contexto aplicado, sendo dividido em seis partes nos quatro tipos de grau de conexão entre os atores, definidas como:

- a) dominante: quando é o emissor de uma conexão com cada um dos outros atores em uma determinada rede ;
- b) dominada: quando num conjunto de atores existe pelo menos um ator dominante e não é o emissor de conexões com outros atores;
- c) semidominante: quando apesar de não ocupar uma posição dominante, é o emissor de uma conexão com um ou vários atores ;
- d) subdominante: quando na ausência de um ator dominante e apesar de não se encontrar numa posição semidominada é o emissor numa conexão com um ou vários atores;
- e) subdominada: quando um conjunto de atores sem ator na posição dominante, não é o emissor de nenhuma conexão com outro ator e é o destinatário de uma conexão proveniente de pelo menos outro ator ;
- f) isolada: quando um conjunto de atores, não são nem um emissor nem um destinatário de qualquer conexão com outro ator (LEMIEUX; OUIMET, 2004).

Em grafos não orientados, a distinção das posições ocupadas dos atores se dá por meio da noção de centralidade. Métricas essas definidas por Freeman (1977) como:

- a) centralidade de grau (*degree centrality*): reflete atividade relacional direta do ator. Mede o número de conexões diretas de cada ator em um grafo. De acordo com essa medida, o ator que ocupar a posição mais central em um grafo é aquele que possui o maior número de conexões diretas;
- b) centralidade de proximidade (*closeness centrality*): é uma medida que assenta na distância geodésica, ou seja, no comprimento do caminho mais curto que liga dois atores. No objetivo de medir a capacidade de autonomia ou de independência dos atores. Trata-se deste modo, de medir o afastamento, ao invés da proximidade. A hipótese consiste em afirmar que

quanto mais um ator se encontra afastado dos outros atores, mais autônomo será no que se refere as suas escolhas de ação. O grau de afastamento de um ator é medido procedendo à soma das distâncias geodésicas que o ligam aos outros atores da rede;

- c) centralidade de intermediaridade (*betweenness*): é uma medida de importância da posição de intermediariedade ocupada pelos atores de um grafo. Define a medida de um ator de um grafo de modo a assegurar um papel de coordenação e de controle. Consiste em afirmar que um ator se encontra numa posição intermediária, ou seja, encontra-se numa situação em que os atores têm de passar por ele para chegar a outros atores.

2.4 TEORIAS EXPLICATIVAS SOBRE PROPAGAÇÃO

As análises correlacionadas comentadas anteriormente servem para descrever relações entre os atores sociais. Teorias explicativas que usam esses dados tentam explicar como os atores se comportam na sociedade. Também podem ser aplicadas em GS, formando um conjunto de proposições de diferentes níveis de generalidade.

O ciberespaço e as ferramentas de comunicação possuem particularidades a respeito dos processos de interação. Há uma série de fatores diferenciais. O primeiro deles é que os atores não se dão imediatamente a conhecer. Não há pistas da linguagem não verbal e da interpretação do contexto da interação. É tudo construído pela mediação do computador. O segundo fator relevante é a influência das possibilidades de comunicação das ferramentas utilizadas pelos atores. Há multiplicidade de ferramentas que suportam essa interação e o fato de permitirem que a interação permaneça mesmo depois do ator estar desconectado do ciberespaço. Esse fato permite, por exemplo, o aparecimento de interações assíncronas (RECUERO, 2009).

O Capital Social é um fator importante para o entendimento do comportamento dos atores em um grafo, no qual Cruz (2011) define como: “interesses partilhados pelos autores da rede social [...] normas, valores, instituições e relacionamento compartilhados que permitem a cooperação dentro ou entre os diferentes grupos sociais [...]”.

Segundo Recuero (2009) o conceito, entretanto é variado, e não há uma concordância, entre os estudiosos, sobre qual vertente deve ser seguida. O que se concorda é que o conceito refere-se a um valor constituído a partir das interações entre os atores sociais.

Elas permitem que as redes sejam formadas sem dispêndio de energia para a manutenção dos laços, ou seja, quando um ator social cria um relacionamento virtual, ele precisa ter uma interação no sistema, sendo através de uma interação reativa no caso do FB (aceitar ou não a “amizade”), ou ainda começar a seguir uma pessoa no caso do TW.

No entanto, essas redes, ao contrário das RS ditas off-line abordadas na sociologia, não necessitam de interação para serem mantidas segundo Barabási e Watts (2006), as medidas sociométricas definem uma percepção em cima de uma estrutura estática, enquanto a teoria das redes compreende a dinâmica tratando as redes em uma constante evolução, conforme Tomaél, Alcará e Chiara (2005) ressaltam tais redes têm características como: não linear, descentralizadas, flexível, dinâmica, sem limites definidos e auto organizacional.

Os laços sociais e o tipo de rede social formada através das ferramentas sociais observadas na Internet podem ser auxiliadas pelo capital social. É preciso, assim, estudar não apenas a existência das conexões entre atores nas redes sociais mediadas pelo computador, mas, igualmente, estudar o conteúdo dessas conexões, transformando os dados em informações através do estudo de suas interações, possibilitando a mineração de dados para a extração de padrões de comportamento nas conexões (RECUERO, 2009).

Pela natureza dos sistemas de RSV abordados, uma vez criado o relacionamento a mesma será mantida independentemente da existência de interação entre os agentes. Trata-se, portanto, de uma conexão que não perde força com o tempo, o que, em se tratando de relações sociais, demonstra certa artificialidade nas estruturas sociais apresentadas (RECUERO, 2009).

Na teoria dos laços fortes de Granovetter (1973), são definidos dois graus de ligações, os fortes ou estreitos que unem os parentes, amigos próximos e cônjuges, e os laços fracos que unem parentes distantes ou apenas conhecidos. Essas diferenças se aplicam diretamente a configuração em que a rede é formada. Em relações fortes, assumem normalmente uma forma de cliques⁴ com densidade um (total).

As redes de laços fortes têm, por conseguinte, muito mais tendência para se fecharem sobre si mesmas, do que as redes de laços fracos, as quais tende preferencialmente a abrir-se para o exterior. É o que acontece com a circulação da informação entre os parentes, por um lado, e os conhecidos, por outro. Os parentes, que se veem com frequência, transmitir-

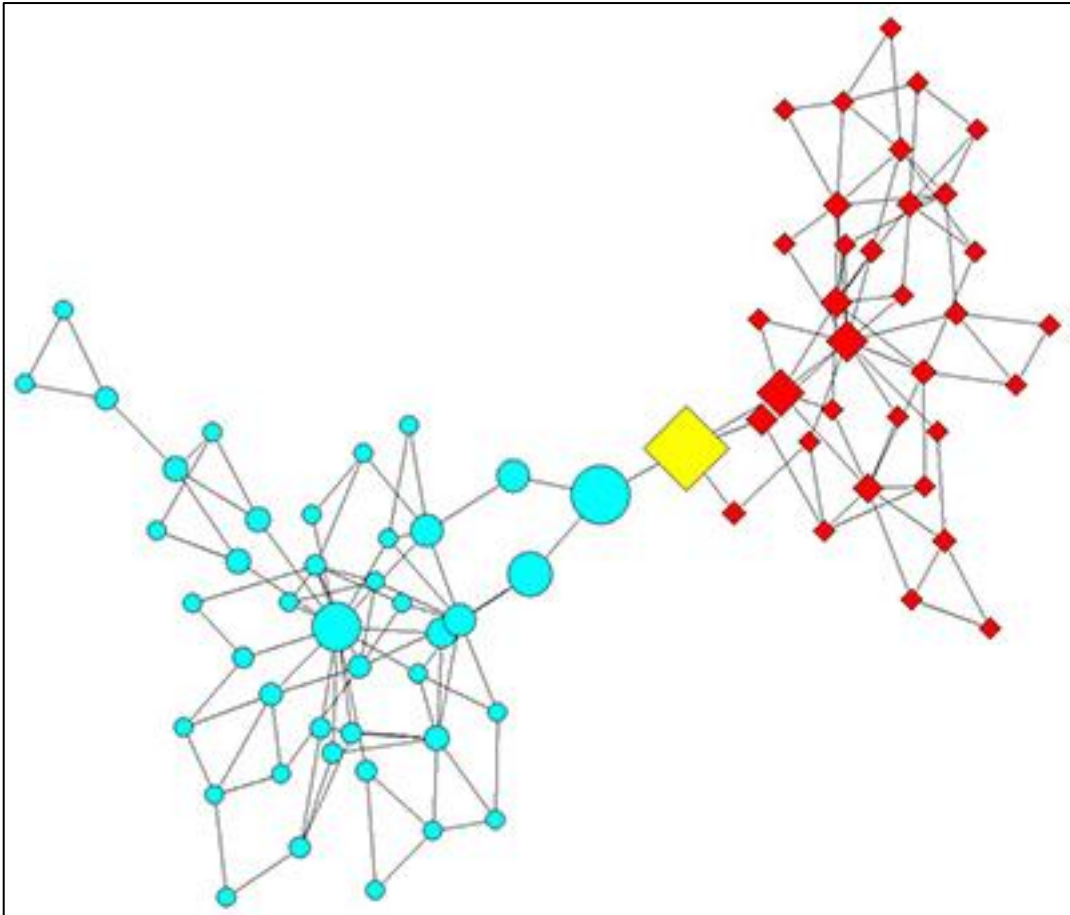
⁴ Na área da matemática da teoria dos grafos, uma clique em um grafo não-orientado é um subconjunto de seus vértices tais que cada dois vértices do subconjunto são conectados por uma aresta. Uma clique em um grafo G é um sub-grafo de G que é completo (BOAVENTURA, 2003).

se-ão menos informação nova do que os conhecidos, que se encontram ocasionalmente (LEMIEUX; OUIMET, 2004).

Na teoria de Burt (1992), buracos estruturais têm uma proposição geral na qual um ator que se beneficia com o fato de estabelecer contatos não tem qualquer conexão entre si, denominada “*TERTIUS GAUDENS*”, encontrando-se numa posição vantajosa, acima dos buracos estruturais. Existe uma distinção entre os buracos estruturais no interior do ambiente interno de um ator e aqueles que existem entre atores no seu ambiente externo. Nas situações de competição ou de conflito, os buracos estruturais no ambiente externo são vantajosos para um ator que enfrenta opositores divididos entre si. Já em um ambiente interno essa afirmativa é errônea uma vez que representa falhas na cooperação que internamente poderão ser exploradas pelos opositores.

Como refere Burt (1992) isso está ligado diretamente à teoria dos laços fortes e laços fracos, cuja existência de buracos estruturais somente ocorre em laços fracos onde a densidade do grafo é baixa. Isso pode ser claramente entendido através da Figura 5, em que se tem dois grupos distintos, o vermelho e o azul, sendo ligado através de um laço fraco com o ator determinado “*TERTIUS GRAUDENS*”, representado pela cor amarela.

Figura 5 - Representação de uma rede com TERTIUS GRAUDENS.



Fonte: DURANT-LAW (2007).

A teoria da grupabilidade discutida por Lemieux e Ouimet (2004) que teve origem nos trabalhos de um psicólogo, Heider 1946 e de um psicólogo social, Festinger 1957, mostrou que embora dois pares de elementos de uma tríade estejam ligados positivamente, consequentemente o par restante vai estar relacionado igualmente.

Essas teorias entre várias outras tentam explicar o comportamento dos atores em um grafo não através de suas propriedades como sexo, idade, profissão entre outras, mas sim na posição ocupada no grafo e seus relacionamentos estabelecidos.

No FB as distinções apresentadas podem ser vistas com uma grande facilidade, por meio do mapeamento dos atores, feita pela configuração de parentesco dos amigos conectados e pela formação de grupos. No TW, a distinção das mensagens de parentescos e amigos, torna-se inviável por não possuir ferramentas que proporcionam essas distinções. As definições de relacionamentos são feitas basicamente por relacionamentos formados através do modelo de seguidores definida por Leblanc (2011).

Outra forma de entender o comportamento dos atores em RSV e através das áreas

de redes complexas ou definidas por alguns de larga-escala, é conceituada por (BARABASÍ, 2003 apud RECUERO, 2009), como sendo um grafo que apresenta uma estrutura topográfica não trivial. Essa abordagem apresenta uma característica muito peculiar em RSV, que é o número exorbitante de atores e relações obtidas nos grafos sociais formados. Recuero (2009) afirma que essas topologias são interessantes para o estudo de vários elementos das redes sociais, tais como os processos de difusão de informações. No entanto, é preciso ter claro que se trata de modelos fixos e que uma mesma rede social pode ter características de vários deles, a partir do momento em que se escolhe limitar a rede.

3 ALGORITMO DE PROPAGAÇÃO

Neste capítulo serão abordadas duas ferramentas que vão auxiliar no entendimento da propagação de mensagens em um grafo. A teoria de Grafo Fator (GF), que proporciona uma visão adequada sobre a passagem de mensagens em uma rede e o algoritmo genérico Soma e Produto (SP), escritas através das seguintes referências: Abrantes(2005), Andrade (2010), Loeliger (2004), Kock (2007), Kschischang, Frey e Loeliger (2001).

Problemas relevantes à estimação em ambientes de incerteza, como em inteligência artificial e processamento de sinais, podem ser reformulados matematicamente por uma função global ou por meio de um problema de marginalização. Uma função global pode ser definida como uma “função custo” que é aplicada em um sistema para obter um valor de medida, como por exemplo: potência, energia, probabilidade e confiança. Resultando em um valor em relação as sua marginalização.

As origens do Grafo Fator residem na teoria da codificação que oferece uma notação atraente para uma grande variedade de problemas de processamento de sinal. Em particular, um grande número de algoritmos práticos para uma grande variedade de problemas de detecção e de estimativa pode ser derivada como algoritmos de propagação de resumo. Os algoritmos derivados desta forma muitas vezes incluem os melhores algoritmos conhecidos anteriormente como casos especiais ou como aproximações óbvias (LOELIGER, 2004).

Algoritmo genérico de transferência de mensagens, usado em decodificação de sinais baseado em grafos, tem duas vertentes que surgiram em comunidades distintas, as quais mais tarde, foram identificadas como sendo mesma: o algoritmo genérico SP e o algoritmo de propagação e crença (*Belief Propagation*) (BP) aplicado na área de inteligência artificial (ABRANTES , 2005).

A aplicação do GF em conjunto com o algoritmo de SP é amplamente usada para representação da passagem de mensagens até a marginalização de um determinado grafo, mesmo que a aplicação do algoritmo de SP não seja dependente do GF o uso do conjunto torna mais fácil o entendimento do problema aplicado, assim Frey e Loeliger (2001) defendem que a maneira em que GF representa a passagem de mensagens em uma rede, a utilização dela se torna um meio pedagógico para a aplicação.

Funções-fatores podem ser entendidas como funções custos associadas aos nós no grafo. O algoritmo SP é o processo sistematizado de obter funções custos simples associadas a cada ramo do Grafo Fator. Desta forma, cada nó recebe como mensagem de um dado ramo

uma função custo simples, para dar continuidade aos cálculos seguintes. Segundo Andrade (2010), quando considerado um ramo e uma direção, é possível definir o fluxo local de mensagens relativo a eles, ou seja, as mensagens anteriores.

3.1 GRAFO FATOR

Através do uso de GF é permitido representar modelos gráficos de sistemas complexos de uma forma consistente, unificada e sistemática. Eles contribuem para derivar uma grande variedade de algoritmos. Um GF é uma representação gráfica de um modelo matemático (KOCH, 2007).

Pode-se dizer que um GF é uma coleção de funções que compartilham uma mesma coleção de variáveis. É um grafo bi-partido⁵ em nós-variáveis e nós-funções. Por convenção os nós-variáveis são representados por círculos e os nós-funções são representados por quadrados.

Existem em um GF como em TG regras e definições que ajudam a interpretar e comprovar a real definição de um gráfico em um GF, como as definições de arestas, conexidade e o número de funções relacionadas por variáveis e a própria possibilidade da fatoração de um Grafo Fator.

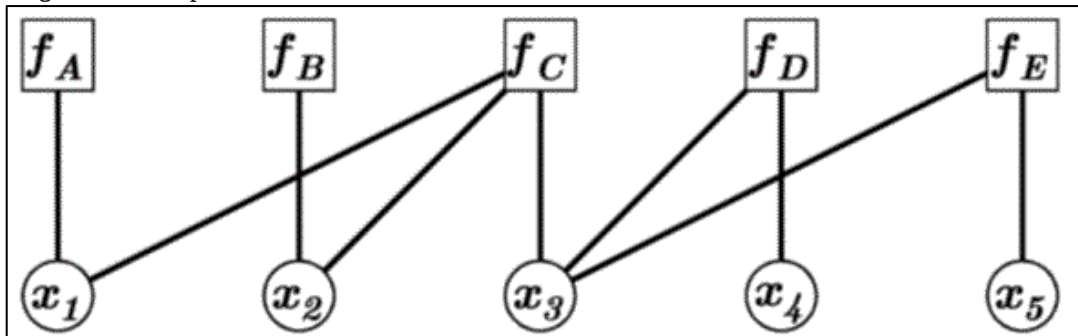
Considerando uma determinada função $g(x_1, x_2, x_3, x_4, x_5)$ a qual admite a fatoração representada pela equação (3), o resultado em um GF é definido pela

Figura 6.

$$g(x_1, x_2, x_3, x_4, x_5) = fa(x_1)fb(x_2)fc(x_1, x_2, x_3)fd(x_3, x_4)fe(x_3, x_5) \quad (3)$$

⁵ É um grafo cujo, vértices podem ser divididas em dois conjuntos distintos U e V tais que toda aresta conectada a um vértice em U deve estar conectada a um vértice em V . Um grafo bipartido é um grafo que não contém qualquer ciclo de comprimento ímpar (BOAVENTURA, 2003).

Figura 6 - Exemplo de um Grafo Fator.



