

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

MARCO ANDREI LAMPERT

ANÁLISE DA RELEVÂNCIA DE MENSAGENS NO TWITTER ATRAVÉS DE UM
SISTEMA MULTI-AGENTE

SÃO LEOPOLDO
2012

Marco Andrei Lampert

ANÁLISE DA RELEVÂNCIA DE MENSAGENS NO TWITTER ATRAVÉS DE UM
SISTEMA MULTI-AGENTE

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre pelo
Programa de Pós-Graduação em Computação
Aplicada da Universidade do Vale do Rio dos
Sinos — UNISINOS

Orientador:
Prof. Dr. Sérgio Crespo Coelho da Silva Pinto

São Leopoldo
2012

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Lampert, Marco Andrei

Análise Da Relevância De Mensagens No Twitter Através De Um Sistema Multi-Agente / Marco Andrei Lampert — 2012.

84 f.: il.; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2012.

“Orientador: Prof. Dr. Sérgio Crespo Coelho da Silva Pinto, Unidade Acadêmica de Pesquisa e Pós-Graduação”.

1. Computação. 2. Sistema multiagente. 3. Redes Sociais. 4. Mensagem - Twitter. I. Título.

CDU 004

Catálogo na Publicação: Bibliotecário Eliete Mari Doncato Brasil — CRB 10/1184

(Esta folha serve somente para guardar o lugar da verdadeira folha de aprovação, que é obtida após a defesa do trabalho. Este item é obrigatório, exceto no caso de TCCs.)

Dedico este trabalho primeiramente a Liliane, minha fiel amiga, companheira e esposa, sem a qual este trabalho não teria iniciado e, por muitas vezes, continuado. Aos meus filhos, Luciano e Lúcia, pelos inúmeros momentos de ausência. Aos meus pais, Danilo e Eunice, que muito valorizaram e incentivaram os meus estudos e a minha formação.

*O homem se torna muitas vezes o que ele próprio acredita que é.
Se insisto em repetir para mim mesmo que não posso fazer uma determinada coisa,
é possível que acabe me tornando realmente incapaz de fazê-la.
Ao contrário, se tenho a convicção de que posso fazê-la,
certamente adquirirei a capacidade de realizá-la, mesmo que não a tenha no começo.*

— MAHATMA GANDHI

AGRADECIMENTOS

Agradeço inicialmente ao grande mestre Creso, meu orientador, pela paciência, sugestões e por ter acreditado na realização desta pesquisa. Agradeço a PROCERGS, pelo apoio proporcionado, em especial no nome dos colegas Sérgio Dalanhol e Joel Darcie. Agradeço os colegas Alex Roehrs, Johnson Verdun, Lucas Braz, Michele Lermen e Tássia Serrão pelas inúmeras trocas de experiências e ajuda técnica. Agradeço o colega Lucas Graebin pelas inúmeras caronas. E agradeço a Linda Summer pela ajuda psicológica em momento decisivo.

Um amigo é alguém com quem se pode pensar em voz alta.

(Ralph Waldo Emerson)

“Se você acha que a educação é cara, tenha a coragem de experimentar a ignorância.”.
(Derek Bok)

RESUMO

O surgimento de novas tecnologias e as inovações em mídias sociais têm alterado a forma como as pessoas se comportam. Destacadamente as Redes Sociais estão cada vez mais inseridas na vida das pessoas. Nunca houve tanto desenvolvimento, penetração, diversificação, dispersão da informação, comunicação em tempo real, compressão do espaço e tempo, concomitante com a pluralidade de perspectivas, definições, análises e de cenários prospectivos sobre os possíveis desdobramentos dos fatos do presente. Convergência está em todo lugar e nunca foi tão fácil atingir um público tão grande. Diante deste cenário exploramos as pesquisas existentes e propomos uma abordagem para analisar a relevância de mensagens do Twitter, monitorando a sua evolução na rede e estabelecendo a influência exercida em um espaço demográfico. Desenvolvemos uma aplicação capaz de fazer esta monitoração, com intuito de verificar e validar o modelo proposto.

Palavras-chave: Computação. Sistema multiagente. Redes Sociais. Mensagem - Twitter.

ABSTRACT

The appearance of new technologies and innovations in social media has changed the way how people behave themselves. Remarkably Social Networks are more and more incorporated in people's lives. There has never been so much development, penetration, diversification, dispersion, real-time communication, compression of space and time, concomitant with a plurality of perspectives, definitions, analysis and prospective scenarios on the possible unfolding of the facts of the present. Convergence is everywhere and has never been so easy to achieve such a large audience. In this scenario we explore the existing research and we propose an approach to analyze the relevance of Twitter messages, monitoring its evolution in the network and establishing the influence in a demographic space. We develop an application able to do this monitoring, with the intent to verify and validate the proposed model.

Keywords: Computation. Multi-agent system. Social Networks. Post - Twitter.

LISTA DE FIGURAS

| | | |
|------------|---|----|
| Figura 1: | Twitter - Grafo Dirigido | 24 |
| Figura 2: | Arquitetura do Twitter | 27 |
| Figura 3: | Postagem de um Tweet (forma antiga) | 28 |
| Figura 4: | Registrando uma Aplicação | 29 |
| Figura 5: | Postagem de um Tweet (forma atual). | 30 |
| Figura 6: | Pesquisando Sobre Determinado Assunto. | 30 |
| Figura 7: | Linha do Tempo. | 31 |
| Figura 8: | Tweets de um Usuário. | 31 |
| Figura 9: | Código Que Obtém Aproximadamente 1% De Todos Os Tweets. | 31 |
| Figura 10: | Exemplos de Tweets. | 32 |
| Figura 11: | Relação entre Agente e Ambiente. | 34 |
| Figura 12: | Nuvens de Palavras das Comunidades e Sub-Comunidades Gruber | 43 |
| Figura 13: | Número de Seguidores (eixo X) e Taxa de Retweet (eixo Y) | 51 |
| Figura 14: | Número de Seguidos (eixo X) e Taxa de Retweet (eixo Y) | 52 |
| Figura 15: | Número de Tweets no Passado (eixo X) e Taxa de Retweet (eixo Y) | 52 |
| Figura 16: | Idade da Conta no Twitter (eixo X) e Taxa de Retweet (eixo Y) | 53 |
| Figura 17: | Exemplo de página do site WhoRetweetedMe | 55 |
| Figura 18: | Espaço de Grafo de Seguidos / Seguidores | 56 |
| Figura 19: | Espaço de Grafo Informação / Retweet | 57 |
| Figura 20: | Espaço Demográfico da Informação versus Usuário | 57 |
| Figura 21: | Arquitetura do Sistema | 58 |
| Figura 22: | Objetivos conforme metodologia MaSE | 58 |
| Figura 23: | Papéis MaSE | 59 |
| Figura 24: | NodeMonitor | 60 |
| Figura 25: | MessageMonitorUser | 60 |
| Figura 26: | MessageMonitorTwitter | 61 |
| Figura 27: | MessageEvaluator | 61 |
| Figura 28: | GetOriginalTweet | 62 |
| Figura 29: | DataAnalyzer | 62 |
| Figura 30: | Diagrama de Entidade Relacionamento | 64 |
| Figura 31: | Principais Agentes do Sistema | 64 |
| Figura 32: | Quantidades de retweets por dia | 70 |
| Figura 33: | Quantidades de retweets após o primeiro dia | 70 |
| Figura 34: | Usuários orbitando em torno de uma mensagem | 71 |
| Figura 35: | Medidas de Centralidade | 72 |
| Figura 36: | Nuvem de Palavras de Todos os Retweets do Usuário 3 | 73 |
| Figura 37: | Nuvem de Palavras de Retweets Únicos do Usuário 3 | 73 |

LISTA DE TABELAS

| | | |
|-----------|---|----|
| Tabela 1: | Ranking dos Desenvolvedores | 44 |
| Tabela 2: | Ranking dos Desenvolvedores Apple | 45 |
| Tabela 3: | Ranking dos Jornalistas | 45 |
| Tabela 4: | Sites mais populares em Tweets e Retweets | 50 |
| Tabela 5: | Hashtags mais populares em Tweets e Retweets | 51 |
| Tabela 6: | Algumas Propriedades dos Usuários Objetos do Estudo | 66 |
| Tabela 7: | Escores das mensagens com mais retweet | 69 |
| Tabela 8: | Hashtags mais populares nos dados obtidos | 74 |
| Tabela 9: | Tabela comparativa entre os principais trabalhos | 76 |

LISTA DE ABREVIATURAS

| | |
|-------|---------------|
| Dr. | Doutor |
| n.a. | não se aplica |
| parc. | parcialmente |
| Prof. | Professor |

LISTA DE SIGLAS

| | |
|------|--|
| API | - Application Programming Interface |
| AUML | - Agent Unified Modeling Language |
| MaSE | - Multiagent Systems Engineering Methodology |
| QPS | - Querys Per Second |
| RSS | - Really Simple Syndication |
| SGBD | - Sistema de Gerenciamento de Banco de Dados |
| SMS | - Short Message Service |
| SQL | - Structured Query Language |
| TI | - Tecnologia da Informação |
| TPS | - Tweets Per Second |
| URL | - Universal Resource Locator |
| UML | - Unified Modeling Language |

SUMÁRIO

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO | 15 |
| 1.1 | Contextualização | 15 |
| 1.2 | Motivação | 16 |
| 1.3 | Questão de Pesquisa | 16 |
| 1.4 | Objetivo | 16 |
| 1.5 | Metodologia | 16 |
| 1.6 | Organização do Trabalho | 17 |
| 2 | ARCABOUÇO TEÓRICO | 19 |
| 2.1 | Redes Sociais | 19 |
| 2.1.1 | Microblogs | 20 |
| 2.1.2 | O Twitter | 22 |
| 2.2 | Agentes de Softwares | 32 |
| 2.2.1 | Conceito de agentes | 32 |
| 2.2.2 | Propriedades dos agentes | 33 |
| 2.2.3 | Agentes e o Ambiente | 34 |
| 2.2.4 | Agentes e a sua Estrutura | 35 |
| 2.3 | Sistemas Multi-Agentes | 36 |
| 2.3.1 | Metodologias da Engenharia de Software Orientada a Agentes | 37 |
| 3 | TRABALHOS RELACIONADOS | 39 |
| 3.1 | Análise do Espaço Demográfico | 39 |
| 3.1.1 | Mapping Communities In Large Virtual Social Networks: Using Twitter Data To Find The Indie Mac Community | 39 |
| 3.2 | Relevância de Mensagens em Redes Sociais | 46 |
| 3.2.1 | Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter | 46 |
| 3.2.2 | Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network | 49 |
| 3.2.3 | WhoReTweetedMe | 53 |
| 4 | TRABALHO PROPOSTO | 56 |
| 4.1 | Visão Geral | 56 |
| 4.2 | Arquitetura do Sistema | 57 |
| 4.3 | Tecnologias e Ferramentas Adotadas | 65 |
| 4.4 | Análise dos Dados Obtidos | 66 |
| 4.4.1 | Abrangência | 66 |
| 4.4.2 | Relevância | 68 |
| 4.4.3 | Sobrevida | 69 |
| 4.4.4 | Órbita | 71 |
| 4.4.5 | Centralidades | 71 |
| 4.4.6 | Assunto | 73 |
| 4.4.7 | Apreciações Adicionais | 74 |
| 5 | ANÁLISE COMPARATIVA ENTRE OS TRABALHOS | 75 |

| | |
|--|-----------|
| 6 CONCLUSÕES | 77 |
| 6.1 Limitações da Análise | 77 |
| 6.2 Trabalhos Futuros | 78 |
| REFERÊNCIAS | 80 |

1 INTRODUÇÃO

“... aprender não é um ato findo. Aprender é um exercício constante de renovação...”

Paulo Freire

Este capítulo introduz a proposta da dissertação contextualizando o domínio envolvido, apresentando as motivações inerentes aos problemas levantados, trazendo a questão de pesquisa a ser respondida, definindo os objetivos a serem atingidos, estabelecendo a metodologia a ser aplicada e apresentando a organização do restante do trabalho.

1.1 Contextualização

Desde que nascemos estamos inseridos em estruturas aonde mantemos relações que se desenvolvem durante toda a vida. Iniciamos na estrutura familiar, passamos para o ambiente escolar e para a comunidade que vivemos e trabalhamos. A própria natureza humana nos liga a outras pessoas e estrutura a sociedade em rede, onde cada um possui uma função e identidade. Agrupamos-nos com semelhantes e estabelecemos relações de trabalho, de amizade, ou de qualquer outro interesse, que se desenvolvem e se modificam (TOMAEL; ALCARA; CHIARA, 2005). Cada uma destas estruturas de convívio em rede podemos chamar de Rede Social.

Partindo deste princípio, o termo redes sociais não é algo inusitado, já existindo antes mesmo do surgimento da *Web* e tendo sido aplicado em diversas áreas da sociologia para a compreensão de diferentes eventos sociais, políticos, econômicos, afetivos, entre outros (WASSERMAN; FAUST, 1994). Organizações antigas como os templários, maçons ou organizações religiosas podem ser consideradas Redes Sociais.

O que é singular, nesta onda de desenvolvimento, é a penetração, sua diversificação, a compressão do espaço e tempo, a dispersão da informação, a comunicação em tempo real, concomitante com a pluralidade de perspectivas, definições, análises e cenários prospectivos sobre os possíveis desdobramentos dos fatos do presente. O surgimento de novas tecnologias e as inovações em mídias sociais têm alterado a forma como as pessoas se comportam. Convergência está em todo lugar e nunca foi tão fácil atingir um público tão grande. Para que seja possível ter uma ideia da dimensão desta questão, atualmente, 96% das pessoas da chamada geração Y (nascidos após 1980) utilizam redes sociais; um em cada oito casais americanos se conheceu através de uma rede social; a televisão levou 13 anos para atingir os 50 milhões de expectadores, enquanto que a rede social Facebook alcançou o dobro da marca, em número de usuários, em apenas nove meses (BRUNO MARCOS GOMES DA SILVA BRUNO; MONTEIRO, 2010), somando hoje, em fevereiro de 2012, mais de 800 milhões de usuários (LINE, 2012). Caso fosse um país, o Facebook seria o 3º mais populoso do mundo e teria população maior que o total da Europa.

1.2 Motivação

Neste contexto, é relevante o papel das redes sociais, que a cada dia estão mais inseridas na vida das pessoas, com utilização maciça e crescente, principalmente para fins pessoais. Por este motivo, no início do surgimento das redes sociais, os gestores de TI (Tecnologia da Informação) das empresas se apressaram em bloquear o acesso a todas as redes sociais, em um trabalho incansável, já que a cada dia mais e mais redes sociais surgem na Internet, sob o pretexto de inibir uma eventual perda de produtividade dos trabalhadores, que poderiam passar a dedicar mais tempo a questões pessoais do que ao seu trabalho.

Esta situação está mudando, a evolução das redes sociais convergiu para a utilização destas ferramentas também para fins profissionais, passando a ser uma importante ferramenta de pesquisa e publicidade, mas nunca deixando de lado aquele caráter de finalidade pessoal da rede social, o que hoje já faz com que a visão dos gestores de TI comece a mudar.

A cada dia é mais comum ver empresas, muitas delas de grande porte, liberando o acesso de seus funcionários às redes sociais, até porque pesquisas já revelam que empresas que permitem o uso de mídias sociais a seus empregados são, em média, 9% mais produtivas do que as que agem no sentido contrário (INC, 2010).

1.3 Questão de Pesquisa

A questão central deste trabalho é: como analisar a relevância de mensagens em uma rede social para que se possa estabelecer a influência exercida em um espaço demográfico?

1.4 Objetivo

Levando em conta, a contextualização, a motivação e questão de pesquisa anteriormente expostos, a proposta de dissertação procura melhorar o entendimento do assunto em questão e também tem por objetivos:

- Analisar a relevância dada a uma mensagem enviada, monitorando a sua evolução na rede;
- Desenvolver uma aplicação capaz de monitorar as mensagens trocadas sobre um determinado assunto ou em um determinado grupo, avaliando a sua importância;
- Verificar e validar os modelos propostos em uma rede social real, no caso o Twitter .

1.5 Metodologia

Para o pleno atendimento dos objetivos estabelecidos nesta proposta a seguinte metodologia foi definida para conduzir os trabalhos.

