



**UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**



JOSÉ AUGUSTO SARAIVA LUSTOSA FILHO

**UM FRAMEWORK BASEADO EM ONTOLOGIAS PARA
ANÁLISE DE REDES SOCIAIS ON-LINE APLICADO À
PROSPECÇÃO DE ATIVISMO DIGITAL NO *TWITTER***

MOSSORÓ – RN

2013

JOSÉ AUGUSTO SARAIVA LUSTOSA FILHO

**UM FRAMEWORK BASEADO EM ONTOLOGIAS PARA
ANÁLISE DE REDES SOCIAIS ON-LINE APLICADO À
PROSPECÇÃO DE ATIVISMO DIGITAL NO *TWITTER***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação – associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Carla Katarina de Monteiro Marques - UERN.

Co-orientador: Prof. Dr. Francisco Milton Mendes Neto - UFERSA.

MOSSORÓ – RN

2013

JOSÉ AUGUSTO SARAIVA LUSTOSA FILHO

**UM FRAMEWORK BASEADO EM ONTOLOGIAS PARA
ANÁLISE DE REDES SOCIAIS ON-LINE APLICADO À
PROSPECÇÃO DE ATIVISMO DIGITAL NO *TWITTER***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: ____ / ____ / ____.

BANCA EXAMINADORA

Prof. Dr. Carla Katarina de Monteiro Marques – UERN
Presidente



Prof. Dr. Francisco Milton Mendes Neto – UFERSA
Co-orientador

Prof. Dr. Rommel Wladimir de Lima – UERN
Membro Interno

Prof. Dr. Henrique Jorge Amorim Holanda – UERN
Membro Interno



Prof. Dr. Alisson Vasconcelos de Brito – UFPB
Membro Externo

Dedico este trabalho aos meus pais,
José Augusto e Gracilene, por todo
sacrifício que fizeram para que eu
tivesse uma boa educação.

AGRADECIMENTOS

Quero agradecer primeiramente a Deus, por ter me dado o dom da vida, saúde, capacidade e sabedoria para superar mais uma etapa em minha vida.

Aos meus pais, José Augusto Saraiva Lustosa Filho e Gracilene Barbosa Lima Lustosa, sem eles eu não estaria aqui hoje, aos meus primos e avós, por sempre acreditarem no meu empenho, darem confiança, paciência e apoio nos momentos difíceis durante esta jornada.

À minha orientadora Carla Katarina de Monteiro Marques e ao meu co-orientador Francisco Milton Mendes Neto, por acreditarem em mim e por todo seu apoio dado à orientação desse trabalho.

Aos professores da banca, pela disponibilidade em avaliar e contribuir com este trabalho e aos colegas e professores do mestrado, no qual pude aprender e crescer com eles.

À CAPES, pelo apoio financeiro concebido no qual viabilizou a realização deste trabalho. Agradeço a todos aqueles que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

RESUMO

Com o advento e expansão da Web, diversos serviços facilitam a busca de informações e interação entre pessoas ao redor do mundo. As redes sociais on-line obtiveram destaque nesse cenário, possibilitando a interação instantânea de milhões de usuários conectados à internet. Opiniões, manifestações, campanhas publicitárias e campanhas políticas também passaram a serem promovidas através das redes sociais on-line. Dessa forma a busca pela análise das atividades de usuários em redes sociais on-line ganharam bastante atenção nos últimos anos. Essa dissertação apresenta um framework baseado em ontologias para análise de redes sociais on-line, extraindo características relacionadas à atividades de usuários em redes sociais on-line. Para uma melhor abordagem na descrição do framework, uma análise experimental de um conjunto de interações de usuários do *Twitter* é feita.

Palavras-Chave: *Twitter*, ativismo digital, framework, redes sociais on-line, ontologia.

ABSTRACT

With the advent and expansion of the Web, many services have facilitated the search for information and interaction between people around the world. Online social networks have gained importance in this scenario, allowing interactions instantaneous of millions of users connected to the internet. Opinions, statements, advertising and political campaigns also began to be promoted through on-line social networks. Thus the search for the analysis of user activity on online social networks have gained much attention in recent years. For a better approach in the description of the framework, an experimental analysis of a set interactions of *Twitter* users is made.

Key-Words: *Twitter*, digital activism, framework, online social network, ontology.

LISTA DE FIGURAS

FIGURA 1 - PROBLEMA DAS PONTES DE KÖNIGSBERG	20
FIGURA 2 - GRAFO DO PROBLEMA DAS PONTES DE KÖNIGSBERG;	20
FIGURA 3 - PÁGINA INICIAL DO TWITTER.....	24
FIGURA 4 - TWEETS COM MENÇÃO A USUÁRIOS.....	25
FIGURA 5 - RESPOSTA A UM TWEET.....	26
FIGURA 6 - ASSUNTOS MAIS CITADOS EM TODO MUNDO EM 4 DE FEVEREIRO DE 2013, ÀS 7 HORAS (GMT +3).....	26
FIGURA 7 - UM EXEMPLO DE UM GRAFO ALEATÓRIO DE ERDÖS E RÉNYI.....	28
FIGURA 8 - A DISTRIBUIÇÃO DA CONECTIVIDADE PARA UMA REDE COM 10.000 VÉRTICES, USANDO UMA PROBABILIDADE $P = 0,2$	28
FIGURA 9 - EXEMPLO DE UMA REDE GERADA PELO MODELO LIVRE DE ESCALA DE BARABÁSI E ALBERT.....	30
FIGURA 10 - CÁLCULO DO COEFICIENTE DE AGRUPAMENTO EM TRÊS CENÁRIOS DIFERENTES.	31
FIGURA 11 - EXEMPLIFICAÇÃO DE ASSORTATIVIDADE EM UMA REDE.....	32
FIGURA 12 - REPRESENTAÇÃO DE VÉRTICES COM MAIOR BETWEENNESS.....	33
FIGURA 13 - REPRESENTAÇÃO DE CENTRALIDADE DE PROXIMIDADE MÁXIMA NA REDE.....	34
FIGURA 14 - ONTOLOGIA DE DEPENDÊNCIA TECNOLÓGICA DE OBJETOS DIGITAIS ..	36
FIGURA 15 - ESPECIFICAÇÃO DE UM RECURSO EM RDF.	40
FIGURA 16 - TELA DE EDIÇÃO DE CLASSES DE ONTOLOGIAS NO PROTÉGÉ.....	43
FIGURA 17 - VISÃO GERAL DA ARQUITETURA DO FRAMEWORK TWEOPY.	45
FIGURA 18 - DIAGRAMA DE SEQUÊNCIA DO TWEOPY.	45
FIGURA 19 - ESTRUTURA HIERÁRQUICA DE CLASSES DA ONTOLOGIA DESENVOLVIDA.....	46
FIGURA 20 – VISUALIZAÇÃO GRÁFICA DA ONTOLOGIA.....	47
FIGURA 21 - ONTOLOGIA DESENVOLVIDA NO PRESENTE TRABALHO.	48
FIGURA 22 - HIERARQUIA DOS MÓDULOS DESENVOLVIDOS NO FRAMEWORK.....	49
FIGURA 23 - PARTE DO CÓDIGO-FONTE DO FRAMEWORK TWEOPY RESPONSÁVEL PELO ARMAZENAMENTO DO GRAFO EM MEMÓRIA.....	49
FIGURA 24 - PARTE DO CÓDIGO-FONTE DO FRAMEWORK TWEOPY CORRESPONDENTE AOS MÉTODOS DESENVOLVIDOS NA CLASSE GRAFO.	50
FIGURA 25 - PARTE DO CÓDIGO-FONTE DO FRAMEWORK TWEOPY CORRESPONDENTE À BUSCA NA ONTOLOGIA DOS MÉTODOS ASSOCIADOS À FINALIDADE DE USO DO FRAMEWORK.....	51

FIGURA 26 - PARTE DO CÓDIGO-FONTE RESPONSÁVEL PELO PRÉ-PROCESSAMENTO DOS DADOS.	51
FIGURA 27 - REDE FORMADA PELA HASHTAG #IPHONE4SDATIM.....	55
FIGURA 28 - REDE FORMADA PELA HASHTAG #MARIMAR.	55
FIGURA 29 - REDE FORMADA PELA HASHTAG #CONTRAOAUMENTO.	55
FIGURA 30 - NÚMERO DE INTERAÇÕES PELO NÚMERO DE USUÁRIOS.....	56
FIGURA 31 - RELAÇÃO ENTRE O VÉRTICE COM MAIOR E MENOR GRAU.....	57
FIGURA 32 - MÁXIMOS VALORES ENCONTRADOS NA PROPRIEDADE BETWEENNESS CENTRALITY.	57

LISTA DE SIGLAS

API - *Application Programming Interface*

OWL - *Ontology Web Language*

RDF – *Resource Description Framework*

UML – *Unified Modeling Language*

URI - *Uniform Resource Identifier*

W3C - *World Wide Web Consortium*

XML - *eXtensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO	13
1.1 OBJETIVOS	16
1.1.1 OBJETIVO GERAL.....	16
1.1.2 OBJETIVOS ESPECÍFICOS	16
1.2 MOTIVAÇÃO	16
1.3 ORGANIZAÇÃO DO TEXTO.....	18
2 ANÁLISE DE REDES SOCIAIS ON-LINE	19
2.1 ASPECTOS HISTÓRICOS DA CONSTRUÇÃO DA ABORDAGEM DE ANÁLISE DE REDES SOCIAIS ON-LINE	19
2.2 CONCEITOS BÁSICOS DE REDES SOCIAIS ON-LINE.....	21
2.3 FUNCIONALIDADES COMUNS DE REDES SOCIAIS ON-LINE.....	22
2.4 TIPOS DE INTERAÇÕES ENTRE USUÁRIOS.....	23
2.5 INTRODUÇÃO AO <i>TWITTER</i>	24
2.5.1 TIMELINE	25
2.5.2 HASHTAGS	25
2.5.3 MENTIONS.....	26
2.5.4 FOLLOWERS.....	27
2.5.5 TRENDING TOPICS	27
2.5.6 RETWEETS	27
2.6 MODELOS DE REDES	28
2.6.1 MODELO DE REDE ALEATÓRIA.....	28
2.6.2 MODELO SMALL WORLD.....	30
2.6.3 MODELO DE REDES LIVRES DE ESCALA	30
2.7 MEDIDAS DE ANÁLISE DE REDES SOCIAIS.....	31
2.7.1 GRAU DOS VÉRTICES	31
2.7.2 COEFICIENTE DE AGRUPAMENTO	31
2.7.3 COMPONENTE	32
2.7.4 DISTÂNCIA MÉDIA E DIÂMETRO	32
2.7.5 ASSORTATIVIDADE.....	33
2.7.6 BETWEENNESS	34
2.7.7 CENTRALIDADE DE PROXIMIDADE	35
2.7.8 PAGERANK.....	35
3 ONTOLOGIAS	36
3.1 CLASSES.....	37
3.2 PROPRIEDADES.....	38
3.3 INSTÂNCIAS.....	38
3.4 RELACIONAMENTOS ENTRE AS CLASSES.....	38

3.5 RESTRIÇÕES E REGRAS.....	38
3.6 CLASSIFICAÇÃO DE ONTOLOGIAS.....	39
3.7 LINGUAGENS PARA REPRESENTAÇÃO	39
3.7.1 RDF	40
3.7.2 RDF <i>Schema</i>	41
3.7.3 OWL – <i>Ontology Web Language</i>	42
3.8 FERRAMENTAS PARA O DESENVOLVIMENTO DE ONTOLOGIAS	43
3.8.1 PROTÉGÉ-FRAMES.....	43
3.8.2 PROTÉGÉ-OWL	44
4 TWEEDY: UM FRAMEWORK BASEADO EM ONTOLOGIAS PARA ANÁLISE DE REDES SOCIAIS ON-LINE	45
4.1 OBJETIVOS	45
4.2 ARQUITETURA.....	45
4.3 CENÁRIOS DE USO.....	53
4.4 IMPLEMENTAÇÃO COMPUTACIONAL	54
5 AVALIAÇÃO EXPERIMENTAL.....	55
6 CONCLUSÕES E TRABALHOS FUTUROS.....	59
REFERÊNCIAS	60

1 INTRODUÇÃO

Com o advento da *Word Wide Web*, ou Web, houve uma ligeira mudança em aspectos do cotidiano das pessoas. O rápido crescimento nas últimas décadas fez da Web a maior e mais popular fonte de dados público, (Liu, 2009). A imensa fonte de dados proporcionada pela web pode facilmente ser acessada e incrementada. As consultas à informação antes da Web eram realizadas, em sua maioria, manualmente através de livros, artigos e outras fontes, o qual impunham restrições de acesso devido à sua composição física. Porém, hoje em dia, muita informação antes restrita a arquivos físicos está disponível na Web.

Com o surgimento de redes sociais on-line a web passou a ser não somente um meio para encontrar informações, mas também um meio de interação entre usuários, que compartilham informação e conhecimento. Através desses canais criados pelas redes sociais on-line, os usuários passaram a se comunicar, expressar suas opiniões e discutir acerca de um determinado assunto com qualquer pessoa do mundo conectada à internet. Atualmente essas redes sociais on-line fazem parte do dia a dia de milhões de pessoas no mundo que criam, buscam e disseminam conteúdo na Web.

Com o crescimento e popularização, as redes sociais on-line têm chamado a atenção de vários pesquisadores ao redor do mundo, que visam reconhecer a relação entre o conteúdo publicado com os acontecimentos das vidas cotidianas dos usuários. Segundo os autores de Sakaki et al. (2010), quando ocorrem terremotos no Japão, diversas mensagens são publicadas logo em seguida no *Twitter*¹ e essas mensagens são utilizadas para desenvolver um modelo para localizar o centro do terremoto e sua trajetória.

Entre 2010 e 2011, uma série de protestos e manifestações contra governos em países do Oriente Médio e do Norte da África, denominada Primavera Árabe, atraiu os olhares de todo o mundo. Através das redes sociais on-line, os protestantes conseguiam facilmente articular encontros, disseminar imagens, vídeos e informações que possibilitavam as manifestações serem mais eficazes e dinâmicas (Burns et al., 2009).

Assim como a primavera árabe, outros protestos em países do Oriente haviam ocorrido através do uso maciço das redes sociais on-line. Em 2009, após o resultado das eleições presidenciais no Irã, surgiu uma série de manifestações motivadas por suspeitas de fraude eleitoral. Os protestos iniciaram a partir de 13 de junho de 2009, tanto na capital

¹ Serviço de *microblogging* no qual seus usuários divulgam textos ou links para outros usuários que possuem conexão.

do Irã quanto em cidade vizinhas e ganharam força com a dinâmica do meio proporcionado pelas redes sociais on-line. O governo em represália aos protestos bloqueou o acesso à maioria das redes sociais on-line. No entanto, a população do Irã, que é em sua maioria jovem e com formação universitária, conseguiu driblar a censura digital imposta pelo governo iraniano (Burns et al., 2009). No *Twitter* a principal *hashtag*² utilizada pelos manifestantes foi *#iranelection*, que foi um dos assuntos mais comentados durante o período das manifestações (*Twitter.a*, 2009).