Fonte: Andrade (2010, p. 13).

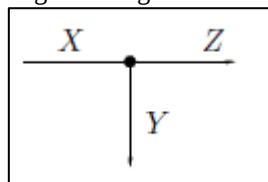
As arestas de um GF têm duas definições distintas. A aresta inteira que liga um nó a outro sendo representado na

Figura 6 entre os nós x_1 e x_3 e a meia-aresta que é definida pela ligação de um nó com um nó-fator sendo representado igualmente entre o x_1 e o f_c .

A conexidade em Grafo Fatores dar-se-á existência de uma conexão entre um nó e um nó-fator, o qual se e apenas se a função correspondente ao nó-fator é uma função da variável correspondente ao nó da extremidade. Por exemplo, a meia-aresta corresponde a x_1 está ligada ao nó-fator f_c correspondente a função $f_c(x_1, x_3)$ cujo função de x_1 .

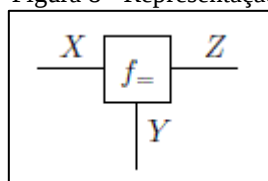
Nenhuma variável pode aparecer em mais de duas funções assim o número de arestas de um nó é restringido a apenas duas. Se três ou mais nó-fator são funções de uma mesma variável, é utilizado um nó de restrição de igualdade que permite assim obedecer esse regra. Esse nó tem a função de validar a configuração da relação entre as três variáveis. Este processo é reconhecido como clonagem de variáveis representada pelas Figura 7 e Figura 8, em que há uma igualdade das variáveis x y z sendo representadas pela função de igualde $f =$.

Figura 7 – Igualdade entre três nós.



Fonte: Kock (2004).

Figura 8 - Representação da função de igualdade.



Fonte: Kock (2004).

3.2 ALGORITMO SOMA E PRODUTO

Segundo Andrade (2010, p. 21) o “algoritmo SP estabelece que qualquer marginalização sobre um Grafo Fator pode ser realizada pelo cálculo recursivo de um conjunto de funções simples (dependentes de variáveis únicas), que podem ser entendidas como mensagens que se propagam ao longo do grafo”.

Uma vez que um dado sistema é modelado por funções-fatores e uma função global é estabelecida, o algoritmo SP é um genérico e pode ser operado sobre um GF sem ciclos, possibilitando a extração das marginais.

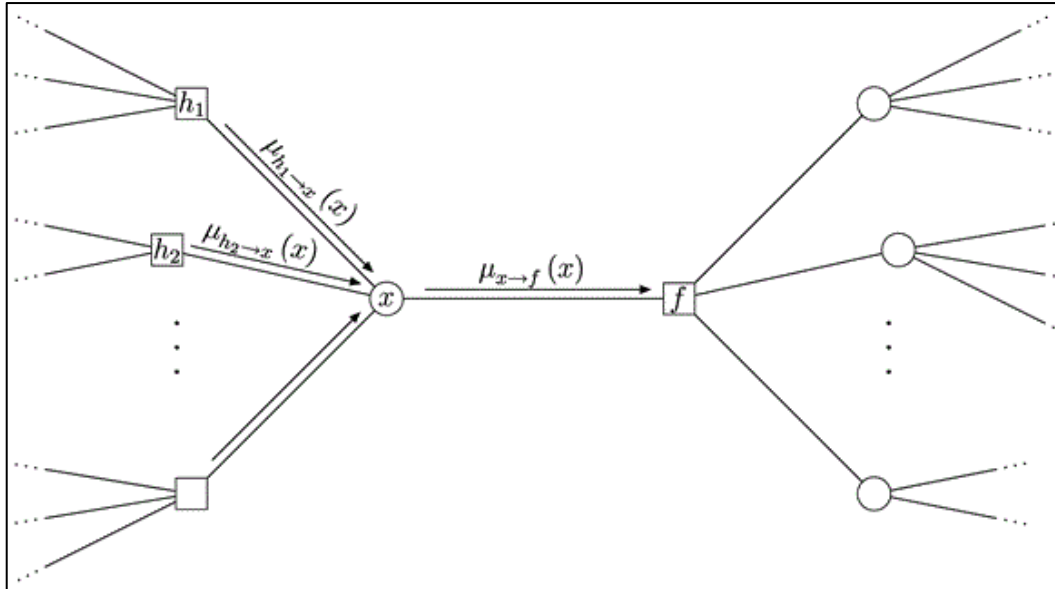
Assim o algoritmo é definido pelo processo sistematizado de obter as funções custo simples associadas a cada caminho de um GF ao contrário do GF cujas funções custo são associadas diretamente aos nós do grafo (ANDRADE, 2010) .

A fórmula da marginalização de uma variável x_i é definida pela equação (4) para uma função de várias variáveis $g(x_1, \dots, x_n)$.

$$marg_{x_i} g(x_1, \dots, x_n) = \sum_{x_1} \dots \sum_{x_{i-1}} \sum_{x_{i+1}} \dots \sum_{x_n} g(x_1, \dots, x_n) \quad (4)$$

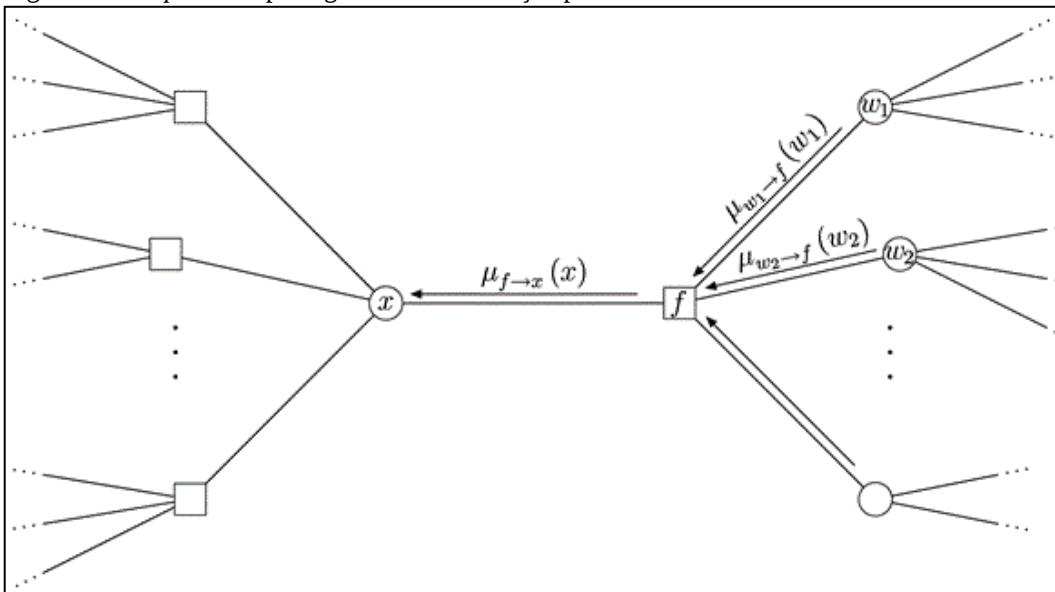
Definindo como x o nó-variável e f como nó-função de um GF representando a função custo da passagem de mensagens entre os nós. São associadas duas mensagens, uma para cada direção, denotando $\mu_{x \rightarrow f}$ para as mensagens enviadas no x para a função f , e $\mu_{f \rightarrow x}$ para as mensagens em direção oposta. Representadas pelas Figura 9 e Figura 10 respectivamente, denotando as funções de x como h além de f , e as variáveis de f como w além de x .

Figura 9 - Esquema de passagem de um nó-variável para um nó-função.



Fonte: Andrade (2010, p. 22).

Figura 10 - Esquema de passagem de um nó-função para um nó-variável.



Fonte: Andrade (2010, p. 22).

As regras de atualizações dos valores das mensagens são dados por:

$$\mu_{x \rightarrow f}(x) = \prod_{\substack{h \in N(x) \\ h \neq f}} \mu_{h \rightarrow x}(x) \quad (5)$$

$$\mu_{f \rightarrow x}(x) = \sum_{\substack{\omega \in N(f) \\ \omega \neq x}} f \prod_{\substack{\omega \in N(f) \\ \omega \neq x}} \mu_{\omega \rightarrow f}(\omega) \quad (6)$$

O resultado da marginal de cada variável é dada pelo produto do valor de todas as mensagens incidentes em seu nó, representadas pela equação (7) .

$$\mu(x) = \prod_{f \in N(x)} \mu_{f \rightarrow x}(x) \quad (7)$$

Em casos particulares, notados nas marginais de um GF, em que se encontram nós-variáveis ou nós-funções, como nós folhas⁶ do grafo, em especial são definidos algumas regras de atualização para a passagens de sinais, representadas, respectivamente, pelas equações (8) e (9).

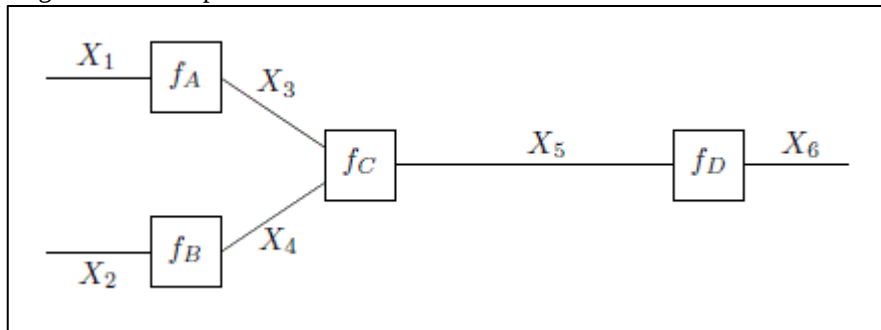
$$\mu_{f \rightarrow x}(x) = 1 \quad (8)$$

$$\mu_{x \rightarrow f}(x) = f \quad (9)$$

Seguindo como exemplo a fatoração do GS representado na Figura 11 têm a função global definida pela equação (10).

$$f(x_1, x_2, x_3, x_4, d_5, x_6) = f_A(x_1, x_2) f_B(x_2, x_4) f_C(x_3, x_4, x_5) f_D(x_5, x_6) \quad (10)$$

Figura 11 - Exemplo Grafo Fator.



Fonte: Kock (2007, p. 66).

Definimos assim que a $f(x_1, x_2, x_3, x_4, d_5, x_6)$ é uma função de probabilidade a

⁶ Denominação dada a nós em um grafo que tenha grau um (1), nós que tenham somente uma aresta, que são conectados a somente a outro nó.

qual engloba o grafo aplicado, tendo as seguintes fatorações: $f_A(x_1, x_2)$, $f_B(x_2, x_4)$, $f_C(x_3, x_4, x_5)$ e $f_D(x_5, x_6)$. A probabilidade marginal aplicada em cima do nó x_5 é determinado por $p(x_5)$ representada pela equação (11) e (12).

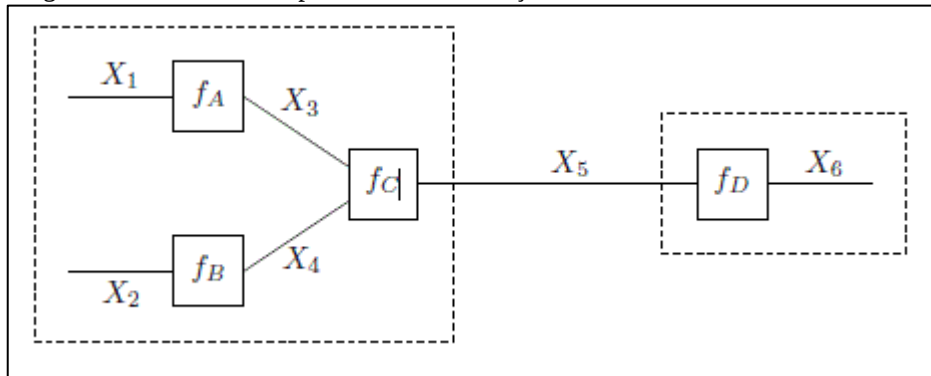
$$f(x_5) = \sum_{x_1, x_2, x_3, x_4, x_6} f(x_1, x_2, x_3, x_4, x_5, x_6) \quad (11)$$

$$f(x_5) = \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_6} f_A(x_1, x_2) f_B(x_2, x_4) f_C(x_3, x_4, x_5) f_D(x_5, x_6) \quad (12)$$

Podendo ser visualizado em cima de GF através da Figura 12 a fatoração representada pela equação (13) :

$$f(x_1, x_2, x_3, x_4, x_5, x_6) = [f_A(x_1, x_2) f_B(x_2, x_4) f_C(x_3, x_4, x_5)] \cdot [f_D(x_5, x_6)] \quad (13)$$

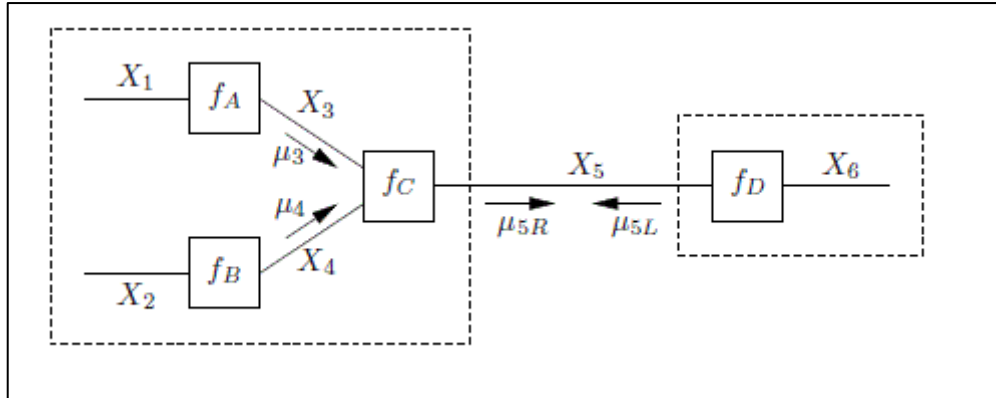
Figura 12 - Grafo Fator representando a fatoração.



Fonte: Kock (2007, p. 67).

Algumas partes desta função são equivalentes às mensagens enviadas entre os nós sendo percebidas pela Figura 13.

Figura 13 - Representação mensagens em um Grafo Fator.



Fonte: Kock (2007, p. 68).

Tendo assim o cálculo da marginal x_5 representada por $p(x_5)$, pelas equações (14), (15), (16), (17) e (18).

$$\mu_1 = \sum_{x_1} f_A(x_1, x_2) \quad (14)$$

$$\mu_2 = \sum_{x_2} f_B(x_2, x_4) \quad (15)$$

$$R = \sum_{x_3} \sum_{x_4} f_C(x_3, x_4, x_5) \cdot \mu_1 \cdot \mu_2 \quad (16)$$

$$\mu_{5L} = \sum_{x_6} f_D(x_5, x_6) \quad (17)$$

$$p(x_5) = \mu_{5R} \cdot \mu_{5L} \quad (18)$$

4 TRABALHOS CORRELATOS

O objetivo deste capítulo é abordar trabalhos, casos de uso e pesquisas relacionadas às áreas de abordagem desta pesquisa, tais como: propagação em grafos, utilização do algoritmo de SP e análise de RS.

4.1 GRAFOS-FATORES E DECODIFICAÇÃO ITERATIVA: NOVAS APLICAÇÕES

Tese de doutorado apresentada na Faculdade de Engenharia Elétrica e de Computação, defendida por Andrade (2010), o qual aborda métodos de estimação probabilística em sistemas de comunicação, usando GF e o algoritmo de SP aplicados em sistemas de comunicação formados por grafos com ciclos. Para o estudo de um ambiente ideal para derivar esquemas de codificação/estimação e interação de modelos típicos de sistemas de comunicação.

Apresentando novos entendimentos e novas observações relevantes na área específica de receptores iterativos generalizados, descrito por grafos-fatores. Tendo como contribuição uma modelagem de decodificação no contexto de GF para o esquema turbo “*stream-oriented*”, com vantagens ao esquema tradicional e a integração da quantização vetorial no contexto de GF.

O qual teve como resultado, a confirmação do uso de GF de modo generalizado enfatizando a melhor utilização dele em relação a outras representações de grafos, como a de Tanner e Forney. O entendimento geral do algoritmo SP aplicado de uma forma genérica a qual opera sobre duas operações algébricas, com espaço suficiente para fornecer novos esquemas de algoritmos de inúmeras variações aplicados em casos particulares.

4.2 REDES SOCIAIS VIRTUAIS DE INFORMAÇÃO SOBRE O AMOR

Dissertação defendida por Cruz (2011), para a obtenção do título de Mestrado em Ciência da Informação apresentada na Universidade Federal de Minas Gerais, cujo estudo levanta nas redes sociais virtuais informações sobre o amor, formadas pelos usuários através da interação no site Orkut, levando em consideração o comportamento, a cultura e as postagens sobre o amor na rede social.

Tendo como problema de pesquisa o seguinte questionamento: “Como se configura o comportamento e a cultura informacional dos usuários de redes sociais virtuais de

informação sobre o amor existentes nas comunidades virtuais do site Orkut?” aplicadas em algumas comunidades do site Orkut, apresentando teorias sociológicas explicando como funciona a interação dos usuários na rede social virtual, com os seguintes tópicos: redes sociais virtuais, redes sociais, elementos de análise e tipos de redes sociais, fluxo de informações em redes sociais virtuais entre outros.

Após uma análise qualitativa de como se configura o comportamento e acultura informacional dos usuários na rede social virtual do site Orkut em cinco comunidades obteve informações importantes defendidas pelas teorias de AERS sobre o seu problema de pesquisa.

4.3 E-CRM E REDES SOCIAIS: UMA INTERAÇÃO POSSÍVEL

É um trabalho apresentando por Giusti (2010) para a conclusão do curso de Bacharel em Ciência da Computação na Universidade do Extremo Sul Catarinense, o qual defende o uso de sistemas de “*Customer Relationship Management*” (CRM) relacionados em RSV para a obtenção de informações relevantes dos usuários possibilitando futuras tomadas de decisões.