- **Atividade 1.** *Elaboração do Problema.* Formulação do problema a ser resolvido com base no conhecimento e análise da matéria em questão;
- **Atividade 2.** *Pesquisa por Trabalhos Correlatos.* Levantamento e análise de publicações relacionadas ao domínio do problema;
- **Atividade 3.** *Fundamentação Teórico-Prática.* Estudos e práticas sobre potenciais tecnologias aplicáveis na solução do problema;
- **Atividade 4.** *Definição de Visão de Solução.* Elaboração de uma proposta com a visão de solução para o problema mapeado;
- **Atividade 5.** *Análise e Projeto da Solução.* Análise e Projeto da Solução com base nas tecnologias elencadas para resolver o problema;
- **Atividade 6.** *Implementação da Solução.* Implementação das estruturas necessárias para resolver o problema mapeado;
- **Atividade 7.** *Verificação e Validação da Solução.* Verificação e validação das implementações no mundo real;
- **Atividade 8.** *Registro dos Resultados Obtidos.* Elaboração da dissertação que registra todos os estudos realizados e resultados obtidos com o desenvolvimento do trabalho.

1.6 Organização do Trabalho

O presente trabalho está disposto em seis capítulos organizados da seguinte forma:

- *Capítulo 1.* Introduce a proposta de dissertação, contextualiza o domínio do problema e apresenta as principais motivações que levaram ao desenvolvimento do presente trabalho. Além disto, são apresentados os objetivos gerais e específicos para o desenvolvimento da dissertação, bem como sua metodologia e organização;
- *Capítulo 2.* É realizada a fundamentação teórica das redes sociais, conceituado microblogs, analisado em detalhe o funcionamento do Twitter. É feito um estudo sobre o desenvolvimento de sistemas orientados a agentes de softwares e sistemas multi-agentes, definindo, contextualizando e apresentando algumas metodologias existentes;
- *Capítulo 3.* São apresentados alguns trabalhos relacionados ao domínio do problema, com foco no Twitter, análise da relevância de mensagens e análise do espaço demográfico através da formação de subgrupos a partir de uma rede social maior;
- *Capítulo 4.* A proposta do trabalho propriamente dita é apresentada, onde temos a arquitetura do sistema, a modelagem do sistema multi-agente e a análise dos dados obtidos;

- *Capítulo 5.* É feita uma análise comparativa entre os trabalhos pesquisados, frente ao trabalho proposto;
- *Capítulo 6.* Finalizando ocorrem as conclusões sobre o presente trabalho, as limitações da análise e indicações para trabalhos futuros.

2 ARCABOUÇO TEÓRICO

Nesta seção vamos introduzir conceitos sobre Redes Sociais, Microblogs, o Twitter e suas principais ferramentas.

2.1 Redes Sociais

Do ponto de vista científico a base teórica para estudo das redes sociais é a da teoria dos grafos, onde um grafo é um conjunto de vértices e um conjunto de arestas que conectam pares de vértices. Conforme (IZQUIERDO; HANNEMAN, 2006), uma rede (grafo orientado) é um conjunto de atores (ou agentes, vértices, nodos, pontos ou nós) que podem ter relacionamentos (ou conexões, arestas, vínculos ou ligações) uns com os outros. As redes podem ter muitos ou poucos atores e vários tipos de relacionamentos entre os pares de atores. No estudo das redes sociais o foco é nas relações e não nos atributos dos atores (FERREIRA, 2010).

Um ponto interessante na evolução do estudo dos grafos é a característica de que, em certos pontos de transição crítica de grafos aleatórios, emergem certas propriedades numa escala temporal muito mais rápida do que a totalidade do processo de desenvolvimento do grafo (ROSA, 2002). Esta característica pode, por exemplo, explicar o sucesso rápido de certos sites da Internet.

Outra característica importante é a utilização dos conceitos de laços fracos (*weak ties*) e laços fortes (*strong ties*), onde se coloca que a manutenção dos laços fracos é mais importante para manter a rede social do que a preocupação com os laços fortes (GRANOVETTER, 1973). Esses laços fracos são muito úteis, por exemplo, para descobrir oportunidades de trabalho. Amigos próximos são menos eficientes em tal quesito porque, em geral, circulam no mesmo meio e têm acesso às mesmas informações.

Um clássico na área é o trabalho de (MILGRAM, 1967), sobre *small world* ("mundo pequeno") e sua abordagem quantitativa que gerou posteriormente a frase: "seis graus de separação", que indica a distância ("separação") entre quaisquer duas pessoas.

Estudos mais recentes, que tiveram seus dados divulgados pela equipe do Facebook, em novembro de 2011, dizem que, entre todos os usuários do Facebook no momento da pesquisa (721 milhões de utilizadores com 69 bilhões de ligações de amizade), havia uma distância média de 4,74 pessoas (BACKSTROM et al., 2011). Também foi constatado que 99,91% dos usuários do Facebook estão interligados, formando uma grande componente conectado (UGANDER et al., 2011).

Segundo estudos realizado pela empresa Sysomos¹ de monitoramento, a distância média no Twitter é 4,67 pessoas e, em média, cerca de 50% das pessoas no Twitter estão a apenas quatro passos de distância um do outro, enquanto quase todo mundo está a cinco passos (CHENG, 2010).

¹<http://www.sysomos.com/>

Em outro trabalho, os pesquisadores demonstraram que a distância média de 1.500 usuários aleatórios no Twitter é 3,435. Eles calcularam a distância entre cada par de usuários usando todos os usuários ativos no Twitter (BAKHSHANDEH et al., 2011).

Ainda sobre o "mundo pequeno", o trabalho sobre redes complexas de (WATTS; STROGATZ, 1998), também é muito importante. Eles partiram do princípio de que apenas uns poucos indivíduos bem colocados podem acelerar dramaticamente o fluxo de informação em uma empresa, agindo como emissários entre andares e departamentos. Quando Watts e Strogatz fizeram as contas, descobriram que são precisos apenas algumas poucas conexões aleatórias, ou atalhos, para fazer um pequeno mundo de um grande. Uma vez transformado, esse mundo não precisa mais de conexões adicionais.

Outro estudo importante é o de (BARABASI, 1999) que diz que há uma ordem dinâmica de estruturação das redes. As redes não seriam igualitárias, ao contrário, alguns vértices seriam altamente conectados enquanto outros teriam poucas conexões. Os vértices ricos seriam os *hubs* e tenderiam a receber mais conexões (lei do "*rich get richer*", isto é, "o rico fica mais rico").

Especificamente abordando redes sociais, (HANNEMAN; RIDDLE, 2005) fazem as primeiras contribuições para o entendimento da subestrutura das redes usando uma abordagem *bottom-up*, ou seja, a partir de membros individuais, considerando suas conexões com outros membros, criam-se micro e macroestruturas.

Já os físicos (GIRVAN; NEWMAN, 2002) propuseram uma abordagem *top-down* para encontrar subestrutura das redes com base na medida de intermediação das extremidades (*edge betweenness*), uma generalização da intermediação dos centros (*betweenness centrality*) de vértices de (FREEMAN, 1979). Esta abordagem avalia cada uma das ligações da rede em termos dos caminhos mais curtos entre pares de atores que corre ao longo da conexão (NEWMAN, 2001), remove a conexão com a mais alta intermediação (*betweenness*), e repete.

2.1.1 Microblogs

Um *blog* (junção do termo *web log* (BLOOD, 2000) é uma espécie de jornal pessoal publicado na internet e que consiste de mensagens postadas, normalmente apresentadas em ordem cronológica invertida, ou seja, a mensagem mais recente aparece primeiro. Microblog é uma forma de *blog* na qual se publica mensagens tipicamente de conteúdo curto, como frases, rápidos comentários, ou *links* para imagens e vídeos (KAPLAN; HAENLEIN, 2011), para ser visto por qualquer pessoa ou por um grupo restrito que pode ser escolhido pelo usuário. Tipicamente estas mensagens curtas são textos com tamanho não superior a 200 caracteres, o que permite a fácil iteração através de uma multidão de vários canais de comunicação, tais como: mensagens de texto de telefones celulares, mensagens instantâneas, e-mail e *web*. Comparado ao *blog* regular, microblogs satisfazem uma necessidade de comunicação para um modo ainda mais rápido. Ao incentivar mensagens mais curtas, diminui-se a exigência de tempo dos usuários e de investimento na geração de conteúdo, sendo este um dos principais fatores de diferenciação dos

blogs em geral. A segunda diferença importante é a frequência de atualização. Em média, um usuário de *blog* prolífico pode atualizar seu *blog* uma vez por dia, por outro lado um usuário de microblog costuma postar várias atualizações em um único dia (JAVA; FININ, 2008).

Microblogs preenchem uma lacuna entre os *blogs* e as mensagens instantâneas, permitindo que as pessoas publiquem mensagens curtas na *web* sobre o que eles estão fazendo, em tempo real (PASSANT et al., 2008). Após os *blogs*, que permitem que as pessoas publiquem seus pensamentos para um público aberto, *podcast* onde as pessoas as registram e *videoblog* onde são disponibilizadas mensagens em vídeo, microblog permite a qualquer um trocar mensagens curtas dentro de sua comunidade ou simplesmente escrever de forma breve para o público da *web* em geral. A simplicidade de publicar tais atualizações curtas, em várias situações, e em uma rede social baseada em assinaturas e mensagens de resposta, torna o microblog um método de comunicação inovador que pode ser visto como um híbrido de mensagens de *blogs*, mensagens instantâneas e atualização de status (PASSANT et al., 2008).

Como exemplo de softwares que proveem este tipo de serviço temos o Twitter ², o Tumblr ³, o Jaiku ⁴, o Posterous ⁵, o FriendFeed ⁶ e o quase extinto Google Buzz ⁷. Adicionalmente, outro exemplo de microblog é o recurso de status de serviços de redes sociais, como o Facebook ⁸, MySpace ⁹, LinkedIn ¹⁰, Google+ ¹¹ e XING ¹². Recentemente, como serviços de microblog ganharam grande popularidade, os usuários adotaram-no para novos propósitos incluindo divulgação de notícias, promoção de pontos de vista políticos, marketing, acompanhamento em tempo real de eventos e como meio de comunicação para perguntas e discussão de questões (SUH et al., 2010) (BOYD; GOLDBER; LOTAN, 2010) (JAVA; FININ, 2008) (WRIGHT, 2009) (ZHAO; ROSSON, 2009).

Microblogs, por serem caracterizados por rápida (quase em tempo real) troca de conhecimento e rápida propagação de novas informações (PASSANT et al., 2008), também são promissores para ambientes corporativos onde podem facilitar a aprendizagem, a comunicação e o conhecimento informal. Seu potencial atualmente é pouco explorado e pode ser comparado ao uso que as empresas faziam dos *wikis* a alguns anos atrás. Empresas como a Socialcast ¹³, Jive ¹⁴, 7Geese ¹⁵, e Yammer ¹⁶ iniciaram a investigar como adaptar microblogs ao ambiente corporativo. (ZHAO; ROSSON, 2009) e (CONVERTINO et al., 2010) investigaram como fazer

²<https://twitter.com/>

³<https://www.tumblr.com/>

⁴<http://www.jaiku.com/>

⁵<https://posterous.com/>

⁶<http://friendfeed.com/>

⁷<http://www.google.com/buzz>

⁸<http://www.facebook.com/>

⁹<http://www.myspace.com/>

¹⁰<http://www.linkedin.com/>

¹¹www.google.com/+

¹²<http://www.xing.com/>

¹³<http://www.socialcast.com/>

¹⁴[http:// http://www.jivesoftware.com/](http://http://www.jivesoftware.com/)

¹⁵<http://www.7geese.com/>

¹⁶<https://www.yammer.com/>

estas adaptações sugerindo que microblogs podem ser adaptados para facilitar a comunicação informal entre colegas nas organizações (SUH et al., 2010).

Microblog é um fenômeno relativamente novo, sendo o Twitter a mais popular e expressiva plataforma de microblog atual, alcançando em poucos anos milhões de usuários. Devido a isto, nossos estudos incidirão sobre este serviço.

2.1.2 O Twitter

O Twitter é uma rede social e serviço de microblog que permite aos usuários enviar e ler mensagens curtas conhecidas como *tweets*, permitindo aos usuários compartilhar e descobrir temas de interesse em tempo real. Os usuários escolhem seguir outros usuários para ganhar atualizações sobre notícias e status. Uma vez criados por um usuário, *tweets* são imediatamente entregues aos seguidores do autor. Para um leitor, *tweets* de todos os usuários que ele segue são reunidos e exibidos em uma lista única, cronologicamente reversa.

Estas mensagens curtas, que podem ter no máximo 140 caracteres, justificam muito do seu sucesso perante outras redes sociais. Isso foi estabelecido para facilitar a publicação usando SMS (Short Message Service - mensagem de texto enviada pelo celular) que, por convenção, tem 160 caracteres. Nos 20 caracteres restantes da mensagem é colocado o nome de usuário do remetente. Quando o serviço foi lançado nos Estados Unidos, em 2006, pouca gente tinha celular com conexão banda larga e o SMS permitia que o usuário atualizasse a sua página no Twitter sem precisar do computador. Muitos usuários do Twitter elogiam o serviço justamente pelo desafio de sintetizar a informação. Facilita muito quem começa a seguir 50, 80, 100 pessoas e tenta acompanhar todas as atualizações que chegam. Como as mensagens são curtas, pode-se verificar rapidamente se alguma coisa interessa ou não. Uma ideia pode ser desenvolvida em um *blog*, sendo enviado pelo Twitter um resumo com o *link*. Como endereços de páginas da *web* geralmente são grandes, a necessidade de se economizar espaço nas mensagens abriu a oportunidade para o desenvolvimento de serviços de redução de *links* - chamados "encurtadores de URL (Universal Resource Locator)". No Brasil, o mais conhecido é o "migre.me".

Outra característica interessante, que não costuma haver em outras redes sociais, é o ato de enviar *retweets*. Desta forma, além de promover relacionamentos, o Twitter incentiva a troca de informações relevantes entre seus seguidores. Com isto visibilidade, reputação, seguidores e notícias relevantes podem ser obtidos em curto espaço de tempo. O Twitter é mais eficiente do que o Google para a busca de informações recentes. O impacto de uma mensagem pode ser avaliado acompanhando a quantidade de vezes que um endereço eletrônico é repassado ou monitorando palavras-chave. Também é possível identificar quem são as pessoas com poder de influência para determinados segmentos de público.

O Twitter fornece uma API ¹⁷ (Application Programming Interface) que permite que aplicativos de terceiros enviem e recebam *tweets* (SUH et al., 2010). Boa parte do sucesso do Twitter

¹⁷<http://apiwiki.twitter.com/>

se deve a esta API de uso fácil, bem como o fato de ser estimulado que o conteúdo seja público. Isto permitiu que muitos dos aplicativos que usam o Twitter tenham sido desenvolvidos por outras empresas que não a do próprio Twitter. Hoje existem bibliotecas disponibilizadas pela comunidade para as seguintes linguagens: ActionScript/Flash, C++, Clojure, Erlang, Java, JavaScript, .NET, Objective-C / Cocoa, Perl, PHP, Python, Ruby e Scala.

O Twitter começou como um experimento em 2006 sendo fundado em 2007 (NOORDHUIS; HEIJKOOP; LAZOVIK, 2010). Seu número de usuários tem crescido significativamente desde então. Em setembro de 2011, tinha cerca de 200 milhões de usuários cadastrados, sendo ativos aproximadamente 100 milhões destes usuários. Diariamente são enviados mais de 230 milhões de *tweets* (TAYLOR, 2011). A CNN ¹⁸, por exemplo, possui duas contas no Twitter, ambas com mais de 3 milhões de seguidores. Com este crescimento explosivo percebe-se como se tornou difícil encontrar informações relevantes entre todas as fontes de difusão de informação. Uma pesquisa pode resultar em um *tweet* anunciando as últimas notícias que, por padrão, tem o mesmo valor de uma adolescente reclamando de seu trabalho escolar. Mineração e análise de dados de redes sociais pode ser difícil por causa da grande quantidade de dados envolvidos, sendo atividades geralmente muito caras, por exigir muitos recursos computacionais (NOORDHUIS; HEIJKOOP; LAZOVIK, 2010).