No Brasil, as redes sociais on-line também se tornaram aliadas na organização e disseminação de informações de movimentos sociais. O primeiro fenômeno expressivo ocorreu em 2008, depois de uma das maiores catástrofes naturais da história do estado de Santa Catarina. Após dias ininterruptos de chuvas intensas, vários rios transbordaram isolando várias cidades e provocando deslizamentos de morros e encostas. Várias ferramentas on-line, como blogs e redes sociais on-line foram utilizadas para mobilizar campanhas de doação de mantimentos e dinheiro às vítimas no estado de Santa Catarina (Recuero, 2009).

Em Teresina, capital do Estado do Piauí, estudantes protestaram no início de 2012 contra o aumento de preço de passagens de ônibus urbanos. As manifestações físicas aconteceram na região central de Teresina e eram organizadas integralmente por meio de redes sociais on-line, como *Twitter* e *Facebook*³. Essas redes sociais on-line supriram, de forma aparentemente satisfatória, o papel da mídia televisiva do Piauí, que se posicionou contra os manifestantes, vetando a veiculação de informações acerca dos protestos (contraOaumento, 2012). No *Twitter*, o movimento podia ser acompanhado pela *hashtag* *#contraoaumento*.

Como exposto nos parágrafos anteriores, pode-se perceber a necessidade de estudos que possam aferir características pertinentes o ativismo digital, o qual é compreendido por protestos e manifestações organizados por serviços on-line que estão modificando profundamente a dinâmica de organização, identidade e mobilização social. Apesar de extensas pesquisas envolvendo dados de *e-mails* e páginas da Web, pouco se conhece sobre dinâmica do meio proporcionado pelas redes sociais on-line. Há pouco conhecimento sólido sobre redes sociais on-line. Dessa forma, alguns questionamentos são pertinentes e objeto de estudo de pesquisadores, como: De que forma os usuários

² Trecho de caracteres, sem o caractere de espaço, geralmente precedido do caractere “#”.

³ Facebook é um site e serviço de rede social operado e de propriedade privada da Facebook Inc.

manifestam suas opiniões e quais são as propriedades relacionadas ao conteúdo e a estrutura topológica da rede formada pela propagação da informação?

Compreender a dinâmica das interações entre os usuários envolvidos em alguma forma de ativismo digital pode criar oportunidades para melhorar o funcionamento dos sistemas de redes sociais on-line, possibilitando a extração de características intrínsecas ao ativismo digital que possam servir como *feedback* na implementação de novas funcionalidades voltadas ao público das redes sociais on-line. Além disso, é importante para diversas aplicações relacionadas à verificação de acurácia de propagandas realizadas através de redes sociais on-line, eficiência de campanhas políticas e na detecção e prevenção de ataques de *spammers*.

Essa dissertação está inserida justamente nesse contexto, onde o objetivo é propor uma ferramenta para possibilitar uma melhor análise das interações entre usuários de redes sociais on-line. A Seção 1.1 apresenta os objetivos do trabalho. A Seção 1.2 discute a motivação para sua realização. A Seção 1.3 mostra como o restante do trabalho está organizado.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

O objetivo principal deste trabalho é o desenvolvimento de um framework baseado em ontologias para análise de redes sociais on-line que possibilite uma fácil análise das interações entre usuários de redes sociais on-line. Dessa forma, o framework tem como fim, automatizar e dinamizar a análise de dados gerados por usuários de redes sociais on-line.

1.1.2 OBJETIVOS ESPECÍFICOS

Para alcançar o objetivo geral deste trabalho, foram considerados os seguintes objetivos específicos:

- Conceber e desenvolver um framework baseado em ontologias que possibilite a análise dos dados coletados do *Twitter*.

- Definir e utilizar uma ontologia no âmbito da problemática abordada anteriormente, que permita a utilização do framework para análise dos dados advindos de redes sociais on-line.
- Extrair e disponibilizar informações a partir de inferências feitas através da utilização do framework sob diferentes conjuntos de dados do *Twitter*.

1.2 MOTIVAÇÃO

Apesar da extensa quantidade de dados gerados diariamente por usuários de redes sociais on-line, ainda há pouco conhecimento sólido acerca do conteúdo desses dados. Existem vários aspectos relevantes e preponderantes que motivam pesquisadores a promover pesquisas utilizando essa larga quantidade de dados advindas de redes sociais on-line, como:

- Sociológico: O início das análises de redes sociais se deu através de estudos realizados por sociólogos, que utilizavam técnicas rústicas, como entrevistas e pesquisas de campo que não serviam de insumo suficiente para obtenção de resultados precisos (Wasserman et al., 1994; Jarry, 1999; Freeman, 2004). Com o avanço da tecnologia e expansão da Web, milhões de usuários passaram a se conectar através de redes sociais on-line, gerando, através de interações entre si, uma quantidade de dados jamais vista anteriormente. Dessa forma, os dados gerados por usuários de redes sociais on-line tornaram-se objeto de estudo de sociólogos que em décadas passadas não dispunham de grande quantidade de dados para conduzir os experimentos a resultados mais precisos. O *Twitter*, por ser uma rede social on-line bastante utilizada na internet e pela facilidade de coleta de dados através de uma API⁴, torna-se atrativo para pesquisadores.
- Político: Em 2008, pela primeira vez na história, o mundo pôde acompanhar, através de redes sociais on-line, as campanhas eleitorais nos Estados Unidos dos candidatos Barack Obama e John McCain. Através do *Twitter* era possível acompanhar notícias, comentários, *links* para imagens e vídeos relacionados às

⁴ Um conjunto de rotinas e padrões estabelecidos por um software para utilização de suas funcionalidades por outros aplicativos.

eleições. Protagonizou-se os maiores índices de comparecimento de eleitores às urnas de todos os tempos nas eleições americanas (Recuero, 2009). No Brasil, durante as eleições presidenciais em 2010, a candidata presidencial Marina Silva teve expressivo apoio à sua candidatura por usuários do *Twitter*, conseguindo a doação de cerca de 170 mil reais de usuários de rede social on-line (TÚLIO COSTA, 2011). (Cha et al. (2010) realizaram um estudo a fim de extrair características acerca do grau de influência de usuários do *Twitter*. No estudo, constatou-se que o conjunto de três propriedades; seguidores, *retweets*⁵ e a quantidade de menções feitas por outros usuários; eram fatores preponderantes na inferência de usuários influentes no *Twitter*. Dessa forma, o trabalho fornecia novas perspectivas para aplicação real de tais experimentos, como melhorar eficácia de campanhas eleitorais realizadas através do *Twitter*.

- Publicidade: Com a expansão e aumento da quantidade de horas dispendidas em redes sociais on-line pelas pessoas, as empresas estão cada vez mais buscando essas redes sociais on-line para promover campanhas publicitárias. O trabalho desenvolvido por Cha et al. (2010), que visa extrair conhecimento acerca dos fatores que determinam o grau de influência de um usuário, possui extrema aplicação no desenvolvimento de campanhas publicitárias, pois possibilita saber quais usuário podem influenciar mais os outros, poupando recursos de tais campanhas.

Para a ciência da computação, a pesquisa apresentada nesse trabalho demonstra, através do desenvolvimento do framework, uma oportunidade para explorar características inerentemente relacionadas às atividades de usuários de redes sociais on-line, possibilitando o aprimoramento de desenvolvimento de novas funcionalidades de redes sociais on-line.

⁵ São tweets que foram replicados, mantendo toda estrutura da mensagem e referência ao autor original

1.3 ORGANIZAÇÃO DO TEXTO

Este trabalho está organizado da seguinte forma. O Capítulo 2 aborda as características de redes sociais, fazendo uma descrição das funcionalidades do *Twitter*, apresenta métricas relacionadas às propriedades estruturais de redes complexas e aborda aspectos referentes às pesquisas de análise de redes sociais. O Capítulo 3 aborda conceitos relativos às ontologias. O Capítulo 4 apresenta o framework desenvolvido, exemplificando suas funcionalidades e seu funcionamento. O Capítulo 5 apresenta um estudo de caso e faz-se uma análise dos dados de interações de usuários do *Twitter*, a fim de se verificar propriedades relacionadas ao ativismo digital. Por fim, o Capítulo 6 conclui esta dissertação, revendo nossas principais contribuições e discussões.

2 ANÁLISE DE REDES SOCIAIS ON-LINE

Este capítulo descreve conceitos associados às redes sociais on-line. A seção 2.1 faz uma breve introdução à formulação matemática de redes sociais on-line, abordando conceitos proeminentes da teoria dos grafos. Já as seções 2.2, 2.3 e 2.4 apresentam conceitos de redes sociais on-line, funcionalidades típicas encontradas em redes sociais on-line e tipos de interação entre usuários. Na seção 2.5, destaca-se as principais funcionalidades do *Twitter*. A seção 2.6 apresenta três tipos de modelos de rede, exemplificando as características associadas à cada modelo. Por fim, a seção 2.7 apresenta medidas de análise de redes sociais on-line utilizadas na implementação do *framework*.

2.1 ASPECTOS HISTÓRICOS DA CONSTRUÇÃO DA ABORDAGEM DE ANÁLISE DE REDES SOCIAIS ON-LINE

A formulação matemática das redes sociais encontra-se no conceito de grafos, desenvolvido pelo matemático Leonhard Paul Euler em meados do século XVIII (Albert, Barabási, 2002; Newman, 2003b). Matematicamente, grafos podem ser definidos como estruturas compostas por um conjunto de vértices que se conectam entre pares, formando arestas. Um grafo é uma representação de uma rede através de pontos e linhas. Leonhard Paul Euler resolveu um famoso problema de sua época, chamado de "Problema das Pontes de Königsberg". Nesse problema, haviam duas ilhas conectadas uma a outra através de uma ponte e conectadas às margens do rio através de seis pontes, conforme a ilustração da Figura 1. O desafio do problema era começar em qualquer uma das quatro áreas e caminhar somente uma vez por cada ponte, retornando ao ponto de saída. Euler provou que não havia solução para o problema, substituindo cada porção de terra por um ponto e cada ponte por arcos unindo os pontos correspondentes, formulando um grafo. O grafo análogo ao desenvolvido por Euler é mostrado na Figura 2, onde os pontos correspondem às porções de terra. Dessa forma, Euler generalizou o problema e desenvolveu um critério para que um grafo pudesse ser atravessado, no qual o grafo deveria estar conectado e cada ponto deveria ser incidente a um número par de arestas. Na Figura 2, apesar do grafo estar conectado, não há nenhum ponto que incida um número par de arestas.

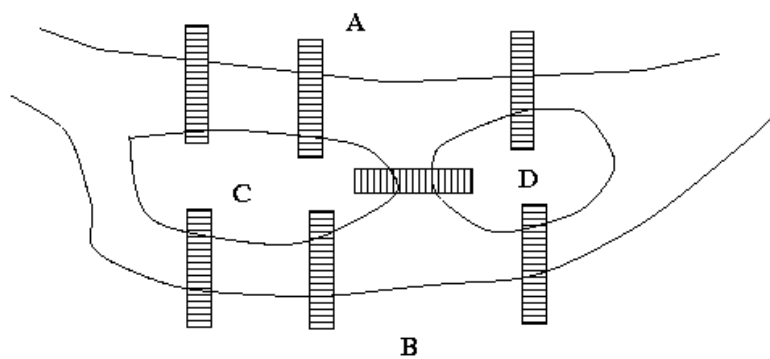


Figura 1 - Problema das Pontes de Königsberg

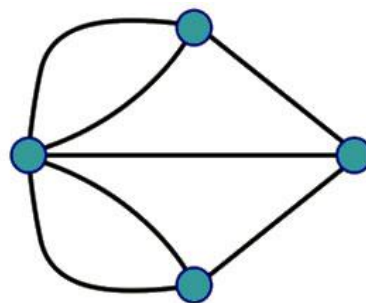


Figura 2 - Grafo do Problema das Pontes de Königsberg;

Os estudos iniciais sobre análise de redes sociais foram desenvolvidos na sociologia. O trabalho de Jacob Levy Moreno (1934) introduziu o conceito de sociogramas para representar as redes de relações entre pessoas na *Hudson School for Girls*. Vários pesquisadores, em meados de 1930, sob o aspecto antropológico, também desenvolveram pesquisas em Harvard buscando características da estrutura global da sociedade. No final da década de 1970, a análise de redes sociais foi reconhecida pelos cientistas sociais e contribuições sob o aspecto computacional tornaram-se indispensáveis ao desenvolvimento desta área de pesquisa, posto que o desenvolvimento de ferramentas capazes de visualizar e computar grandes quantidades de dados tornou possível realizar trabalhos mais abrangentes e precisos.

2.2 CONCEITOS BÁSICOS DE REDES SOCIAIS ON-LINE

As redes sociais on-line são geralmente definidas como grupos de usuários que através de um meio digital interagem entre si promovendo informações. Porém, para fins específicos deste trabalho, utiliza-se a definição de redes sociais on-line adotada por Boyd and Ellison (2007), no qual define-se redes sociais on-line como serviços baseados na web que permitem aos usuários: (i) criar um perfil público ou semipúblico dentro de um sistema limitado; (ii) articular uma lista de outros usuários com quem se deseja compartilhar informação; e (iii) ver e pesquisar sua lista de conexões e aquelas feitas por outros dentro do sistema. Conforme as definições extraídas de (Boyd and Ellison, 2007), há uma ampla variedade de redes sociais on-line que variam conforme seu propósito, no entanto, seguem a definição adotada nesse trabalho. Por exemplo, existem redes sociais on-line para profissionais (por exemplo, *Linked*⁶ e *via6*⁷), amigos (por exemplo, *Orkut*⁸ e *Facebook*⁹), mensagens curtas (por exemplo, *Twitter*¹⁰ e *Laconica*¹¹) e vídeos (por exemplo, *Youtube*¹² e *Vimeo*¹³).

2.3 FUNCIONALIDADES COMUNS DE REDES SOCIAIS ON-LINE

Os tipos de interações que redes sociais on-line promovem estão diretamente relacionados às funcionalidades disponibilizadas pela rede social on-line em questão. A seguir descrevem-se algumas das funcionalidades mais importantes em interações entre usuários de redes sociais on-line:

- Perfis de usuários: Dentre todas as funcionalidades disponibilizadas pelas redes sociais on-line, o perfil de um usuário congrega todas as características individuais de cada membro da rede social on-line, como detalhes demográficos (idade, localização,

⁶ <http://www.linkedin.com>

⁷ <http://www.via6.com>

⁸ <http://www.orkut.com>

⁹ <http://www.facebook.com>

¹⁰ <http://www.Twitter.com>

¹¹ <http://www.laconica.com>

¹² <http://www.youtube.com>

¹³ <http://www.vimeo.com>

sexo, etc.), interesse (música, filmes, etc.) e fotos. Dessa forma, o perfil contém, além da identificação do usuário, características e preferências dos mesmos.