O autor propôs um modelo de integração entre e-CRMs e RSV para recuperação de informações sobre um determinado produto a partir de opiniões extraídas dos consumidores, analisando interações.

Segundo Giusti (2010) os dados sobre a integração dos e-CRMs com as RSV com a aplicação de um software open-source chamado VtigerCRM e a API PHP Twitter Search, com o modelo de integração proposto no trabalho foram positivos.

4.4 OS IMPACTOS DAS REDES SOCIAIS NA COMUNICAÇÃO ORGANIZACIONAL

Monografia apresentada por Carrer (2012) ao Curso de Pós-Graduação em Comunicação Estratégica nas Organizações da Pontifícia Universidade Católica de Minas Gerais para a obtenção do título de Especialista em Comunicação Estratégica nas Organizações apresenta uma leitura do relacionamento da Instituição de Ensino Superior, Universidade do Extremo Sul Catarinense (UNESC), com seus alunos, buscando identificar o relacionamento com os estudantes entre instituição versus acadêmicos, levantando questões norteadoras para o aprimoramento do uso das redes sociais na comunicação e organização na instituição, por meio de questionários com os acadêmicos e profissionais da comunicação da

universidade.

A pesquisa teve como resultado, a clara dominância das redes sociais Facebook e Twitter no uso cotidiano dos acadêmicos e profissionais da comunicação, tornando as redes sociais um canal de aproximação, ultrapassando os limites físicos da universidade e atingindo o domicílio dos usuários. Embora algumas barreiras foram levantadas, como o bloqueio do acesso à rede social Facebook por meio da rede interna da universidade, os acadêmicos informaram que as redes sociais têm seu papel principalmente para a atualização de novidades da universidade, divergente das ideias dos funcionários de comunicação que afirmaram que o canal é utilizado também para fins de responder dúvidas dos alunos.

4.5 SOCIAL INFLUENCE ANALYSIS IN LARGE-SCALE NETWORKS

Trabalho apresentado por Tang, Sun e Yang (2009) na conferência anual de ACM SIGKDD, principal fórum internacional para pesquisadores de mineração de dados e profissionais da academia, ocorrido em Paris (França). O Trabalho propõe um modelo de propagação denominado *Topical Affinity Propagation* (TAP), para modelar a influência em redes sociais de larga escala. Apresentando 2 perguntas, as quais ajudam a analisar a influência: “O que são os nós representativos sobre um determinado assunto?” e “Como identificar as influências sociais de nós vizinhos de um nó particular?”. Demonstrando a real eficácia do algoritmo TAP em grandes conjuntos de dados.

A conclusão do trabalho obteve resultados positivos no uso do algoritmo proposto por eles sobre um modelo de Map-Reduce, por meio da aplicação três diferentes conjuntos de dados, contribuindo com a constatação de influências de peritos em velocidade e tempo de resposta.

4.6 SOCIODRAMA E O TESTE SOCIOMÉTRICO

Trabalho de conclusão de curso apresentado por Dias (2012) para a obtenção do grau de Psicólogo pela Universidade do Extremo Sul Catarinense na linha de pesquisa em Saúde e Processos Psicossociais. O trabalho constatou em constituir as ferramentas que ajudem na forma de compreensão das relações interpessoais. Coletando dados de uma pesquisa sociométrico aplicada a sete pessoas.

Foi levantada a importância do teste sociométrico, pontuando que a ferramenta

produzida é de extrema importância para o auxílio do movimento psicodinâmico do grupo, levando a um melhor entendimento em um grupo terapêutico.

5 PROPAGAÇÃO DE INFORMAÇÃO: AVALIAÇÃO DO ALGORITMO SP

Esta pesquisa buscou avaliar a aplicabilidade do uso do algoritmo SP para calcular pontos de propagação de informação em redes sociais. Para atingir o objetivo desta pesquisa foram efetuados: uma pesquisa bibliográfica referente á AERS, em se identificou teorias e características que definem o comportamento das pessoas em RS e no entendimento do algoritmo SP para o cálculo da propagação das informações junto ao GF, o qual auxiliou no entendimento do relacionamento entre os atores.

5.1 METODOLOGIA

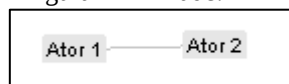
Dentro das linhas de pesquisa do curso de Ciência da Computação, este projeto se insere na área de Teoria de Grafos, Sistemas de Informação e Inteligência Artificial.

As etapas de desenvolvimento da pesquisa foram: levantamento bibliográfico, construção de grafos de representação de RS, a aplicação do algoritmo SP nos grafos sociais para a identificação de pontos de propagação da informação e por último a avaliação dos dados obtidos da aplicação do algoritmo junto ás teorias explicativas na área de AERS.

5.2 GRAFOS DE REPRESENTAÇÃO

Segundo Recuero (2005) a unidade básica das redes sociais definidas pelos sociólogos eram as díades, relações entre duas pessoas seriam a menor estrutura relacional em uma sociedade. Relação essa, definida pela Figura 14 onde se pode fazer o entendimento das relações mais simples de dois dos três tipos de redes sociais definidas por Leblanc (2011). A de conexão cujo grafo é formado por uma conexão não direcionada e a de seguidor em que são definidas conexões direcionadas.

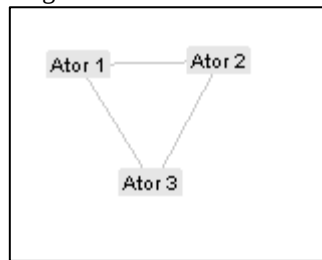
Figura 14 – Díade.



Fonte: Do autor.

Um segundo grafo a ser estudado representando pela Figura 15, é constituído de uma tríade, em que a rede têm uma forma triangular, possuindo dois (2) atores que tenham relacionamento com um ator distinto.

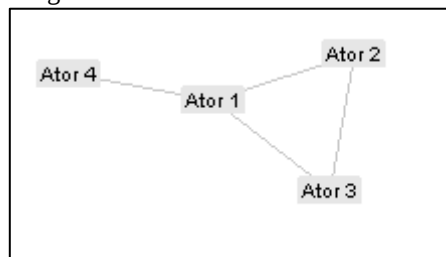
Figura 15 – Tríade.



Fonte: Do autor.

Outro grafo um pouco mais complexo proposto por Lemimeux e Ouimet (2004), compreende na relação entre quatro atores, contendo uma díade e uma tríade, podendo ser aplicado os três tipos de redes definidos por Leblanc (2011) apresentado na Figura 16.

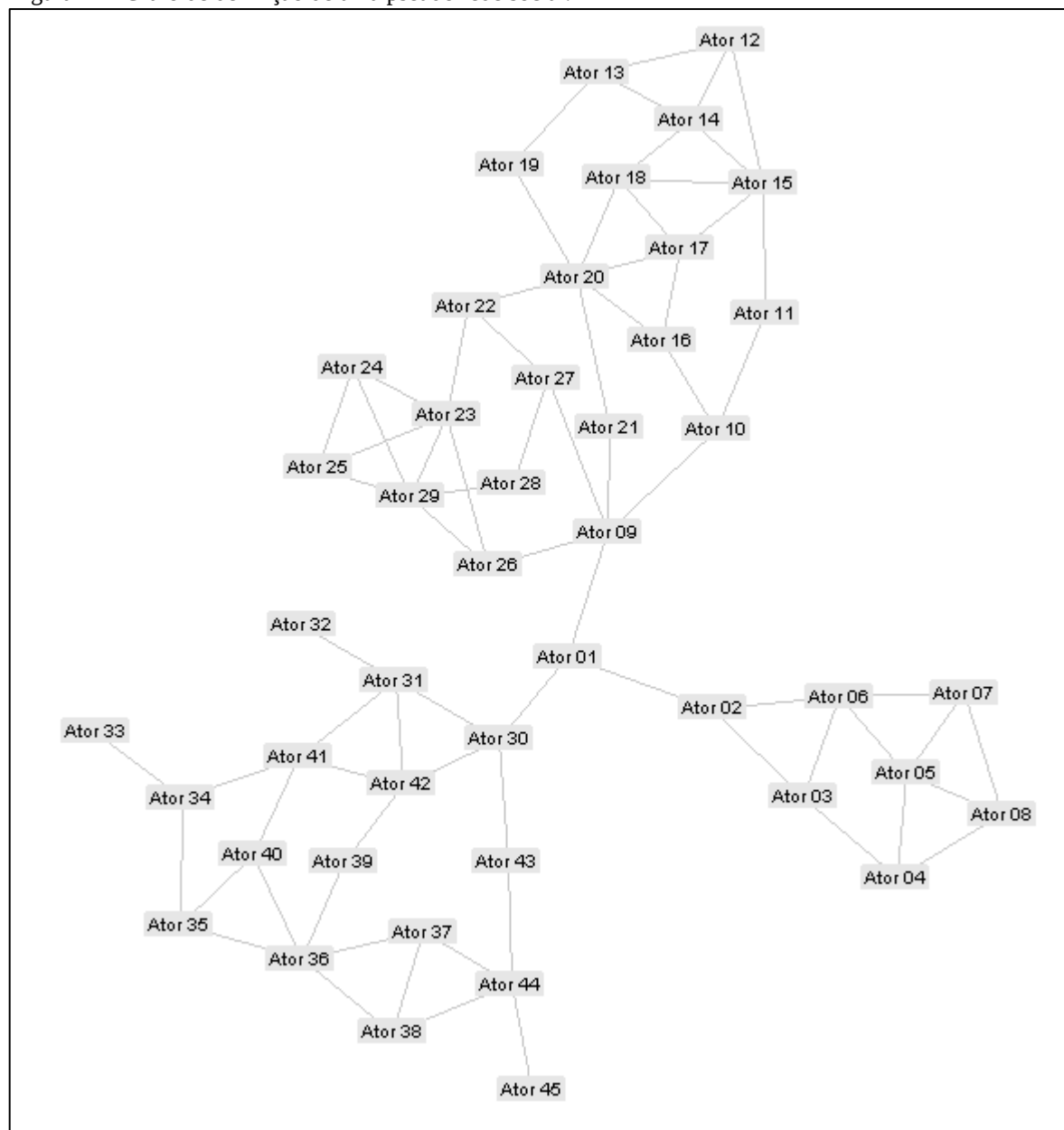
Figura 16 - Grafo 3.



Fonte: Do autor.

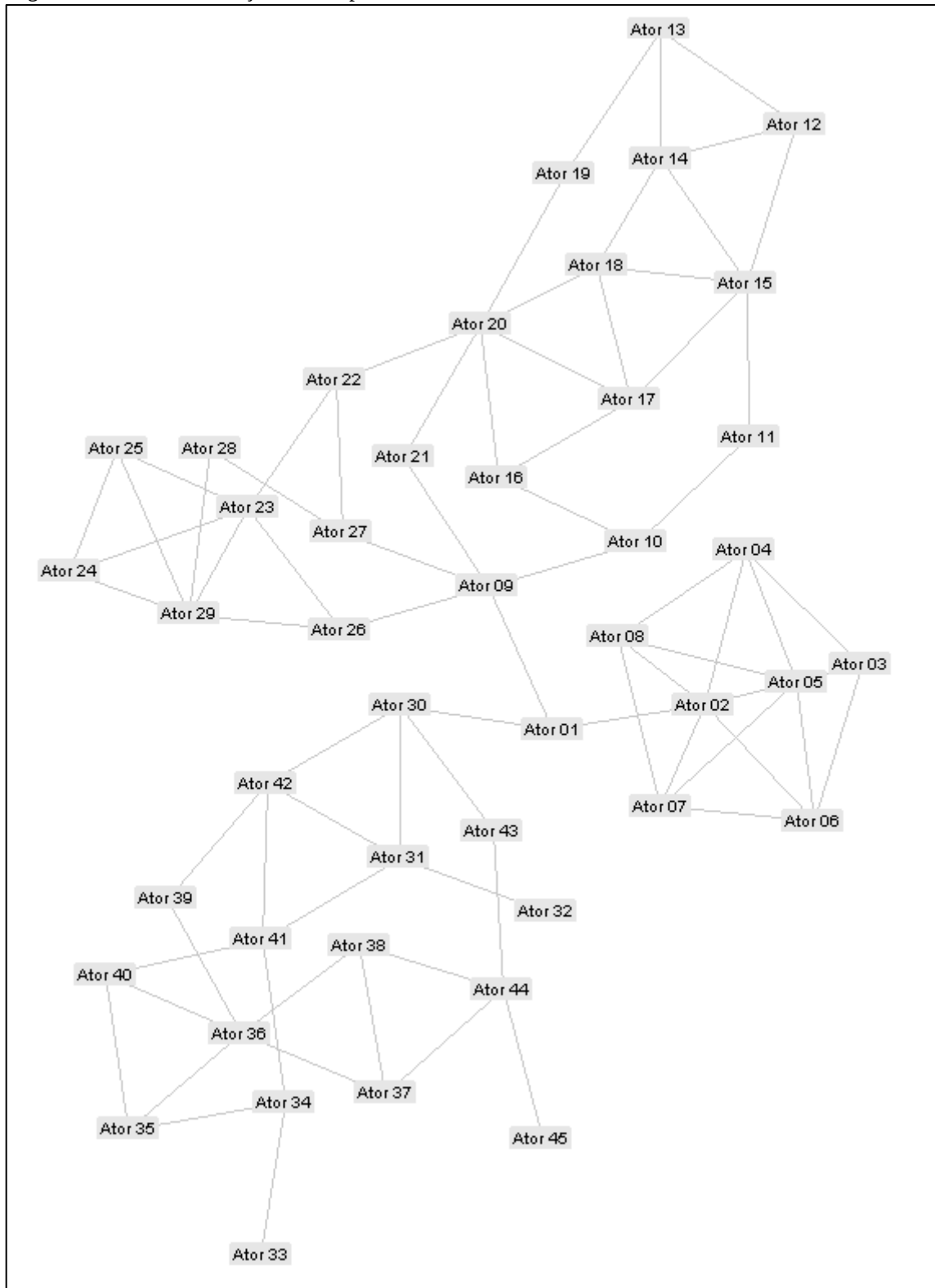
E por último são propostos dois grafos definidos pelas Figura 17 e Figura 18, os quais apresentam quarenta e cinco (45) atores, com as várias características e teorias defendidas no capítulo 2.3 CONSTRUÇÕES DAS RELAÇÕES NAS REDES SOCIAIS e no capítulo 2.4 TEORIAS EXPLICATIVAS SOBRE PROPAGAÇÃO, os quais contém somente uma pequena diferença entre os relacionamentos contidos no sub-grafo mantidos pelos atores 2 a 8, o número de conexões do Ator 2.

Figura 17 - Grafo de definição de uma pseudo rede social.



Fonte: Do autor.

Figura 18 - Grafo de definição de uma pseudo rede social 2.



Fonte: Do autor.

5.3 ALGORITMO SOMA E PRODUTO

Para a aplicação do algoritmo de Soma e Produto (SP) junto a um Grafo Fator (GF), foi utilizada a implementação do algoritmo contido do *framework* Mallet (2012), junto com a implementação do algoritmo Junction Tree.

Em geral as regras de propagação das mensagens do algoritmo SP são recursivas e localmente aplicadas em GF envolvendo somente as variáveis e funções vizinhas. Portanto a obtenção das marginais por meio de grafos com ciclos⁷ é obtida somente por meio de técnicas de ordem de propagação de mensagens ou usando algum critério de parada, defendida por Andrade (2010).

Em redes sociais a formação de ciclos é muito comum, sendo notada através da obtenção de cliques e a formação de grupos. Assim, o surgimento de ciclos para o cálculo de propagação em uma RS torna-se inevitável. Na utilização do *framework* Mallet, a classe responsável para o cálculo das marginais (JunctionTreeInferencer) é responsável também pela determinada solução desse problema.

A propagação da informação é definida pelo no-fator em um GF, o qual é definido pela chamada do método `addFactor` de uma instância da classe `FactorGraph` o qual é passada as duas variáveis, é uma matriz que corresponde à função custo da passagem de mensagens entre as duas variáveis.

No *framework* Mallet é definido o número de hipóteses de cada variável no construtor do objeto (ex: `'new Variable(2)'`), sendo que o valor mínimo possível para estas hipóteses é 2, compreendido como negação e afirmação da passagem das mensagens na RS. Portanto a matriz responsável pela função custo têm que ter, no mínimo, uma dimensão de 2x2. Em que 0 corresponde à negação e 1 à afirmação. Cada ligação dirigida de um ator a outro é quantificada por uma matriz de probabilidade condicional exemplificado demonstrada pela Tabela 1, em que são demonstrados na terceira coluna valores hipotéticos do custo das mensagens.

Tabela 1 - Matriz de probabilidade condicional.

Ator 0	Ator 1	Resultado
0	0	0.7
0	1	0.3
1	0	0.0
1	1	1.0

⁷ Um ciclo num grafo é um caminho de um elemento u para o próprio u . Dizemos que um grafo tem ciclos se existe pelo menos um nó com um ciclo. Um grafo é uma árvore se é conexo e não tem ciclos Andrade (2010).

Fonte: Do autor.

5.3.1 Framework Mallet

Mallet (*Machine Learning for Language Toolkit*) é um *framework* mantido pelo departamento de Ciência da Computação da Universidade de Massachusetts Amherst (UMASS), feito na linguagem Java para processamento de linguagem estatística natural, classificação de documentos, *clustering*, modelagem tópico, extração de informação e outras máquinas para o aprendizado da aplicação em textos. O framework tem um add-on chamado GRMM (Graphical Models in Mallet) usado para calcular a inferência em geral, modelos gráficos e formação de *Conditional Random Fields* (CRFs) com estrutura gráfica arbitrária.