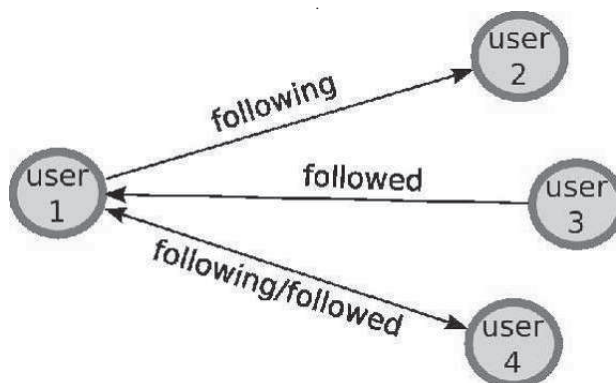
Uma forma de abordar a questão da grande quantidade de informações foi feita por (PERERA et al., 2010), onde se estuda o comportamento temporal de mensagens que chegam em uma rede social. A modelagem do processo de chegada de *tweets* no Twitter pode ser aplicado para prever o comportamento do usuário em redes sociais. Foi desenvolvido um software que utiliza a API do Twitter para coletar os *tweets* enviados para usuários específicos, com a data e hora exata dos *tweets*, o que permitiu caracterizar o tempo de chegada entre *tweets* e o número de *retweets*. Os estudos de (PERERA et al., 2010) indicam que o processo de chegada de novos *tweets* para um usuário pode ser modelado como um Processo de Poisson, enquanto o número de *retweets* seguem uma distribuição geométrica.

Conforme revela a figura 1, o Twitter é diferente da maioria das redes sociais por ser um grafo dirigido: pode-se optar por receber atualizações de um usuário, mas isso não significa que o inverso é verdadeiro. O Twitter é menos focado na amizade e mais no interesse: o usuário tem mais chances de estar ligado a outro usuário, porque ele está interessado no que esse usuário tem para dizer e menos por serem amigos na vida real. Também é diferente pelo fato de que você pode estar em contato com alguém que você não está conectado no gráfico, através de pesquisa de todos os *tweets* recentes para uma determinada palavra-chave (NOORDHUIS; HEIJKOOP; LAZOVIK, 2010).

O Twitter tem sido uma fonte de atividade não apenas para usuários, mas também para outros grupos de interesse. Neste contexto, há um crescente número de aplicações no Twitter que melhoram a sua usabilidade. Um exemplo é uma aplicação *user-friendly* que permite às

¹⁸www.cnn.com

Figura 1: Twitter - Grafo Dirigido



Fonte: (PERERA et al., 2010)

peças agilizar tweets: o TweetDeck ¹⁹. Existe outro conjunto de aplicações que não são para melhorar a usabilidade do Twitter, mas são de análises da base de dados do Twitter, como o Twitscoop ²⁰, que fornece os tópicos principais (*top trending*).

Vale salientar também alguns outros estudos como o de (KWAK et al., 2010) que analisou em grande escala as características topológicas do Twitter e seu poder como um novo meio de compartilhamento de informações. Da linha de tempo pública do Twitter (JAVA; FININ, 2008) recolheram 1.348.543 os *tweets* criado por 76.177 usuários, examinando as propriedades topológicas e geográficas do Twitter. Eles identificaram algumas categorias de uso, tais como: *chats* diários, conversas, compartilhamento de informação, URLs e divulgação de novidades jornalísticas (SUH et al., 2010). Talvez o uso mais popular seja para os usuários informarem os outros e se expressarem. (NAAMAN; BOASE; LAI, 2010) examinou o conteúdo de 3.379 *tweets* públicos manualmente, descobrindo que 80% das mensagens dos 350 usuários que estudou postam notícias relativas a si mesmos ou seus pensamentos, em vez compartilhar notícias de uso geral. O Twitter também tem sido usado politicamente por candidatos em campanhas políticas. Após a eleição Iraniana de 2009, os manifestantes usaram o Twitter como uma ferramenta de mobilização e como um método de comunicação com o mundo exterior (SUH et al., 2010).

Em (MARASANAPALLE et al., 2010) estuda-se a utilização do Twitter para a coleta de informações úteis sobre programas de TV. Tradicionalmente, muitas empresas, incluindo as empresas de mídia têm usado o Twitter para publicidade. Empresas de mídia têm contas no Twitter onde transmitem as últimas notícias de sua programação, detalhes do próximo show, *links* para clipes e as últimas notícias para os seus seguidores. Outro uso comum do Twitter (não necessariamente usado por empresas de mídia) é no sentido de obter *feedback* dos clientes. O Twitter é uma importante fonte de dados em tempo real, para reações a determinados eventos ao vivo, como uma conferência técnica, um jogo de esportes, um comício político ou um show

¹⁹<http://www.tweetdeck.com/>

²⁰<http://www.twitscoop.com/>

de música. Ele permite que os participantes expressem suas opiniões enquanto o evento se desenrola e, portanto, tem o potencial de coleta de reações imediata. Estas informações podem ajudar os produtores de conteúdo a descobrir os aspectos que podem ser melhorados, compreender as preferências do público, extrair inteligência competitiva dos concorrentes e entender as tendências da audiência.

Os pesquisadores também examinaram como construir ferramentas em cima do Twitter. Por exemplo, (JANSEN et al., 2009) investigaram como o Twitter é usado para compartilhar opiniões dos consumidores sobre marcas. (RAMAGE; DUMAIS; LIEBLING, 2010) e (BERNSTEIN et al., 2010) propuseram ferramentas para agrupar *tweets* em tópicos para suportar navegação rápida (SUH et al., 2010).

2.1.2.1 Retweets

Quando um usuário encontra um *tweet* interessante, escrito por um outro usuário do Twitter, e quer compartilhá-lo com seus seguidores, ele pode enviar um *retweet*, do *tweet*, copiando a mensagem, normalmente adicionando um indicador de *retweet*, sendo os mais comuns RT ou Via, seguido do nome do autor original em formato de @ usuário. Por exemplo, "**RT @WSJ: Breaking: The U.S. economy added 120,000 jobs in November. The unemployment rate fell to 8.6%: <http://t.co/aSwhokfY>**". Frequentemente os usuários adicionam mais conteúdo ou modificam ligeiramente o original quando enviam *retweets*. Esta prática se tornou tão difundida, que o Twitter acrescentou um recurso em 2009 para permitir aos usuários facilmente enviarem *retweets* com um clique. Enviar *retweets* tornou-se o mecanismo-chave para a difusão de informação no Twitter. Portanto, é importante explorar como o envio de *retweets* funciona para entender como a informação é difundida na rede Twitter. A construção do modelo de enviar *retweets* pode levar a otimização da difusão da informação que naturalmente ocorre na rede.

(ZARRELLA, 2009a) diz acreditar que *retweets* são um dos desenvolvimentos mais importantes na comunicação moderna. O advento da *web* fez os memes ²¹ se espalharem mais rápido, serem expostos a mais pessoas e remover muita das restrições impostas pelos limites da memória humana. Segundo ele o mundo real é feito de informações "contagiosas", que pulam de pessoa para pessoa, e estas informações não são difundidas porque são boas, mas se espalham por causa de algum outro gatilho ou conjunto de gatilhos. Ainda segundo (ZARRELLA, 2009a), nos tempos atuais temos outra mudança igualmente importante: a observação. Podemos comparar milhões de ideias virais para descobrir a forma de construção do contágio. *Retweets* podem parecer uma pequena ideia, e são em alguns aspectos. Mas essa ideia de pequeno porte é a primeira janela real de como as ideias são transmitidas de pessoa para pessoa. Podemos estudar os traços linguísticos, as características tópicas, a dinâmica epidemiológica, e as interações de rede sociais que ocorrem quando uma pessoa espalha um meme. Não somente pode esta informação nos ajudar a criar *tweets* mais relevantes, mas muitas das lições aprendidas

²¹http://pt.wikipedia.org/wiki/Memes_de_Internet

através dos *retweets* podem ser aplicáveis a ideias virais em outras mídias. Com esta ideia, de entendermos o funcionamento de ideias contagiosas, que se propagam, é que criamos o sistema que será descrito no próximo capítulo.

2.1.2.2 Arquitetura do Twitter

A infraestrutura do Twitter iniciou usando Ruby on Rails no *front-end* (camada de apresentação), MySQL como *back-end* (banco de dados) e C, Scala e Java para a camada de regras de negócio (AVRAM, 2009). Listas reversas usando estruturas de dados B-tree foram construídas no MySQL para alavancar transações simultâneas, suportando indexação concorrente e pesquisas. O dimensionamento era feito particionando o índice entre vários bancos de dados e replicando o *front-end* Rails (TWITTERSEARCH, 2011).

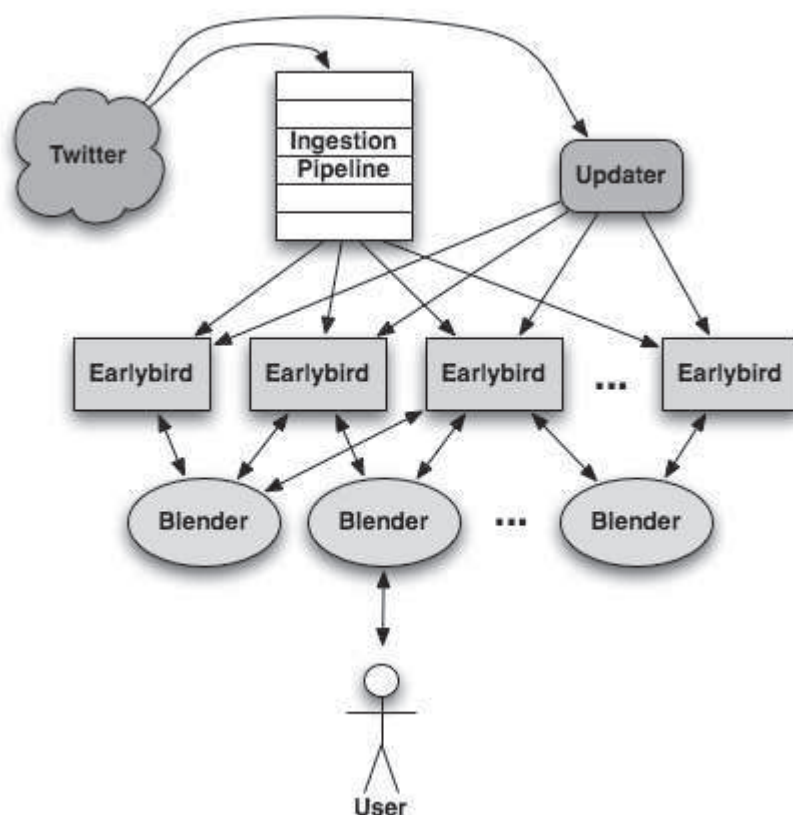
Em 2008 o Twitter manipulava uma de média 20 TPS (*tweets per second*) e 200 QPS (*queries per second*). Em outubro de 2010, quando o MySQL foi substituído pelo Earlybird, o sistema estava lidando com 1.000 TPS e 12.000 QPS, em média. O Earlybird, uma lista reversa em tempo real, baseado no Lucene²², forneceu um melhor desempenho do que o MySQL para pesquisas em tempo real, duplicando a eficiência da memória e permitindo a flexibilidade de adicionar filtragem por relevância (TWITTERSEARCH, 2011).

Em abril de 2011, Ruby on Rails foi removido da infra-estrutura e no seu lugar foi lançado um substituto, chamado Blender, feito em JAVA, que melhorou a latência de busca em três vezes e aumentou a taxa de transferência(*throughput*) em dez vezes. No final do mês de maio de 2011, o Twitter estava indexando uma média de 2.200 TPS e servindo 18.000 QPS (1.6 bilhões de consultas por dia) (TWITTERSEARCH, 2011).

A figura 2 mostra esta arquitetura. Cabe salientar que o *Ingestion Pipeline* proporciona a atualização das informações estáticas, como por exemplo, informações sobre o usuário e a linguagem do texto do *tweet*. Já o *Updater* atualiza continuamente as informações dinâmicas no Earlybird, como por exemplo as interações dos usuários com os *tweets*, que chegam ao longo do tempo.

²²<http://lucene.apache.org/>

Figura 2: Arquitetura do Twitter



Fonte: <http://engineering.twitter.com/2011/05/engineering-behind-twitters-new-search.html>

2.1.2.3 Bibliotecas do Twitter para Java: o Twitter4J

O *link* oficial da API do Twitter ²³ nos fornece indicações de três bibliotecas em Java para uso:

- Twitter API ME: <http://www.twapime.com/>
- Scribe: <https://github.com/fernandezpablo85/scribe-java#readme>
- Twitter4J: <http://twitter4j.org/>

Além desta identificamos na internet mais duas:

- Java-Twitter: <http://code.google.com/p/java-twitter/>
- JTwitter: <http://www.winterwell.com/software/jtwitter.php>

Neste trabalho optamos por usar a API Twitter4J por abranger o núcleo da API e todos os apetrechos normais de projetos *open source*: repositório dos códigos fontes disponível publicamente, documentação *on-line*, listas de discussão ativas e versões novas liberadas recentemente.

²³<http://dev.twitter.com/> ou <http://dev.twitter.com/pages/libraries>

Comparada com as outras APIs, observamos que a cobertura da API de serviços do Twitter4J é mais completa, a lista de discussão é mais ativa, a documentação é melhor e as atualizações (*releases*) são mais frequentes (PIETRI, 2009). Na internet, também, ela é mais usada e possui muito mais resultados nas buscas, inclusive com exemplos e tutoriais.

A facilidade de uso e a rapidez com que as versões iniciais da API permitiam postar um *tweet* foram um dos motivos do sucesso do Twitter. Segue a figura 3 mostrando como era possível fazê-lo até novembro de 2010:

Figura 3: Postagem de um Tweet (forma antiga)

```
import twitter4j.*;

public class TwitterTeste {

    public static void main(String[] args) throws TwitterException {

        Twitter twitter = new Twitter("username", "password");
        Status status = twitter.updateStatus("Escreva sua mensagem aqui.");

    }
}
```


Fonte: elaborado pelo autor

Hoje já não é mais tão simples. Primeiro é necessário registrar a aplicação no Twitter em <http://twitter.com/apps/new>, conforme mostra a figura 4:

Figura 4: Registrando uma Aplicação

twitter Página Inicial Perfil Encontrar Pessoas Configurações Ajuda Sair

Register an Application

Application Icon: 
Tamanho máximo de 700k. JPG, GIF, PNG.

Application Name:

Description:

Application Website:
Where's your application's home page, where users can go to download or use it?

Organization:

Website:
The home page of your company or organization.

Application Type: Client Browser
Does your application run in a Web Browser or a Desktop Client?
Browser uses a Callback URL to return to your App after successfully authentication.
Client prompts your user to return to your application after approving access.

Callback URL:
Where should we return to after successfully authentication?

Default Access type: Read, Write, & Direct Messages Read & Write Read-only
What type of access does your application need? Note: @Anywhere applications require read & write access.

Use Twitter for login: Yes, use Twitter for login
Does your application intend to use Twitter for authentication?

wordvid word

Não consegue ler isto?
Obter duas palavras novas:

Escute um conjunto de palavras

Desenvolvido por reCAPTCHA.

Welcome to the Developer Beta of the Twitter Application Platform! We're just getting started, but we thought we'd start releasing components that will help you, the developers, connect your users with the world, **right now.**

For starters, we're allowing you to both register your application here, as well as providing an improved Authentication System, OAuth. To read more about how this help both you and your users, please visit <http://oauth.net>.

Enjoy! And please report any bugs or general feedback to api@twitter.com.

Ajuda

© 2011 Twitter Sobre o Twitter Contato Blog Status Recursos API Empresas/Negócios Ajuda Empregos Termos Privacidade

Fonte: <http://twitter.com/apps/new>

Feito o registro são fornecidas uma *ConsumerKey* e uma *ConsumerSecret*. A partir destas informações, seguindo as orientações contidas no site oficial do Twitter ²⁴ ou no *blog* da EMC Computing ²⁵, obtém-se um *Tokem* e um *TokemSecret*. De posse destas quatro informações, *ConsumerKey*, *ConsumerSecret*, *Tokem* e *TokemSecret*, através dos comandos da figura 5, a aplicação tem acesso atualmente a API do Twitter.

Figura 5: Postagem de um Tweet (forma atual).

```
String consumerKey = "*****!";
String consumerSecret = "*****!";
String twitterToken = "*****!";
String twitterTokenSecret = "*****!";

Twitter twitter = new TwitterFactory().getInstance();
twitter.setOAuthConsumer(consumerKey, consumerSecret);
AccessToken accessToken = new AccessToken(twitterToken, twitterTokenSecret);
twitter.setOAuthAccessToken(accessToken);

Status status = twitter.updateStatus("Estou usando a API do Twitter !!!");
```

Fonte: elaborado pelo autor

Na figura 6 temos um exemplo de código onde se faz pesquisa sobre um determinado assunto no Twitter.