- **Atualização de usuário:** Com o intuito de promover o compartilhamento de informações entre usuários, essa funcionalidade disponibiliza uma lista com as mais recentes atividades dos amigos de um determinado usuário. Essa funcionalidade expõe alterações no perfil, compartilhamento de *links*, fotos e vídeos para os participantes da rede social on-line que possuem conexão entre si.
- **Comentários:** Essa funcionalidade é a forma mais comum de interação entre pares de usuários. Também pode vir agregadas a fotos, vídeos e a qualquer outro tipo de conteúdo que possa ser compartilhado. Como exemplo, os vídeos do *Youtube* que podem receber comentários, e as fotos de usuários do *Facebook* que estão suscetíveis a comentários conforme a política de privacidade definida pelo emissor da foto na rede social on-line.
- **Aplicativos:** Numa estratégia para atrair e fidelizar mais usuários, várias redes sociais on-line disponibilizam uma API para que desenvolvedores independentes possam integrar seus aplicativos às redes sociais on-line. Com essa inovação, desenvolvedores de aplicativos são capazes de desenvolver aplicações que promovam mais interação entre os usuários de uma rede social on-line. Atualmente a rede social on-line *Facebook* possui mais de sete milhões de aplicativos, com mais de 500 milhões de pessoas utilizando (Statistics Facebook, 2012).
- **Sugestão de amigos:** Adotada nos últimos anos, essa funcionalidade visa promover novas conexões entre usuários de redes sociais on-line. Várias técnicas de predição de novos usuários são utilizadas no desenvolvimento dessa funcionalidade. O trabalho desenvolvido por (Xing Xie, 2010) propôs um framework para recomendação de usuários em redes sociais on-line, caracterizando o interesse do usuário em duas dimensões: contexto (localização, horário) e conteúdo publicado pelos usuários.

2.4 TIPOS DE INTERAÇÕES ENTRE USUÁRIOS

As interações em redes sociais on-line envolvem necessariamente dois ou mais usuários, podendo ocorrer de várias formas, como explicitadas a seguir (Benevenuto et al., 2010):

- Interações textuais: Essa forma de interação é a mais comum e ocorre principalmente por meio de texto. Pode ocorrer de forma privada, visível apenas entre dois indivíduos envolvidos na interação, ou pode ser visível para os demais membros da rede social on-line. Comentários também figuram como interações textuais.
- Fotos: Interações envolvendo fotos ocorrem através de marcação de usuários em fotos de amigos. Algumas redes sociais on-line, como *Orkut* e *Facebook*, oferecem o recurso de armazenamento de fotos e marcação de usuários.
- Vídeos: As interações por vídeos ocorrem através de conversações on-line, por meio de vídeo. Recentemente algumas redes sociais on-line como, por exemplo, o *GooglePlus*¹⁴ e o *Facebook* implementaram essa forma de interação no qual os usuários podem iniciar conversas on-line através de vídeo.
- Interações de aplicativos: Além das formas tradicionais de interações supracitadas, aplicativos desenvolvidos por terceiros também promovem interações entre usuários de redes sociais on-line. Como exemplo, o *Facebook* promove interações que envolvem aplicativos, notificando os usuários sobre aplicativos que seus amigos tenham instalado ou utilizado.

2.5 INTRODUÇÃO AO TWITTER

A rede social on-line *Twitter* atraiu a atenção não só de milhões de usuários na Web, mas também de diversos pesquisadores espalhados por laboratórios de pesquisa no mundo. O *Twitter* possibilita que cada usuário escreva mensagens de no máximo 140 caracteres, disponibilizando informações para outros usuários. A interação entre os usuários no *Twitter* se dá através de mensagens e recebe o nome de *tweet*¹⁵, o qual é agregado em ordem cronológica na página inicial do usuário que o criou, assim como fluirá pela rede de usuários permitidos. A estrutura formada pela rede social on-line está fundamentada em usuários seguidos e usuários seguidores, ou seja, cada usuário do serviço de *microblogging* tem a possibilidade de seguir outros usuários, obtendo assim, os *tweets* dos usuários que estão sendo seguidos. Não necessariamente um usuário X, que

¹⁴ <http://plus.google.com>

¹⁵ Mensagem de até 140 caracteres publicadas por usuários do *Twitter*

segue um grupo de usuários do *Twitter*, vai ser seguido pelos mesmos. A estrutura resultante da rede de vínculos entre os usuários do *Twitter* é um grafo direcionado, onde os vértices são os usuários e as relações de seguidores e os seguidores correspondem às arestas. Ao momento em que um determinado termo torna-se bastante utilizado em *tweets* num determinado espaço de tempo, o mesmo aparece na seção *Trending Topics*¹⁶ no *Twitter*, indicando que os *tweets* que mencionam o termo estão sendo amplamente divulgados no momento. Uma convenção que se tornou amplamente utilizada foi o *retweet*, no qual usuários podem repassar o *tweet* de um usuário que estão seguindo para todos os seus seguidores, mantendo a estrutura intacta assim como a referência ao autor do *tweet*. Dessa forma as mensagens podem ser transmitidas de usuários seguidos para usuários seguidores através de *retweets*. Logo em seguida, faz-se uma breve descrição das principais funcionalidades do *Twitter*.

2.5.1 TIMELINE

A funcionalidade *TIMELINE* contém, em ordem cronológica, os *tweets* e *retweets* dos usuários que o mesmo está seguindo. Os *tweets* publicados pelo próprio usuário também aparecem na *TIMELINE* (*Twitter.a*, 2012), conforme a Figura 3.

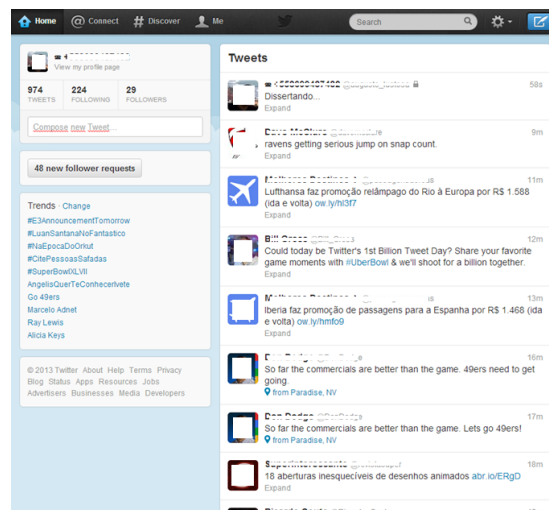


Figura 3 - Página inicial do Twitter

¹⁶ Assuntos que se tornaram populares no *Twitter*.

2.5.2 HASHTAGS

As *hashtags* correspondem a trechos de caracteres, sem o caractere de espaço, geralmente precedidos do caractere "#". A existência de *hashtags* nas mensagens facilita a pesquisa de *tweets* referentes a um determinado tipo de assunto (Twitter.b, 2012). Um exemplo de uso de *hashtags* ocorreu durante os protestos no Irã, em 2009. Apesar do governo iraniano não ter permitido o acesso internacional da imprensa ao processo eleitoral, os eleitores disponibilizavam informações referentes ao processo eleitoral e às manifestações envolvidas, principalmente através da *hashtag* "#iranelection" (Burns et al., 2009).

2.5.3 MENTIONS

Um usuário pode referenciar outros usuários no corpo de seus *tweets*. O usuário referenciado, além de ser notificado, possui acesso todos *tweets* ao qual foi mencionado. A marcação de um usuário em um *tweet* corresponde ao seu nome de usuário precedido do caractere de "@" (Twitter.c, 2012). A Figura 4 ilustra menções a um usuário e a Figura 5 uma resposta a um *tweet*.



Figura 4 - Tweets com menção a usuários.

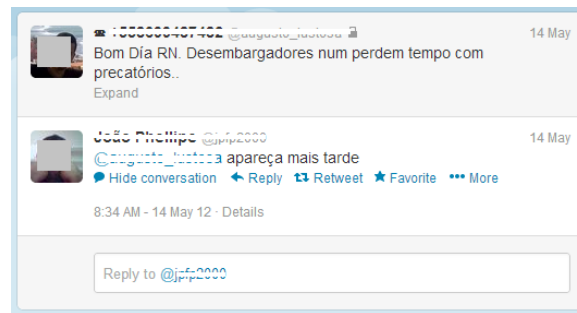


Figura 5 - Resposta a um tweet.

2.5.4 FOLLOWERS

Followers são os membros do *Twitter* que o usuário permite visualizar seus *tweets*. Ou seja, são os perfis autorizados a receber *tweets* de um determinado usuário. Não necessariamente a relação é mútua. A estrutura de rede formada por *followers* e usuários seguidos corresponde a um grafo direcionado (*Twitter.d*, 2012).

2.5.5 TRENDING TOPICS

O algoritmo que compõe a funcionalidade que provê os *TRENDING TOPICS* identifica os assuntos que são amplamente citados em *tweets* num determinado espaço de tempo, ao invés de assuntos que já são populares há algum tempo ou diariamente, para prover os assuntos mais citados durante o dia (*Twitter.e*, 2012). A Figura 6 ilustra os *Trending Topics* com os dez assuntos mais populares em todo mundo.

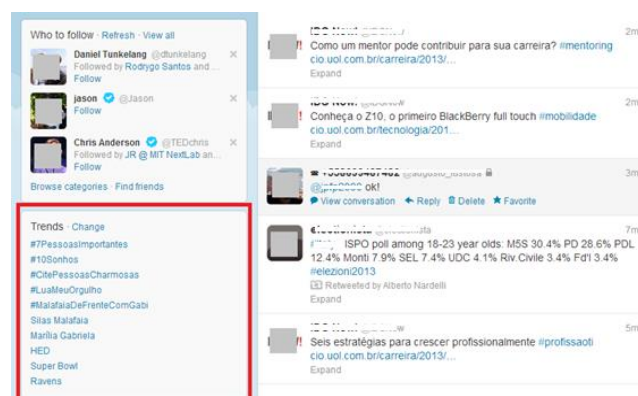


Figura 6 - Assuntos mais citados em todo mundo em 4 de fevereiro de 2013, às 7 horas (GMT +3).

2.5.6 RETWEETS

Os *retweets*, também conhecidos como "RT", correspondem a *tweets* que foram replicados mantendo toda estrutura da mensagem e referência ao autor original. É um recurso que facilita aos usuários compartilhar, com todos os seus seguidores, um determinado *tweet* de forma simples e dinâmica (Twitter.f, 2012).

2.6 MODELOS DE REDES

Diversos trabalhos foram realizados na tentativa de explicar propriedades e características intrinsecamente relacionadas a redes como um todo (ALBERT, BARABÁSI, 2002; BOCCALETTI et al., 2006; COSTA et al., 2007; NEWMAN, 2003b). Diante do avanço provocado por esses trabalhos, foram desenvolvidos modelos de análise de redes. Watts (2003) propôs que em novos estudos de redes que não deveria ser apenas levado em consideração a estrutura da rede no qual suas propriedades estavam fixada no tempo, mas também a dinâmica da rede, já que ela está evoluindo com o tempo. Dessa forma, a possível contribuição para os estudos relacionados a redes sociais está no fato de perceber a estrutura da rede não como fixa e determinante, mas sujeita a mudanças no tempo e espaço. Dessa forma são apresentados, nas seções subsequentes, três modelos de redes.

2.6.1 MODELO DE REDE ALEATÓRIA

Em meados de 1959, Paul Erdos e Alfred Rényi, dois matemáticos húngaros, consideraram os grafos como objetos estocásticos, ao invés de analisá-los de forma estritamente determinística, como fazia até então a matemática discreta e a sociologia. Dessa forma, os Erdos e Rényi sugeriram um modelo de rede baseado em ligações aleatórias, que ficou conhecido como grafos aleatórios de Erdos e Rényi (Erdos; Rényi, 1959, 1960, 1961). Esses tipos de grafos iniciam com N nós sem qualquer conexão entre si e a cada passo dois nós são escolhidos aleatoriamente e conectados com uma

probabilidade fixa p , sendo cada par de vértices considerado apenas uma vez. Dessa forma, todas as ligações possuem uma mesma probabilidade de acontecerem, ou seja, a rede formada tem uma estrutura estritamente homogênea. A Figura 7 ilustra um exemplo de rede aleatória. A distribuição da conectividade para essas redes, quando N é grande e a conectividade média é mantida constante, tende à distribuição de Poisson (Figura 8).

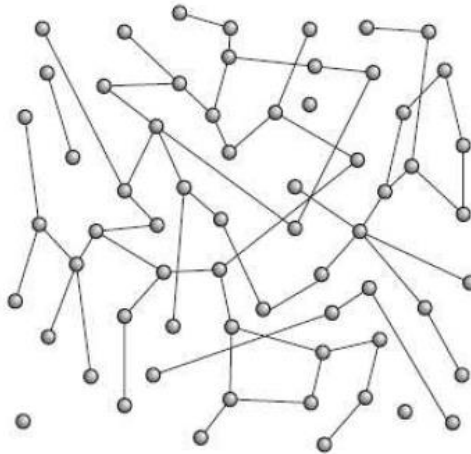


Figura 7 - Um exemplo de um grafo aleatório de Erdős e Rényi.

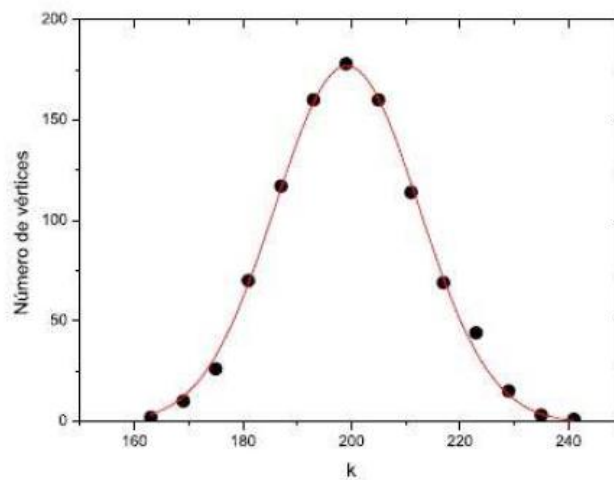


Figura 8 - A distribuição da conectividade para uma rede com 10.000 vértices, usando uma probabilidade $p = 0,2$.

2.6.2 MODELO SMALL WORLD

O conceito associado às redes *small-world* ficou amplamente conhecido após experimento realizado em (Milgram, 1967). O experimento consistia de um grupo de voluntários que tentavam enviar uma carta para uma pessoa através de outras pessoas que eles conheciam. Milgram enviou cartas a várias pessoas que explicavam que ele estava querendo atingir uma pessoa específica nos EUA e que o destinatário deveria repassar a carta para alguém que ele achasse que poderia levar a carta o mais próximo do destinatário final, ou entregá-la diretamente se conhecesse o destinatário final. Antes de repassar a carta, o remetente adicionava seu nome ao fim da carta para que se pudesse ter o controle do caminho percorrido pela carta. Das cartas entregues ao destinatário final, o número médio de passos requeridos para o alvo foi 6, resultado que ficou conhecido como princípio dos seis graus de separação.

Uma rede pode ser considerada *small-world* se ela tiver duas propriedades básicas: coeficiente de agrupamento alto e diâmetro pequeno (Watts, 1999). Essas duas propriedades foram identificadas em várias redes, como a Web (Albert et al., 1999; Broder et al. 2000); redes de colaboração científica (Newman 2001, 2004), no qual os vértices são representados pelos autores e as arestas as coautorias de artigos; redes de atores de filmes (Amaral et al., 2000), onde os vértices são os atores e as arestas os filmes; e redes sociais on-line (Mislove et al., 2007), onde verificou-se a existência de propriedades *small-world* em quatro redes sociais on-line (*LiveJournal*, *Flickr*, *Orkut* e *Youtube*).