O *framework* pode ser obtido pelo site <http://mallet.cs.umass.edu/>, sendo de código livre, liberado sobre a Licença Pública Comum (CPL), a qual permite usar o código para fins de pesquisa ou comercial, o que confirma a utilização do mesmo.

A Figura 19 demonstra a utilização do *framework* para calcular as marginais de um grafo, seguida pela Tabela 3 que explica individualmente o uso de cada classe.

Figura 19 - Código Exemplo Framework Mallet.

```
Variable variable1 = new Variable(2);
variable1.setLabel("Ator 1");
Variable variable2 = new Variable(2);
variable2.setLabel("Ator 2");

FactorGraph factorGraph = new FactorGraph();
factorGraph.addFactor(variable1, variable2,
    new double[]{.1, .5, .5, 1});

Inferencer inf = new JunctionTreeInferencer();
inf.computeMarginals(factorGraph);

for (Variable variable : new Variable[]{variable1, variable2}) {
    Factor ptl = inf.lookupMarginal(variable);
    for (AssignmentIterator it = ptl.assignmentIterator();
        it.hasNext(); it.next()) {
        int outcome = it.indexOfCurrentAssn();
        System.out.println(variable + " " + outcome
            + " " + ptl.value(it));
    }
    System.out.println();
}
```

Fonte: Do autor.

Tabela 2 - Classes utilizadas do *framework* Mallet.

Classe	Pacote	Descrição
Variable	<i>cc.mallet.grmm.ty</i>	Representa os nós-variáveis de um GF, e o argumento do construtor é o número de valores compreendidos
FactorGraph	<i>cc.mallet.grmm.types</i>	Representa os nós-variáveis de um GF, e o argumento do construtor é o número de valores compreendidos, e a função <i>addFactor</i> adiciona o relacionamento dos atores com sua devida função custo, representada por uma matriz do tipo primitivo <i>double</i> .
JunctionTreeInferencer	<i>cc.grmm.inference.abstractinferencer</i>	Responsável por calcular a inferência de uma instância de <i>FactorGraph</i> o qual abstrai a implementação do algoritmo Junction Tree.

Fonte: Do autor.

5.3.2 Algoritmo Junction Tree

Para a solução do cálculo de propagação de uma RS contendo ciclos, defendida por Barber (2012) cujo *framework* Mallet disponibiliza a utilização do algoritmo Junction Tree.

Jordan (1998) e Barber (2012) defendem que o algoritmo Junction Tree, seja utilizado para transformar grafos orientados ou não orientados, em grafos triangulares, formando uma estrutura em árvore, cujo o algoritmo SP pode seja aplicado sem nenhum

problema.

A execução dos procedimentos do algoritmo pode ser resumida em sete procedimentos:

- a) moralizar (grafos orientados apenas): quando o grafo em questão é orientado, ele transforma o grafo em um não orientado para possibilitar a transformação do grafo em um de tipo de árvore. Esta etapa pode ser desconsiderada nos casos em que o grafo em questão já é um não orientado;
- b) triangulação do grafo: para cada ciclo maior que 4 se possuir 2 arestas, é adicionada uma aresta até a obtenção geral de sub-grafos com 3 arestas;
- c) forma da árvore de junção: transformar o grafo em árvore duplicando se necessário às arestas;
- d) atribuir os potenciais para os cliques de árvore de junção e inicializar o separador potencial para a unidade: em grafos-fatores os valores das funções-custo deverão ser redefinidos, pelo fato da alteração da estrutura da rede ;
- e) selecionar um clique de raiz (arbitrária): é selecionado um nó-variável que será a raiz da árvore;
- f) realizar a passagem de mensagens com a absorção do nó raiz até atualizações passar ao longo dos dois sentidos de cada link na árvore de junção: isso é feito para garantir a convergência da árvore em casos de ter feito a primeira etapa de transformação de grafos orientados;
- g) ler os potenciais cliques marginais da árvore de junção: depois de realizar a passagem de mensagens para toda a árvore, obtém-se a função-custo de cada nó para obter assim a função geral do sistema para possibilitar o cálculo das marginais.

Segundo Jordan (1998) se algum nó-variável estiver em mais de um clique, estará contida em mais de um agrupamento no grafo em árvore. A escolha da árvore se torna imaterial exceto por considerações de eficiência computacional, podendo perder algumas propriedades de independência condicional na distribuição do grafo original pela adição de novas arestas.

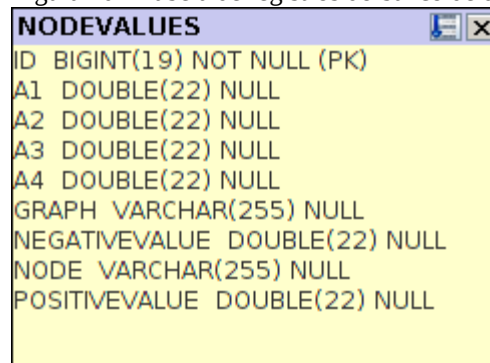
5.4 RESULTADOS OBTIDOS

A avaliação do uso do algoritmo SP na identificação dos pontos de propagação em uma rede social foi obtida por meio da aplicação nos grafos definidos no capítulo 5.2 que foram confrontados com as teorias de AERS abordadas na fundamentação teórica.

Para o cálculo da propagação de informação entre todos os atores, foram aplicados valores que variam de 0.0 a 1, para cada um dos quatros possíveis resultados da matriz. Tratando-se de um algoritmo de probabilidade, não necessariamente a soma dessas possíveis hipóteses tem que ser igual a um, no caso cem por cento (100%), assim Pearl (1988) define que cada nó pode conter um conjunto de hipóteses mutualmente exclusivas, em que o resultado final sofrerá mudanças pela normalização do resultado.

Para cada autor foram gerados 14641 registros das possíveis hipóteses, por meio da combinação dos quatro valores possíveis da matriz de probabilidade condicional. Os resultados foram gravados em um banco de dados MYSQL, devido o alto custo computacional envolvido para o cálculo das hipóteses. Foi criada uma tabela contendo nove colunas, sendo: uma para a chave primária, outras quatro (4) representando os valores entre 0.0 a 1 de cada linha da matriz, outro para a identificação do grafo de aplicação, outra para identificar o ator e por fim uma do resultado de propagação, representada pela Figura 20.

Figura 20 - Tabela de registros do banco de dados



NODEVALUES								
ID	BIGINT(19)	NOT NULL	(PK)					
A1	DOUBLE(22)	NULL						
A2	DOUBLE(22)	NULL						
A3	DOUBLE(22)	NULL						
A4	DOUBLE(22)	NULL						
GRAPH	VARCHAR(255)	NULL						
NEGATIVEVALUE	DOUBLE(22)	NULL						
NODE	VARCHAR(255)	NULL						
POSITIVEVALUE	DOUBLE(22)	NULL						

Fonte: Do autor.

A Figura 21 representa valores de alguns cálculos executados, em que foi feito a partir de uma classe Java exemplificado pela figura Figura 22.

Figura 21 - Exemplo dos dados gerados

ID	A1	A2	A3	A4	GRAPH	NEGATIVE...	NODE	POSITIVEVALUE
5974726	0	0	0	0.5	11 GRAPH TERTIUS	0	Autor: 0	1
5974727	0	0	0	0.8	11 GRAPH TERTIUS	0	Autor: 0	1
5974728	0	0	0	0.1	11 GRAPH TERTIUS	0	Autor: 0	1
5974729	0	0	0	0.4	11 GRAPH TERTIUS	0	Autor: 0	1
5974730	0	0	0	0.3	11 GRAPH TERTIUS	0	Autor: 0	1
5974731	0	0	0	0.2	11 GRAPH TERTIUS	0	Autor: 0	1
5974732	0	0	0	0.7	11 GRAPH TERTIUS	0	Autor: 0	1
5974733	0	0	0	0	11 GRAPH TERTIUS	<null>	Autor: 0	<null>
5974734	0	0	0	1	11 GRAPH TERTIUS	0	Autor: 0	1
5974735	0	0	0	0.9	11 GRAPH TERTIUS	0	Autor: 0	1
5974736	0	0	0	0.6	11 GRAPH TERTIUS	0	Autor: 0	1
5974737	0	0.1	0.1	0.2	11 GRAPH TERTIUS	0.00192	Autor: 0	0.99808
5974738	0	0.1	0.1	0.3	11 GRAPH TERTIUS	0.00005	Autor: 0	0.99995
5974739	0	0.1	0.1	0.9	11 GRAPH TERTIUS	0	Autor: 0	1
5974740	0	0.1	0.1	1	11 GRAPH TERTIUS	0	Autor: 0	1
5974741	0	0.1	0.1	0.6	11 GRAPH TERTIUS	0	Autor: 0	1
5974742	0	0.1	0.1	0.4	11 GRAPH TERTIUS	0	Autor: 0	1
5974743	0	0.1	0.1	0.5	11 GRAPH TERTIUS	0	Autor: 0	1
5974744	0	0.1	0.1	0.8	11 GRAPH TERTIUS	0	Autor: 0	1
5974745	0	0.1	0.1	0.1	11 GRAPH TERTIUS	0.09091	Autor: 0	0.90909

Fonte: Do autor.

Figura 22 - Classe de representação da tabela do banco de dados para o cálculo na linguagem Java

```
import java.io.Serializable;
import javax.persistence.*;

@Entity
public class NodeValues implements Serializable {

    @Id
    @GeneratedValue(strategy = GenerationType.SEQUENCE)
    private Long id;
    @Column
    private String graph;
    @Column
    private String node;
    @Column
    private Double a1;
    @Column
    private Double a2;
    @Column
    private Double a3;
    @Column
    private Double a4;
    @Column
    private Double negativeValue;
    @Column
    private Double positiveValue;

    getters setter
}
```

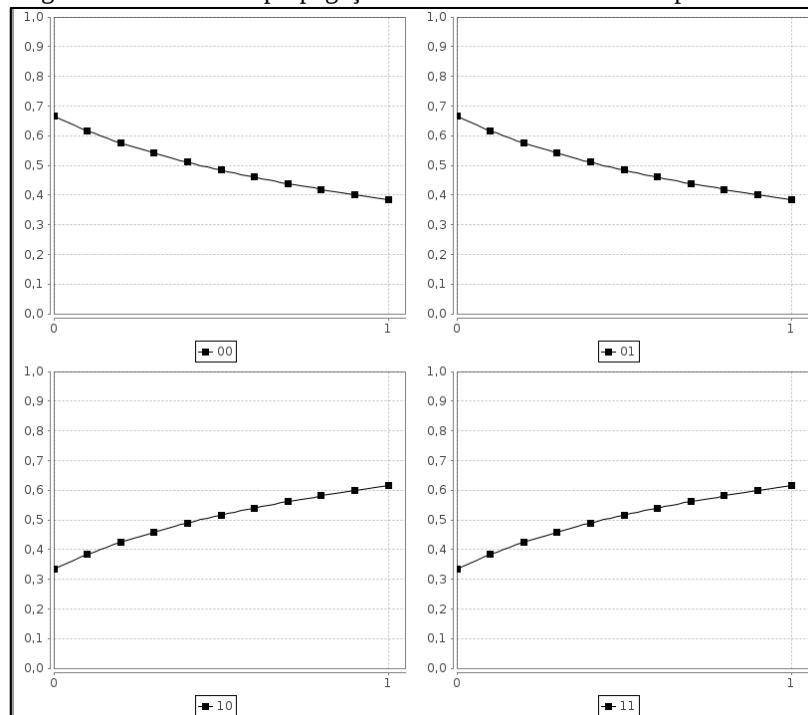
Fonte: Do autor.

Os resultados são demonstrados por meio de gráficos, mostraram a curva de

propagação dos atores de acordo com cada combinação de valores proposta pela matriz de probabilidade condicional, onde o eixo X representa o valor da função definida pela matriz e o eixo Y o valor de propagação obtido.

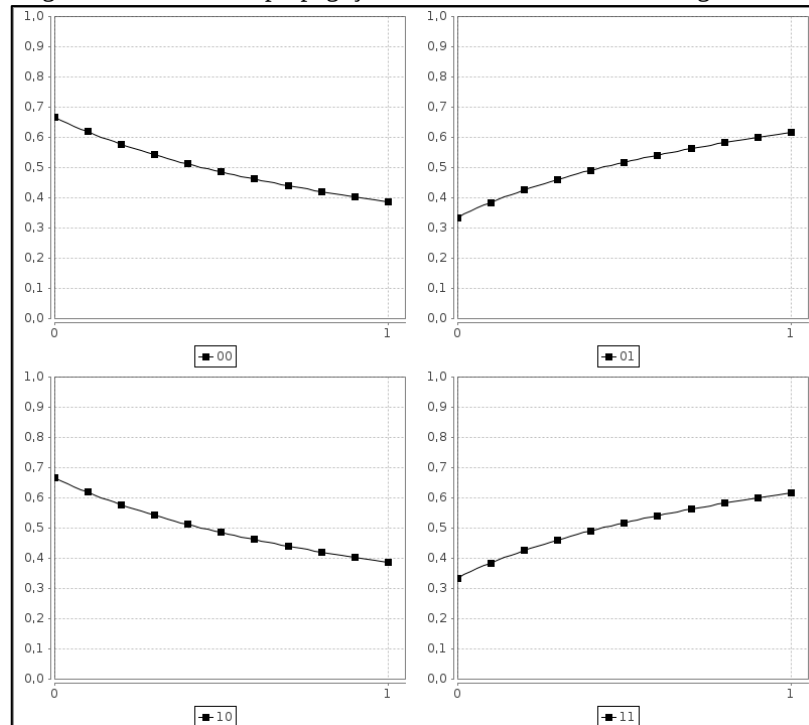
Em um teste de propagação em uma díade exemplificada pela Figura 14 os gráficos gerados para representação da propagação dos Atores 1 e 2 são demonstrados nas Figura 23 e Figura 24, respectivamente. Na observação da linha dos gráficos, percebe-se uma clara distinção na direção das mensagens, quando a mensagem de negação é iniciada pelo Ator 1 nas duas primeiras linhas cujo valor do ator é zero (0) na matriz, o resultado de propagação tende a cair quanto maior o valor resultado, o inverso acontece nas duas últimas linhas da matriz cujo valor é um. Na observação da propagação do Ator 1 isso também é claramente observado e a linha de propagação da primeira e da última linha da matriz é mantida igual, invertendo somente a segunda e terceira linha, onde acontece a inversão do resultado de propagação.

Figura 23 - Gráfico de propagação de uma Díade iniciado no primeiro ator.



Fonte: Do autor.

Figura 24 - Gráfico de propagação de uma Díade iniciado no segundo ator.



Fonte: Do autor.

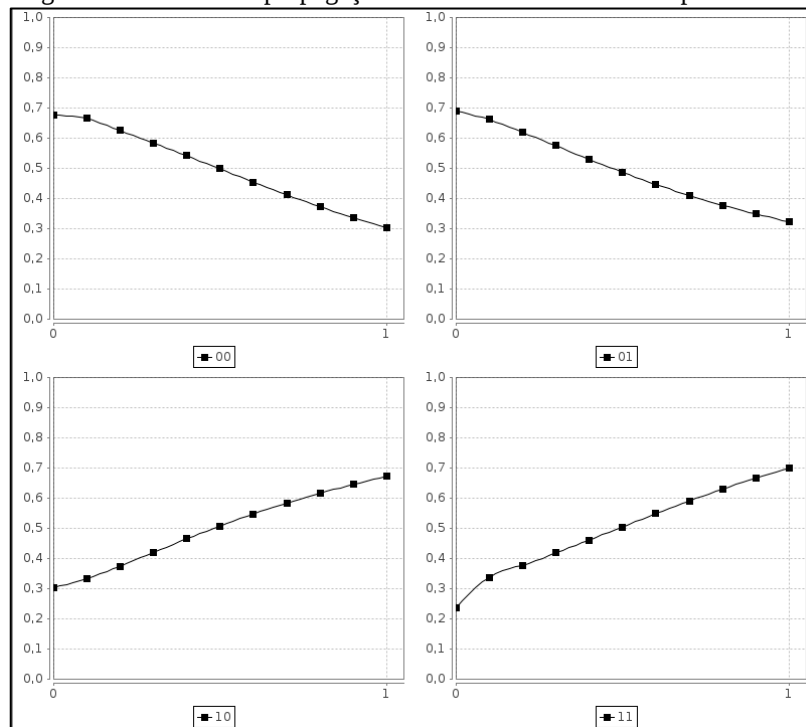
Cruzando a propagação dos dois atores com seus respectivos valores na matriz, pode-se dividir os resultados obtidos para distinguir o ambiente em que eles vão se adequar corretamente. Avaliando pela divisão de tipos de conexão definidas por Leblanc (2011), quando os valores da matriz forem iguais, a conectividade é definida pelo modelo de Conexão (FB), e quando se divergem, pode ser definida pelo modelo de Seguidor (TW).

Aplicando o algoritmo SP em uma tríade representada pela Figura 15, comprova-se a aplicação do mesmo. Por meio de uma análise em cima das características de orientação, densidade e conexidade, mesmo que o grafo apresente características claras de uma rede não direcionada e conexa, fazendo uma análise empírica, pode-se obter resultados de outros sub-grafos, retirando e/ou adicionando direções a ligações dos atores. Esta técnica pode ser explanada, fazendo uma seleção prévia dos resultados a serem analisados da matriz de probabilidade condicional.

Nas classificações de conexidade é definido como conexo se todos os atores forem dominantes, desde quando observar o grafo com conexões não direcionadas. Se observar o grafo com conexões direcionadas pode-se aplicar as classificações de não-conexo, obtendo uma díade a partir de um sub-grafos deixando um ator sem qualquer conexão com os outros atores podendo classificar como quase-fortemente-conexo, podendo observar também atores dominantes e dominados de acordo com a direção das conexões.