Figura 6: Pesquisando Sobre Determinado Assunto.

```
String info = "UNISINOS";

Query query = new Query(info);
QueryResult resultadoBusca = null;
try {
    resultadoBusca = twitter.search(query);
    List <Tweet> listaTweets = resultadoBusca.getTweets();
    Iterator <Tweet> itTweet = listaTweets.iterator();
    while (itTweet.hasNext()) {
        Tweet tweet = itTweet.next();
        System.out.println(tweet.getFromUser() + " - " + tweet.getText());
        System.out.println("-----");
    }
} catch (TwitterException e) {
    e.printStackTrace();
}
```

Fonte: elaborado pelo autor

²⁴<https://dev.twitter.com/pages/xauth>

²⁵<http://consultingblogs.emc.com/nileeshabojjavar/archive/2010/03/18/twitter4j-oauth-generating-the-access-token.aspx>

Na figura 7 temos um exemplo de código onde se mostra a linha do tempo do usuário do aplicativo do Twitter.

Figura 7: Linha do Tempo.

```
System.out.println("O que meus amigos estão fazendo:");
List<Status> linhaDoTempo = twitter.getHomeTimeline();
for (Status status : linhaDoTempo) {
    System.out.println(status.getUser());
    System.out.println(status.getText());
    System.out.println("-----" + i++);
}
}
```

Fonte: elaborado pelo autor

Na figura 8 temos um exemplo de código onde se mostra os *tweets* de um usuário do Twitter.

Figura 8: Tweets de um Usuário.

```
String pessoa = "screspo";
linhaDoTempo = twitter.getUserTimeline(pessoa);
System.out.println("O que " + pessoa + " está fazendo:");
for (Status status : linhaDoTempo) {
    System.out.println(status.getUser());
    System.out.println(status.getText());
}
}
```

Fonte: elaborado pelo autor

Na figura 9 temos um exemplo de código onde se obtém aproximadamente 1% de *tweets* do Twitter.

Figura 9: Código Que Obtém Aproximadamente 1% De Todos Os Tweets.

```
StatusListener listener = new StatusListener() {

    public void onStatus(Status status) {
        grava (++i + " " + status.getCreatedAt()
            + " @" + status.getUser().getScreenName()
            + " - " + status.getText());
    }

    twitterStream.addListener(listener);
    twitterStream.sample();
}
```

Fonte: elaborado pelo autor

Este código executou das 16:46:10 de 30/04/2011 até 20:59:33 de 03/05/2011, obtendo 4.166.813 *tweets*. A figura 10 mostra os primeiros e os últimos cinco *tweets* deste período.

Figura 10: Exemplos de Tweets.

```

1 Sat Apr 30 16:46:10 BRT 2011 @dulcemavera - y se daño de nuevo la tarjeta de mi cámara :@ #fail
2 Sat Apr 30 16:46:10 BRT 2011 @Anniehutzler - Watching the Departed(: It's great so far<3
3 Sat Apr 30 16:46:10 BRT 2011 @jirinabot - ...,(=?^??^=?),...
4 Sat Apr 30 16:46:10 BRT 2011 @jimmarcon - coloquei uma música e todo mundo se animou pra sair hoje! pelo visto,
    estudos ficou na história!
5 Sat Apr 30 16:46:10 BRT 2011 @DJSoulMUTT - Was just about to start givin away some free GAME....Had to hit that
    delet button, HA!

    ....
    ....
    ....

4166809 Tue May 03 20:59:33 BRT 2011 @ladibujante - @Dulcesita92 Si che la verdad q el ya,no se puede tirar un
    pedo aparte quien le cree a esa ? en verdad es una parabra contra la otra,q se yo
4166810 Tue May 03 20:59:33 BRT 2011 @gasliespasion - @benjaminamadeo Que grosso que sos Amadeo , admiro lo
    que estas haciendo , contestando a todo mundo , besos!
4166811 Tue May 03 20:59:33 BRT 2011 @VictorNogueiira - @CarlaMartins Hahahaha ai ai, e qué ter ? danadinha *-
    UAHUAHA
4166812 Tue May 03 20:59:33 BRT 2011 @d_mone_y305 - RT @mkeefor3: I'm better then Mario Chambers!
4166813 Tue May 03 20:59:33 BRT 2011 @heloliveira - RT @caodadepressao: Ela veio falar contigo. "Desculpa, te
    confundi com outra pessoa".

```

Fonte: elaborado pelo autor

2.2 Agentes de Softwares

A tecnologia de agentes é uma área de pesquisa relativamente nova e que tem atingido um alto nível de reconhecimento na comunidade científica internacional, devido ao potencial que possui para o projeto de sistemas complexos e sistemas situados em ambientes dinâmicos.

Esses tipos de sistemas são tipicamente difíceis de tratar usando concepções tradicionais de sistemas computacionais. A ideia de empregar agentes de software e delegar a eles certas atividades foi introduzida por pesquisadores que perceberam as grandes oportunidades desta promissora área (BORDINI, 2001).

2.2.1 Conceito de agentes

Um agente de software é definido como uma entidade de software autônoma que tem capacidades, interage com seu ambiente e adapta seu estado e comportamento com base nesta interação, a fim de atingir seus objetivos específicos. Este ambiente de interação pode ser constituído de máquinas, de humanos, e de outros agentes de software em vários ambientes e através de várias plataformas (DELOACH; WOOD, 2001).

Os agentes representam um tipo de abstração de software. Eles são autônomos, percebem seu ambiente, respondem às mudanças do ambiente e possuem condutas baseadas em metas. Muito mais do que ser definido em termos de atributos e métodos, um agente é definido pelo

seu comportamento. Ele não recebe simplesmente um conjunto de entradas que são processadas e geram uma saída, mas processa um conjunto de entradas segundo um propósito específico e é em função desse propósito que ele adapta o seu comportamento.

Durante o desenvolvimento de um sistema orientado a agentes, o foco do projetista passa a ser o comportamento de cada agente, avaliando as tarefas que o agente deve cumprir para atingir seu objetivo. A utilização da tecnologia de agentes tem uma importância muito grande em um mundo competitivo como o atual, pois os usuários são liberados de realizar tarefas repetitivas ou que possam ser automatizadas para fazerem outras mais importantes.

Um agente de software normalmente não é encontrado sozinho em uma aplicação, mas frequentemente forma uma organização com outros agentes, constituindo assim o que é denominada aplicação multi-agente. Uma aplicação multi-agente geralmente tem diversos tipos de agentes de software, como agentes de informação, agentes usuários e agentes de interface.

2.2.2 Propriedades dos agentes

Para que possamos diferenciar um agente de um programa convencional são necessárias algumas propriedades. Essas por sua vez são características comportamentais que um agente pode ter para alcançar seus objetivos. A seguir estão listadas algumas definições para as principais propriedades de agentes (WOOLDRIDGE, 2002). Em geral, autonomia, interação e adaptação são consideradas fundamentais em agentes de software.

- *Autonomia*: Um agente é capaz de agir sem intervenção externa direta, seja humana, ou de outro agente, e possui seu próprio controle podendo aceitar ou recusar uma mensagem;
- *Interação*: Um agente comunica-se com o ambiente e outros agentes;
- *Adaptação*: Um agente adapta e modifica seu estado de acordo com as mensagens recebidas do ambiente;
- *Aprendizagem*: Um agente pode aprender baseado em experiências anteriores, enquanto reage e interage com seu ambiente;
- *Mobilidade*: Um agente é capaz de mover-se de um ambiente para outro;
- *Colaboração ou Sociabilidade*: Um agente pode cooperar com outros agentes para alcançar seus objetivos e os do sistema também;
- *Reatividade*: Um agente é capaz de reagir a mudanças ocorridas no ambiente;
- *Proatividade*: Um agente é capaz de tomar iniciativas que visem alcançar seus objetivos, quando necessário.

Os agentes podem apresentar todas estas características e por isso podemos definir um agente como sendo uma entidade computacional que age no lugar de outra entidade de forma autônoma, realiza suas ações com um nível alto de pró-atividade e possui um nível de aprendizagem, cooperação e mobilidade.

2.2.3 Agentes e o Ambiente

A Figura 11 mostra o conceito apresentando a relação dos agentes com o seu ambiente. Pode-se perceber que o arco de entrada é caracterizado por sensores que percebem o ambiente, enquanto que, o arco de saída é caracterizado por ações que transformam o ambiente (WOOLDRIDGE, 2002).

Figura 11: Relação entre Agente e Ambiente.



Fonte: (WOOLDRIDGE, 2002)

A variedade de ambientes existentes é muito grande, porém pode-se identificar um número limitado de dimensões, das quais os ambientes podem ser organizados em categorias.

Tais dimensões determinam o projeto apropriado de agentes e a aplicabilidade de cada uma das principais técnicas de implementação. A seguir são listadas as categorias de dimensões de ambientes (NORVIG, 2004):

- *Completamente Observável versus Parcialmente Observável.* Um ambiente é completamente observável quando os agentes tem acesso ao estado completo do ambiente em todo instante de tempo. Por outro lado, um ambiente é parcialmente observável quando há ruídos, sensores imprecisos ou porque simplesmente não é possível conhecer algum estado;
- *Determinístico versus Estocástico.* O ambiente é determinístico quando seu próximo estado é completamente determinado pelo estado atual e pela ação do próprio agente, caso contrário ele é considerado estocástico. Em princípio, um agente não precisa se preocupar com a incerteza em um ambiente completamente observável e determinístico, mas um ambiente parcialmente observável pode parecer estocástico, principalmente se este for complexo, o que dificulta o controle dos aspectos não observados;
- *Episódico versus Sequencial.* Em ambiente episódico a experiência do agente é dividida em episódios atômicos, onde cada episódio consiste na percepção do agente e na execução

de uma ação. Cada ação ou decisão é totalmente independente, ou seja, não é influenciado ou influencia outras ações ou decisões. Já em um ambiente sequencial cada ação ou decisão tem impacto nas ações subsequentes do agente;

- *Estático versus Dinâmico.* Se um ambiente puder se alterar enquanto o agente está deliberando, este ambiente é dinâmico, caso contrário, ele é estático. Ambientes estáticos são mais fáceis de manipular, pois o agente não precisa continuar observando o mundo enquanto decide por uma determinada ação, enquanto que ambientes dinâmicos perguntam constantemente ao agente o que ele deseja fazer;
- *Discreto versus Contínuo.* A distinção entre ambiente discreto e contínuo é definida pelo modo que tempo é tratado, pelas percepções e pelas ações do agente. Um ambiente discreto possui um número finito de estados, percepções e ações, enquanto que, em um ambiente contínuo não é possível enumerar os estados, percepções ou ações envolvidas;
- *Agente Único versus Multi-Agente.* A diferença entre ambientes de um único agente e multi-agentes está na quantidade de agentes que interagem com o ambiente e entre si. Um ambiente de único agente possui um agente específico o qual tem percepções e realiza ações para um determinado fim. Por outro lado, em um ambiente multiagente os agentes, além de possuírem percepções e ações específicas, interagem entre si para atingir um objetivo comum.

2.2.4 Agentes e a sua Estrutura

Outra questão importante, a considerar no desenvolvimento de agentes, é a sua estrutura, ou seja, como estes são construídos e organizados internamente. Pode-se dizer que a estrutura de um agente é a união de uma arquitetura e um programa. O programa é a implementação da função que mapeia percepções e ações. Já a arquitetura são os sensores e atuadores que possibilitam interagir com o ambiente. A seguir, são listados os quatro tipos básicos de programas de agentes que incorporam a essência da maioria dos programas inteligentes (NORVIG, 2004).

- *Agentes Reativos Simples.* É o tipo mais simples de agentes, os quais selecionam ações com base na percepção atual, ignorando o restante do histórico de percepções;
- *Agentes Reativos Baseados em Modelos.* O agente mantém estados internos dependentes do histórico de percepções, refletindo pelo menos algum dos estados não observáveis a partir do estado atual. Este conhecimento sobre os possíveis estados é chamado de modelo do mundo;
- *Agentes Baseados em Objetivos.* O fato de conhecer os estados possíveis nem sempre é suficiente para decidir o que fazer. É importante que o agente além do conhecer o

universo de estados também tenha alguma informação sobre seus objetivos que descreva as situações desejáveis;

- *Agentes Baseados em Utilidade.* Somente os objetivos não são suficientes para gerar agentes de alta qualidade. Uma função de utilidade mapeia um estado, ou uma sequência de estados, em um valor que descreve o grau de "felicidade" do agente, assim este pode tomar suas decisões com base neste grau, buscando sempre os estados mais funcionais.

2.3 Sistemas Multi-Agentes

Sistemas multi-agentes são sistemas constituídos de múltiplos agentes que interagem ou trabalham em conjunto de forma a realizar um determinado conjunto de tarefas ou objetivos. Esses objetivos podem ser comuns a todos os agentes ou não. Os agentes dentro de um sistema multi-agente podem ser heterogêneos ou homogêneos, colaborativos ou competitivos, ou seja, a definição dos tipos de agentes depende da finalidade da aplicação que o sistema multi-agente está inserido (WOOLDRIDGE, 2002).

Os sistemas multi-agentes com agentes reativos são constituídos por um grande número de agentes. Estes são bastante simples, não possuem inteligência ou representação de seu ambiente e interagem utilizando um comportamento de ação e reação. A inteligência surge conforme ocorrem as interações entre os agentes e o ambiente. Ou seja, os agentes não são inteligentes individualmente, mas o comportamento é global.

Já os sistemas multi-agentes constituídos por agentes cognitivos são geralmente compostos por uma quantidade bem menor de agentes se comparado aos sistemas multi-agentes reativos. Estes, conforme a definição de agentes cognitivos, são inteligentes e contêm uma representação parcial de seu ambiente e dos outros agentes. Podem, portanto, comunicar-se entre si, negociar uma informação ou um serviço e planejar uma ação futura. Em geral os agentes cognitivos são dotados de conhecimentos, competências, intenções e crenças que lhes permite coordenar suas ações visando à resolução de um problema ou a execução de um objetivo.

Para (WOOLDRIDGE, 2002), algumas razões para o crescimento do interesse em pesquisas com sistemas multi-agentes são:

- A capacidade de fornecer robustez e eficiência;
- A capacidade de permitir interoperabilidade entre sistemas legados;
- A capacidade de resolver problemas cuja especialidade ou controle é distribuído.

Apesar de muitas vantagens, os sistemas multi-agentes ainda possuem muitos desafios e dificuldades tanto em relação a projeto quanto implementação. Segundo (WOOLDRIDGE, 2002) estes desafios podem ser resumidos em apenas três:

- *Comunicação:* Como ela deve ser realizada entre os agentes e que tipo de protocolos usar;

- *Interações*: Como vão ocorrer as interações entre os agentes e que linguagem eles devem usar para interagirem entre si e combinar seus esforços;
- *Coordenação*: Como garantir essa coordenação entre os agentes para que haja uma coerência na solução do problema ou objetivo que estão tentando resolver.

O desenvolvimento de sistemas multi-agentes é uma área em franca expansão devido a sua adaptação a resolução de problemas de natureza distribuída . Devido a isso, a pesquisa em sistemas multi-agentes está interessada principalmente em observar como eles coordenam seu conhecimento, planos, objetivos e crenças com o objetivo de tomar ações ou resolver problemas.

2.3.1 Metodologias da Engenharia de Software Orientada a Agentes

Algumas metodologias voltadas ao projeto de sistemas orientados a agentes surgiram ao longo dos tempos, sendo que as mais discutidas na literatura atualmente são: MaSE, Prometheus, Tropos e Gaia. Essas metodologias foram selecionadas baseadas em vários fatores, como a abrangência na engenharia de software orientada a agentes, a difusão na comunidade de pesquisadores e os recursos disponíveis pelas metodologias como documentação detalhada e de fácil acesso e suporte à ferramenta. A seguir, serão expostas algumas das características de cada uma destas metodologias:

- **MaSE (Multiagent Systems Engineering Methodology)**: É uma metodologia orientada a agentes onde os agentes não são considerados como autônomos e pró-ativos. Eles são processos de software simples que interagem uns com os outros para atingir a finalidade geral do sistema (DELOACH; WOOD, 2001). O objetivo principal da MaSE é guiar o projetista através de todo o ciclo de vida do software, independentemente de qualquer arquitetura de agente, linguagem de programação ou sistema de troca de mensagens. A metodologia MaSE consiste da fase de análise e da fase de projeto e possui uma ferramenta denominada AgentTool2 que auxilia no processo de modelagem;
- **Prometheus**: É uma metodologia na engenharia de software orientada a agentes para especificar, projetar e implementar sistemas de agentes (WIKIKOFF, 2005). O principal objetivo no desenvolvimento desta metodologia é ter um processo definido, que possa ser usado no desenvolvimento de sistemas de agentes e que possa ser ensinado aos profissionais da área e aos estudantes que não possuam conhecimento em agentes. Prometheus suporta o desenvolvimento de agentes inteligentes que possuem objetivos, crenças, planos e eventos. A metodologia utiliza o conceito de capacidade, a qual pode ser composta por planos, eventos, crenças e outras capacidades que dão habilidades específicas ao agente;
- **Tropos**: É uma metodologia de desenvolvimento de software orientado a agentes, criada por um grupo de pesquisadores de várias universidades do Brasil, do Canadá e da Itália (BRESCIANI et al., 2004). A metodologia usa a noção de agente em seus conceitos, tais

como objetivos, planos e tarefas durante as fases do desenvolvimento do software e adota uma abordagem de refinamento de passos, utilizando um conjunto específico de operadores de transformação. Diferentemente das demais metodologias apresentadas, Tropos tem um foco maior na fase de requisitos iniciais, onde os conhecimentos e as necessidades dos clientes são identificados e analisados. A metodologia consiste das fases de requisitos iniciais, requisitos finais, projeto arquitetural, projeto detalhado e implementação;

- **Gaia:** É uma metodologia de uso geral, aplicável à uma gama de domínios de aplicação. Para Gaia, um sistema multi-agente é visto como uma organização composta de vários papéis. A metodologia provê um *framework* conceitual que permite ao analista sistematicamente refinar requisitos de maneira tal que estes possam ser diretamente implementados, detalhando papéis, interações, responsabilidades, permissões e propriedades na fase de análise (WOOLDRIDGE, 2002). Apesar de ser uma metodologia de uso geral, Gaia é voltada aos aspectos organizacionais e fornece um formalismo que dá uma ênfase muito maior à etapa de análise, sendo que a etapa de projeto provê apenas um detalhamento mais refinado, produto do trabalho já realizado.