2.6.3 MODELO DE REDES LIVRES DE ESCALA

No trabalho de Barabasi and Albert (1999), demonstrou-se que algumas redes possuíam uma ordem na dinâmica de estruturação, com características bem específicas. Uma dessas características, a conexão preferencial, resume-se na tendência de um novo nó se conectar com outro nó do grafo que possui alto grau de conexões. Essa característica implica em redes com poucos nós altamente conectados, denominados *hubs*, e muitos nós com poucas conexões. Redes com essas características são denominadas livres de escala

devido à representação matemática da rede. As redes livres de escala são aquelas em que a distribuição de seus graus segue a Lei de Potência. De acordo com o apresentado em Newman (2003), essas redes têm sido observadas em vários sistemas, como, por exemplo, na web, redes de citações de artigos científicos e redes de metabolismo. A Figura 9 ilustra um exemplo de rede livre de escala.

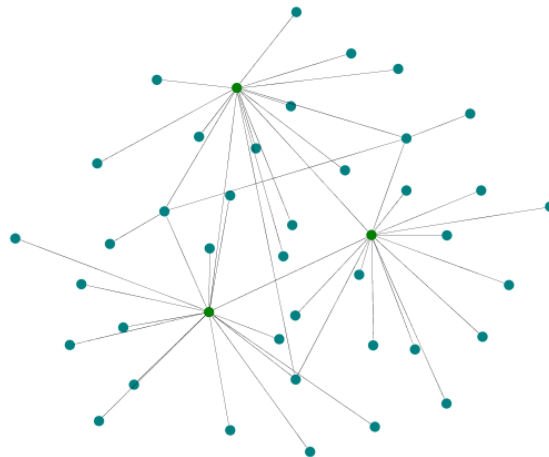


Figura 9 - Exemplo de uma rede gerada pelo modelo livre de escala de Barabási e Albert.

2.7 MEDIDAS DE ANÁLISE DE REDES SOCIAIS

2.7.1 GRAU DOS VÉRTICES

Em um grafo não orientado, o grau de um vértice pode ser obtido pelo número de conexões que o mesmo faz com os outros vértices. Já em grafos direcionados, essa definição também é válida, no entanto pode-se ter dois semigraus, o de entrada e o de saída. Dessa forma, a soma desses dois semigraus constitui o grau do vértice. Para redes direcionadas, é comum analisar as distribuições dos graus dos nós em ambas as direções, isto é, a distribuição do grau de entrada e a distribuição do grau de saída.

2.7.2 COEFICIENTE DE AGRUPAMENTO

O coeficiente de agrupamento (*clustering coefficient*) de um nó é a razão entre o número de arestas existentes entre seus vizinhos e o número máximo de arestas possíveis entre esses vizinhos. Por exemplo, na Figura 10 temos três cenários. No primeiro cenário, todos os vizinhos do nó mais escuro estão conectados, portanto o valor do coeficiente de agrupamento corresponde a um. Já no segundo cenário, há apenas uma conexão entre os nós vizinhos do nó escuro e três possíveis conexões entre os mesmos. Portanto o coeficiente de agrupamento para o segundo cenário é $1/3$. No último cenário, como não existem conexões entre os vizinhos, o valor do coeficiente de agrupamento para o nó mais escuro é zero.

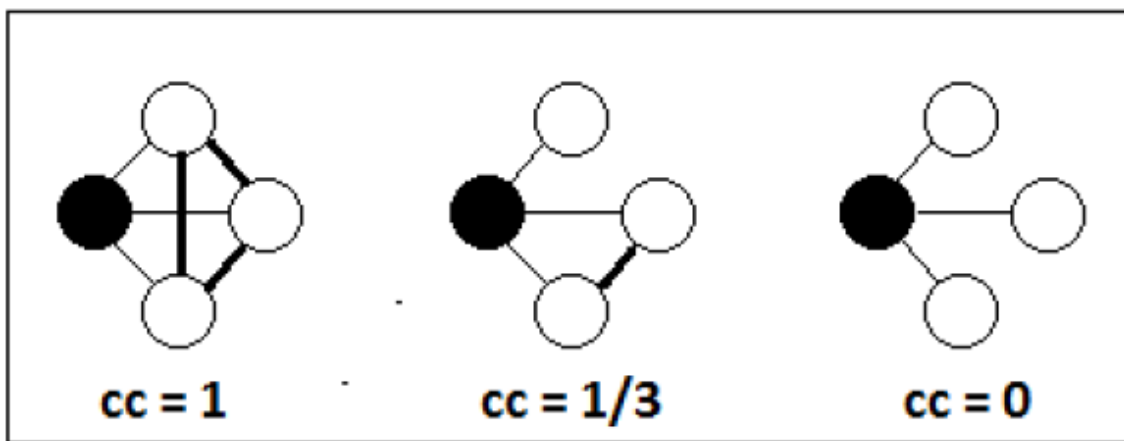


Figura 10 - Cálculo do coeficiente de agrupamento em três cenários diferentes.

2.7.3 COMPONENTE

Um componente de um grafo corresponde a um conjunto de nós, onde cada nó possui um caminho para todos os outros nós do conjunto. Para grafos direcionados, um componente é chamado de fortemente conectado (SCC - *Strongly Connected Component*) quando existe um caminho direcionado entre cada par de nós do conjunto. Um componente é fracamente conectado (WCC - *Weakly Connected Component*) se o caminho é não direcionado.

2.7.4 DISTÂNCIA MÉDIA E DIÂMETRO

A distância média de um grafo corresponde ao número médio em todos os caminhos mínimos existentes entre todos os pares de nós do grafo. Frequentemente a distância média é calculada apenas no componente fortemente conectado para grafos direcionados ou no componente fracamente conectado para grafos não direcionados. Outra métrica bastante utilizada é o diâmetro do grafo. O diâmetro é definido como a distância do maior caminho mínimo existente no grafo e, em geral, é também abordada somente para nós dos componentes fortemente e fracamente conectados.

2.7.5 ASSORTATIVIDADE

A assortatividade é uma medida típica de redes sociais on-line. Uma rede demonstra propriedades assortativas quando nós com muitas conexões tendem a se conectar com nós que também possuem muitas conexões. Para caracterizar a assortatividade de uma rede, calcula-se o grau médio de todos os nós vizinhos dos nós de grau k , dado $k_{nn}(k)$. A assortatividade ou disassortatividade é estimada avaliando os valores de $k_{nn}(k)$ em função de k , conforme demonstrado na Figura 11.

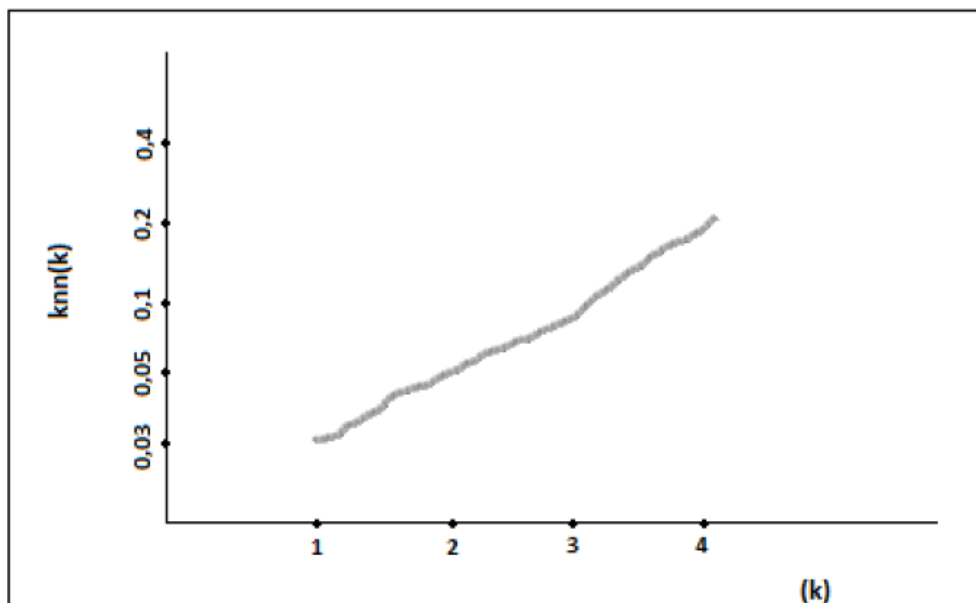


Figura 11 - Exemplificação de assortatividade em uma rede.

A Figura 11 mostra que para valores crescentes de $knn(n)$ em função de k , o gráfico resultante é crescente. Portanto, demonstrando assortatividade. A diassortatividade é caracterizada por um gráfico decrescente.

2.7.6 BETWEENNESS

A medida betweenness está inerentemente relacionada à centralidade dos nós ou arestas de uma rede. O *betweenness* $B(e)$ de uma aresta e é definido como o número de caminhos mínimos entre todos pares de nós em um grafo que passam por e . Caso existam múltiplos caminhos mínimos entre um par de nós, cada caminho recebe um peso, de tal forma que a soma dos pesos seja um. O *betweenness* de uma aresta indica a importância dessa aresta no grafo em termos de sua localização. Vértices com maior *betweenness* fazem parte de um número maior de caminhos mínimos e, portanto, são mais importantes para a estrutura do grafo.

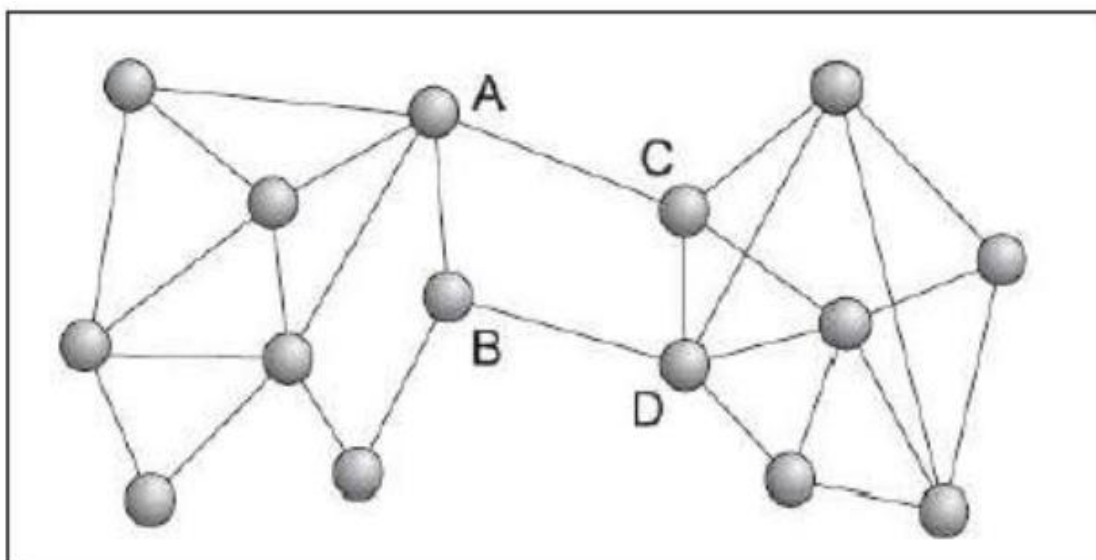


Figura 12 - Representação de vértices com maior betweenness.

Na Figura 12, os vértices A, B, C e D possuem maior centralidade de intermediação (*betweenness centrality*), pois participam da maioria dos menores caminhos da rede.

2.7.7 CENTRALIDADE DE PROXIMIDADE (CLOSENESS CENTRALITY)

Esta propriedade está inversamente relacionada com a distância entre nós de uma rede. Ela representa o quão próximo um nó se encontra em relação aos demais nós da rede. A centralidade de proximidade é obtida pelo inverso da soma das distâncias de um nó para os outros possíveis nós. Ou seja, quanto mais próximo este nó estiver dos demais, mais aguçada estará esta propriedade. Nós com essa propriedade em destaque podem interagir mais rápido com os demais, sendo, portanto, teoricamente, mais produtivos no compartilhamento de informações com os outros nós da rede. A Figura 13 mostra que o nó B, localizado no centro, é adjacente a todos os outros nós. Portanto o mesmo possui uma centralidade de proximidade máxima.

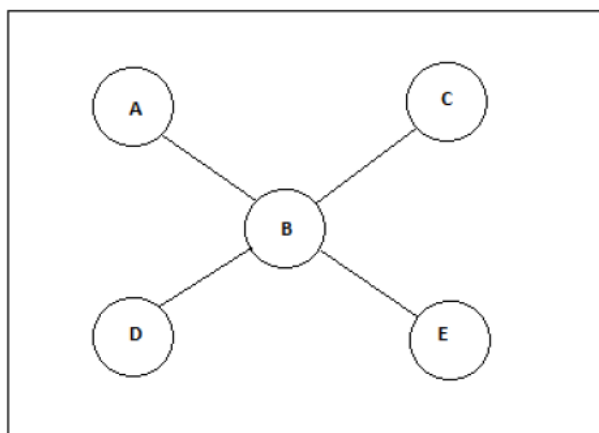


Figura 13 - Representação de centralidade de proximidade máxima na rede.

2.7.8 PAGERANK

PageRank é um algoritmo que estima um peso numérico para cada nó com o propósito avaliar a importância relativa do mesmo em um grafo. O algoritmo foi inicialmente proposto em (Page et al., 1998) para ordenar resultados de busca da Google. Conforme o algoritmo, um nó tem mais importância num grafo se o mesmo possui um considerável grau de entrada. No contexto desenvolvido na Google, páginas que eram

mais referenciadas em outros sites na web possuíam maior importância, refletindo em melhor posicionamento na máquina de busca.

3 ONTOLOGIAS

Este capítulo aborda conceitos referentes à ontologias, destacando as regras e restrições, tipos de linguagem para representação de ontologias e ferramentas para o desenvolvimento de ontologias.

Ontologias, de acordo com Daconta (2003), definem palavras em comum e conceitos utilizados para descrever e representar uma área do conhecimento, padronizando seus significados. Dessa forma, ontologia é um modelo de dados que define e representa conceitos e seus relacionamentos, geralmente dentro de um domínio. Também pode ser definida de forma mais genérica, uma vez que pode ser aplicada a diversos domínios. Para Gruber (1993), uma ontologia é uma especificação formal de uma conceitualização. Na filosofia, o termo “ontologia” é uma disciplina que estuda os seres ou a existência das coisas. Como ontologia é utilizada em diversas áreas de pesquisa, existem inúmeras variações de sua conceitualização.

Em Ciência da Computação, ontologias são modelos de representação que descrevem um domínio em termos e parâmetros formais, definindo conceitos com seus respectivos relacionamentos, restrições e demais regras, de tal forma que sua estrutura pode ser facilmente analisada por programas de computador. Diversos sistemas e aplicações utilizam ontologias para compartilhar e reaproveitar conhecimento de um determinado domínio, como um assunto na área de educação, comércio ou saúde.