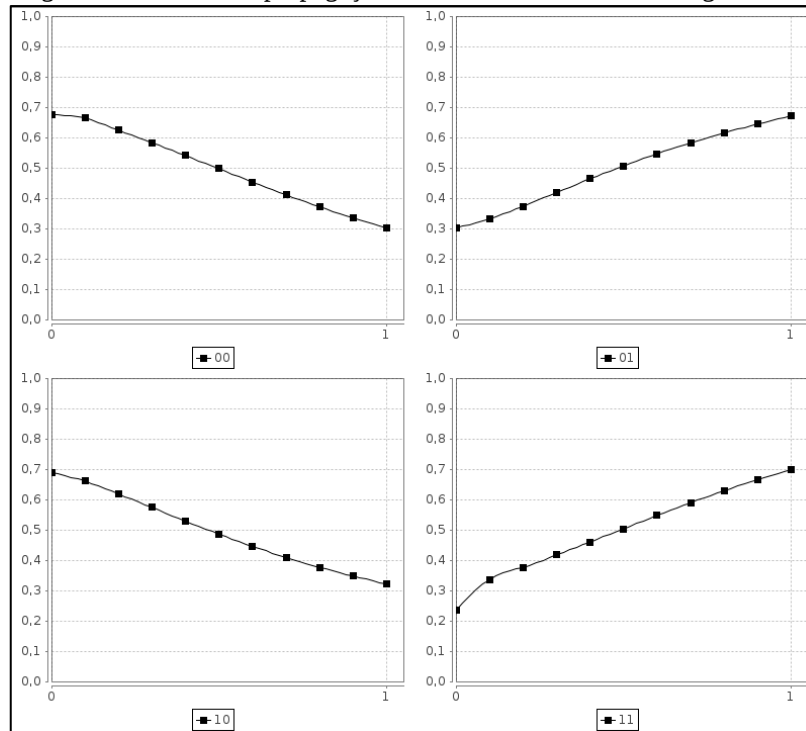
A relação da densidade com o fator de propagação pode ser provada através das conexões orientadas ou não, tendo uma propagação proporcional com a da conexidade do grafo, cujas possíveis conexões variam de acordo com a conexidade da Tríade. Isso pode ser comprovado a partir dos gráficos demonstrados pelas Figura 25 e Figura 26, em que foi obtido o mesmo comportamento de propagação definida na matriz de probabilidade condicional da propagação de uma díade, destacando que pico de propagação da tríade foi de $\text{marg}(\text{ator1}) = 0.7$ quando o da díade foi de $\text{marg}(\text{ator1}) = 0.6$. A díade teve um crescimento uniforme em razão aos valores da função de propagação e a tríade teve um crescimento rápido nos intervalos em que os valores variam entre 0 e 0.1.

Figura 25 - Gráfico de propagação de uma Tríade iniciado no primeiro ator.



Fonte: Do autor.

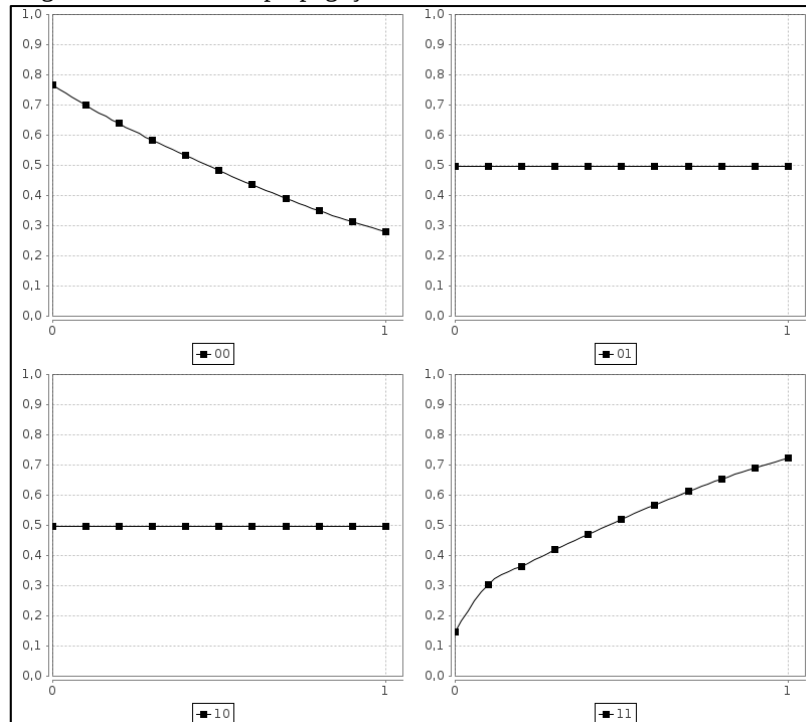
Figura 26 - Gráfico de propagação de uma Tríade iniciado no segundo ator.



Fonte: Do autor.

Um problema encontrado para o cálculo de propagação foi a definição da orientação em relação à matriz, o estudo de uma díade, com uma única aresta em sua devida matriz. O resultado foi coerente, porém quando aplicado em uma tríade o resultado não foi o esperado, se observar na Figura 27. Os valores da matriz referentes ao tipo de **conexão seguidor** (01 e 10) mantiveram uma media de propagação uniforme de acordo com a variação dos valores do coeficiente de propagação da matriz.

Figura 27 - Gráfico de propagação de uma Tríade iniciado no terceiro ator.



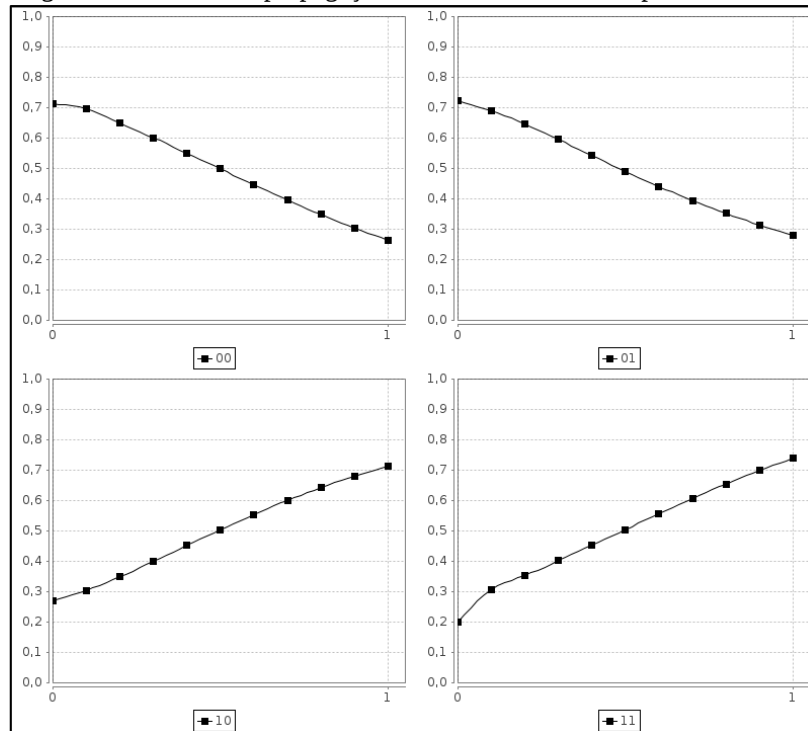
Fonte: Do autor.

Na adição de um ator no grafo analisado em uma Tríade anteriormente, representado pela Figura 15, é feita uma análise em que a gama de características relativamente é maior e mais próxima da realidade de uma RSV; como: densidade, conexidade e a posição dos atores. Sendo proposto o grafo da Figura 16, extraído a partir de um sub-grafo definido por Lemieux e Ouimet (2004), o qual ele defende que há uma conexidade quase forte. Analisando os gráficos de propagação das Figura 28,

Figura 29, Figura 30 e Figura 31, os quais representam o valor médio de propagação dos Atores 1, 2, 3 e 4 respectivamente, segue a análise referente de cada ator.

Analisando os valores do Ator 1, percebe-se que o valor médio máximo de propagação é maior que os demais, destacando que ele é o único ator na rede que possui uma ligação direta com o resto dos atores. Pode se definir assim, que ele é um ator dominante em relação aos demais.

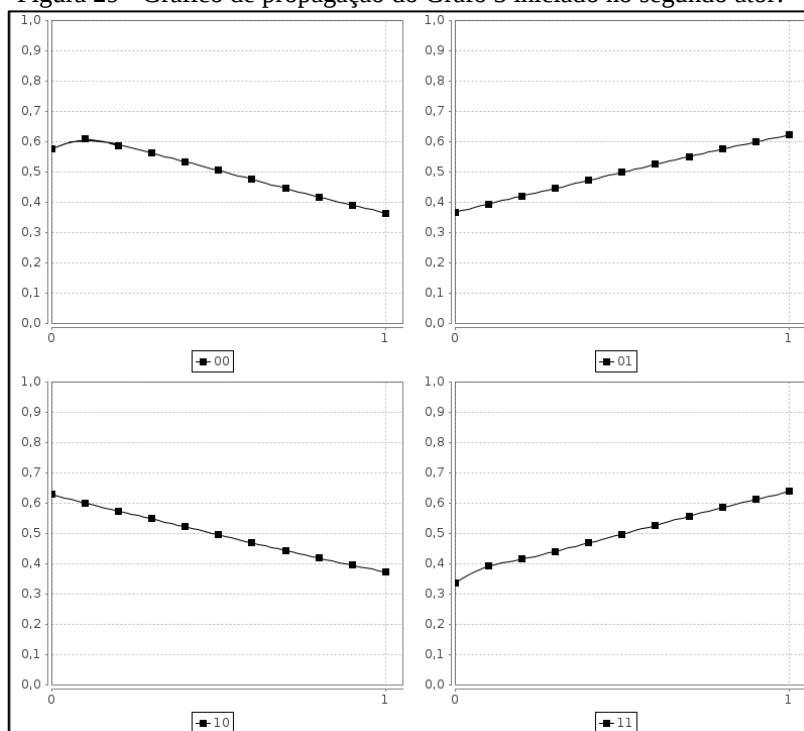
Figura 28 - Gráfico de propagação do Grafo 3 iniciado no primeiro ator.



Fonte: Do autor.

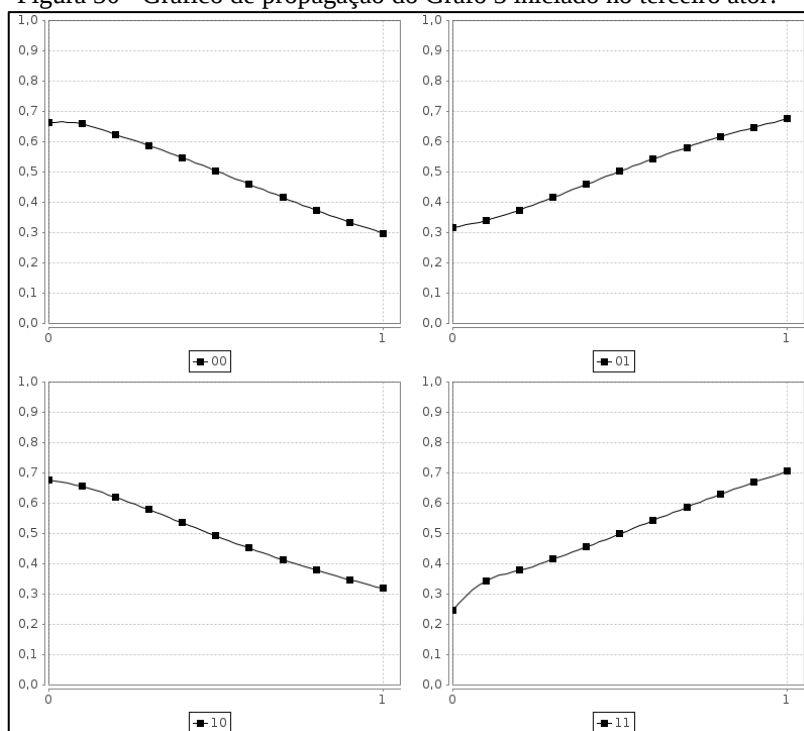
Os Atores 2 e 3 tiveram um resultado parecido devido o compartilhamento das mesmas conexões e ligações com atores em comum. Como resultado as variações se explicam possivelmente devido ao problema da aplicação da matriz pelo *framework* explicada na avaliação da tríade.

Figura 29 - Gráfico de propagação do Grafo 3 iniciado no segundo ator.



Fonte: Do autor.

Figura 30 - Gráfico de propagação do Grafo 3 iniciado no terceiro ator.

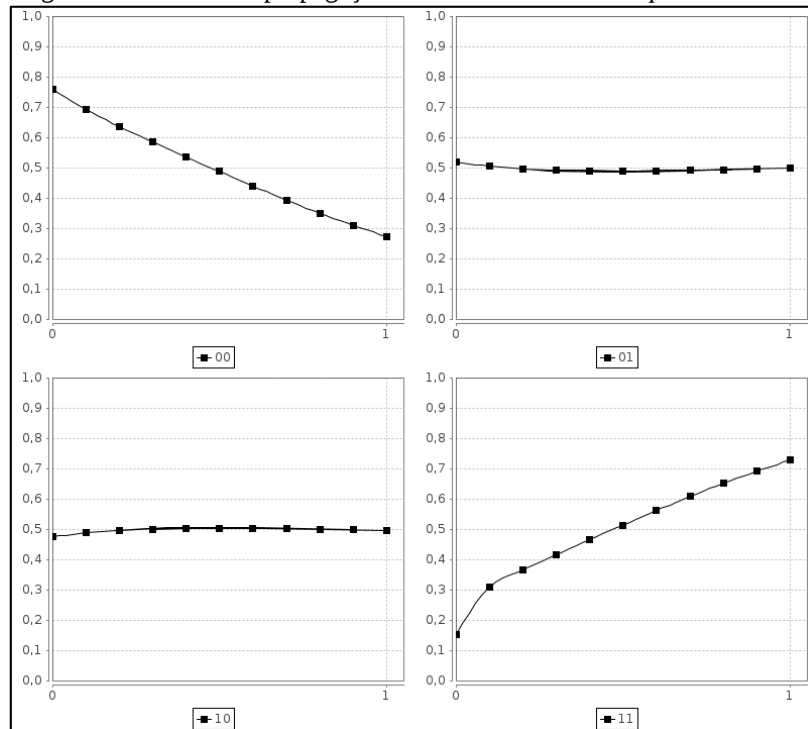


Fonte: Do autor.

Observando os valores do Ator 4, percebe-se que ele têm um comportamento

parecido com o do Ator 1, porém com uma pequena diferença entre a curva de propagação, levando à conclusão que o Ator 1 é o “*TERTIUS GAUDENS*” da rede.

Figura 31 - Gráfico de propagação do Grafo 3 iniciado no quarto ator.



Fonte: Do autor.

Em uma visão mais ampla, parecida com uma RSV real, foi proposto o grafo representado pela Figura 17 em que se pode se observar todas as características levantadas anteriormente, com o apoio das teorias levantadas no capítulo 2.4 em que foram desvendadas as teorias dos laços fortes e fracos de Granovetter (1973), Buracos Estruturais de Burt (1992), Grupalidade discutida por Lemieux e Ouimet (2004) e sobre o Capital Social levantada pela Recuero (2009). Os dados obtidos por meio das médias de propagação de cada ator no grafo estão representados em ordem decrescente pela Tabela 3.

Tabela 3 - Tabela de valores médios de propagação do grafo representado pela figura 17.

Posição	Ator	Valor	Posição	Ator	Valor
1	45	0,4137746510639890	24	6	0,3193527003278820
2	32	0,4049852668682750	25	3	0,3184197199008640
3	33	0,3907886212794540	26	4	0,3181165515141020
4	21	0,3683538938815530	27	35	0,3176456748721130
5	8	0,3626067167871740	28	37	0,3148336580025150
6	43	0,3488993328446070	29	13	0,3127165022505810
7	42	0,3485488511038640	30	24	0,3125386665688710
8	28	0,3477912268857850	31	40	0,3121598952135880
9	38	0,3477789660006340	32	10	0,3112681700884990
10	39	0,3468144102009440	33	16	0,3062188451894900
11	29	0,3439606932525870	34	22	0,3055503500659840
12	19	0,3366892974773660	35	18	0,3046918823159830
13	7	0,3347877904736870	36	30	0,3039722843490440
14	11	0,3333225778635900	37	5	0,3008681907577490
15	1	0,3328455035171820	38	31	0,2999101389537000
16	25	0,3301008163243890	39	14	0,2911679678195860
17	26	0,3283418172921760	40	17	0,2886346941926520
18	44	0,3263112895264270	41	9	0,2882931740860430
19	12	0,3238090652015380	42	36	0,2850342004557160
20	34	0,3220124398813990	43	15	0,2787736869980950
21	41	0,3216788860288010	44	23	0,2780512615812120
22	2	0,3211801698065680	45	20	0,2748421521382990
23	27	0,3197898579822370			

Fonte: Do autor.

Tendo uma visão de acordo com a teoria dos Buracos Estruturais de Burt, é observado um buraco estrutural mantido pelo Ator 1 entre a $f(Ator2, \dots, Ator8)$, $f(Ator9, \dots, Ator29)$ e $f(Ator30, \dots, Ator45)$, notou-se que ele não teve um melhor ponto de propagação, obtendo o décimo quinto lugar na tabela. Porém devido a o seu importante papel no grafo de unir os três sub-grafos citados, tornou-se assim um grafo **conexo**, ele pode ser considerado de grande importância para a devida propagação. Criando uma competição de propagação nos três sub-grafos mantidos pelo “TERTIUS GAUDESIN”, os Atores 2, 9 e 30 não necessariamente tiveram uma melhor propagação que os seus concorrentes.