Além das metodologias de desenvolvimento de software orientado a agentes citadas acima, existe a linguagem de modelagem AUML (Agent Unified Modeling Language) que é uma extensão da linguagem de modelagem UML (Unified Modeling Language), proposta por (ODELL; PARUNAK; BAUER, 2000), concebida para dar suporte ao desenvolvimento de sistemas orientados a agentes. Essa extensão visa exclusivamente estender a linguagem UML de modo a dar suporte à especificação de protocolos de interação entre agentes, não se preocupando em fornecer artefatos para todo o ciclo de vida do desenvolvimento de um sistema orientado a agentes. Os diagramas AUML são utilizados em diversas metodologias orientadas a agentes com o objetivo de fornecer uma semântica semi-formal e intuitiva através de uma notação gráfica amigável para o desenvolvimento de protocolos de interação entre agentes.

3 TRABALHOS RELACIONADOS

Neste capítulo são apresentados os principais trabalhos e ferramentas encontradas cujas funcionalidades foram analisadas e estudadas a fim de entender os seus benefícios, com o intuito de ajudar na confecção desta dissertação.

3.1 Análise do Espaço Demográfico

Na ampla pesquisa bibliográfica que foi feita, foram encontrados alguns trabalhos analisando o espaço demográfico de redes e a formação de sub-redes em rede sociais, como em (MEETEREN; POORTHUIS; DUGUNDJI, 2010), (OH; YEOM; AHN, 2011), (PHITHAK-KITNUKON; DANTU, 2010), (ALANI et al., 2003), (SATHIK M.M.; RASHEED, 2009) e (SHENQING, 2010). Dentre estes escolhemos por apresentar o trabalho de (MEETEREN; POORTHUIS; DUGUNDJI, 2010), devido o mesmo ser sobre o Twitter e seguir uma metodologia detalhadamente explicada, sendo inclusive fonte de inspiração para este trabalho.

3.1.1 Mapping Communities In Large Virtual Social Networks: Using Twitter Data To Find The Indie Mac Community

Neste relevante trabalho (MEETEREN; POORTHUIS; DUGUNDJI, 2010), é analisada uma abordagem multi-método que delinea uma comunidade real no Twitter. O ponto de partida é uma comunidade de desenvolvedores independentes ("*Indie*") que criam software para o Macintosh e para o iPhone e são muito ativos no Twitter, aonde sustentam interações entre pares (socializam), trocam informações técnicas e fazem marketing. O público avaliado através da API do Twitter é usado para fazer prospecção de dados em uma rede constituída por milhões de arestas. Um rápido algoritmo guloso é usado para detectar subgrafos dentro desta grande rede, tendo dois objetivos:

- isolar e mapear a comunidade *Indie* através do Twitter e
- descrever a estrutura social da comunidade analisando uma série de medidas de rede.

Uma questão adicional levantada no trabalho foi o papel dos formadores de opinião. Ter seu software aprovado pela comunidade pode ser de muita valia em termos econômicos. A literatura sobre a indústria cultural tem enfatizado o papel dos formadores de opinião neste contexto (CAVES, 2000). Um formador de opinião poderia ser definido como um ator intermediário que tem poder, muitas vezes simbólico, e utiliza esse poder, por exemplo, para ajudar os autores selecionados a se tornarem economicamente bem sucedidos. Ao longo da história, jornalistas, críticos de arte e artistas famosos desempenharam papéis importantes na geração de novos talentos famosos: como pode ser encontrado na indústria da música (CAVES, 2000) ou no campo literário francês do século 19 (BOURDIEU, 1996). Na comunidade *Indie* acontece o

mesmo onde estas pessoas muitas vezes funcionam como uma ponte entre o grupo de desenvolvedores e o primeiro nível de usuários críticos. A pessoa que recorrentemente surge como um importante formador de opinião é o jornalista de tecnologia John Gruber, que mantém o *blog* *Daring Fireball*¹. Gruber, que tem formação em ciência da computação, é considerado como um importante conhecedor de software e indústria de tecnologia, tanto por desenvolvedores *Indie* como pelo público. O RSS (*Really Simple Syndication*) *feed* do seu *blog* tem mais de 150 mil assinantes e o site tem um número estimado de 1,3 milhões de visualizações por semana. Ele também é significativo pelo estatística do seu Twitter, em 16 de maio de 2009, ele tinha 27191 seguidores, enquanto seguia apenas 311 pessoas.

Um dos problemas ao pesquisar uma comunidade virtual é como delimitar a população. Tendo sido John Gruber identificado como um exemplo de um importante formador de opinião, pode-se supor que a maioria dos desenvolvedores que fazem parte da comunidade *Indie* e que usam o Twitter podem ser encontradas dentro de um grau de separação de John Gruber. Uma vez que a população foi definida, pode-se usar medidas de centralidade da rede para analisar a estratificação da comunidade de desenvolvedores *Indie*.

3.1.1.1 Primeira Fase: Coleta de Dados

A rede Twitter é uma rede dirigida. Os usuários podem seguir outros usuários, mas esta ligação não tem que ser recíproca. No Twitter cada usuário tem "seguidores" (*out-degrees*) e "amigos" (*in-degrees*). John Gruber foi originalmente usado como ponto de partida da rede. Cada usuário que estava dentro de 2 graus de Gruber era incluído. Na prática, isso significa que todos que os amigos (cerca de 300 usuários) e seguidores (cerca de 27.000 usuários) de Gruber, e todos os amigos e seguidores que seguem ou que sejam amigos de Gruber. Para cada vértice na rede, metadados foram armazenados, incluindo nome, localização, descrição, data de adesão, fonte dos *tweets* (cliente Twitter) e o número total de *tweets* postado.

3.1.1.2 Corte

Esta primeira etapa de coleta de dados resultou em uma enorme rede com 4 milhões de vértices e 28 milhões de arestas. Cortando a rede de Gruber em 1 grau de separação, o resultado foi uma rede com 27.218 vértices e 1,5 milhões de arestas. Foram excluídos também todos os usuários que seguem mais de 600 pessoas, por considerar impossível um ser humano acompanhar mais de 600 pessoas, ou por estas contas possivelmente serem institucionais, o que produziu uma rede consideravelmente menor com cerca de 22.000 vértices e 440.000 arestas.

¹<http://daringfireball.net>

3.1.1.3 Detecção de Comunidades

Usando um software estatístico, uma biblioteca para redes sociais e rápidos algoritmos gulosos foram obtidos 5 subgrafos, contendo respectivamente 5577, 8780, 445, 6.172 e 780 vértices.

3.1.1.4 Comparação com Dados de Campo Qualitativos

Comparando com dados de campo qualitativos (MEETEREN, 2009), revelou-se que todos os entrevistados foram encontrados no subgrafo 4 (6172 vértices). Isso dá razão suficiente para acreditar que subgrafo 4 contém o grupo de desenvolvedores *Indie*. Medidas de centralidade de rede (grau de entrada, grau de saída, proximidade, intermediação e autovetor) são assim calculados para este subgrafo. Da análise dos vértices centrais, torna-se claro que, apesar da rede de Gruber dar uma notável imagem da comunidade *Indie*, algumas peças importantes faltaram. Portanto, foram introduzidos vértices adicionais e os mesmos dados para eles, como havia sido feito para Gruber.

3.1.1.5 Segunda Fase: Nova Coleta de Dados

São coletados os dados dos 15 vértices mais centrais da comunidade e 7 vértices adicionais com base em resultados qualitativos de trabalho de campo. Como resultado, 23 vértices (incluindo John Gruber) são utilizados como um novo ponto de partida para a coleta de dados repetindo o mesmo procedimento descrito para a primeira fase.

3.1.1.6 Novo Corte

Em seguida foi aplicado o mesmo procedimento de corte nesta rede estendida para obter valores gerenciáveis. Só são selecionados novos vértices que estão dentro de um grau de separação. Isso resultou em uma rede de aproximadamente 52.000 vértices. Então foram selecionados somente os vértices que seguem menos de 600 outros utilizadores, com base nos mesmos pressupostos do primeiro corte. O resultado foi uma rede final para análise que consistiu de 40.512 vértices e 1.023.317 arestas.

3.1.1.7 Nova Detecção de Comunidades

Executando novamente o rápido algoritmo guloso, nesta nova rede, obteve-se uma identificação mais precisa da comunidade de desenvolvedores *Indie* do que baseada na rede de Gruber sozinho. Foram obtidos 5 subgrafos, contendo respectivamente 13.978, 11.522, 591, 1.428 e 12.669 vértices. A triangulação com dados qualitativos mostra que a detecção produziu resultados perspicazes, estando todos os entrevistados agora no subgrafo cinco ($N = 12.669$), que foi

posteriormente chamado de "a comunidade *Indie*."

3.1.1.8 Comparação da Rede Total com a Rede Indie

Para as sub-redes geradas, foram criadas nuvens de palavras, conforme a figura 12, usando a frequência das palavras no campo "descrição" dos vértices, onde as palavras com maior frequência são maiores do que aquelas palavras que aparecem menos. Com isto obtêm-se uma visão sobre a essência de cada comunidade rapidamente. A nuvem total da rede mostra um reflexo dos interesses dos leitores do *blog* de Gruber. Nela podemos discernir claramente várias palavras-chaves relativas a criatividade na indústria e tecnologia, o que é de se esperar de um guru da tecnologia criativa. Palavras-chave como fotografia, design, música, *web*, tecnologia, mídia, geek e desenvolvedor se destacam.

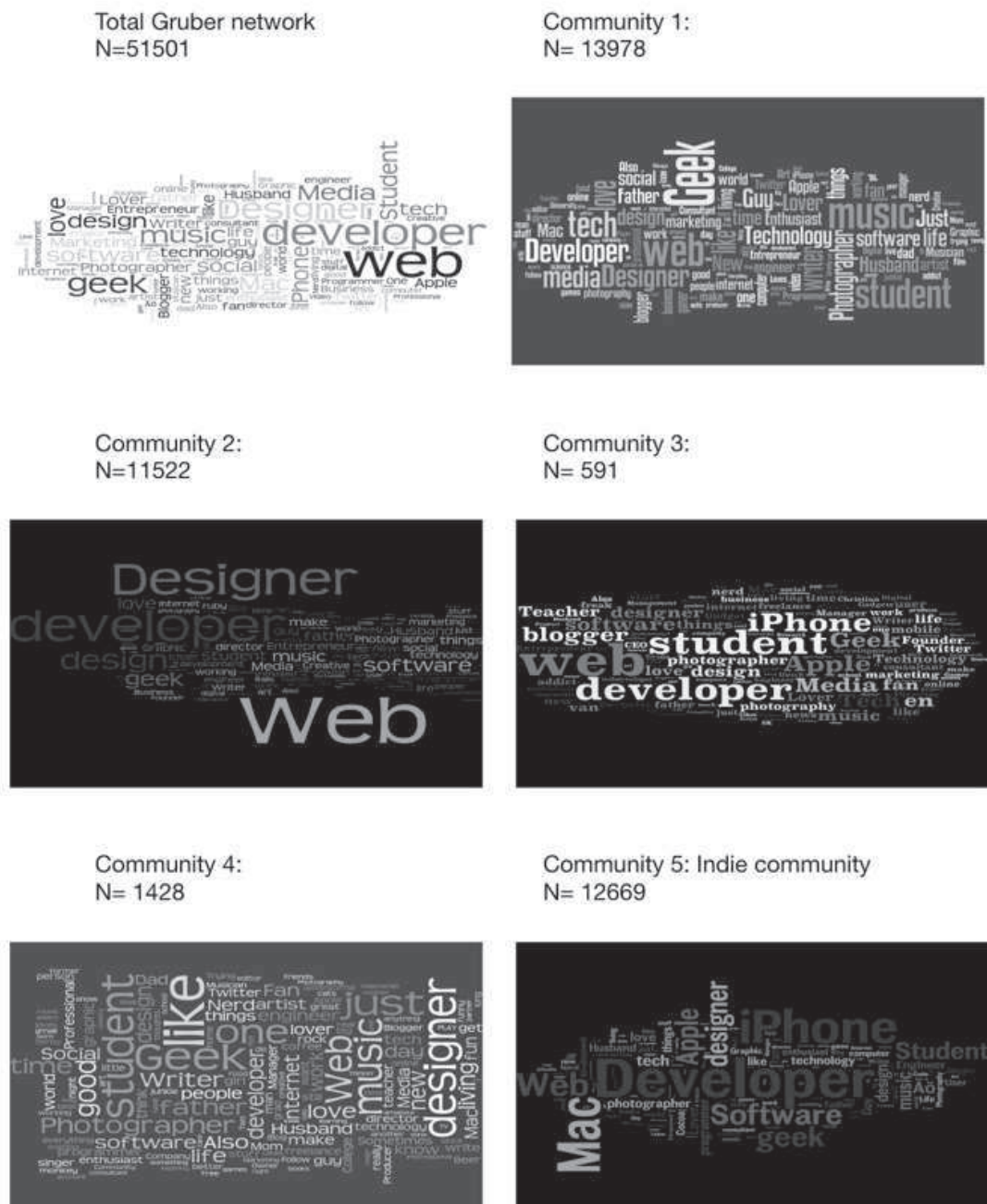
Em contraste, a comunidade cinco, a *Indie*, parece capturar o interesse dos desenvolvedores de software. Palavras-chave como: Mac, desenvolvedor, iPhone, software e Apple se destacam. Isso já é uma reflexão qualitativa que o trabalho de detecção na comunidade cinco fez um bom trabalho de isolamento da comunidade pertinente em toda a rede, lembrando que, para tanto, não foram usadas informações do que foi escrito, só as arestas. A comunidade três parece se sobrepor a comunidade 5, mas com uma maior variedade. A comunidade dois é interessante porque parece captar outro campo específico da audiência de Gruber: *web*, desenvolvimento e concepção dominam esta sub-rede. Estas palavras indicam um interesse para temas como *web*, desenvolvimento e *design*, que é um campo distinto de desenvolvimento de software para Mac / iPhone. Por último, as comunidades um e quatro mostram uma coleção mais variada de palavras-chave. Elas capturam palavras de outras áreas criativas, como música, fotografia e mídia e mostram um maior grau de palavras-chave de "consumidor", como ventilador, estudante, *geek* e entusiasmo.