As ontologias estruturam-se e funcionam através de:

- Classes;
- Instâncias;
- Relacionamentos entre classes;
- Restrições e regras;
- Propriedades.

3.1 CLASSES

Em ontologias, classes estão associadas a conceitos de um determinado domínio e são organizadas em uma taxonomia. Por exemplo, em uma ontologia sobre dependência tecnológica de objetos digitais têm-se as classes Hardware, Software, Formato_Arquivo, Midia_Armazenamento e Sistema_Leitor. A Figura 14 apresenta uma ontologia sobre dependência tecnológica de objetos digitais.

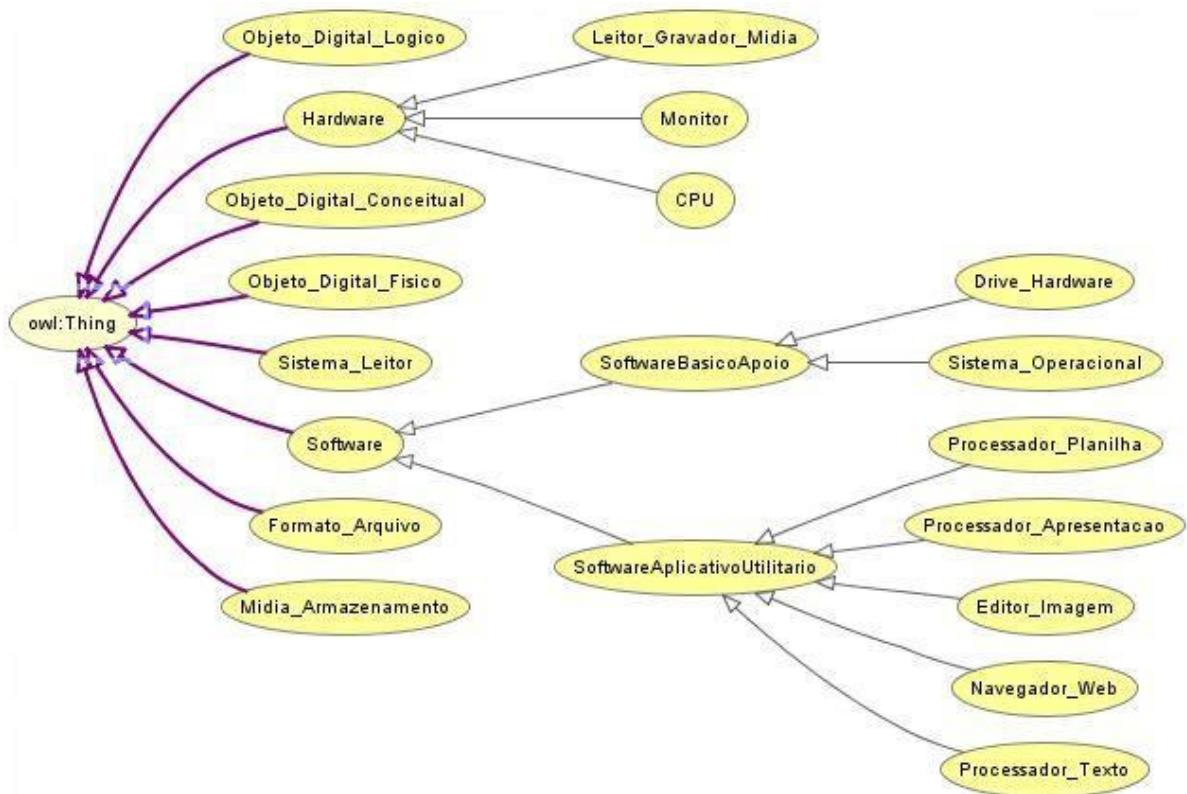


Figura 14 - Ontologia de dependência tecnológica de objetos digitais

A hierarquia de classes na definição de uma ontologia possibilita a definição de relacionamentos de generalização e especialização. Por exemplo as classes Monitor e

CPU são do tipo *Hardware* e por isso herdam todas as propriedades de *Hardware*, que é sua superclasse. Em uma taxonomia típica não existe a possibilidade de herança múltipla, ou seja, uma classe não pode ser filha de duas classes ao mesmo tempo. A maior parte dos sistemas de representação de conhecimento permite herança múltipla na definição da hierarquia de classes (NOY, 2001).

3.2 PROPRIEDADES

Também conhecidas como atributos, as propriedades são utilizadas para descrever características e relacionamentos entre classes. No exemplo de classes sobre dependência tecnológica de objetos digitais, podem existir atributos como espaço de armazenamento, versão do software, tipos de tecnologias de hardware e tipos de materiais de monitor. O tipo de uma propriedade indica quais valores associados a cada propriedade podem ser utilizados. Os tipos mais comuns são *texto*, *número*, *lógico*, *pré-definido* ou *instância*. A cardinalidade de uma propriedade define a quantidade de valores que ela pode ter. Essa definição é semelhante à cardinalidade aplicada aos modelos Entidade-Relacionamento de banco de dados relacionais (MARTINS, 2007).

3.3 INSTÂNCIAS

Instancias são representações reais de classes. Enquanto classes definem coisas genéricas, as instâncias de uma determinada classe representam indivíduos específicos daquela classe (MARTINS, 2007). Um exemplo de instância de uma classe Monitor da Figura 14 é o Monitor LED. Por meio da taxonomia da ontologia sabe-se que Monitor LED é um monitor e que difere de outros equipamentos tecnológicos.

3.4 RELACIONAMENTOS ENTRE AS CLASSES

Uma relação estabelece um tipo de interação entre conceitos de um determinado domínio. As relações podem ser binárias, temporais (implicam em uma ordem cronológica) ou topológica (MARTINS, 2007).

3.5 RESTRIÇÕES E REGRAS

As restrições, também conhecidas como axiomas, complementam os relacionamentos de generalização e especialização criados pela taxonomia, modelando verdades que sempre se cumprem na realidade representada pela ontologia (Becker, 2006). A definição de axiomas em ontologias é tão expressiva que pode ser utilizada para sua classificação, proposta por Sowa (1997):

- *Ontologias Formais*: Possuem suas características restringidas por definições ou axiomas expressados em lógica.
- *Ontologias Terminológicas*: Os relacionamentos e conceitos não necessitam ser totalmente especificados através de axiomas e definições que determinam as condições necessárias para sua utilização.

3.6 CLASSIFICAÇÃO DE ONTOLOGIAS

Segundo Breitman (2005), as ontologias podem ser classificadas quanto ao espectro semântico, à generalidade e quanto ao tipo de informação que representam. De acordo com Lassila e McGuinness (2001), para ser considerada uma ontologia, esta deve possuir:

1. Vocabulário de termos compreensivo e finito;
2. Relacionamentos hierárquicos entre classes bem definidos;
3. Interpretação estritamente bem definida para que não haja ambiguidade entre classes e relacionamentos;
4. Especificação das propriedades nas classes;
5. Inclusão de instâncias de classes;
6. Definição de restrições;
7. Especificação de classes disjuntas;
8. Especificação de relacionamentos lógicos arbitrários entre termos.

Segundo Guarino (1998), a classificação dos tipos de ontologias divide-se em:

- **genéricas:** Descrevem conceitos mais gerais, de caráter abstrato e aplicável a vários outros domínios. São desenvolvidas com o intuito de serem reutilizadas na definição de uma ontologia de domínio;
- **de domínio:** Aplicável a um domínio de conhecimento, descrevendo conceitualizações particulares. É o tipo de ontologia mais comumente desenvolvida.
- **de tarefas:** Descrevem o vocabulário associado a uma tarefa ou atividade, independente do domínio;
- **de aplicação:** Descrevem conceitos associados a uma tarefa ou atividade, contudo, os conceitos são dependentes do domínio e da tarefa representados.

3.7 LINGUAGENS PARA REPRESENTAÇÃO

Para representar o conhecimento de forma que este seja utilizado por sistemas computacionais, é estritamente necessária a utilização de uma linguagem formal para construção de ontologias. Muitas linguagens foram desenvolvidas para representação formal de ontologias, baseadas em frames, lógica descritiva, lógica de primeira ordem ou XML (*eXtensible Markup Language*), porém com diferentes expressividades e propriedades computacionais. Nas subseções seguintes, apresentam-se três linguagens de representação.

3.7.1 RDF

RDF é uma aplicação de *XML* definido pelo *W3C* e sua principal função é prover um modelo formal de dados e sintaxe para codificar metadados que podem ser processados por computadores (Breitman, 2005). Esta linguagem foi proposta pelo *W3C* no ano 2000 e fornece as primitivas básicas para a criação de ontologias, incluindo relacionamentos de generalização de classes. Conforme Mcfarlane (2004), três exemplos de categorias de tipos de informação são: conteúdo, dados e fatos; cada uma processada de maneira diferente. A informação de conteúdo geralmente é processada como um todo, dados são processados por partes e fatos são itens de dados em forma de instruções. Com *RDF* é possível identificar recursos e suas respectivas propriedades. A Figura 15 ilustra

um trecho de código correspondente à identificação de recursos através de uma URI (*Uniform Resource Identifier*).

```
<Description about="http://www.augusto.com/estudo/docs" open="true"/>
```

Figura 15 - Especificação de um recurso em RDF.

A codificação ilustrada pela Figura 15 exemplifica que existe um recurso chamado “http://www.augustolustosa.com/estudo/docs” e este possui uma propriedade “open” cujo valor é “true”. Os recursos podem ser qualquer tipo de coisa com um identificador e as propriedades descrevem as características e relacionamentos entre recursos.

3.7.2 RDF Schema

Segundo Breitman (2005), o *RDF Schema* não fornece as classes e propriedades propriamente ditas, mas um *framework* no qual é possível descrevê-las. As classes em *RDF Schema* são muito semelhantes ao conceito de classes em programação orientada a objetos. *RDF Schema* é uma extensão semântica do *RDF* e seu objetivo é fornecer uma linguagem para descrição de vocabulários a partir do *RDF*. A linguagem especifica uma série de classes fundamentais e a partir dessas classes é realizada a modelagem dos conceitos, propriedades e seus relacionamentos. Essas classes são:

- ***rdfs:Resource***: Classe de todos os recursos, ou seja, todas as descrições em RDF são recursos;
- ***rdfs:Class***: É a classe mãe de todas outras classes;
- ***rdfs:Literal***: Classe de todos os valores literais;
- ***rdfs:Datatype***: Classe de todos os tipos de dados definidos na especificação de RDF;

- ***rdf:Property***: É a classe das propriedades RDF. *rdf:Property* é uma instância de *rdfs:Class*.

Segundo Becker (2006), um déficit no desenvolvimento de ontologias com RDF está no fato de não possuir recursos para modelagem de axiomas e descrição da semântica referente aos conceitos e relações. Outra limitação é a inexistência de expressões que representem conectivos lógicos, negação, disjunção e conjunção. Diante destas limitações, foi desenvolvida uma nova camada acima de RDF e *RDF Schema* para conseguir um nível de expressividade semântica mais significativa, a OWL (Breitman, 2005).

3.7.3 OWL – *Ontology Web Language*

A OWL é uma linguagem criada pelo W3C a partir da linguagem DAML+OIL, sendo desenvolvida com base nas necessidades da Web Semântica (OWL, 2013). Ontologias desenvolvidas na linguagem OWL devem possuir as seguintes características (OWL, 2013):

- Ser distribuídas através de diferentes sistemas;
- Escalabilidade conforme necessidades da Web;
- Compatibilidade com padrões da Web referentes à acessibilidade e internacionalização.

A linguagem OWL adiciona novas formas de expressão de vocabulário nas classes e propriedades, a partir das existentes na camada *RDF*; portanto OWL depende das primitivas em camadas anteriores, se utilizando das mesmas. Existem três tipos de tipos de OWL, variando conforme o grau de complexidade, conforme OWL (2013):

- *OWL Lite*: suporta ontologias simplificadas que precisam de uma classificação hierárquica e restrições simples. Por exemplo, suporta restrições de cardinalidade

podendo só permitir valores de 0 ou 1. A *OWL lite* também proporciona um rápido caminho de migração para léxicos e outras classificações.

- *OWL DL*: suporta maior expressividade e poder computacional. É indicada a usuários que necessitam máxima expressividade sem perder a completude computacional e decidibilidade. A *OWL DL* inclui todos os construtores da linguagem *OWL*, mas eles podem ser utilizados sob certas restrições.
- *OWL Full*: Indicada para usuários que necessitem máxima expressividade e a flexibilidade sintática do *RDF*, sem garantias computacionais. Por exemplo, uma classe *OWL Full* pode ser tratada simultaneamente como um conjunto de indivíduos (instâncias) e como um indivíduo propriamente dito. *OWL Full* permite uma ontologia expandir o significado do vocabulário pré-definido.

3.8 FERRAMENTAS PARA O DESENVOLVIMENTO DE ONTOLOGIAS

O desenvolvimento de ontologias não é uma tarefa trivial, já que torna explícito um conhecimento relacionado a um contexto. Dessa forma, várias ferramentas foram desenvolvidas para auxiliar o desenvolvimento de ontologias, tais como: WebOnto (DOMINGUE, 1998), WebODE (CORCHO et al., 2005), OntoKEM (RAUTENBERG et al., 2010) e Protégé (NOY; McGUINNESS, 2001).

A utilização da ferramenta Protégé tem se destacado por facilitar a construção de modelos de domínio e aplicações baseadas em conhecimento por meio de ontologias. A ferramenta Protégé foi inicialmente desenvolvida pelo Stanford Medial Informatics (SMI), um grupo de pesquisa interdisciplinar do Departamento de Medicina de Stanford University School of Medicine. O SMI reúne cientistas que criam e validam métodos para aquisição, representação e gestão de conhecimento e dados na área de saúde (PROTÉGÉ, 2000). A ferramenta Protégé suporta dois tipos de modelagem de ontologias: Protégé-Frames e Protégé-OWL.

3.8.1 PROTÉGÉ-FRAMES

Esse editor é usado para a modelagem de ontologias usando frames. Suporta a construção e armazenamento das ontologias modeladas além de possuir um vasto conjunto de elementos de interface que facilitam a criação de formulários amigáveis para entrada de informação na ontologia (PROTÉGÉ, 2000).

3.8.2 PROTÉGÉ-OWL

O Protégé-owl é uma extensão para o desenvolvimento de ontologias para a Web semântica, definidas em OWL. É integrado ao motor de inferência JENA e permite (PROTÉGÉ, 2000):

- Visualizar e editar classes, propriedades;
- Executar motores de inferência;
- Editar indivíduos em OWL;
- Carregar ontologias RDF e OWL.

A Figura 16 apresenta a tela de edição de classes no Protégé, com a classe Pizza selecionada e as classes especializadas de Pizza.