Na retirada do Ator 1 definido como o “TERTIUS GAUDENS” os valores estão representados pela Tabela 4, foi percebido um aumento nas posições de propagações dos Atores 2, 9 e 30 que no grafo anterior, eles obtiveram as posições 22, 41 e 36

respectivamente. No segundo grafo apresentado, eles se localizaram nas posições 6, 34 e 18. Este teste prova novamente a teoria de Burt em que ele explica que os atores conectados ao TERTIUS GAUDES não têm necessariamente uma vantagem sobre dos demais atores em seu sub-grafo.

Tabela 4 - Tabela de valores médios de propagação do grafo representado pela figura 17 com a ausência do ator 1.

Posição	Ator	Valor	Posição	Ator	Valor
1	45	0,4165957403761250	23	41	0,3265031906065210
2	32	0,4097270374237270	24	34	0,3255528230911160
3	33	0,3935788654805580	25	12	0,3249899329345170
4	8	0,3751595510832360	26	27	0,3216758484450620
5	21	0,3697618166419300	27	35	0,3214716369905660
6	2	0,3634646348422470	28	37	0,3187230644869790
7	43	0,3555664426281350	29	40	0,3160716595444540
8	42	0,3530101933545430	30	10	0,3149070197977580
9	38	0,3514084204018180	31	5	0,3147057654725350
10	39	0,3509603471064330	32	13	0,3139114487676690
11	28	0,3495108124790940	33	24	0,3137644988495130
12	7	0,3481106891183440	34	9	0,3103720266610090
13	29	0,3453603220072980	35	16	0,3086155695158350
14	19	0,3378742410079040	36	22	0,3074221185915500
15	11	0,3357877116290420	37	31	0,3067384536772730
16	3	0,3352276453500290	38	18	0,3060588569045490
17	6	0,3334373179767400	39	14	0,2924631260321610
18	30	0,3334002058089560	40	17	0,2899589700600300
19	4	0,3328004035582470	41	36	0,2889366633442950
20	25	0,3313664992959710	42	15	0,2800710912070430
21	44	0,3304219017645360	43	23	0,2798029502849850
22	26	0,3297492585019720	44	20	0,2762903212600800

Fonte: Do autor.

No confronto das posições dos atores do sub-grafo pertencente ao Ator 2 nos dois grafos aplicados, percebe-se que em um ranking individual dos atores, o ator 2 era o que mantinha a terceira posição levando os seus vizinhos (Atores 6 e 3) na quarta e quinta posição, no grafo sem a ligação do Ator 1. Somente o Ator 2 mudou a sua posição, ficando em segundo lugar, sendo explicado novamente pela teoria de Burt.

Na alteração do sub-grafo proposta pelo Figura 18, sobre o ponto de vista do “TERTIUS GAUDENS”, foram realizados os mesmos testes aplicados anteriormente. Os valores de propagação dos atores são representados pela Tabela 5 e Tabela 6.

Tabela 5- Tabela de valores médios de propagação do grafo representado pela figura 18.

Posição	Ator	Valor	Posição	Ator	Valor
1	45	0,4137257069644160	24	7	0,3144027839506370
2	32	0,4046623018879600	25	13	0,3127171524350360
3	33	0,3906940146707940	26	24	0,3126235011508520
4	21	0,3683914274115000	27	40	0,3120562236897760
5	42	0,3485165560646270	28	10	0,3109827172902840
6	43	0,3484917515023500	29	6	0,3100136336176240
7	28	0,3477565453011690	30	3	0,3088695894206500
8	38	0,3476737694881480	31	16	0,3060706222397800
9	39	0,3466165480690520	32	22	0,3055260434244350
10	29	0,3439468270869240	33	18	0,3046922730012980
11	8	0,3417583885357160	34	30	0,3020239401324630
12	19	0,3366934768249390	35	31	0,2994472331154660
13	11	0,3331248752559560	36	4	0,2931876708967050
14	1	0,3306467572199470	37	14	0,2911629404924680
15	25	0,3301290852388500	38	17	0,2886418917330320
16	26	0,3284024781739420	39	9	0,2869721427336490
17	44	0,3261362726884620	40	36	0,2848955866098860
18	12	0,3238104054727530	41	15	0,2787732967541580
19	34	0,3219302402535310	42	23	0,2779474814476080
20	41	0,3214274917377590	43	5	0,2772749323769400
21	27	0,3198007819642230	44	20	0,2748233546146280
22	35	0,3174851493988890	45	2	0,2732480567117780
23	37	0,3146875938444920			

Fonte: Do autor.

Tabela 6 - Tabela de valores médios de propagação do grafo representado pela figura 18 com a ausência do ator 1.

Posição	Ator	Valor	Posição	Ator	Valor
1	45	0,4165957403759720	23	35	0,3214716369896280
2	32	0,4097270374237270	24	6	0,3196776719587730
3	33	0,3935788654802450	25	3	0,3195576760683030
4	21	0,3697649180039190	26	37	0,3187230644852860
5	43	0,3555664426281360	27	40	0,3160716595425380
6	42	0,3530101933545430	28	10	0,3149057791441940
7	38	0,3514084203988670	29	13	0,3139114285954280
8	39	0,3509603471064330	30	24	0,3138278316053140
9	8	0,3509392153125670	31	9	0,3103720279181020
10	28	0,3495016817902030	32	16	0,3086169310691930
11	29	0,3453603211299430	33	22	0,3074440991013150
12	19	0,3378763554642330	34	31	0,3067384536772730
13	11	0,3357803487034710	35	18	0,3060530967199500
14	30	0,3334002058089560	36	4	0,3034124918157730
15	25	0,3313733030628560	37	2	0,2987936089306590
16	44	0,3304219017661340	38	14	0,2924557089432680
17	26	0,3297418333333380	39	17	0,2899669829868670
18	41	0,3265031906138420	40	36	0,2889366633442950
19	34	0,3255528230905530	41	5	0,2872479696682590
20	12	0,3249899361085000	42	15	0,2800400892947080
21	7	0,3237613004261550	43	23	0,2798735375910650
22	27	0,3216674576140830	44	20	0,2762692923762560

Fonte: Do autor

Confrontando os resultados obtidos com os do grafo anterior, não houve mudanças significativas em relação aos atores dos sub-grafos dos Atores 9 e 30. O Ator 1 subiu somente uma posição, não possuindo nenhum efeito considerável.

Portanto se analisar a sub-rede a qual o Ator 2 pertence, os dados obtidos tanto do grafo que possuir um ator unindo o buraco estrutural quanto o que forma um grafo não conexo, os atores não tiveram alteração no ranking de propagação relevante, tendo o mesmo comportamento dos testes anteriores.

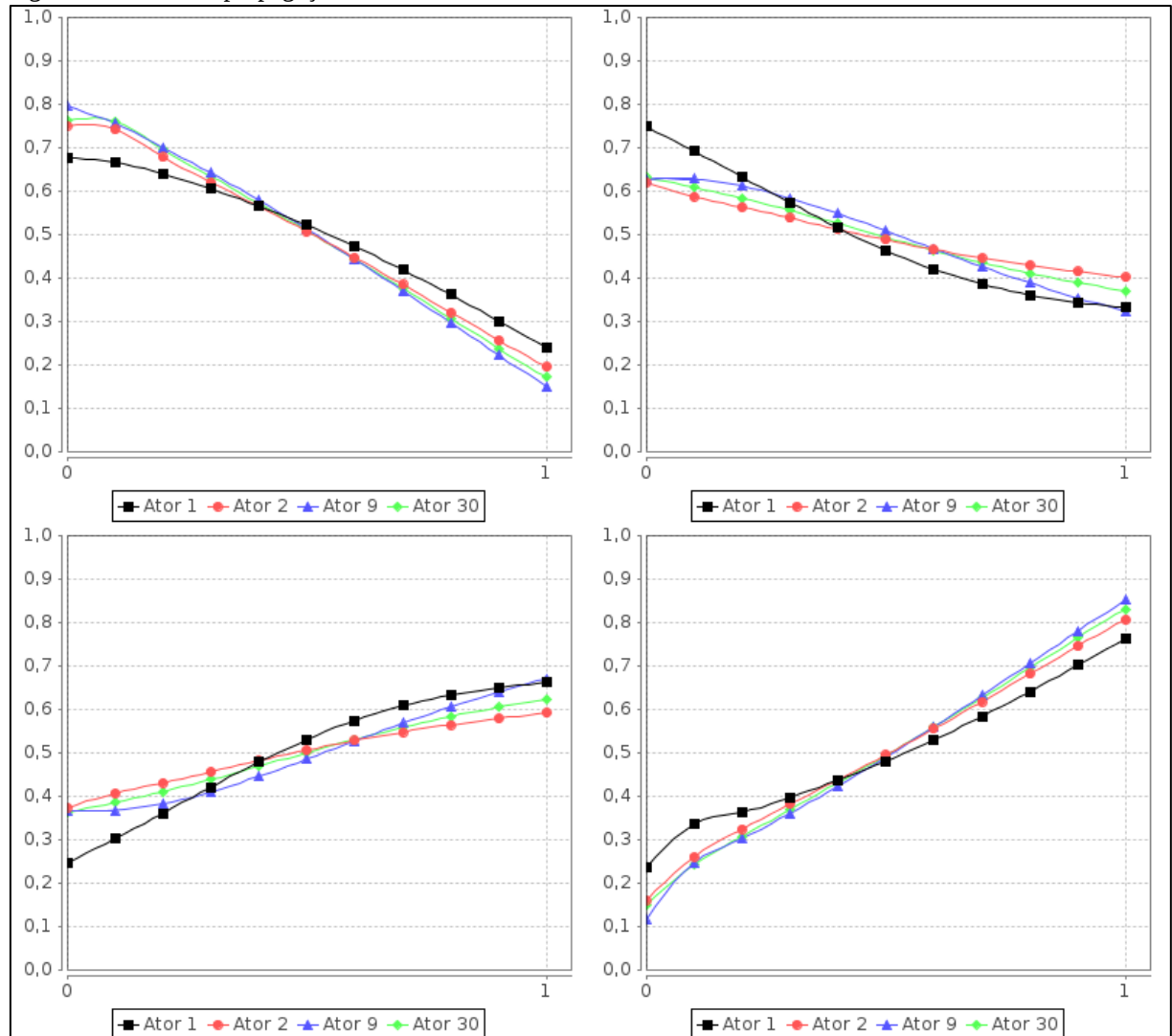
Comparando os resultados dos grafos representados pelas figuras 17 e 18, notou-se que a relação da densidade e o valor de propagação não são proporcionais, tem-se esta afirmação supostamente pelo algoritmo SP funcionar sobre a marginalização de uma rede, lembrando que para a adequada aplicação do algoritmo em um GS a retirada de relações que possam ocasionar ciclos de mensagens é feita pela transformação do grafo em uma estrutura em árvore, pelo algoritmo citado anteriormente denominado Junction Tree. Deixando assim

uma questão a se analisar em futuros trabalhos, o qual se propõe comparar a profundidade da árvore formada com o valor de propagação calculado pelo algoritmo SP.

Se observar os valores dos Atores 45, 33 e 32 em ambas as tabelas, não houve mudança de posição no ranking, tendo somente uma leve mudança nos valores de propagação supostamente devido à alteração do valor de propagação do Ator 30, o qual fazia a função de ponte para as mensagens trafegarem de dentro para fora do grafo quando o mesmo era conexo.

Por meio de uma análise mais profunda, igualmente feita nos gráficos anteriores, a Figura 32 demonstra as curvas de propagação do Ator 1 e de seus vizinhos, os atores 2, 9 e 30. É possível observar que quando o tipo de ligação entre os atores é a de conexão, o “TERTIUS GAUDENS” não obteve uma curva de crescimento tão acentuada quanto os demais, obtendo uma faixa de valor de propagação menor que o restante dos atores. Após o cruzamento dos valores resultantes, ocorrido perto do valor médio de propagação, é claramente observada a influência que os vizinhos obtiveram com a ligação feita através do Ator 1.

Figura 32 - Grafo de propagação do Ator 1 e seus vizinhos.



Fonte: Do autor.

A teoria dos Laços Fortes de Granovetter (1973), pela qual são defendidos dois tipos de laços (fracos e fortes); pode ser aplicados ao estudo em questão, por meio dos valores da função custo, representados por no-função na representação da rede social em Grafo Fatores, pelo qual se entende que o valor resultante da função custo está ligado proporcionalmente a definição dos laços sociais. Tendo uma ligação direta com o Capital Social defendida por Recuero (2009). Essa teoria pode ser aplicada também com a função custo, que ao contrario do estudo feito nesse trabalho, aplicou duas hipóteses de relacionamento entre os atores (negativo e positivo), um número não definido pode ser aplicado, em que cada um irá representar um segmento dentro do capital social.

Algumas dificuldades encontradas na aplicação do algoritmo foram à definição dos relacionamentos, direção e os valores da função custo. Não foi encontrada nenhuma

bibliografia clara em relação a este assunto, então foi proposta uma definição enquanto a direção das mensagens por meio do estudo da implementação do algoritmo SP aplicado junto a GF no *framework* Mallet. Para conseguir o resultado final do trabalho por meio do confronto dos resultados obtidos do algoritmo junto às teorias de AERS, um número exorbitante de resultados teve que ser gerado, fator esse já explicado anteriormente. Esse problema foi contornado com a análise dos resultados, em que resultados foram mostrados em forma de gráficos, minerando os dados obtidos por meio de medias de cada resultado de propagação, separados em relação a cada possível interpretação de resultado das passagens de mensagens entre os atores, gerando quatro gráficos de propagação para cada ator analisado.

Os resultados obtidos foram positivos de acordo com as teorias levantadas pela sociologia abordada neste trabalho. A aplicação do *framework* teve algumas dificuldades pela falta de conhecimento do assunto, não supridas totalmente pelo levantamento bibliográfico, o qual foi de difícil acesso ao conteúdo, devido ao déficit de conteúdo. Em contra partida fora do país essa área possui um melhor alcance onde foi obtido grande parte das bibliografias.

6 CONCLUSÃO

As RSV provocaram uma profunda mudança na forma de relacionamento entre as pessoas e alteraram o fluxo e a forma de propagar a informação na sociedade. As RS agrupam pessoas por suas afinidades e as informações são difundidas a partir desses agrupamentos. A compreensão das possibilidades nas composições das RS foi apresentada neste trabalho, a partir da ideia de grafos (nós-pessoas, conexões-relacionamentos) e a difusão da informação nesse meio foi embasada na teoria da Análise Estrutura de Redes Sociais da sociologia.

A busca de uma possível solução na identificação de pontos de propagação de informação em RS motivou este trabalho. O principal objetivo foi a avaliação do uso do algoritmo SP na identificação de pontos a partir da confrontação dos resultados obtidos na sua aplicação e as proposições com relação à propagação de informações defendidas pelas teorias de Análise Estrutura de Redes Sociais. Constatou-se que as proposições teóricas do estudo de caso são confirmadas. Como os resultados apresentados e discutidos na pesquisa evidenciando que a conectividade influencia fortemente na propagação da informação.

Constata-se que os objetivos específicos são satisfatoriamente respondidos pela pesquisa. A partir do levantamento das técnicas utilizadas na sociologia em Análise Estrutura de Redes Sociais, foi possível descrevê-las e identificar elementos chaves para a compreensão da propagação de informação na sociedade em diversos meios. Levando a uma compreensão paralela às RSV, com suas devidas alterações propostas no decorrer da fundamentação.

O algoritmo Soma e Produto, aplicado juntamente com Grafo Fatores, obteve um bom resultado para a compressão das passagens de mensagens entre os atores, facilitando o entendimento da aplicação do algoritmo para resolver o problema proposto neste trabalho, na identificação dos pontos melhores de propagação.

Confrontando as teorias sociológicas com os dados obtidos por meio da aplicação do algoritmo, houve resultados positivos e convergentes com a sociologia, a partir dos grafos propostos que simulavam as RS e continham elementos compositivos propostos pelas mesmas.

6.1 TRABALHOS FUTUROS

Como sugestão para trabalhos futuros pode-se citar o levantamento do comportamento, estrutura e construção da matriz de probabilidade condicional em relação aos

dados reais, existentes em laços sociais reais na implementação do algoritmo. Com dados extraídos por meio de técnicas de mineração de dados em RSV, extraindo dados como: troca de mensagens, relações em comum e todas as outras comparações levantadas pela sociologia já explicadas anteriormente, possibilitará a obtenção de uma matriz que irá se aproximar ao comportamento real, formará o caminho por onde as mensagens passarão no grafo fator. O trabalho confirmará a aplicação do algoritmo em casos reais, abrindo um leque de áreas em que o resultado auxiliará em tomadas de decisões.