3.1.1.9 Análise da Comunidade Indie

Foram usadas as seguintes medidas de centralidade da rede para a análise:

- **Grau de entrada (in-degree):** O número de outros atores (vértices) na rede que um determinado ator segue (recebe informações);
- **Grau de Saída (out-degre):** O número de outros atores na rede que seguem um determinado ator (envio de informações). Atores (vértices) com um valor alto de grau de saída tendem a ser muito populares dentro da comunidade *Indie*;
- **Proximidade (closeness):** Medida para indicar a distância total de um vértice para os outros vértices na rede (FREEMAN, 1979). No trabalho apresentado foi usada uma medida invertida, ou seja, um vértice de maior pontuação é o mais próximo em relação a todos os outros vértices na rede;

Figura 12: Nuvens de Palavras das Comunidades e Sub-Comunidades Gruber



Fonte: (MEETEREN; POORTHUIS; DUGUNDJI, 2010)

- **Intermediação (betweenness):** Indica quantas vezes um vértice cai no caminho mais curto entre dois outros vértices (FREEMAN, 1979);
- **Autovetor ou vetor característico (eigenvector):** Medida destinada a medir o "poder" de posição dentro de uma rede social (BONACICH, 1987). Vértices com uma pontuação alta no ranking autovetor são conectados a outros vértices com uma centralidade autovetor alta. Assim, a centralidade autovetor é uma potente ferramenta para identificar o

"núcleo" de uma comunidade, em termos de conectividade de rede. Autovetor foi utilizada como ordem de classificação primária no trabalho, por ser a forma mais potente de harmonizar o status de um desenvolvedor Indie em relação ao seu grupo de pares. Além disso, um controle qualitativo sobre o assunto na comunidade Indie reconheceu que o ranking por centralidades autovetor era muito plausível e significativo.

Além das centralidades, foram inferidas qualitativamente a ocupação dos vértices para os vinte vértices de maior ranking em cada ranking. Sete diferentes "ocupações" foram estabelecidas, sendo três discutidas na análise:

- **Desenvolvedor:** uma categoria genérica de engenheiros de software e desenvolvedores que não trabalham para a Apple
- **Apple:** as pessoas que eram funcionários da Apple Inc
- **Jornalista:** jornalistas de tecnologia ou os mantenedores de *blogs* populares

Análises parciais podem ser vistas nas tabelas 1, 2 e 3.

Tabela 1: Ranking dos Desenvolvedores

| Eigenvector | In-degree | Out-degree | Closeness | Betweenness |
|-------------|-----------|------------|-----------|-------------|
| 1 | 114 | 8 | 4 | 4 |
| 2 | 17 | 13 | 8 | 2 |
| 3 | 78 | 12 | 7 | 8 |
| 4 | 98 | 1 | 1 | 1 |
| 5 | 1138 | 4 | 3 | 13 |
| 6 | 2 | 20 | 11 | 3 |
| 7 | 962 | 6 | 2 | 6 |
| 8 | 531 | 11 | 9 | 15 |
| 9 | 5 | 36 | 26 | 9 |
| 10 | 196 | 21 | 17 | 42 |
| 11 | 74 | 25 | 21 | 11 |
| 13 | 107 | 1007 | 19 | 23 |
| 14 | 1915 | 1707 | 15 | 38 |
| 15 | 77 | 523 | 61 | 95 |
| 16 | 518 | 1355 | 14 | 39 |

Fonte: (MEETEREN; POORTHUIS; DUGUNDJI, 2010)

3.1.1.10 Considerações

Os quadros não foram totalmente detalhados por questões de privacidade e problemas legais em potencial. Nos vinte vértices com maiores grau de saída, vértices com alta popularidade, pois a maioria das pessoas opta por seguir esses vértices, estão contas institucionais

Tabela 2: Ranking dos Desenvolvedores Apple

| Eigenvector | In-degree | Out-degree | Closeness | Betweenness |
|--------------------|------------------|-------------------|------------------|--------------------|
| 12 | 208 | 22 | 20 | 29 |
| 17 | 105 | 37 | 42 | 41 |
| 20 | 171 | 48 | 23 | 96 |
| 21 | 66 | 42 | 40 | 30 |
| 22 | 42 | 80 | 28 | 31 |
| 31 | 590 | 43 | 141 | 211 |
| 35 | 1130 | 41 | 82 | 238 |

Fonte: (MEETEREN; POORTHUIS; DUGUNDJI, 2010)

Tabela 3: Ranking dos Jornalistas

| Eigenvector | In-degree | Out-degree | Closeness | Betweenness |
|--------------------|------------------|-------------------|------------------|--------------------|
| 19 | 2054 | 15 | 13 | 70 |
| 25 | 1351 | 9 | 10 | 20 |
| 27 | 495 | 16 | 16 | 37 |
| 29 | 101 | 49 | 108 | 27 |
| 53 | 419 | 45 | 371 | 53 |

Fonte: (MEETEREN; POORTHUIS; DUGUNDJI, 2010)

que não foram pegas no corte dos 600 seguidores da seção 3.1.1.2, jornalistas e alguns "visíveis"desenvolvedores. No entanto, os jornalista e contas institucionais possuem valores para autovetor mais baixos. Outra constatação interessante é que os jornalistas possuem valores baixos no ranking de intermediação, enquanto que se destacam no ranking de proximidade.

Quando foram examinados os desenvolvedores, foi verificado que o maior grau de classificação em termos de centralidades autovetor geralmente têm alto grau de saída, proximidade e intermediação. Curiosamente, os últimos 4 desenvolvedores da figura 1 não se enquadram neste padrão. Estes desenvolvedores podem ser considerados "na multidão", são pessoas que possuem um status elevado, mas não são tão publicamente conhecidas como os outros.

Essa percepção é reforçada quando foram olhados os dados dos trabalhadores da Apple, que possuíam altas taxas de centralidade autovetor, mas com valores relativamente fracos para os demais escores. A lógica por trás disso é que os desenvolvedores da Apple estão supostamente menos inclinados a se tornarem "famosos"na comunidade Indie.

3.1.1.11 Conclusões sobre o Trabalho

Os resultados da análise da nuvem de palavras e da comparação das centralidades mostrou que é possível isolar sub-comunidades significativas, que partilham interesses comuns dentro, de uma rede maior, através de um rápido algoritmo guloso.

O trabalho também constatou que é possível avaliar a extensão em que laços *online* refletem os laços sociais no "mundo real", fazendo inferências em grandes redes sociais. As nuvens de palavras diferentes nos subgrafos abrem oportunidades para pesquisa de mercado e marketing direto. Pode-se acompanhar as mudanças na composição das comunidades e pode-se avaliar as mensagens no Twitter, para verificar a eficácia de uma campanha de marketing. Ou pode-se identificar os vértices centrais de um movimento como o ocorrido no Irã, mesmo que estes não sejam visíveis para o público em geral.

E por fim, o trabalho sugere como estudo futuro a investigação do conteúdo dos *tweets* e a dispersão de *retweets*, analisando a evolução e comportamento ao longo do tempo. Esta sugestão acabou por se tornar boa parte da proposta desta dissertação.

3.2 Relevância de Mensagens em Redes Sociais

Certamente um dos melhores indicadores para considerar o impacto de algo postado, seja o ato de mensurar o repasse de uma informação. Neste sentido o estudo do envio de *retweets* é de suma importância. Dos três trabalhos desta seção, os dois primeiros tratam desta questão específica: *retweets*. O terceiro e último trabalho desta seção aborda a relevância como ela foi considerada neste trabalho, qual seja: a determinação de quais usuários podem ter contato, serem sensibilizados por uma informação.

3.2.1 Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter

O artigo (BOYD; GOLDBER; LOTAN, 2010) aborda uma série de questões relevantes, descreve as diversas convenções e examina as práticas do ato de enviar *tweets* e *retweets*.

No Twitter conversas dirigidas geralmente envolvem o uso da sintaxe "usuário @" para se referir aos outros e endereçar mensagens. Um subconjunto de participantes também usa *hashtags* (#'s) para marcar o tópico de um *tweet*, a fim de que outros possam seguir conversas centradas em um assunto em particular. E uma terceira convenção é conhecida como enviar *retweets* que, estruturalmente falando, é o equivalente do Twitter ao encaminhamento de e-mails, onde usuários postam mensagens originalmente postadas por outros. Esta convenção não tem gramática uniforme, embora "RT @ usuário mensagem" seja uma forma padrão. Já que o Twitter limita os *tweets* a 140 caracteres, mensagens que sofrem *retweets* devem frequentemente ser modificadas para acomodar anotações adicionais destinadas a indicar que a mensagem é um *retweet*.

Enquanto enviar *retweets* pode simplesmente ser visto como o ato de copiar e retransmitir,

a prática contribui para que conversas sejam compostas por uma interação pública de vozes que dão origem a um contexto conversacional compartilhado. Por esta razão, alguns dos mais visíveis participantes do Twitter enviam *retweets* de mensagens de outros e procuram que suas mensagens sofram *retweets*. Isso inclui usuários de todos os tipos mas principalmente comerciantes, celebridades e políticos. Uma vez que a estrutura do Twitter dispersa conversas através de uma rede de atores interligados, em vez de restringir a conversa dentro de espaços delimitados ou grupos, muitas pessoas podem falar rapidamente sobre um tópico em particular, de forma que os outros têm a sensação de estarem rodeados por uma conversa, apesar de talvez não serem um contribuinte ativo. O fluxo de mensagens fornecidas pelo Twitter permite que os indivíduos fiquem ciente periféricamente, sem participar diretamente. Enviar *retweets* traz novas pessoas para uma determinada *thread*, convidando-as a se envolver, sem diretamente endereçá-las. Enviar *retweets* pode ser entendido tanto como uma forma de difusão da informação, como meio de participar de uma conversa difusa. Espalhar *tweets* não serve só para receber mensagens de novas audiências, mas também para validar e se envolver com os outros.

3.2.1.1 Sintaxe para enviar *retweets*.

Não há consenso acerca da sintaxe para enviar *retweets*, embora a formulação inicial tenha sido "RT @ Utilizador ABC", onde o usuário de referência é o autor original e ABC é o conteúdo do *tweet* original. As seguintes sintaxes também são usadas para marcar *retweets*: "RT: @ ", "Retweeting @ ", "retweet @ ", "(via @)", "RT (Via @)", "thx @", "HT @" e "r @". Os usuários têm diferentes razões para a escolha destes estilos diferentes. Alguns são padrões em softwares de terceiros. Outros usam diferentes convenções dependendo do que eles pretendem alcançar. Cada estilo também tem uma convenção diferente para localizá-lo no *retweet*, que influencia as práticas de algumas pessoas. Embora a maioria destas aparecem antes do conteúdo, "(via @ Utilizador)" e "thx usuário" geralmente vêm no final. Alguns participantes também adicionam conteúdo adicional quando enviam *retweets*, quer seja antes da mensagem, entre parênteses, ou no final da mensagem. Outra questão relevante, levantada pelo estudo, é que URLs, em especial as que são reduzidas, têm a sua própria assinatura e a sua reutilização sugere uma referência a um *retweet*.

3.2.1.2 Por que as pessoas enviam *retweets*?

Enviar *retweets* não é uma prática universalmente adotada no Twitter e aqueles que enviam *retweets* não são necessariamente representante de todos os tipos de usuários do Twitter. Há muitos incentivos diferentes para usar o Twitter (JAVA et al., 2007) e aqueles que o estão usando para "conversa diária" tem menos probabilidade de sofrerem *retweets* do que aqueles que estão tentando se envolver em conversas ou compartilhar informações. Entre a sub-população que faz *retweet*, há diversas motivações para o fazer. Uma lista não exaustiva das motivações, com base

nas respostas a um questionário, incluem:

- Para aumentar ou espalhar *tweets* a novos públicos;
- Para entreter ou informar um público específico;
- Para comentar sobre o *tweet* de alguém, com adição de novos conteúdos;
- Para começar uma nova conversa;
- Para mostrar-se presente como um ouvinte visível;
- Para publicamente concordar com alguém;
- Para validar os pensamentos dos outros;
- Como um ato de amizade, lealdade ou por homenagem, chamando a atenção, por vezes, através de um pedido de *retweet*;
- Para reconhecer ou se referir a pessoas menos populares, ou com conteúdo menos visível;
- Para ganho próprio, seja para ganhar seguidores ou reciprocidade de participantes mais visíveis;
- Para salvar *tweets* para acesso futuro pessoal.

Enquanto alguns enviam *retweets* de valor, outros usuários reclamam de *tweets* com motivações egoístas. Enviar *retweets* pode ser tanto uma ferramenta de comunicação produtiva, como um ato egoísta de pessoas que procuram atenção.

3.2.1.3 O que as pessoas enviam em *retweets*?

O que as pessoas enviam em *retweets* também varia, embora esteja fortemente ligado às razões do porquê elas enviam *retweets*. Ao perguntar às pessoas o que elas enviam em *retweets*, alguns relataram que preferem enviar *retweets* de material sensível ao tempo e notícias de última hora. Mas há divergências quanto a que tipo de material sensível ao tempo vale a pena espalhar. Notícias tendem a sofrer *retweets* na forma de *links* para artigos da mídia. *Links* de muitos tipos são enviados, incluindo material que é de interesse permanente para amigos ou de tópicos relevantes para aqueles interessados em um tema específico.

3.2.1.4 Enviar *retweets* para os outros.

Enquanto alguns usuários enviam *retweets* com conteúdo de interesse geral, outros estão mais preocupados com o público para quem vão enviar os *retweets*. Ao escolher o que enviar em *retweets*, estes participantes muitas vezes pensam explicitamente em quem segue os seus

tweets. Embora os usuários do Twitter possam acessar uma lista de quem os segue, isso não é necessariamente o seu público real. Os participantes devem lidar com um público imaginado, assim como se faz quando se utiliza outros meios de comunicação social (BOYD, 2008).

Quando os participantes optam por enviar *retweets*, pode haver uma sobreposição entre o seu público em potencial e o público em potencial do autor, mas é improvável que a pessoa que enviou o *retweet* saiba qual sobreposição possa haver. Comparar seguidores em potencial e amigos, pode não ser necessariamente o melhor método de medição de quem está prestando atenção a quem. (HUBERMAN; ROMERO; WU, 2008) aponta que usuários que respondem a uma mensagem, usando *@reply*, mostram uma forma mais significativa de união.

3.2.1.5 *Retweets* para a ação social.

Muitos *retweets* parecem encorajar diferentes tipos de "ação social". Alguns têm pedidos graves em si mesmo, tais como chamadas para protesto ou doações. Essas chamadas podem ser bastante eficazes. Centenas de usuários enviaram *retweets* da mensagem "**RT @StopAhmadi Bring down Khomeini's website**" com um *link* para o site. Em seguida o site falhou.

Outros *retweets* de ação social envolvem a demonstração de uma decisão coletiva. Alguns deste *tweets* contem o texto, "RETWEET THIS UNTIL IT TRENDS". Ou seja, as pessoas que enviam *retweets* tentam fazer o tópico popular o suficiente para aparecer na página de "tópico de tendências" do Twitter e ser, como consequência, transmitido para um grande número de usuários do Twitter. Este é um exemplo de tentativa de uso do Twitter, para manipular o próprio sistema do Twitter, a fim de efetuar uma ação social.

3.2.1.6 Conclusões do Trabalho

O estudo é importante porque descreve as diversas variações na prática de enviar *retweets* de mensagens no Twitter e as maneiras pelas quais estilos variados procuram manter a autoria e fidelidade na conversação, no formato restrito de 140 caracteres. Explica porque as pessoas enviam *retweets*, o que as pessoas enviam em *retweets* e enviam *retweets* para quem.

3.2.2 Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network

O artigo (SUH et al., 2010) examina uma série de características que podem afetar a taxa de *retweets*, através da identificação de fatores em um conjunto de 74 milhões de *tweets* analisados. Os estudos de URLs, *hashtags*, número de seguidores, número de pessoas seguidas (amigos), idade da conta no Twitter e o número de *tweets* enviados são os principais apresentados.

3.2.2.1 URL

Os estudos do trabalho mostraram que 21,1% dos *tweets* tem pelo menos uma URL no seu texto. Examinando somente *retweets*, obteve-se 28,4%. Ou seja, um *tweet* com URLs é mais provável de ser reenviado. O resultado difere um pouco dos estudos de (ZARRELLA, 2009a) que constatou que 18,96% dos *tweets* contêm uma URL e 56,69% de *retweets* incluem uma URL. O trabalho também identificou os sites mais populares conforme pode ser verificado na tabela 4.

Tabela 4: Sites mais populares em Tweets e Retweets

| Rank | Domain | In Tweet | In Retweet | Retweet Rate |
|------|---|----------|------------|--------------|
| 1 | http://twitpic.com | 793680 | 129692 | 1.47 |
| 2 | http://myloc.me | 533082 | 121950 | 2.05 |
| 3 | http://www.facebook.com | 481349 | 55186 | 1.03 |
| 4 | http://www.youtube.com | 475509 | 79404 | 1.50 |
| 5 | http://formspring.me | 455377 | 2566 | 0.05 |
| 6 | http://www.twitlonger.com | 349760 | 236435 | 6.06 |
| 7 | http://tweetphoto.com | 258049 | 49676 | 1.73 |
| 8 | http://youtu.be | 196557 | 7508 | 0.34 |
| 9 | http://twitcam.com | 159684 | 2187 | 0.12 |
| 10 | http://twitter.com | 144002 | 39127 | 2.44 |
| 14 | http://foursquare.com | 90328 | 1763 | 0.18 |
| 19 | http://www.flickr.com | 47181 | 7599 | 1.44 |
| 20 | http://mashable.com | 43722 | 17778 | 3.65 |
| 25 | http://news.bbc.co.uk | 36286 | 6103 | 1.51 |
| 29 | http://www.nytimes.com | 31339 | 9035 | 2.59 |

Fonte: (SUH et al., 2010)

3.2.2.2 Hashtag

Hashtag é simplesmente uma palavra-chave precedida por um caractere "#" (por exemplo, #nowplaying). *Hashtags* são frequentemente usados como uma palavra-chave para um tópico. Esta característica facilita a busca de *tweets*, através da consulta dos *hashtags*.