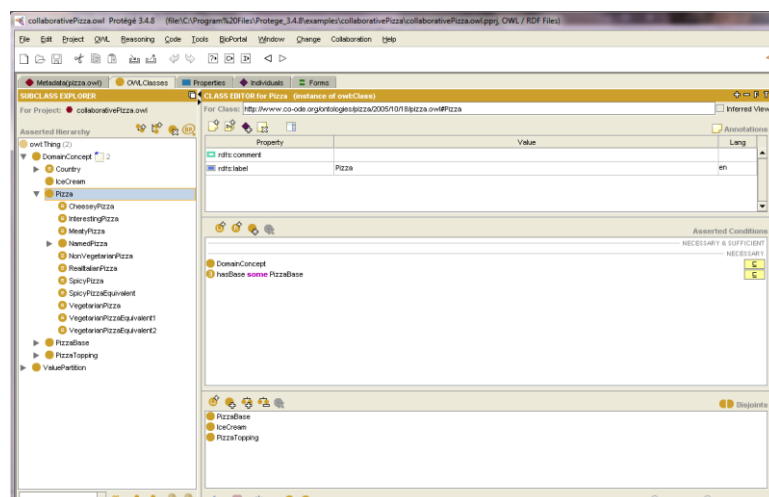


Figura 16 - Tela de edição de classes de ontologias no Protégé.

4 TWEETPY: UM FRAMEWORK BASEADO EM ONTOLOGIAS PARA ANÁLISE DE REDES SOCIAIS ON-LINE

Este capítulo descreve os objetivos do framework Tweepy e a sua arquitetura. Nas Seções 4.1 e 4.2 aborda-se os objetivos e a arquitetura do framework. Já a Seção 4.3 apresenta cenários de uso, onde são apresentados os diferentes contextos em que o framework pode ser utilizado. Por fim, a Seção 4.4 apresenta as ferramentas necessárias para o desenvolvimento do framework.

4.1 OBJETIVOS

O objetivo do framework Tweepy é prover uma infraestrutura computacional para exploração de características intrinsecamente relacionadas à estrutura topológica do grafo formado pela propagação de informação no *Twitter*. Especificamente o framework deve permitir:

1. Carregar grafos em formato de arquivos de texto;
2. Processar os grafos carregados em memória e retornar os valores do processamento;
3. Suportar mudanças nos objetivos de processamento, através de mudanças na ontologia.

O framework proposto foi implementado usando a linguagem de programação Python (Python, 2013), com a utilização de ontologias, utilizando a linguagem OWL. Cada componente do framework que analisa propriedades da estrutura topológica do grafo é encapsulado em classes distintas, o que permite uma melhor depuração de código-fonte.

4.2 ARQUITETURA

A Figura 17 apresenta uma visão geral da arquitetura do framework Tweepy, onde o framework retorna um conjunto de valores correspondentes ao cálculo das métricas que

estão relacionadas à finalidade da requisição, ou seja, no momento da instanciação da classe principal do framework, é passada a finalidade e uma referência ao conjunto de dados a ser analisado. A finalidade está relacionada ao conjunto de métricas a serem analisadas. Para cada finalidade há um conjunto de métricas definidas na ontologia. Dessa forma, verifica-se na ontologia os métodos correspondentes à finalidade requerida e calcula-se as métricas.

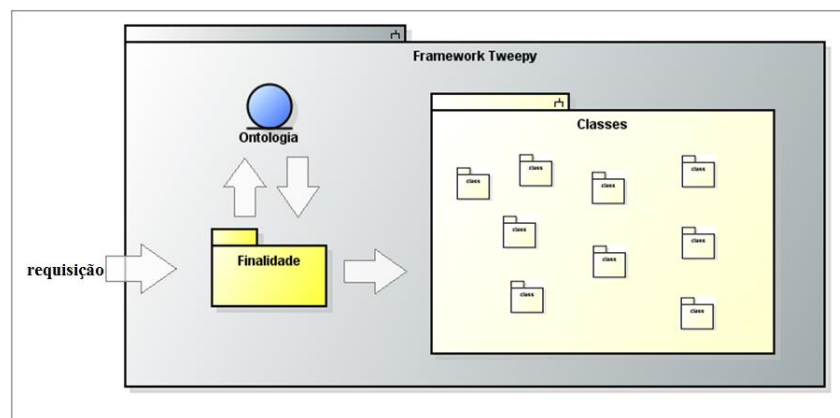


Figura 17 - Visão geral da arquitetura do framework Tweepy.

No diagrama de sequência (UML, 2013) apresentado na Figura 18, é possível visualizar melhor a interação, desde a requisição até o retorno dos valores das métricas calculadas.

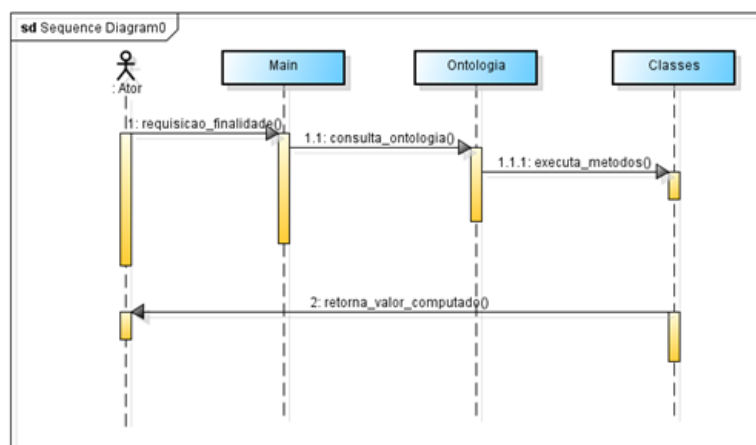


Figura 18 - Diagrama de sequência do Tweepy.

O diagrama apresentado na Figura 18 mostra desde a requisição do usuário, no momento em que faz uma requisição ao framework passando a finalidade de utilização que irá gerar uma consulta à ontologia e execução dos métodos associados à finalidade, até o retorno dos valores das métricas calculadas pelo framework. A ontologia criada possui as classes correspondentes aos diferentes fins de utilização do framework, conforme Figura 19.

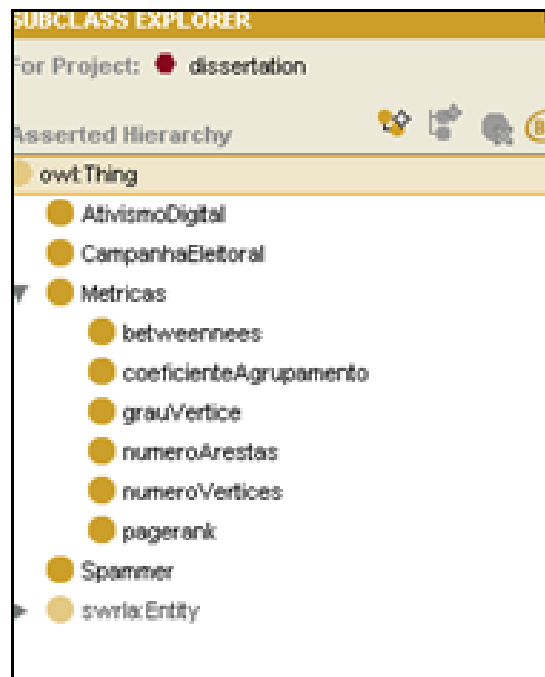


Figura 19 - Estrutura hierárquica de classes da ontologia desenvolvida.

Como pode ser visto na Figura 19, para cada classe existem propriedades de objetos, que representam os relacionamentos entre dois objetos. Dessa forma, para cada classe correspondente a um fim de utilização do framework, existem relacionamentos com outras classes que possuem as métricas desejadas.

A Ontologia apresentada na Figura 20 foi construída no ambiente Protégé 4.1. Após estudo, este ambiente, com seus diversos plug-ins, se mostrou mais adequado para a construção da ontologia. Inicialmente fez-se uma análise de todo o domínio, onde foi possível identificar as métricas mais importantes e relevantes para o contexto das

finalidades. Após a análise, os elementos foram inseridos na ontologia, utilizando o ambiente Protégé. Na Figura 20 os principais elementos da ontologia são mostrados.

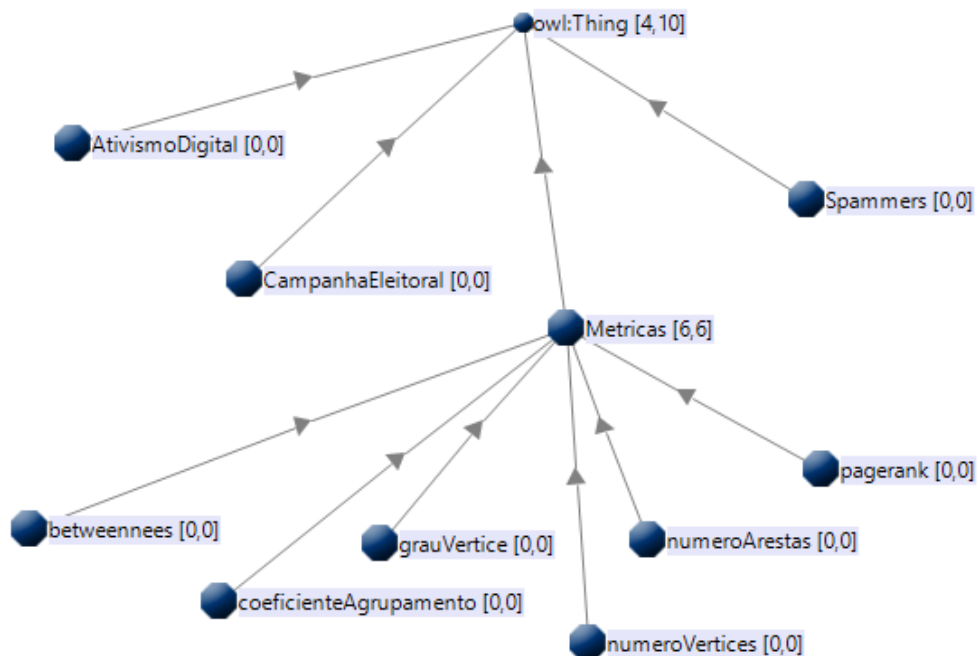


Figura 20 – Visualização gráfica da ontologia.

A Figura 21 ilustra o código desenvolvido para testes do framework, utilizando a linguagem OWL, onde foram definidas as classes e propriedades.

```

1  <?xml version="1.0"?>
2  <rdf:RDF
3      xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
4      xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
5      xmlns:swrla="http://swrl.stanford.edu/ontologies/3.3/swrla.owl#"
6      xmlns="http://www.owl-ontologies.com/Ontology1352354647.owl#"
7      xmlns:swrl="http://www.w3.org/2003/11/swrl#"
8      xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
9      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
10     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
11     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
12     xmlns:owl="http://www.w3.org/2002/07/owl#"
13     xmlns:swvrl="http://swvrl.stanford.edu/ontologies/built-ins/3.4/swvrl.owl#"
14     xml:base="http://www.owl-ontologies.com/Ontology1352354647.owl">
15     <owl:Ontology rdf:about="">
16         <owl:imports rdf:resource="http://swvrl.stanford.edu/ontologies/built-ins/3.4/swvrl.owl"/>
17         <owl:imports rdf:resource="http://swrl.stanford.edu/ontologies/3.3/swrla.owl"/>
18     </owl:Ontology>
19     <owl:Class rdf:ID="CampanhaEleitoral"/>
20     <owl:Class rdf:ID="AtivismoDigital"/>
21     <owl:Class rdf:ID="betweennees">
22         <rdfs:subClassOf>
23             <owl:Class rdf:ID="Metricas"/>
24         </rdfs:subClassOf>
25     </owl:Class>
26     <owl:Class rdf:ID="pagerank">
27         <rdfs:subClassOf rdf:resource="#Metricas"/>
28     </owl:Class>
29     <owl:Class rdf:ID="numeroArestas">
30         <rdfs:subClassOf rdf:resource="#Metricas"/>
31     </owl:Class>
32     <owl:Class rdf:ID="Spammer"/>
33     <owl:Class rdf:ID="numeroVertices">
34         <rdfs:subClassOf rdf:resource="#Metricas"/>
35     </owl:Class>
36     <owl:Class rdf:ID="grauVertice">
37         <rdfs:subClassOf rdf:resource="#Metricas"/>
38     </owl:Class>
39     <owl:Class rdf:ID="coeficienteAgrupamento">
40         <rdfs:subClassOf rdf:resource="#Metricas"/>
41     </owl:Class>
42     <owl:ObjectProperty rdf:ID="CampanhaEleitoralhasMetricas">
43         <rdfs:range>
44             <owl:Class>
45                 <owl:unionOf rdf:parseType="Collection">
46                     <owl:Class rdf:about="#numeroVertices"/>
47                     <owl:Class rdf:about="#numeroArestas"/>
48                     <owl:Class rdf:about="#grauVertice"/>
49                 </owl:unionOf>
50             </owl:Class>
51         </rdfs:range>
52         <rdfs:domain rdf:resource="#CampanhaEleitoral"/>
53     </owl:ObjectProperty>
54     <owl:ObjectProperty rdf:ID="SpammerhasMetricas">
55         <rdfs:domain rdf:resource="#Spammer"/>
56         <rdfs:range>
57             <owl:Class>
58                 <owl:unionOf rdf:parseType="Collection">
59                     <owl:Class rdf:about="#numeroVertices"/>
60                     <owl:Class rdf:about="#coeficienteAgrupamento"/>
61                     <owl:Class rdf:about="#grauVertice"/>
62                 </owl:unionOf>
63             </owl:Class>
64         </rdfs:range>
65     </owl:ObjectProperty>
66     <owl:ObjectProperty rdf:ID="AtivismoDigitalhasMetricas">
67         <rdfs:domain rdf:resource="#AtivismoDigital"/>
68         <rdfs:range>
69             <owl:Class>
70                 <owl:unionOf rdf:parseType="Collection">
71                     <owl:Class rdf:about="#numeroVertices"/>
72                     <owl:Class rdf:about="#betweennees"/>
73                     <owl:Class rdf:about="#numeroArestas"/>
74                     <owl:Class rdf:about="#coeficienteAgrupamento"/>
75                     <owl:Class rdf:about="#pagerank"/>
76                     <owl:Class rdf:about="#grauVertice"/>
77                 </owl:unionOf>
78             </owl:Class>
79         </rdfs:range>
80     </owl:ObjectProperty>
81     <rdf:Description rdf:about="http://www.w3.org/2002/07/owl#Thing">
82         <rdfs:comment xml:lang="pt"><lt;p style="margin-top: 0">>#xD;
83             Ontologia utilizada pelo framework de analise de redes sociais baseado #xD;
84             em ontologias: TweePy #xD;
85         </p>></rdfs:comment>
86     </rdf:Description>
87 </rdf:RDF>
88
89 <!-- Created with Protege (with OWL Plugin 3.4.8, Build 629) http://protege.stanford.edu -->
90

```

Figura 21 - Ontologia desenvolvida no presente trabalho.

Já na Figura 22 é apresentada a hierarquia de módulos desenvolvidos no framework.