REFERÊNCIAS

- ABRANTES, Sílvio A.. **Descodificação iterativa de códigos LDPC por transferência de mensagens em grafos de factores**. 2005. 34 f. Argito (1) - Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, 2005.
- ADAR, Eytan; ADAMIC, Lada A. 2005. **Tracking information epidemics in blogspace**. In: Web Intelligence, 4, Compiegne, 2005. Anais... Compiegne. Acesso em 04/11/2012, disponível em: <<http://www.cond.org/blogs2.html>>.
- ALEXA, The Web Information Company. **Facebook.com**. Disponível em: <<http://www.alexa.com/siteinfo/facebook.com>>. Acesso em: 04 nov. 2012.
- _____, The Web Information Company. **Twitter.com**. Disponível em: <<http://www.alexa.com/siteinfo/twitter.com>>. Acesso em: 04 nov. 2012.
- ANDRADE, Alexandre de. **Grafos-factores e Decodificação Iterativa: Novas Aplicações**. 2010. 131 f. Tese (Doutorado) - Curso de Engenharia Elétrica e de Computação, Unicamp, Campinas - Sp, 2010.
- BARABASI, Albert-laszlo; WATTS, J. Duncan ; NEWMAN M. E. J. **The structure and dynamics of networks**. Princeton, Nj: Princeton University Press, 2006. 588 p.
- BARBER, David. **Bayesian Reasoning and Machine Learning**. New York: Cambridge University Press, 2012.
- BURT, Ronald S.. **Structural Holes: The Social Structure of Competition**. London: Harvard University Press, 1992.
- BERNERS-LEE (2007). **Giant Global Graph**. Disponível em: <<http://dig.csail.mit.edu/breadcrumbs/node/215>>. Acesso em: 24 jun. 2012.
- BOAVENTURA NETTO, P.O. **Grafos : teoria, modelos, algoritmos**. 3. ed. rev. e ampl. São Paulo: Edgard Blücher, 2003. 314 p.
- CARRER, Davi. **OS IMPACTOS DAS REDES SOCIAIS NA COMUNICAÇÃO ORGANIZACIONAL: uma leitura do relacionamento de uma Instituição de Ensino Superior com seus alunos**. 2012. Monografia (Especialista em Comunicação Estratégica Nas Organizações.) - Departamento de Pós-graduação em Comunicação Estratégica Nas Organizações, Pontifícia Universidade Católica De Minas Gerais, Belo Horizonte, 2012.
- CRUZ, Ruleandson do Carmo. **REDES SOCIAIS VIRTUAIS DE INFORMAÇÃO SOBRE AMOR: COMPORTAMENTO E CULTURA INFORMACIONAL DE USUÁRIOS DO ORKUT**. 2011. 320 f. Dissertação (Mestrado) - Ufmg, Belo Horizonte, 2011.
- DIAS, Lucas De Lima. **SOCIODRAMA E O TESTE SOCIOMÉTRICO: UMA FERRAMENTA PARA GRUPOS SOCIAIS**. 2012. 63 f. Trabalho de Conclusão de Curso

(Psicólogo) - Curso de Psicologia, Universidade do Extremo Sul Catarinense, Criciúma, 2012.

DURANT-LAW, Graham (2007). **Who is Tertius Gaudens?** Disponível em: <<http://www.durantlaw.info/who-tertius-gaudens>>. Acesso em: 24 jun. 2012.

FREEMAN, L, C., 1977, A Set of Measures of Centrality Based upon Betweenness, *Sociometry*, 40:35-41

GIUSTI, Tiago de Faveri. **E-CRM E REDES SOCIAIS: UMA INTERAÇÃO POSSÍVEL**. 2010. 100 f. Bacharel - Curso de Ciência da Computação, Unesc, Criciúma, 2010.

GRANOVETTER, Mark, 1973. "**The Strength of Weak Ties.**" *American Journal of Sociology*, 78 (May): 1360-1380.

JORDAN, Michael Irwin; DIVISION, North Atlantic Treaty Organization. Scientific Affairs. **Learning in Graphical Models**. Kluwer Academic: Kluwer Academic, 1998. (NATO ASI series: Behavioural and social sciences).

KLEINBERG, J. 2007. **Cascading behavior in networks: Algorithmic and economic issues**. In: N. NISAN; T. ROUGHGARDEN; E. TARDOS; V. VAZIRANI (eds.), *Algorithmic Game Theory*. New York, Cambridge University Press. Acesso em: 30 out. 2012 , disponível em: <<http://www.cs.cornell.edu/home/kleinber/agtbook-ch24.pdf>>.

KOCK, Volker Maximillian, **A Factor Graph Approach to Model-Based Signal Separation**, ETH Zurich, 2007

KSCHISCHANG, Frank R.; FREY, Brendan J. ; LOELIGER, Hans-Andrea , **Factor Graphs and the Sum-Product Algorithm**, IEEE TRANSACTIONS ON INFORMATION THEORY, 2001

LEBLANC, Jonathan. **Programing Social Application**. United States Of America: O'reilly Media, 2011.

LEMIEUX, Vicent; OUIMET, Mathieu. **Análise Estrutural das Redes Sociais**. Lisboa: Instituto Piaget, 2004. 130 TRADUCAO DE UM LIVRO EM FRANCES

LOELIGER, Hans-Andrea, **An Introduction to Factor Graphs**, IEEE SIGNAL PROCESSING MAGAZINE, 2004.

NEWMAN, Mark; BARABÁSI, Albert-lászló; WATTS, Duncan J.. **The Structure and Dynamics of Networks**. New Jersey: Princeton University Press, 2006.

MALLET. **Machine Learning for Language Toolkit**. Disponível em: <<http://mallet.cs.umass.edu/>>. Acesso em: 30 out. 2012.

PEARL, Judea. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**. Los Angeles: Morgan Kaufmann, 1988.

RECUERO, Raquel . **Redes Sociais na Internet: Considerações Iniciais**. E Compós, v. 2,

2005.

_____, Raquel. **Redes Sociais na Internet**. Porto Alegre - Rs: Meridional, 2009.

_____, Raquel; ZAGO, Gabriela. “RT, por favor”¹: considerações sobre a difusão de informações no Twitter. **Revista Fronteiras: Estudos Midiáticos**, São Leopoldo, n. , p.69-81, 01 ago. 2010.

RUSSELL, Matthew A. **Mineração de Dados na Web Social**. São Paulo - Sp: Novatec, 2011
 TRADUCAO DE UM LIVRO EM INGLES

TANG, Jie; SUN, Jimeng; YANG, Chi Wang And Zi. **Social Influence Analysis in Large-scale Networks**. Tsinghua University, China, p.807-816, 2009.

TOMAÉL, Maria Inês; ALCARÁ, Adriana Rosecler; CHIARA, Ivone Guerreiro Di. Das redes sociais à inovação.Ci. Inf., Brasília, p.92-104, maio/ago 2005.

APÊNDICE(S)

APÊNDICE A – Artigo

Análise de pontos de Propagação de Informação em Redes Sociais com o uso do algoritmo Soma e Produto

João Miguel. O. F¹

¹Departamento de Ciências da Computação – Universidade do Estrumo Sul Catarinense (UNESC)

Criciúma – SC – Brazil

l3ug1m@gmail.com

Abstract. *With the advances in communications, specially the internet, the thirst for information by the human society is increasing by websites like Facebook and Twitter. The study of information propagation in virtual social networks involves sociological, mathematical and computational questions. This paper has the objective to evaluate the application of the sum product algorithm in Graph Factor on the identification of information propagation points. the results were satisfactory, concluding that the probabilistic data generated by the application can support in the identification of propagation points and serve as support in decision making.*

Resumo. *Com o avanço dos meios de comunicação, em especial da internet, a ambição de informações da sociedade humana tem tido um aumento considerável por meio de sites como Facebook e o Twitter. O estudo da propagação das informações em redes sociais virtuais envolve questões sociológicas, matemáticas e computacionais. Este trabalho teve como objetivo avaliar a aplicação do algoritmo de Soma e Produto em Grafo Fator na identificação de pontos propagação de informações. Os resultados foram satisfatórios, concluindo que os dados probabilísticos gerados pela aplicação podem dar suporte na identificação de pontos de propagação e servir de apoio à tomada de decisões.*

1. Introdução

A propagação da informação em redes sociais virtuais é notória podendo ser observada pela observação das mensagens transmitidas na timeline dos usuários do Facebook ou Twitter. Apesar de se reconhecer a importância das redes sociais na difusão da informação, o comportamento da propagação e a função dos atores da rede e seus agrupamentos na difusão ainda é pouco estudado. Como se formam estes agrupamentos, como se estruturam, como se conectam? Como ocorre a propagação de informação e os pontos de maior probabilidade de ocorrência? Há relação entre o tipo de conexão e a difusão da informação? A qualidade do laço social na conexão influencia no grau de propagação? Estes são alguns dos questionamentos ainda pouco abordados em pesquisas e estudos científicos. Porém, são pontos importantes que começam a ser discutido por pesquisadores, como Recuero (2005), Cruz (2011). A compreensão desses processos auxilia a entender como as redes sociais podem propagar informações e até mesmo boatos, como também possibilita planejamento de ações para influenciar esses processos.

Este trabalho tem como tema principal o estudo da propagação das informações em redes sociais, levantando questões sociológicas e matemáticas, assim formando uma base de conhecimento de apoio para o trabalho. A construção e estruturação da rede social tem

subsídio na Teoria dos Grafos (TG) utilizada para representação da rede e suas conexões. O entendimento do comportamento e características de atores, conectividade, laços sociais e difusão da informação é realizado a partir das teorias da área de Análise Estrutural das Redes Sociais (AERS) abordando questões como comportamento, transmissão de informações e relacionamentos na sociedade.

O foco principal da investigação em AERS é dar subsídio na validação da influência da conectividade da rede na difusão de informações. Partindo do princípio que quando mais conectada é a rede, quanto mais visíveis são as possibilidades de interação entre seus atores, maiores as chances de que uma informação seja propagada.

Com aplicação do algoritmo Soma e Produto junto à definição das redes por meio do Grafo Fator se obtém uma possível correlação dos elementos compositivos de uma rede social com o levantamento dos processos para o estudo da passagem de mensagem.

Com a aplicação do algoritmo SP em grafos busca-se uma possível solução na identificação de pontos de propagação de informação em RS. Os grafos que representam redes sociais fictícias são formados por nós (pessoas e seus perfis nas RS), ligações entre os nós (relações entre os participantes da rede e o grau de ligação). A avaliação do uso do algoritmo para o objetivo proposto se dá a partir da confrontação dos resultados obtidos na sua aplicação e as proposições com relação à propagação das informações defendidas pelas teorias de Análise Estrutura de Redes Sociais.

2. Redes Sociais

A tecnologia está presente e tem um papel fundamental no trajeto percorrido das mudanças na vida social atualmente, percebendo assim que as redes sociais virtuais provocaram uma profunda mudança na forma de relacionamento. Recuero (2009) define Redes Sociais Virtuais (RSV) “como agrupamentos complexos instituídos por interações sociais apoiadas em tecnologias digitais de comunicação”, o qual disponibiliza um vasto meio de fornecimento de informações do ciberespaço. Esta definição ajuda a entender “a criação das estruturas sociais; suas dinâmicas, tais como a criação de capital social e sua manutenção, a emergência da cooperação e da competição; as funções das estruturas e, mesmo, as diferenças entre os variados grupos e seu impacto nos indivíduos.”.

As RSV passaram de novidades a fenômeno global nos últimos anos. No primeiro bimestre de 2010, o Facebook ultrapassou a Google em número de visitas as suas páginas. De uma maneira indefensável, este dado indica que a internet se tornou um meio onde a convivência social, com a ajuda de ferramentas, se tornou mais utilizada comparado ao uso de ferramentas de pesquisa. Isto indicando, uma escala considerável que sites de RSV estão satisfazendo, alguns anseios humanos muito básicos, mudando bruscamente a forma como vimos nossas vidas sociais dentro e fora da internet (RUSSEL, 2011).

Leblanc (2011) define que atualmente as RSV contém três modelos de relacionamentos (seguidor, conexão e grupo), conteúdo individualmente propriedades de alcance e complexidades diferentes, os quais estão presentes nas RSV em estudo. O modelo seguidor utilizado pela RSV Twitter, é definida onde uma ator “segue”, tem uma relação com outro, e que não necessariamente a relação oposto tem que ser correspondida. Nos modelos de conexão e grupo é a RSV Facebook que utiliza-os, sendo observado o modelo de conexão nas conexões de amizade e no de grupo por ferramentas como Grupos de Amigos e Eventos. Assim as duas RSV citadas, abrangem todos os tipos de relacionamentos definidos por Leblanc, tornando o estudo completo com a visão defendida por ele em cima dos tipos de relações.

Em ciências humanas a explicação dos fenômenos sociais se dá através do estudo das relações entre os atores, sendo individuais ou coletivamente em redes publicas. A análise casual tenta estabelecer as determinantes pontuando a importância de cada uma, já a análise descritiva tenta mostrar a divisão entre as etapas da evolução. Sendo que a análise feita em ações costuma ser mais racional por ter um foco em políticas públicas levando diretamente nos interesses dos atores (LEMIEUX; OUIMET, 2004).

A chamada Análise Estrutural das Redes Sociais é normalmente usada para explicar a forma das relações entre os atores sociais, que tentam explicar as ligações entre os atores no âmbito do estruturalismo, linguístico e antropológico, dentre outras maneiras. Estes estudos podem ser divididos basicamente em dois níveis; o primeiro sendo aplicado na análise sobre uma visão de regularidade e organização dando uma explicação descritiva e explicativa sobre o processo, já a de segundo nível trata de uma análise estrutural, formando na maioria das vezes um resultado vindo de uma análise totalmente empírica (LEMIEUX; OUIMET, 2004).

Em ciências humanas as análises correlacionais são feitas através do comportamento e propriedades dos atores em um grafo, onde é definida por Aguiar (2007, apud Cruz (2011)) como: dados relacionais, dados de atributos. Para a compreensão de uma rede tem que se definir os elementos compositivos para assim ter início a caracterização de suas propriedades. Estes conceitos fundamentais de uma RS são definidos por Wasserman e Faust (1994, apud CRUZ, 2011) por meio de sete (7) elementos compositivos: ator, laço, díade, tríade, sub-grupo, grupo, relação.

Os laços sociais definem que as conexões entre os atores, são formadas por meio das interações na sociedade. As conexões formam o principal foco de estudo em RS, percebendo que através delas que se apoiam a explicação de todo o estudo da estrutura de uma RS.

A orientação ou não das relações entre os nós, têm um papel importante para a compreensão das relações, formando um entendimento paralelo entre a sociologia e a TG. Nas relações orientadas, denominadas na TG como arcos ou arestas, permitem-se ter mais de um canal de comunicação entre os atores, podendo ser bidirecional ou unidirecional. Já o comportamento da transmissão das mensagens em relações não orientadas definidas como arcos ou arestas em TG, leva à irrelevância a direção da comunicação no relacionamento em comum entre os atores. A orientação ou não das mensagens em grafos sociais está ligada diretamente ao estudo da importância de um ator na rede em estudo (LEMIEUX; OUIMET 2004, BOAVENTURA 2003).

A análise de passagens de um ponto a outro em redes sociais é discutido em TG como caminho, designando uma sequência de relações orientadas; já em sociologia isto é chamada de cadeia. Tendo em vista um nó ou ator para o estudo, em sociologia o termo utilizado em grafos com raiz é dito com fonte, e o final do caminho dito em grafos como nó folha, em sociologia entendesse como alvo, e todos os atores que estão fazendo a ligação são chamados de intermediários. O ator dominante é que iniciou o processo nos casos de grafos onde tenha orientação; ao contrário de redes que não possuem orientação, onde todos os atores serão considerados dominantes (LEMIEUX; OUIMET, 2004).

Outras características que são aplicadas em redes sociais para o entendimento do comportamento são a de conexidade que é aplicada na possibilidade de passagem de um vértice a outra em um grafo através de ligações existentes, sendo caracterizadas como: a) não conexo; b) quase-fortemente-conexo; c) semi-fortemente-conexo; d) fortemente-conexo, a posição dos atores onde é aplicado a definição em grafos orientados onde é distinguido os atores como a) dominante; b) dominado; c) semidominado; d) subdominante; e)

subdominado; f) isolado, e em grafos que não tem orientação: a) centralidade de grau; b) centralidade de proximidade; c) centralidade de intermediariedade.

Essas análises correlacionadas servem para descrever relações entre os atores sociais. Teorias explicativas que usam esses dados tentam explicar como os atores se comportam na sociedade. Também podem ser aplicadas em GS, formando um conjunto de proposições de diferentes níveis de generalidade.

O Capital Social é um fator importante para o entendimento do comportamento dos atores em um grafo, no qual Cruz (2011) define como: “interesses partilhados pelos autores da rede social [...] normas, valores, instituições e relacionamento compartilhados que permitem a cooperação dentro ou entre os diferentes grupos sociais [...]”.

Na teoria dos laços fortes de Granovetter (1973), são definidos dois graus de ligações, os fortes ou estreitos que unem os parentes, amigos próximos e conjugues, e os laços fracos que unem parentes distantes ou apenas conhecidos. As redes de laços fortes têm, por conseguinte, muito mais tendência para se fecharem sobre si mesmas, do que as redes de laços fracos, as quais tende preferencialmente a abrir-se para o exterior.

A teoria de Burt (1992), conhecida como buracos estruturais, tem uma proposição geral no qual um ator que se beneficia com o fato de estabelecer contatos que não têm qualquer conexão entre si, denominada “TERTIUS GAUDENS”, se encontrando numa posição vantajosa, acima dos buracos estruturais. Nas situações de competição ou de conflito, os buracos estruturais no ambiente externo são vantajosos para um ator que enfrenta opositores divididos entre si. Já em um ambiente interno esta afirmativa é errônea uma vez que representam falhas na cooperação que internamente poderão ser exploradas pelos opositores.

Na teoria da grupabilidade discutida por Lemieux e Ouimet (2004), que teve origem nos trabalhos de um psicólogo, Heider 1946 e de um psicólogo social, Festinger 1957, mostraram que embora dois pares de elementos de uma tríade estejam ligados positivamente, consequentemente o par restante vai estar relacionado igualmente.

Essas teorias entre várias outras tentam explicar o comportamento dos atores em um grafo não através de suas propriedades como sexo, idade, profissão entre outras mais sim na posição ocupada no grafo e seus relacionamentos estabelecidos. No FB as distinções apresentadas podem ser vista com uma grande facilidade, através do mapeamento dos atores, feita pela configuração de parentesco dos amigos conectados e pela formação de grupos. No TW a distinção das mensagens de parentescos e amigos, se torna inviável por não possuir ferramentas que proporcionam essas distinções. As definições de relacionamentos são feita basicamente por relacionamentos formados através do modelo de seguidores definida por Leblanc (2011).

3. Algoritmo de Propagação

Problemas relevantes à estimação em ambientes de incerteza, como em inteligência artificial e processamento de sinais, podem ser reformulados matematicamente por uma função global ou traves de um problema de marginalização. Uma função global pode ser definida como uma “função custo” que é aplicada em um sistema para obter um valor de medida, como por exemplo: potência, energia, probabilidade e confiança. Resultando em um valor em relação as sua marginalização.