Os estudos do trabalho revelaram que 10,1% de *tweets* possuem pelo menos um *hashtag* em seu texto, enquanto 20,8% dos *retweets* contêm *hashtags*. Esta observação indica que um *tweet* com *hashtags* é mais provável de sofrer *retweet*. A tabela 8 identifica os *hashtags* mais populares do trabalho.

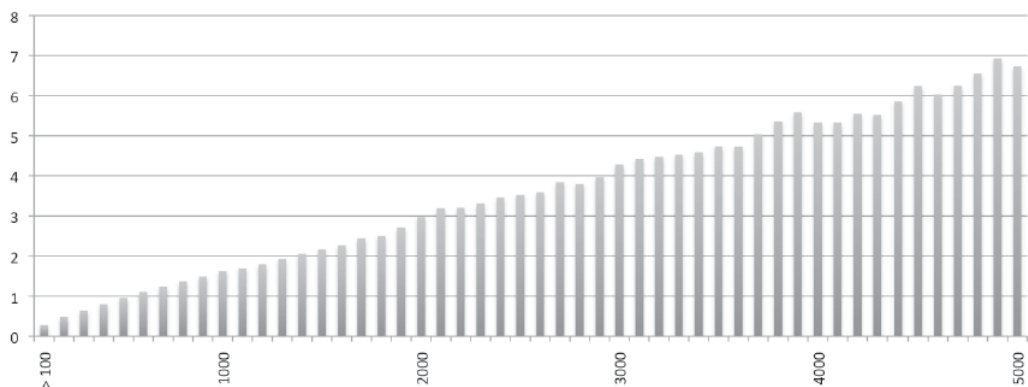
Tabela 5: Hashtags mais populares em Tweets e Retweets

| Rank | Hashtag | In Tweet | In Retweet | Retweet Rate |
|------|------------------|----------|------------|--------------|
| 1 | #nowplaying | 355147 | 29846 | 0.75 |
| 2 | #ff | 224760 | 62331 | 2.49 |
| 3 | #jobs | 124728 | 2173 | 0.16 |
| 4 | #fb | 87959 | 10994 | 1.12 |
| 5 | #tinychat | 67225 | 273 | 0.04 |
| 6 | #vouconfessarque | 51578 | 43628 | 7.59 |
| 7 | #fail | 49248 | 9759 | 1.78 |
| 8 | #tcot | 47394 | 18527 | 3.51 |
| 9 | #1 | 47373 | 9124 | 1.73 |
| 10 | #followfriday | 39986 | 11170 | 2.51 |

Fonte: (SUH et al., 2010)

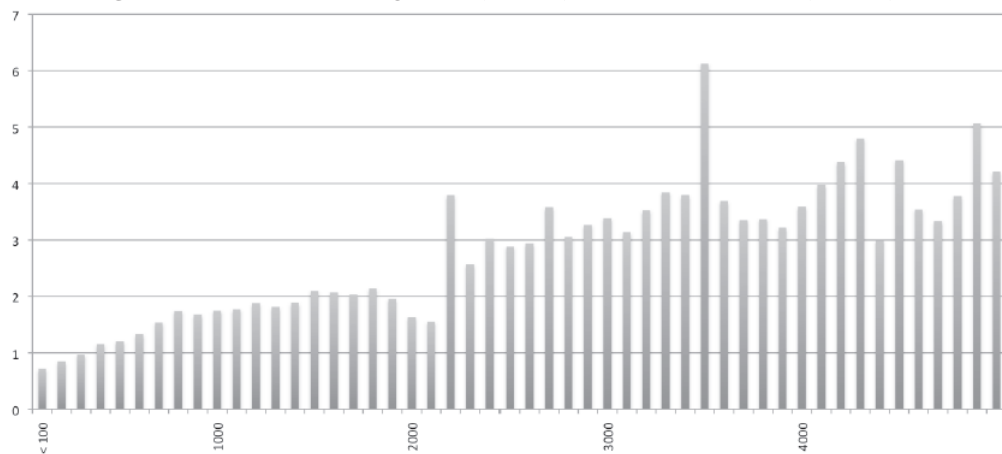
3.2.2.3 Seguidores e Seguidos

Conforme a figura 13, os estudos do trabalho mostram uma relação muito forte entre o número de seguidores e a taxa de retweet. Em outras palavras, é fácil intuir que, quanto maior é a audiência, maior a probabilidade de retweet.

Figura 13: Número de Seguidores (eixo X) e Taxa de Retweet (eixo Y)

Fonte: (SUH et al., 2010)

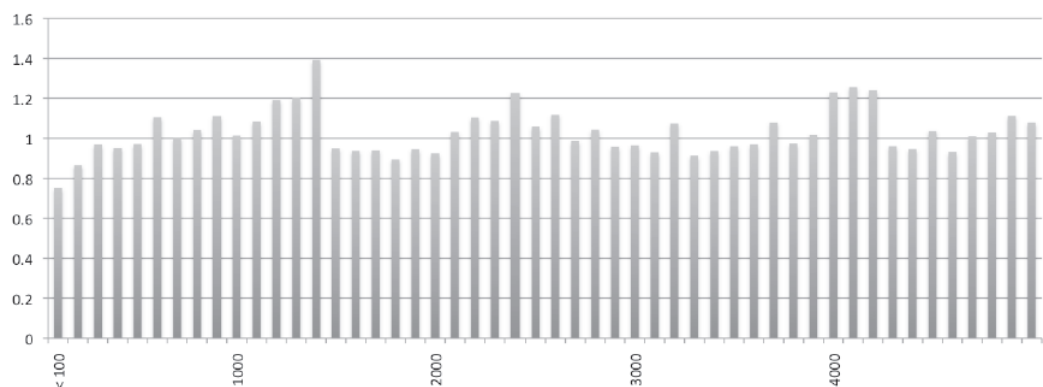
Da mesma forma, quanto maior o número de usuários que são seguidos, maior a probabilidade de retweet. No entanto, pode se verificar na figura 14, que a relação não é tão forte como com os seguidores. Este resultado sugere que quanto mais fontes o usuário segue, mais interessantes os seus *tweets* são, provavelmente devido à diversidade de opinião e informação dos itens consumidos.

Figura 14: Número de Seguidos (eixo X) e Taxa de Retweet (eixo Y)

Fonte: (SUH et al., 2010)

3.2.2.4 Tweets Anteriores

O trabalho também analisou a relação entre *tweets* postados por um usuário do Twitter, desde que a sua conta foi criada, e a taxa de *retweets*. Com esta característica o estudo não foi capaz de encontrar um padrão significativo, as taxas de *retweet* não mostram nenhum padrão óbvio, aparentemente não há uma relação forte conforme mostra a figura 15.

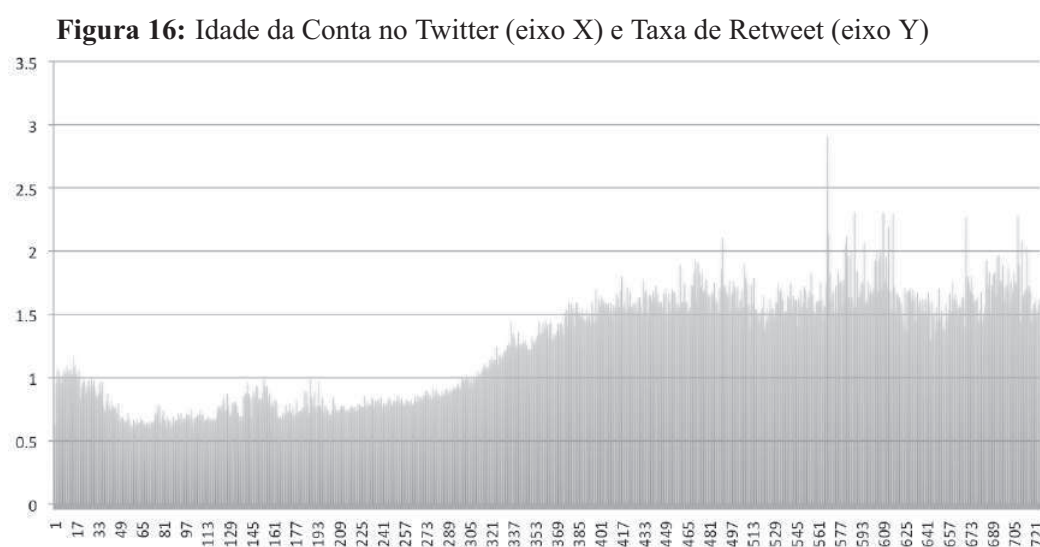
Figura 15: Número de Tweets no Passado (eixo X) e Taxa de Retweet (eixo Y)

Fonte: (SUH et al., 2010)

Além do número total de *tweets*, também foi investigada a relação da taxa de *retweet* com o número médio de *tweets* diários, que é calculado pelo número de *tweets* total dividido pelo número de dias desde que o autor de um *tweet* criou a conta no Twitter. Da mesma forma, o resultado das análises mostraram que não há qualquer relação significativa entre a média número de *tweets* diários e a taxa de *retweet*.

3.2.2.5 Idade da Conta no Twitter

A figura 16 mostra várias taxas de *retweet* dependendo da idade das contas do Twitter. Como mostrado no gráfico, a antiguidade dos usuários do Twitter tem uma relação significativa com a taxa de *retweet*. Os *tweets* feitos por usuários do Twitter, que criaram suas contas há mais de 300 dias atrás, mostra uma taxa de *retweet* superior à média. Enquanto os *tweets* de autoria de usuários novos do Twitter apresentam baixa taxa de *retweet* em geral. É interessante observar o aumento na taxa de *retweet* dos usuários do Twitter que criaram suas contas muito recentemente (menos de 30 dias), resultando numa curva parecida com um U.



Fonte: (SUH et al., 2010)

3.2.2.6 Conclusões do Trabalho

De uma forma global, os resultados do trabalho acima sugerem que a taxa de *retweets* possuem uma relação muito estreita com o contexto de rede social dos autores e a informação de conteúdo e valor contido nos *tweets*.

3.2.3 WhoReTweetedMe

Em agosto de 2011, (ZARRELLA, 2009b) lançou uma página na internet com uma ferramenta, em versão considerada beta, a WhoReTweetedMe (ZARRELLA, 2011). Nesta página, somente através do envio de uma URL de uma postagem recente, entre um dia e duas semanas de existência, se obtém um relatório contendo a quantidade de *retweets* da URL, a quantidade média de seguidores do usuários que enviaram *retweets* da URL, um número sobre o potencial de alcance da URL, um gráfico com a linha do tempo dos *retweets* da URL e uma lista dos 20 usuários mais influentes que enviaram *retweets* da URL. A figura 17 mostra uma visão parcial

de uma página com a referida pesquisa.

Nenhuma documentação foi encontrada relativa a implementação do WhoReTweetedMe. No entanto conseguimos reproduzir o resultado das consultas simplesmente fazendo uma pesquisa com a URL, através da API do Twitter. Para cada *retweet* recebido como resposta, buscamos, do usuário que enviou o *retweet*, a quantidade de seguidores, mantendo em memória as informações dos 20 usuários com maior quantidade de seguidores.

A importância deste trabalho reside no fato em ser o único que encontramos a calcular a relevância de uma mensagem da mesma forma que propomos, qual seja: calculando a quantidade possível de usuários a serem atingidos por uma mensagem do Twitter. Também foi o único local em que há algum detalhamento sobre o tempo de sobrevivência de um *tweet*.

Por outro lado, encontramos uma série de limitações. Ao colocar a consulta em uma URL pública, a limitação de acesso a API do Twitter fica muito forte, pois ocorre concorrência entre todos os usuários que acessam a referida consulta, sendo facilmente atingido os limites impostos pelo Twitter. Outra restrição importante é que, da forma que a tela é apresentada, podemos observar somente vinte usuários. Por fim, a restrição de somente permitir consulta a *tweets* que possuam URL, limita desnecessariamente a ferramenta.

3.2.3.1 Conclusões do Trabalho

Ainda que de forma muito simplificada, e com muitas restrições, o trabalho acima busca a relevância de mensagens no Twitter através da informação sobre o potencial alcance da mensagem.

Figura 17: Exemplo de página do site WhoReTweetedMe

WhoTweetedMe.com

Contagious Content Analytics from 

The Science of Social Media: Aug 23rd @2PM ET
#SMSCI Register Now!

URL:

<http://veja.abril.com.br/noticia/economia/empresas-de-internet-decretam-o-fim-do-curriculo>

Your Email Address: (Optional)

[Analyze This URL](#) [Privacy Policy](#)

WhoReTweetedMe.com will analyze a URL and show you its most influential retweeters, potential reach and timeline. Due to Twitter API limitations, it works best on blog post URLs that are between 1 day and 2 weeks old.

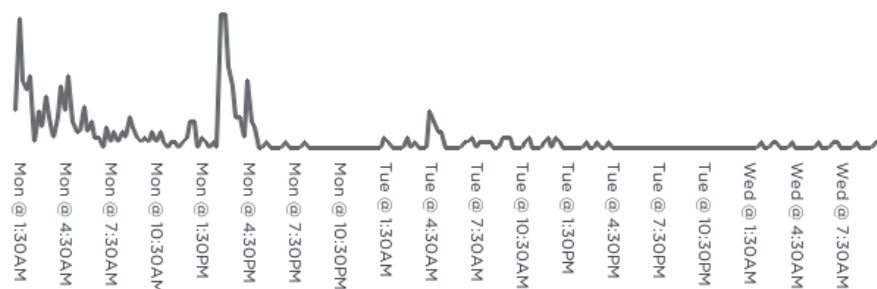
Drag this link to your bookmarks to add a bookmarklet: [Analyze in WhoReTweetedMe](#)

[Tweet This Page!](#)

Quick Stats:

ReTweets: **421** Avg. Followers: **185** Potential Reach: **1,926,183**

ReTweet Timeline:



20 Most Influential ReTweeters:

| User | Tweet | Followers | |
|--|--|-----------|-----------------------|
|  VEJA | Empresas de internet decretam o fim do currículo no Brasil http://t.co/hvEG0kiJ 02/13/12 01:35 | 1,431,563 | Thank |
|  AdrianoJocafe | Empresas de internet decretam o fim do currículo no Brasil: http://t.co/d3AFqN6p 02/13/12 04:51 | 74,929 | Thank |
|  djrobsonmichel | Empresas de internet decretam o fim do currículo no Brasil http://t.co/RfDmOTHU 02/13/12 11:01 | 53,951 | Thank |

Fonte: <http://whoretweetedme.com/>

4 TRABALHO PROPOSTO

O presente capítulo apresenta o modelo proposto para o desenvolvimento de um ambiente para análise da abrangência de mensagens no Twitter. Para tal, desenvolveu-se um ambiente multi-agentes para automatizar a inspeção dos atores e objetos envolvidos na comunicação, possibilitando definir a relevância e espectro de cobertura de determinado *tweet*. O ambiente proposto é composto pelas seguintes capacidades:

- mapear o processo de difusão de mensagens no Twitter
- verificar a abrangência de mensagens
- analisar a velocidade de propagação e esgotamento do assunto
- consolidar e reportar os resultados obtidos

Basicamente procura-se responder as perguntas: Que impacto a mensagem que enviei causou? Uma vez tendo sido enviado um *tweets*, como saber para quem ele foi? Como saber qual a abrangência? Quanto tempo a mensagem "sobreviveu" na rede? Que usuários orbitaram em torno da mensagem?