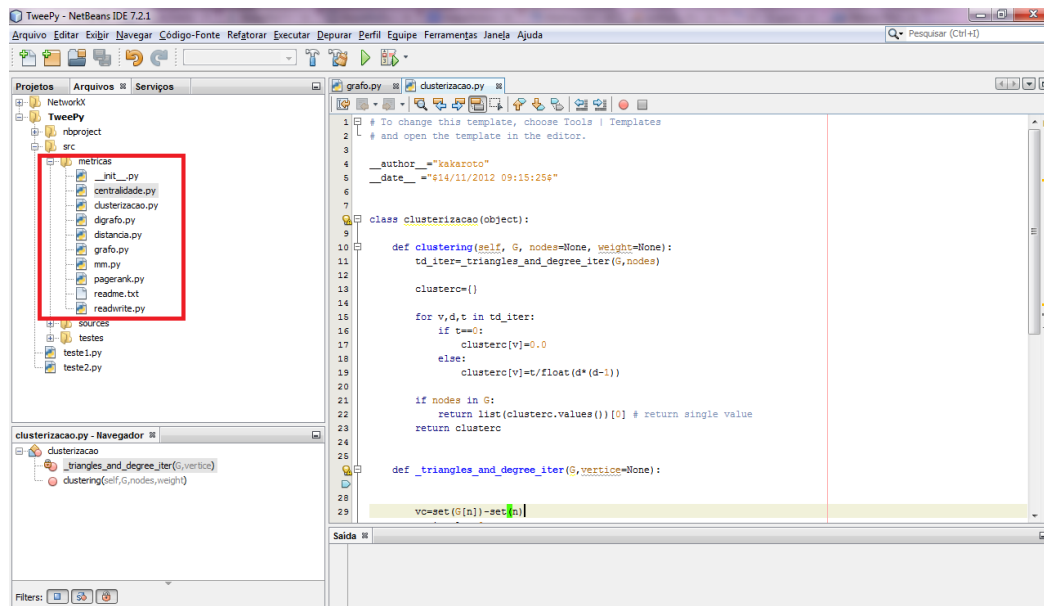


Figura 22 - Hierarquia dos módulos desenvolvidos no framework.

Como pode ser visto na Figura 22, para cada métrica a ser utilizada, há um módulo específico contendo classes e métodos responsáveis por computar a métrica desejada. Já a Figura 23 mostra o trecho de código-fonte responsável pela estrutura de armazenamento, em memória, do grafo coletado do *Twitter*.

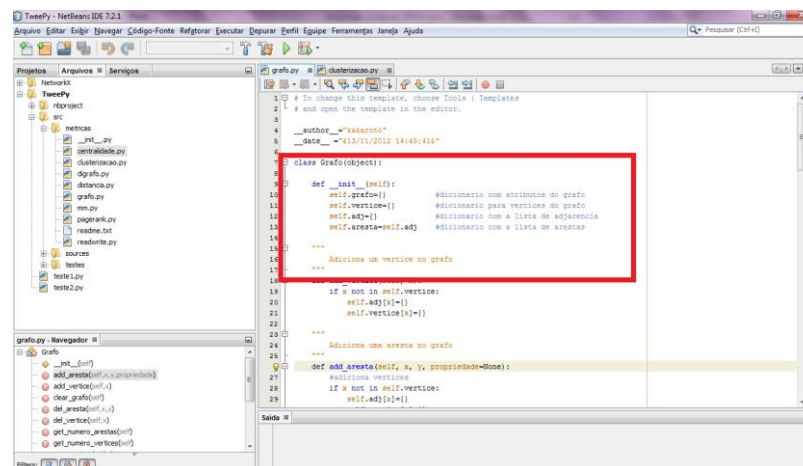


Figura 23 - Parte do código-fonte do framework Tweepy responsável pelo armazenamento do grafo em memória.

A Figura 23 ilustra a classe Grafo, onde se define, através da estrutura de dados *dicionário*, nativa da linguagem *Python*, dicionários onde serão armazenados os vértices, arestas, atributos do grafo e a lista de adjacência (linhas 9 a 13). Ainda na classe Grafo, são definidos métodos de inserção e exclusão de vértices e arestas; obtenção de vizinhos de um vértice; verificação de existência de arestas e vértices; e obtenção do grau de um vértice, conforme pode ser visto na Figura 24.

```

18 def add_vertice(self, x):
19     if x not in self.vertice:
20         self.adj[x]={}
21         self.vertice[x]={}
22 def add_aresta(self, x, y, propriedade=None):
23     #adiciona vertices
24     if x not in self.vertice:
25         self.adj[x]={}
26         self.vertice[x]={}
27     if y not in self.vertice:
28         self.adj[y]={}
29         self.vertice[y]={}
30     #adiciona arestas
31     data=self.adj[x].get(y,propriedade)
32     self.adj[x][y]=data
33     self.adj[y][x]=data
34 def del_vertice(self, x):
35     try:
36         lista_keys=list(self.adj[x].keys())
37         del self.vertice[x]
38     except Error:
39         print "O vertice nao esta contido no grafo"
40
41     for i in lista_keys: #deleta todas arestas que contem (1..10,x)
42         del self.adj[i][x]
43     del self.adj[x] #deleta toda a aresta (x, 1..10)
44
45 def del_aresta(self, x, y):
46     try:
47         del self.adj[x][y]
48         if x!=y:
49             del self.adj[y][x]
50     except Error:
51         print "A aresta nao esta contida no grafo"
52 def clear_grafo(self):
53     self.vertice.clear()
54     self.adj.clear()
55     self.grafo.clear()
56 def verifica_existe_vertice(self, x):
57     try:
58         return x in self.vertice
59     except:
60         return False
61 def verifica_existe_aresta(self, x, y):
62     try:
63         return y in self.adj[x]
64     except:
65         return False
66 def get_vizinhos(self,x):
67     try:
68         return list(self.adj[x])
69     except:
70         return False
71
72 def grau(self, x):
73     if x in self.vertice:
74         return len(self.adj[x])

```

Figura 24 - Parte do código-fonte do framework TweepPy correspondente aos métodos desenvolvidos na classe Grafo.

A Figura 25 ilustra o trecho de código responsável pela busca, na ontologia, dos métodos associados à finalidade da requisição do framework.

```
46
47 def requisicao_finalidade(self, finalidade):
48     rede = file("ontologia-dissertacao.owl")
49     soup=BeautifulSoup(rede.read())
50     search = soup.findAll('owl:ObjectProperty',{'rdf:ID':finalidade})
51     lista= search.find('rdf:about')
52     return lista
53
```

Figura 25 - Parte do código-fonte do framework Tweepy correspondente à busca na ontologia dos métodos associados à finalidade de uso do framework.

Na Figura 25, o método *requisição finalidade* recebe a finalidade que o usuário do framework deseja (linha 47), carrega em memória a ontologia (linha 49), faz uma busca na ontologia, onde existe uma *tag* "owl:ObjectProperty" com atributo "rdf:ID" possuindo a *string* passada como parâmetro no método (linha 50). O retorno da busca é utilizado, conforme ilustrado na linha 51, para extrair os nomes correspondentes aos métodos associados à finalidade desejada.

```
1  fille = open("tweets.txt","r")
2  grafo = open("grafo.txt","w")
3
4  lines = fille.readlines()
5
6  i=0
7  a=0
8  usuario1=""
9  usuario2=[]
10 while i < len(lines):
11     usuario1=lines[i]
12     i=i+2
13     tweet = lines[i]
14     x=0
15     while x < len(tweet):
16         if tweet[x] == '@':
17             a=x
18             usuario=""
19             while tweet[a] != " ":
20                 usuario=usuario+tweet[a]
21                 a=a+1
22             usuario2.append(usuario)
23             x=x+1
24     for t in usuario2:
25         grafo.write(usuario1+" "+t+"\n")
26
27     usuario2=[]
28     i=i+2
29
30 fille.close()
31 grafo.close()
```

Figura 26 - Parte do código-fonte responsável pelo pré-processamento dos dados.

O código ilustrado na Figura 26 é responsável pelo pré-processamento dos dados advindos do *Twitter*. Esse código tem a função de extrair, em meio ao conjunto de *tweets* coletados, a rede de menções entre usuários do *Twitter*. Dado um conjunto de *tweets*, extrai-se o nome do usuário que publicou o *tweet* e verifica-se no corpo da mensagem, usuários mencionados. O processo de extração de usuários mencionados se dá através da localização do caractere '@' seguido de caracteres diferentes do caractere ' '(linhas 16 e 19 na Figura 26). Dessa forma, se um usuário *X* publica um *tweet* citando dois ou mais usuários, a saída do processamento haverá dois ou mais arestas representadas num arquivo no formato de texto, que representa o grafo a ser analisado pelo framework.

4.3 CENÁRIOS DE USO

Os seguintes cenários apresentam, de forma resumida, as possibilidades de utilização do framework Tweepy:

C1: As redes sociais on-line possibilitaram, em grande escala, o registro de diversos aspectos intrínsecos ao comportamento humano. Alguns desses comportamentos podem ser compreendidos como formas de promoção de manifestações na Web, representadas na forma de ativismo digital. Portanto, a utilização do framework possibilitaria a inferência de algumas características ligadas à formas de ativismo digital.

C2: Através da utilização do framework, é possível obter informações pertinentes à acurácia de campanhas publicitárias no *Twitter*, verificando quais usuários estão mais conectados com os outros na rede.

C3: Com a utilização do framework, pode-se caracterizar possíveis usuários que utilizam o *Twitter* para disseminar *spam*, através da inferência de algumas propriedades intrínsecas às atividades de *spammers*.

4.4 IMPLEMENTAÇÃO COMPUTACIONAL

Para implementação e testes do framework *TweePy* foram utilizadas as seguintes ferramentas:

- Linguagem Python versão 2.5 (Python, 2013): Linguagem de programação de propósito geral, frequentemente utilizada como linguagem de *script* para aplicações via Web. Toda parte funcional do framework foi desenvolvida em Python.
- Eclipse versão 4.2 (ECLIPSE, 2013): Ambiente integrado de desenvolvimento, em inglês IDE (*Integrated Development Enviroment*), usado para criação do framework. É amplamente extensível, com o uso de *plugins*.
- Protege-OWL Editor versão 3.4.8 (PROTEGE, 2013): Editor de ontologias OWL, extensível através de *plugins*.

5 AVALIAÇÃO EXPERIMENTAL

Em continuidade ao capítulo anterior, onde foi feita uma descrição conceitual da arquitetura do framework, este capítulo explora o funcionamento do framework proposto nesta dissertação, sob três conjuntos de dados do *Twitter*, sob contextos bem distintos. Os dados foram coletados a partir da *Search API* (TWITTER SEARCH, 2013) armazenados em arquivo no formato de texto. Posteriormente foi feito um pré-processamento dos dados com código ilustrado na Figura 26. Assume-se que *tweets* que contenham a mesma *hashtag*, num espaço curto de tempo, estão agregados a um contexto comum. Cada *hashtag* escolhida para coleta possuía contextos bem distintos, conforme descrito a seguir:

- *#Iphone4SdaTIM*: Essa *hashtag* surgiu em decorrência de uma promoção da empresa de telefonia móvel TIM que sorteava um aparelho celular para usuários que publicassem *tweets* com a *hashtag* "*#Iphone4SdaTIM*".
- *#marimar*: Essa *hashtag* foi publicada durante a exibição de uma novela na TV, cuja personagem se chamava Marimar.
- *#contraOaumento*: Essa *hashtag* surgiu em decorrência de uma série de manifestações na cidade Teresina, capital do Piauí, onde estudantes protestavam contra o aumento no valor cobrado pelas passagens de ônibus na cidade.

As Figuras 27, 28 e 29 ilustram o grafos formados pelos três conjuntos de dados. Os vértices dos grafos correspondem ao nome identificador dos usuários do *Twitter* e as arestas às interações entre os usuários. Os gráficos foram gerados através da ferramenta NodeXL usando o algoritmo Fruchterman-Reingo.

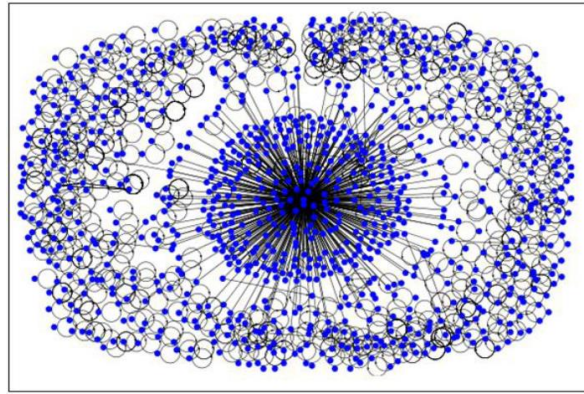


Figura 267 - Rede formada pela hashtag #Iphone4SdaTIM.

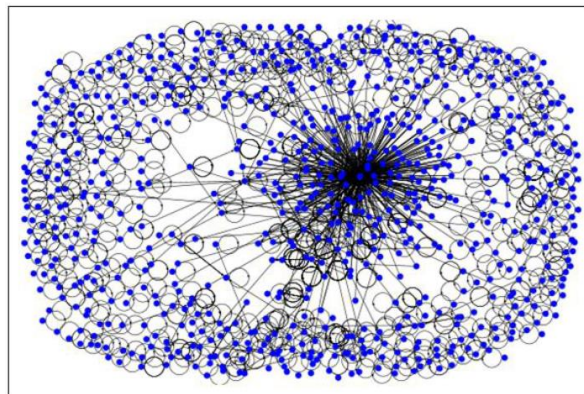


Figura 278 - Rede formada pela hashtag #Marimar.

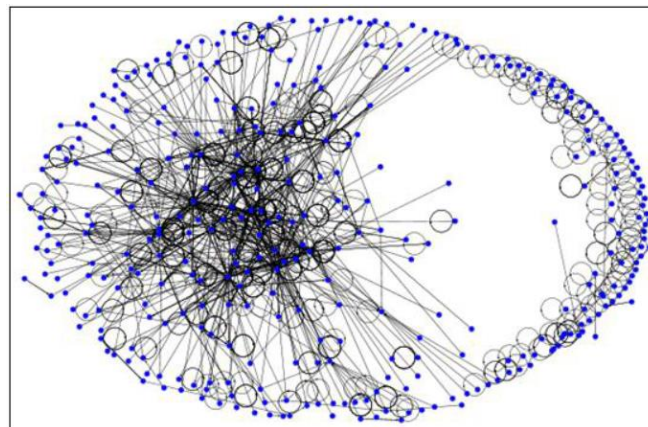


Figura 289 - Rede formada pela hashtag #contraOaumento.

A Figura 30 apresenta um gráfico e uma tabela que explicitam a relação entre a quantidade de usuários extraída dos *tweets* coletados com a quantidade de interações entre os mesmos.

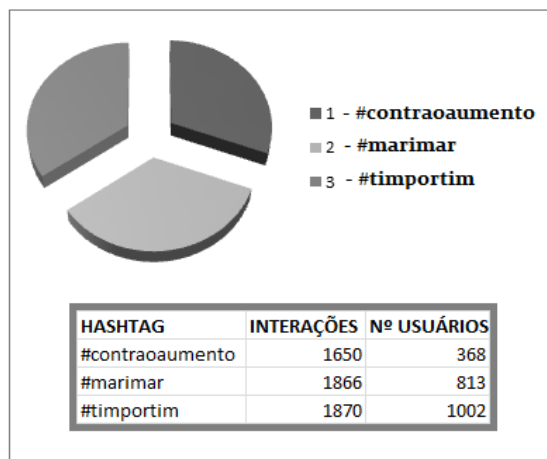


Figura 29 - Número de interações pelo número de usuários.