As origens do Grafo Fator residem na teoria da codificação, que oferece uma notação atraente para uma grande variedade de problemas de processamento de sinal. Em

particular, um grande número de algoritmos práticos para uma grande variedade de problemas de detecção e de estimativa pode ser derivada como algoritmos de propagação de resumo. Os algoritmos derivados desta forma muitas vezes incluem os melhores algoritmos conhecidos anteriormente como casos especiais ou como aproximações óbvias (LOELIGER, 2004).

A aplicação do GF em conjunto com o algoritmo de SP é amplamente usada para representação da passagem de mensagens até a marginalização de um determinado grafo, mesmo que a aplicação do algoritmo de SP não seja dependente do GF o uso do conjunto torna mais fácil o entendimento do problema aplicado, assim Frey e Loeliger (2001) defendem que a maneira em que GF representa a passagem de mensagens em uma rede, a utilização dela se torna um meio pedagógico para a aplicação.

Funções-fatores podem ser entendidas como funções custos associadas aos nós no grafo. O algoritmo SP é o processo sistematizado de obter funções custos simples associadas a cada ramo do Grafo Fator. Desta forma, cada nó recebe como mensagem de um dado ramo uma função custo simples, para dar continuidade aos cálculos seguintes. Segundo Andrade (2010), quando considerado um ramo e uma direção, é possível definir o fluxo local de mensagens relativo a eles, ou seja, as mensagens anteriores.

4. Propagação de informação: Avaliação do Algoritmo SP

Esta pesquisa buscou avaliar a aplicabilidade do uso do algoritmo SP para calcular pontos de propagação de informação em redes sociais. Para atingir o objetivo desta pesquisa foram efetuados: uma pesquisa bibliográfica referente à AERS, onde se identificou teorias e características que definem o comportamento das pessoas em RS e no entendimento do algoritmo SP para o cálculo da propagação das informações junto ao GF, o qual auxiliou no entendimento do relacionamento entre os atores.

Para a aplicação do algoritmo de Soma e Produto (SP) junto a um Grafo Fator (GF), foi utilizado a implementação do algoritmo contido do framework Mallet (2012), junto com a implementação do algoritmo Junction Tree.

Em geral as regras de propagação das mensagens do algoritmo SP são recursivas e localmente aplicadas em GF envolvendo somente as variáveis e funções vizinhas. Portanto a obtenção das marginais por meio de grafos com ciclos é somente obtida através de técnicas de ordem de propagação de mensagens, ou usando algum critério de parada, defendida por Andrade (2010).

Em redes sociais a formação de ciclos é muito comum, sendo notada através da obtenção de cliques, e a formação de grupos. Assim o surgimento de ciclos para o cálculo de propagação em uma RS torna-se inevitável. Na utilização do framework Mallet, a classe responsável para o cálculo das marginais (JunctionTreeInferencer) já é responsável para a solução desse problema.

No framework Mallet é definido o número de hipóteses de cada variável no construtor do objeto (ex: 'new Variable(2)'), sendo que o valor mínimo possível para estas hipóteses é 2, sendo compreendido como negação e afirmação da passagem das mensagens na RS. Portanto a matriz responsável pela função custo têm que ter no mínimo uma dimensão de 2x2. Aonde 0 corresponde à negação e 1 a afirmação. Cada ligação dirigida de um ator a outro é quantificada por uma matriz de probabilidade condicional exemplificada demonstrada pela Tabela 1, em que são demonstrados na terceira coluna um valores hipotéticos do custo das mensagens.

Para o cálculo da propagação de informação entre todos os atores, foram aplicados valores que variam de 0.0 a 1, gerado 14641 registros das possíveis hipóteses, por meio da combinação dos quatro valores possíveis da matriz, para cada um dos quatros possíveis resultados da matriz. Se tratando de um algoritmo de probabilidade não necessariamente a soma dessas possíveis hipóteses tem que ser igual a um, no caso cem por cento (100%), assim Pearl (1988) define que cada nó pode conter um conjunto de hipóteses mutualmente exclusivas, em que o resultado final sofrera mudanças pela normalização do resultado.

Segundo Recuero (2005) a unidade básica das redes sócias definidas pelos sociólogos eram as díades, relações entre duas pessoas seriam a menor estrutura relacional em uma sociedade. Com a aplicação do algoritmo percebe-se uma clara distinção na direção das mensagens, quando a mensagem de negação é iniciada pelo primeiro ator, nas duas primeiras linhas aonde o valor do ator é zero (0) na matriz, o resultado de propagação tende a cair quanto maior o valor resultado, o inverso acontece nas duas últimas linhas da matriz aonde o valor é um. Na observação da propagação do segundo ator isto também é claramente observado aonde a linha de propagação da primeira e ultima linha da matriz é mantida igual, invertendo somente a segunda e terceira linha, onde acontece a inversão do resultado de propagação.

Cruzando a propagação dos dois atores com seus respectivos valores na matriz, pode se dividir os resultados obtidos para distinguir o ambiente aonde eles vão se adequar corretamente. Avaliando pela divisão de tipos de conexão definidas por Leblanc (2011), quando os valores da matriz forem iguais, a conectividade é definida pelo modelo de Conexão (FB), e quando se divergem, pode ser definida pelo modelo de Seguidor (TW).

Um segundo grafo a ser estudado é constituído de uma tríade, onde a rede têm uma forma triangular, possuindo dois (2) atores que tenham relacionamento com um ator distinto. Por meio de uma análise em cima das características de orientação, densidade e conexidade, mesmo que o grafo apresente características claras de uma rede não direcionada e conexa, fazendo uma análise empírica, pode se obter resultados de outros sub-grafos, retirando e/ou adicionando direções a ligações dos atores. Esta técnica pode ser explanada, fazendo uma seleção prévia dos resultados a serem analisados em cima da matriz de probabilidade condicional. Se observar o grafo com conexões direcionadas podemos aplicar as classificações de não-conexo, obtendo uma díade a partir de um sub-grafos deixando um ator sem qualquer conexão com os outros atores podendo classificar como quase-fortemente-conexo, podendo observar também atores dominantes e dominados de acordo com a direção das conexões. A relação da densidade com o fator de propagação pode ser provada através das conexões orientadas ou não, tendo uma propagação proporcional com a da conexidade do grafo, onde as possíveis conexões variam de acordo com a conexidade da Tríade. Como resultado de destaque foi obtido com a observação da linha de propagação da Tríade, em que o pico alcançado foi $\text{marg}(\text{ator1}) = 0.7$ quando que o da Díade foi de $\text{marg}(\text{ator1}) = 0.6$. Quando a Díade teve um crescimento uniforme em razão aos valores da função de propagação a Tríade teve um crescimento rápido nos intervalos em que os valores variam entre 0 e 0.1.

Outro grafo um pouco mais complexo proposto por Lemieux e Ouimet (2004), compreende na relação entre quatro atores, contendo uma díade e uma tríade. Analisando os valores do primeiro ator, percebe-se que o valor médio máximo de propagação é maior que os demais, destacando que ele é o único ator na rede em que possui uma ligação direta com o resto dos atores. Pode se definir assim, que ele é um ator dominante em relação aos demais.

Os segundo e terceiro atores tiveram um resultado parecido devido o compartilhamento das mesmas conexões e ligações com atores em comum. Como resultado as variações se explicam possivelmente devido ao problema da aplicação da matriz pelo

framework explicada na avaliação da tríade. Observando os valores do quarto percebe-se que ele têm um comportamento parecido com o do Ator 1, porém com uma pequena diferença entre a curva de propagação, levando a conclusão que o Ator 1 é o “TERTIUS GAUDENS” da rede.

Um problema encontrado para o cálculo de propagação foi a definição da orientação em relação à matriz, o estudo feito em cima de uma díade, onde contém uma única aresta com sua devida matriz. O resultado foi coerente, porém quando aplicado em uma tríade o resultado não foi o esperado, se observar na Figura 24. Os valores da matriz referentes ao tipo de conexão seguidor (01 e 10) mantiveram uma media de propagação uniforme de acordo com a variação dos valores de coo eficiência de propagação da matriz.

E por último é proposto um grafo que contém quarenta e cinco atores, onde possui as várias características e teorias defendidas na sociologia. Em uma visão mais ampla, parecida com uma RSV real, foi proposto o grafo representado pela Figura 17 aonde pode se observar todas as características levantadas anteriormente, com o apoio das teorias levantadas no capítulo 2.4 em que foram levantadas as teorias dos laços fortes e fracos de Granovetter (1973), Buracos Estruturais de Burt (1992), Grupalidade discutida por Lemieux e Ouimet (2004) e sobre o Capital Social levantada pela Recuero (2009). Os dados obtidos por meio das medias de propagação de cada ator no grafo está representado em ordem decrescente pela Tabela 3.

Tendo uma visão de acordo com a teoria dos Buracos Estruturais de Burt, é observado um buraco estrutural mantido pelo Ator 1 entre a $f(\text{Ator}2, \dots, \text{Ator}8)$, $f(\text{Ator}9, \dots, \text{Ator}29)$ e $f(\text{Ator}30, \dots, \text{Ator}45)$, notou-se que ele não teve um melhor ponto de propagação, obtendo o décimo quinto lugar na tabela. Porém devido o seu importante papel no grafo que foi o de unir os três sub-grafos citados, tornando assim um grafo conexo, ele pode ser considerado de grande importância para a devida propagação. Criando uma competição de propagação nos três sub-grafos mantidos pelo “TERTIUS GAUDENS”, os Atores 2, 9 e 30 não necessariamente tiveram uma melhor propagação que os seus concorrentes.

Na retirada do Ator 1 definido como o “TERTIUS GAUDENS” os valores estão representados pela Tabela 4, foi percebido um aumento nas posições de propagações dos Atores 2, 9 e 30 onde no grafo anterior, eles obtiveram as posições 22, 41 e 36 respectivamente no segundo grafo apresentado eles se localizaram nas posições 6, 34 e 18. Este teste prova novamente a teoria de Burt onde ele explica que os atores conectados ao TERTIUS GAUDENS não têm necessariamente uma vantagem em cima dos demais atores em seu sub-grafo.

No confronto das posições dos atores do sub-grafo pertencente ao Ator 2 nos dois grafos aplicados, percebe-se que em um ranking individual dos atores, o ator 2 era o que mantinha a terceira posição levando os seus vizinhos (Atores 6 e 3) na quarta e quinta posição, no grafo sem a ligação do Ator 1. Somente o Ator 2 mudou a sua posição, ficando em segundo lugar, sendo explicado novamente pela teoria de Burt.

Confrontando os resultados obtidos com o do grafo anterior, não teve mudanças significantes em relação aos atores dos sub-grafos dos Atores 9 e 30. O Ator 1 subiu somente uma posição, não possuindo nenhum efeito considerável.

Portando se analisar a sub-rede em que o Ator 2 pertence, os dados obtidos tanto do grafo que possuir um ator unindo o buraco estrutural quanto o que forma um grafo não conexo, os atores não tiveram alteração no ranking de propagação relevante, tendo o mesmo comportamento dos testes anteriores.

Por meio de uma análise mais profunda, igualmente feita nos gráficos anteriores, é possível observar que quando o tipo de ligação entre os atores é a de conexão, o “TERTIUS GAUDENS” não obteve uma curva de crescimento tão acentuada quanto os demais, obtendo uma faixa de valor de propagação menor que o restante dos atores. Após o cruzamento dos valores resultantes, ocorrida perto do valor médio de propagação, é claramente observado a influência que os vizinhos obtiveram com a ligação feita através do Ator 1

A teoria dos Laços Fortes de Granovetter (1973), o qual é defendido dois tipos de laços (fracos e fortes); pode ser aplicados no estudo em questão, por meio dos valores da função custo, representados por no-função na representação da rede social em Grafo Fatores, no qual se entende que o valor resultante da função custo esta ligada proporcionalmente a definição dos laços sociais. Tendo uma ligação direta com o Capital Social defendida por Recuero (2009). Essa teoria pode ser aplicada também com a função custo, que ao contrario do estudo feito neste trabalho que se aplicou duas hipóteses de relacionamento entre os atores (negativo e positivo), um número não definido pode ser aplicado, onde cada um irá representar um segmento dentro do capital social.

Algumas dificuldades encontradas na aplicação do algoritmo foram à definição dos relacionamentos, direção e os valores da função custo. Não foi encontrada nenhuma bibliografia clara em relação a este assunto, onde foi proposta uma definição enquanto a direção das mensagens por meio do estudo da implementação do algoritmo SP aplicado junto a GF no framework Mallet. Para conseguir o resultado final do trabalho por meio do confronto dos resultados obtidos do algoritmo junto às teorias de AERS, um número exorbitante de resultados teve que ser gerando, fator esse já explicado anteriormente. Este problema teve que ser contornado fazendo análise dos resultados, onde se obteve uma alternativa de demonstrar os resultados em forma de gráficos, minerando os dados obtidos através de medias de cada resultado de propagação, separados em relação a cada possível interpretação de resultado das passagens de mensagens entre os atores, gerando quatro gráficos de propagação para cada ator analisado.

Os resultados obtidos foram positivos de acordo com as teorias levantadas pela sociologia abordadas neste trabalho. A aplicação do framework teve algumas dificuldades pela falta de conhecimento do assunto, não supridas totalmente pelo levantamento bibliográfico, o qual foi de difícil acesso ao conteúdo, por ter um déficit grande de conteúdo gerado pelo âmbito acadêmico nacional pelo fato do assunto ser muito novo. Em contra partida fora do país esta área possui um melhor alcance, onde foi obtido á maioria das bibliografias, e que o trabalho se fundamentou para conseguir os resultados obtidos.

5. Conclusão

As RSV provocaram uma profunda mudança na forma de relacionamento entre as pessoas e alteraram o fluxo e a forma de propagar a informação na sociedade. As RS agrupam pessoas por suas afinidades e as informações são difundidas a partir desses agrupamentos. A compreensão das possibilidades nas composições das RS se deu neste trabalho, a partir da ideia de grafos (nós-pessoas, conexões-relacionamentos) e a difusão da informação neste meio, foi embasada na teoria da Análise Estrutura de Redes Sociais, da sociologia.

A busca de uma possível solução na identificação de pontos de propagação de informação em RS motivou este trabalho. O principal objetivo foi à avaliação do uso do algoritmo SP na identificação de pontos a partir da confrontação dos resultados obtidos na sua aplicação e as proposições com relação à propagação de informações defendidas pelas teorias de Análise Estrutura de Redes Sociais. Ao que se constatou que as proposições teóricas do

estudo de caso são confirmadas. Como os resultados apresentados e discutidos na pesquisa evidenciando que a conectividade influencia fortemente na propagação da informação.

Constata-se que os objetivos específicos são satisfatoriamente respondidos pela pesquisa. A partir do levantamento das técnicas utilizadas na sociologia em Análise Estrutura de Redes Sociais, foi possível descrevê-las e identificar elementos chaves para a compreensão da propagação de informação na sociedade em diversos meios. Levando a uma compreensão paralela às RSV, com suas devidas alterações propostas no decorrer da fundamentação.

O algoritmo Soma e Produto, aplicado juntamente com Grafo Fatores, obteve um bom resultado para a compressão das passagens de mensagens entre os atores, facilitando o entendimento da aplicação do algoritmo para resolver o problema proposto neste trabalho, na identificação dos pontos melhores de propagação.

Confrontando as teorias sociológicas com os dados obtidos por meio da aplicação do algoritmo, obtiveram-se resultados positivos e convergentes com a sociologia, a partir dos grafos propostos que simulavam as RS e continham elementos compositivos propostos pelas mesmas.

References

- ANDRADE, Alexandre de. Grafos-fatores e Decodificação Iterativa: Novas Aplicações. 2010. 131 f. Tese (Doutorado) - Curso de Engenharia Elétrica e de Computação, Unicamp, Campinas - Sp, 2010.
- BURT, Ronald S.. Structural Holes: The Social Structure of Competition. London: Harvard University Press, 1992.
- BOAVENTURA NETTO, P.O. Grafos : teoria, modelos, algoritmos. 3. ed. rev. e ampl. São Paulo: Edgard Blücher, 2003. 314 p.
- CRUZ, Ruleandson do Carmo. REDES SOCIAIS VIRTUAIS DE INFORMAÇÃO SOBRE AMOR: COMPORTAMENTO E CULTURA INFORMACIONAL DE USUÁRIOS DO ORKUT. 2011. 320 f. Dissertação (Mestrado) - Ufmg, Belo Horizonte, 2011.
- GRANOVETTER, Mark, 1973. "[The Strength of Weak Ties](#)." American Journal of Sociology, 78 (May): 1360-1380.
- LEBLANC, Jonathan. Programing Social Application. United States Of America: O'reilly Media, 2011.
- LEMIEUX, Vicent; OUIMET, Mathieu. Análise Estrutural das Redes Sociais. Lisboa: Instituto Piaget, 2004. 130 TRADUCAO DE UM LIVRO EM FRANCES
- LOELIGER, Hans-Andrea, An Introduction to Factor Graphs, IEEE SIGNAL PROCESSING MAGAZINE, 2004.
- MALLET. Machine Learning for Language Toolkit. Disponível em: <<http://mallet.cs.umass.edu/>>. Acesso em: 30 out. 2012.
- PEARL, Judea. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Los Angeles: Morgan Kaufmann, 1988.
- RECUERO, Raquel . Redes Sociais na Internet: Considerações Iniciais. E Compós, v. 2, 2005.
- _____, Raquel. Redes Sociais na Internet. Porto Alegre - Rs: Meridional, 2009.

RUSSELL, Matthew A. Mineração de Dados na Web Social. São Paulo - Sp: Novatec, 2011
TRADUCAO DE UM LIVRO EM INGLES.