Na Figura 30, observa-se que o grafo formado pela rede de menções de usuários que publicaram *tweets* com a *hashtag* "#contraoamento" possui uma quantidade inferior a 400 vértices, enquanto possui quase o mesmo número de arestas que os outros grafos oriundos das outras *hashtags*. Dessa forma constata-se que, devido ao contexto ao qual a *hashtag* "#contraoamento" estava inserida, houveram muito mais interações entre usuários.

Verificou-se que o grafo formado pelos *tweets* com a *hashtag* "#Iphone4SdaTIM" possui um único vértice com grau de entrada superior aos maiores graus de entrada encontrados nos outros grafos, evidenciando uma polarização de interações com um único usuário da rede, conforme pode ser visto na Figura 31.

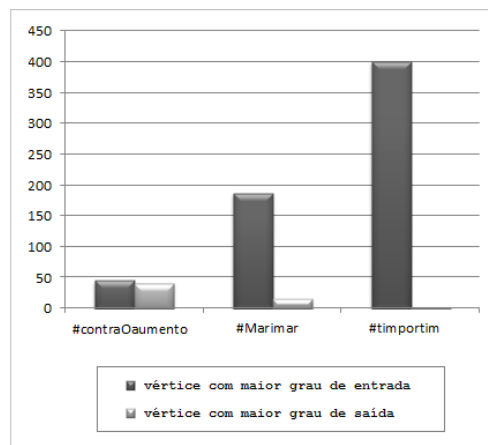


Figura 31 - Relação entre o vértice com maior e menor grau.

Como apresentado na Figura 31, ressalta-se o fato do grafo formado pela *hashtag* "#contraOaumento" possuir uma singela variação entre o vértice com maior grau de entrada e o vértice com maior grau de saída, refletindo em um grafo com maior raio.

A Figura 32 apresenta os máximos valores encontrados na propriedade *betweenness centrality* para cada grafo. Constata-se que o grafo formado pela *hashtag* # *Iphone4SdaTIM* possui um vértice com a propriedade *betweenness centrality* maior que nos outros grafos, refletindo em um vértice no qual passam muitos caminhos mínimos entre pares de nós.

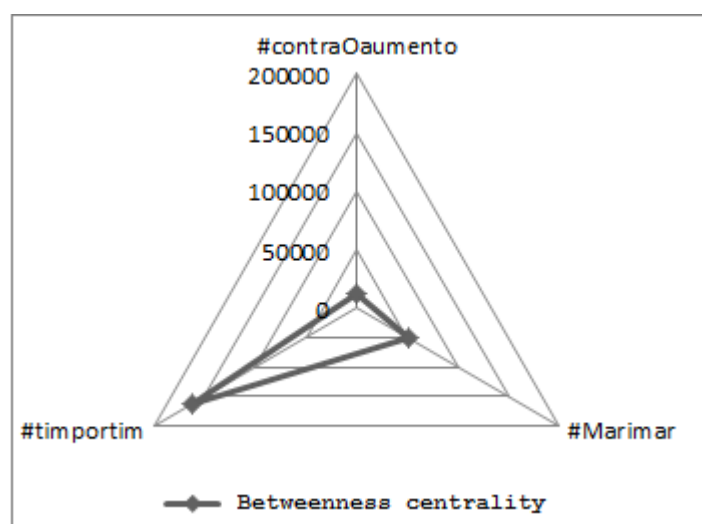


Figura 302 - Máximos valores encontrados na propriedade betweenness centrality.

O conjunto de dados relacionado à *hashtag* #Iphone4SdaTIM originou-se de uma campanha publicitária na Web. Já a *hashtag* #Marimar surgiu em decorrência da reexibição de uma novela em um canal brasileiro de televisão, ou seja, houve uma influência externa ao *Twitter* para que usuários publicassem *tweets* com a *hashtag*. Por fim, a *hashtag* #contraOaumento surgiu no âmbito das manifestações de estudantes contra o aumento nas passagens de ônibus na cidade de Teresina, onde os manifestantes se organizaram através das redes sociais on-line, publicando informações e imagens acerca do protesto.

Dessa forma, esse trabalho apresentou o framework baseado em ontologias, exemplificou o seu uso através de um estudo de caso, onde foram analisados três conjuntos de dados.

6 CONCLUSÕES E TRABALHOS FUTUROS

Com o rápido crescimento da Web, vários aspectos do cotidiano das pessoas foram modificados. Consultas a informações ficaram mais rápidas e dinâmicas. A partir do advento das redes sociais on-line, a Web passou a ser um meio pelo qual usuários passaram a se comunicar, expressar suas opiniões e discutir sobre um determinado assunto. Dessa forma, as redes sociais on-line passaram a fazer parte do cotidiano de milhões de usuários no mundo. Entender a dinâmica das interações entre os usuários em redes sociais on-line pode criar oportunidades para melhorar o funcionamento dos sistemas de redes sociais on-line, possibilitando a extração de características que possam servir como *feedback* na implementação de novas funcionalidades voltadas ao público das redes sociais on-line. Além disso, é importante para diversas aplicações relacionadas a propagandas, campanhas políticas e detecção de comportamento malicioso.

Nesse contexto, este trabalho apresenta um framework que propicia uma análise de dados de redes sociais on-line sob diferentes contextos, através do uso de ontologias em sua composição. Dessa forma, há como se verificar, computacionalmente, o propósito de *tweets* publicados. Analisaram-se três conjuntos de dados e verificou-se que o conjunto de dados com características de ativismo digital possuía propriedades diferentes dos outros dois conjuntos de dados, evidenciando computacionalmente algumas propriedades que podem servir de parâmetros para aferir ativismo digital no *Twitter*.

Apesar do trabalho realizado, há algumas melhorias que podem ser realizadas. Assim, como trabalhos futuros, pretende-se:

- Desenvolver um módulo para visualização dos grafos.
- Desenvolver uma plataforma distribuída para processamento de grandes quantidades de dados.
- Desenvolver um algoritmo para eliminar mensagens coletadas que sejam *spam*.
- Expandir o framework para outros contextos em redes sociais on-line.

REFERÊNCIAS

Albert, R., Jeong, H., and Barabasi, A. (1999). Diameter of the world wide web. *Nature*, 401:130--131.

ALBERT, R.; BARABÁSI, A.-L. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, v. 74, n. 1, p. 47-97, 2002.

Amaral, A., Scala, A., Barthelemy, M., and Stanley, E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Science (PNAS)*, 97(21):11149--11152.

Barabasi, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, pages 286--509. 6, 8

BECKER, Júnior. Ontologia Terminológica para apoio a ferramentas de Recuperação de informações e de Text Mining. 2006. Projeto de Diplomação (Bacharelado em Ciências da Computação) – Instituto de Ciências Exatas e Tecnológicas (ICET), Centro Universitário FE-EVALE, Novo Hamburgo.

BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D.-U. Complex networks: Structure and dynamics. *Physics Reports*, v. 424, p. 175-308, 2006.

Boyd, D. and Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1-2).

BREITMAN, Karin Koogan. Web semântica: A internet do futuro. Rio de Janeiro: LTC, 2005.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000). Graph structure in the web. *Computer Networks*, 33:309--320.

Burns A.; Eltham B. (2009). *Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operation in Iran's 2009 election crisis*. In: Communications Policy & Research Forum 2009, 19th-20th November 2009, University of Technology, Sydney.

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. Measuring User Influence in *Twitter*: The Million Follower Fallacy. In Proceedings of the Int'l AAAI Conference on Weblogs and Social Media (ICWSM).

ContraOaumento. Teresina tem novos protestos contra aumento da passagem de ônibus, 2012. <http://g1.globo.com/jornal-nacional/noticia/2012/01/teresina-tem-novas-protestos-contr-aumento-da-passagem-de-onibus.html>

CORCHO et al. Building legal ontologies with Methondology and WebODE. Law and the Semantic Web, Springer-Verlag, Mar, 2005.

COSTA, L. F.; RODRIGUES, F. A.; TRAVIESO, G.; VILLAS BOAS, P. R. Characterization of complex networks: A survey of measurements. *Advances in Physics*, v. 56, n. 1, p. 167-242, 2007.

DACONTA, Michael C.; OBRST, Leo J.; SMITH, Kevin T. The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. Wiley Publishing: 2003. 281p.

Disponível em: <http://www.dia.fi.upm.es/~ocorcho/documents/LawSemWeb2004_CorchoEtAl.pdf>. Acesso em: 12 jan. 2013.

DOMINGUE, J. Tadzebao and WebOnto: discussing, browsing and editing ontologies on the web. In: Proceedings of the 11th Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Canada, 1998. Disponível em: <<http://kmi.open.ac.uk/publications/pdf/kmi-98-12.pdf>>. Acesso em: 5 fev. 2013.

ECLIPSE. 2013. “Eclipse - an open development platform”. Disponível em: <http://>

ECLIPSE. www.eclipse.org. Acesso em 07 fev. 2013.

ERDÖS, P.; RÉNYI, A. On random graphs. *Publicationes Mathematicae*, v. 6, p. 290–297, 1959.

ERDÖS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, v. 5, p. 17–61, 1960.

ERDÖS, P.; RÉNYI, A. On the strenght of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, v. 12, p. 261–267, 1961.

F. Benevenuto, Jussara, M. Almeida, Altigran. S. Silva. Coleta e Análise de Grandes Bases de Dados de Redes Sociais On-line, 2010.

Freeman L. C. THE DEVELOPMENT OF SOCIAL NETWORK ANALYSIS: A STUDY IN THE SOCIOLOGY OF SCIENCE. Empirical Press Vancouver, BC Canada, 2004.

GRUBER, THOMAS R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5: 199-220, 1993.

GUARINO, Nicola. Formal Ontology and Information Systems. In: Proceedings of the FOIS'98. Formal Ontology in Information Systems, Trento, 1998.

Jarry Richardson R. Pesquisa Social: Métodos e Técnicas. Edidora Atlas S.A, 1999.

L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford University, 1998.

LASSILA, ORA; MCGUINNESS, DEBORA. The Role of Frame-Based Representation on the Semantic Web. In Electronic Transactions on Artificial Intelligence, 2001. Volume 6, Number 005: 2001-03-07. Disponível em: <<http://www-ksl.stanford.edu/people/dlm/etai/etai-abstract.html>>. Acesso em: 5 Fev. 2013.

Liu, B. (2009). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, 1st ed. 2009. corr. 2nd printing edição.

M. Cha, H. Haddadi, F. Benevenuto, K. P. Gummadi. Measuring User Influence in *Twitter*: The Million Follower Fallacy. 4th Int'l AAAI Conference on Weblogs and Social Media, 2010.

MARTINS, Fernando. ONTOLOGIAS NA ANÁLISE DE BLOGS. 2007. Trabalho de conclusão de curso. Projeto de Diplomação (Bacharelado em Ciências da Computação) – Instituto de Ciências Exatas e Tecnológicas (ICET), Centro Universitário FE-EVALE, Novo Hamburgo.

MCFARLANE, Nigel. Rapid Application Development with Mozilla. Prentice Hall: 2004. 770p.

Milgram, S. (1967). The small world problem. *Psychology Today*, 2:60--67.

Mislove, A., Marcon, M., Gummadi, K., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of on-line social networks. In Proceedings of the ACM SIGCOMM Conference on Internet Measurement (IMC), pages 29--42.

Newman, M. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Science (PNAS), 101(1):5200—5205, 2004.

Newman, M. E. J. The Structure and Function of the Complex Network. SIAM Review, v. 45, n. 2, p. 167-256, 2003b.

Newman, M. The structure of scientific collaboration networks. Proceedings of the National Academy of Science (PNAS), 98(2):404—409; 2001.

NOY, N. F.; MCGUINNESS, D. L. Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory, Technical Report KSL-01-05, 2001.

OWL. Web Ontology Language OWL/W3C Semantic Web Activity. 2013. Disponível em: <<http://www.w3.org/2004/OWL/>>. Acesso em: 5 Fev. 2013.

PROTÉGÉ. Ontology editor and knowledge-base framework, 2000. Disponível em: <<http://protege.stanford.edu/>>. Acesso em: 3 fev. 2013.

PYTHON, 2013. "Python Programming Language". Disponível em: <http://www.python.org/>. Acesso em 07 fev. 2013

RAUTENBERG, S.; TODESCO, J. L.; GAUTHIER, F. A. O. Processo de desenvolvimento de ontologias: uma proposta e uma ferramenta. Revista Tecnologia, Fortaleza, v.30, n.1, p. 133-144, 2009.

Recuero R. REDES SOCIAIS NA INTERNET. EDITORA SULINA, 2009.

Sakaki, T.; Okazaki, M. & Matsuo, Y. Earthquake shakes *Twitter* users: realtime event detection by social sensors. Em Proceedings of the 19th international conference on World wide web, WWW '10, pp. 851--860, New York, NY, USA. ACM.

Statistics Facebook. <http://www.facebook.com/press/info.php?statistics>, 2012.

Túlio Costa, C. O Papel da Internet na Conquista dos Votos de Marina Silva. INTERESSE NACIONAL, 2011.

TWITTER SEARCH. Using the Twitter Search API. <https://dev.twitter.com/docs/using-search>. Acessado em Março/2013.

Twitter.a (2009). Top *Twitter* Trends of 2009. <http://blog.Twitter.com/2009/12/top-Twitter-trends-of-2009.html>. Acessado em Janeiro/2013.

Twitter.a. What is a Timeline? <https://support.Twitter.com/articles/49309-what-are-hashtags-symbols>. Acessado em Janeiro/2013.

Twitter.b. What Are Hashtags ("#" Symbols)?. <https://support.Twitter.com/articles/49309-what-are-hashtags-symbols>. Acessado em Janeiro/2013.

Twitter.c. What are @Replies and Mentions? <https://support.Twitter.com/articles/14023-what-are-replies-and-mentions>. Acessado em Janeiro/2013.

Twitter.d. My Follow Count Shows more Followers than my Followers Page!. <https://support.Twitter.com/articles/123002-my-follow-count-shows-more-followers-than-my-followers-page>. Acessado em Janeiro/2013.

Twitter.e. About Trending Topics. <https://support.Twitter.com/articles/101125-about-trending-topics>. Acessado em Janeiro/2013.

Twitter.f. What Is Retweet? (RT). <https://support.Twitter.com/articles/77606-what-is-retweet-rt>. Acessado em Janeiro/2013.

UML (2013). Unified Modeling OMGlanguage. Disponível em <http://www.uml.org/>, acessado em 01/2013.

Wasserman S., Faust K., and Iacobucci D. Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences). Cambridge University Press, 1994.

Watts, D. (1999). Small Worlds: the Dynamics of Networks Between Order and Randomness. Princeton University Press.

WATTS, D. J. Six Degrees: The Science of a Connected Age. New York: W. W. Norton &Company, 2003.

Xing Xie. Potential Friend Recommendation in On-line Social Network. 2010 IEEE/ACM International Conference on Green Computing and Communications & 2010 IEEE/ACM International Conference on Cyber, Physical and Social Computing.