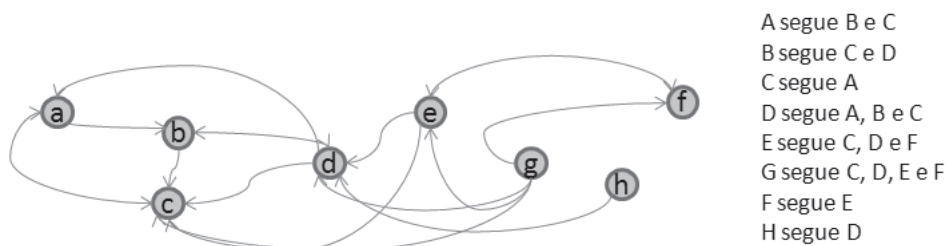
Para tanto, no contexto deste trabalho, assumem-se as seguintes definições:

- *Relevância (ou Abrangência)*: é a quantidade de pessoas que receberam a mensagem do Twitter na sua linha do tempo
- *Espaço Demográfico (ou Escopo)*: são as pessoas, ou grupo de pessoas, que estão envolvidas no envio de uma mensagem, ou grupo de mensagens, do Twitter

4.1 Visão Geral

Inicialmente a mensagem vai para todos os seguidores, portanto deverá ser analisada a lista de pessoas que seguem o autor da mensagem. A figura 18 dá uma ideia deste processo.

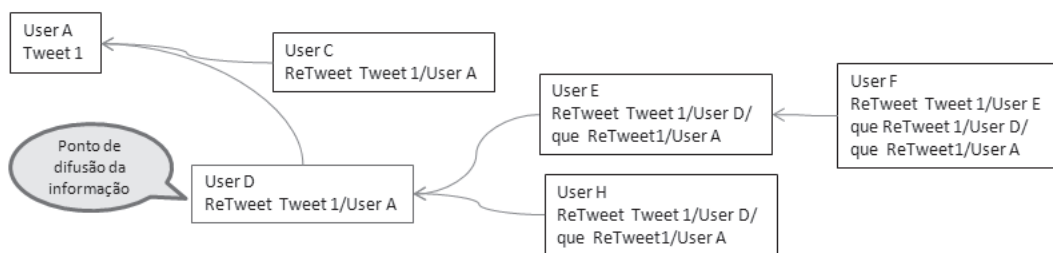
Figura 18: Espaço de Grafo de Seguidos / Seguidores



Fonte: elaborado pelo autor

Conforme a figura 18 uma mensagem postada por A é em princípio somente lida por C e D, portanto esta seria a abrangência da mensagem. Exceto se C ou D enviarem *retweets* desta mensagem. Caso D envie *retweet* desta mensagem, os vértices B, E, G e H também passam a lê-la e a abrangência triplica em números de vértices, pois a abrangência de C, D passa a ser acrescida dos vértices B, E, G e H. Observe na figura 19 o que acontece se os vértices C, D, E, H e F enviarem *retweets* da mensagem.

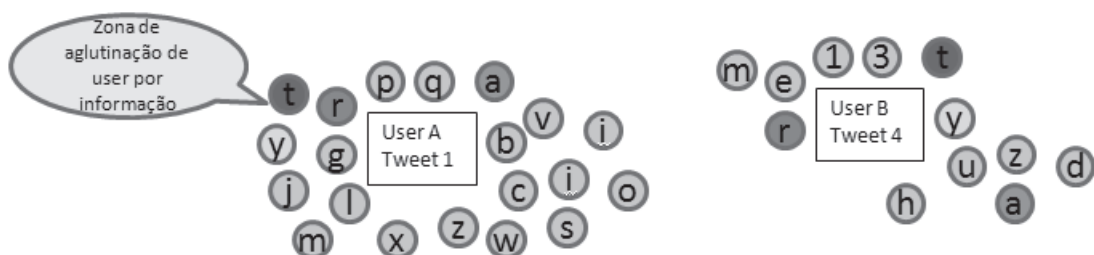
Figura 19: Espaço de Grafo Informação / Retweet



Fonte: elaborado pelo autor

O ambiente proposto deve avaliar estas situações, e outras semelhantes, com base no conhecimento que tem nos processos da rede social e explicitar o espaço demográfico da informação, a abrangência, informando os usuários (vértices) que foram sensibilizados pela informação, a exemplo da figura 20.

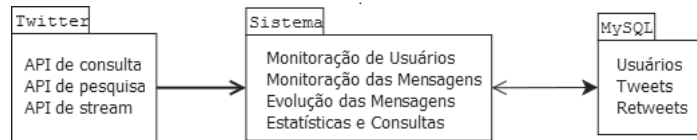
Figura 20: Espaço Demográfico da Informação versus Usuário



Fonte: elaborado pelo autor

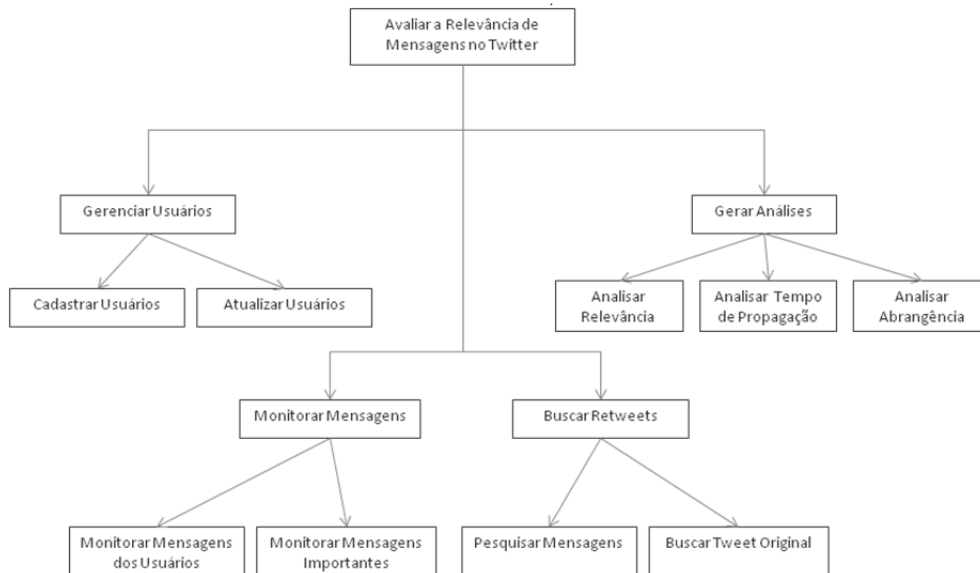
4.2 Arquitetura do Sistema

A solução adotada parte da busca de informações através das APIs do Twitter, por um conjunto de agentes de software com a função de implementar os objetivos traçados, armazenando as informações em banco de dados MySQL. A figura 21 dá uma ideia desta arquitetura.

Figura 21: Arquitetura do Sistema

Fonte: elaborado pelo autor

Para modelar o comportamento dos agentes usamos a metodologia MaSE. Conforme (DELOACH; WOOD, 2001), a primeira fase desta metodologia é a descoberta dos objetivos, baseado na especificação inicial do sistema, que resulta em um conjunto de metas estruturadas de forma hierárquica, conforme mostrado na figura 22.

Figura 22: Objetivos conforme metodologia MaSE

Fonte: elaborado pelo autor

O objetivo principal do sistema é analisar a relevância de mensagens no Twitter. Para tanto desmembrou-se o objetivo principal em quatro responsabilidades distintas, descritas a seguir:

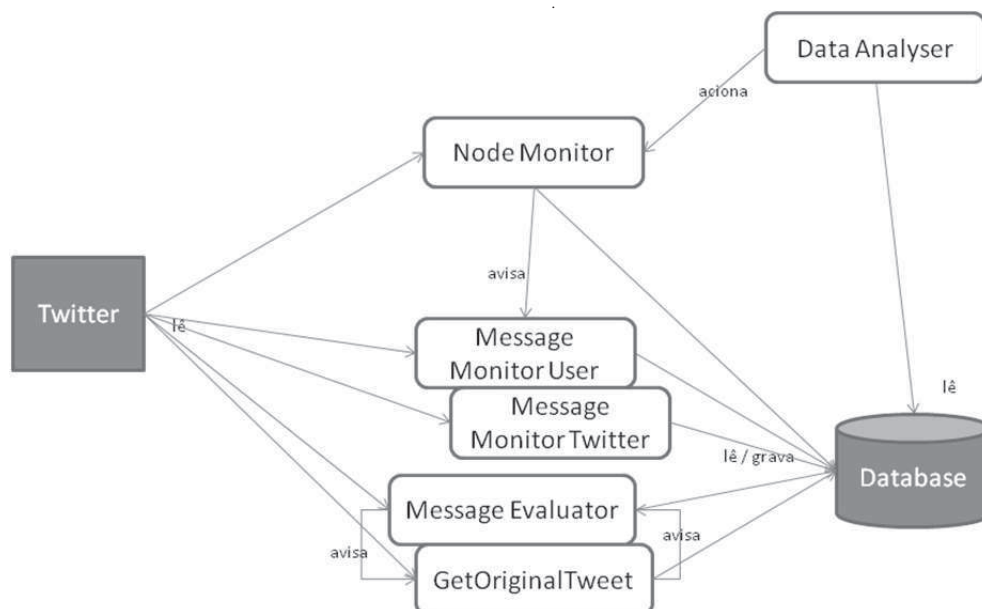
- Gerenciar Usuários: Objetivo responsável por gerenciar a inclusão e atualização das informações dos usuários tais como quantidade de mensagens enviadas, quantidade de seguidores e quantidade de seguites;
- Monitorar Mensagens: Objetivo responsável por monitorar novas mensagens dos usuários cadastrados, bem como monitorar mensagens de todo o Twitter consideradas altamente relevantes, seja porque a mensagem possui alto número de *retweets*, seja porque o assunto é considerado relevante;
- Buscar Retweets: Objetivo responsável por pesquisar os *retweets* de uma mensagem,

armazenando-os para posterior análise. Também é responsável por buscar o *tweet* original quando identificar que a mensagem analisada já é resultado de um retweet;

- Gerar Análises: Objetivo responsável por consolidar e reportar os resultados obtidos, fornecendo as informações sobre o processo de difusão de mensagens, sua relevância, abrangência e velocidade de propagação.

O próximo passo na metodologia MaSE é transformar esta hierarquia de objetivos em papéis (DELOACH; WOOD, 2001). Os papéis definem as classes de agentes e representam os objetivos da ferramenta durante a etapa de projeto. Assim, cada objetivo é garantido por um papel desempenhado por um agente, conforme a figura 23.

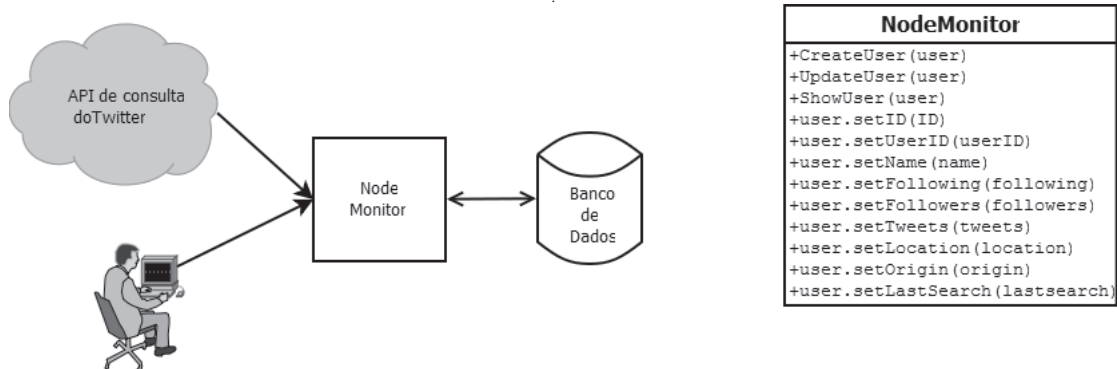
Figura 23: Papéis MaSE



Fonte: elaborado pelo autor

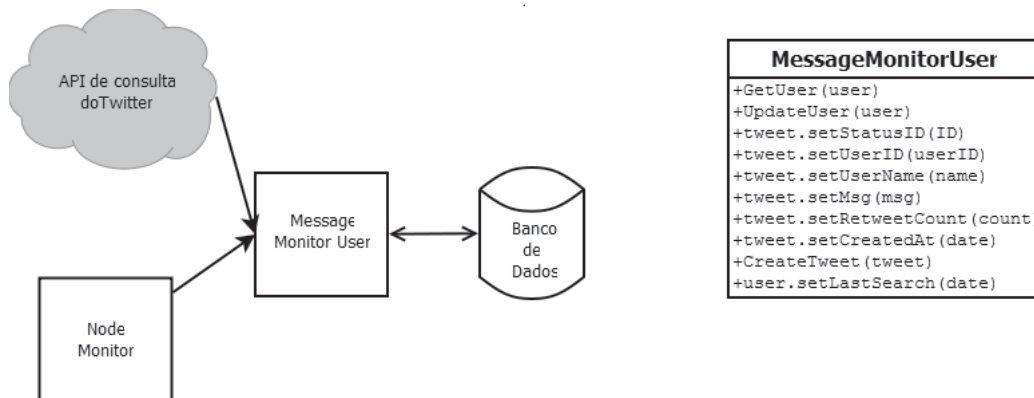
- NodeMonitor: O agente NodeMonitor, conforme figura 24, é o agente que monitora as atividades dos autores das mensagens, ou dizendo de outra forma, os vértices do grafo da rede social. É responsável por acrescentar usuários que participarão da monitoração e manter atualizadas as informações dos mesmos, tais como *tweets* feitos, números de seguidores e de seguidos. É responsável por avisar o agente Message Monitor, via informações no banco de dados, que novos usuários devem ter suas mensagens analisadas.
- MessageMonitorUser: O agente MessageMonitorUser, conforme figura 25, é o agente responsável por verificar novos *tweets* enviados pelos usuários cadastrados. Para isto acessa a linha do tempo do Twitter de todos os autores de mensagens cadastradas e atualizadas pelo NodeMonitor, gravando no banco de dados os novos *tweets* encontrados.

Figura 24: NodeMonitor



Fonte: elaborado pelo autor

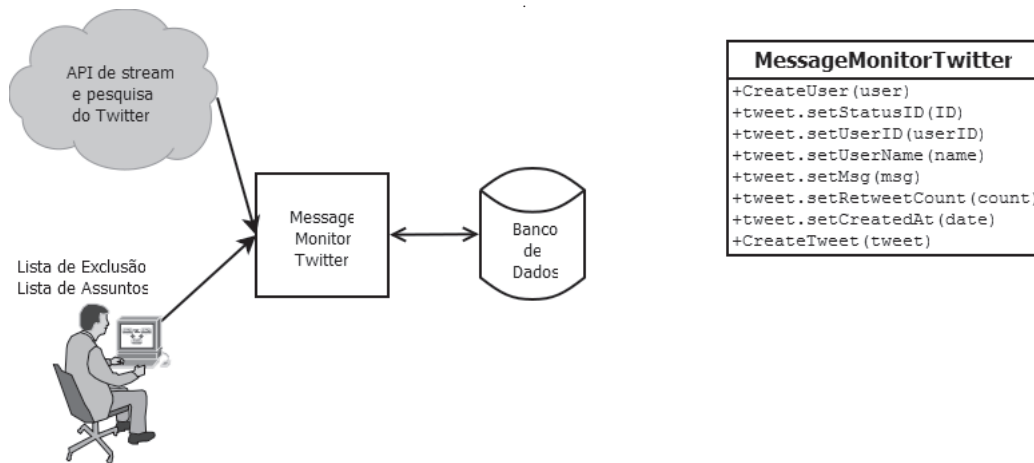
Figura 25: MessageMonitorUser



Fonte: elaborado pelo autor

- **MessageMonitorTwitter:** O agente MessageMonitorTwitter, conforme figura 26, é o agente responsável por receber mensagens aleatórias do Twitter, verificando quais destas mensagens são candidatas a monitorar. Para isto é usada a API de streaming do Twitter, que fornece aproximadamente um por cento da quantidade total de *tweets*. No banco de dados são gravados os *tweets* com assuntos escolhidos ou com mais *retweets*, portanto considerados mais relevantes e conseqüentemente passíveis de análise. Para este agente foi criado uma "lista de exclusão" de termos cujos *tweets* devem ser desprezados, pois algumas poucas personalidades e *tweets* com anúncios ou promoções de vendas representavam aproximadamente oitenta por cento dos *tweets* que seriam analisados.
- **MessageEvaluator:** O agente MessageEvaluator, conforme figura 27, é o agente responsável por monitorar a evolução e relevância de uma mensagem até o seu fenecimento. Para isto periodicamente acessa todos os *tweets* cadastrados para análise, e faz uma pesquisa no Twitter com o conteúdo do *tweet* analisado. Esta pesquisa não é feita analisando a linha do tempo de cada seguidor do usuário do Twitter, pois para alguns usuários a quantidade de seguidores é enorme e os limites permitidos para a consulta a linha do tempo

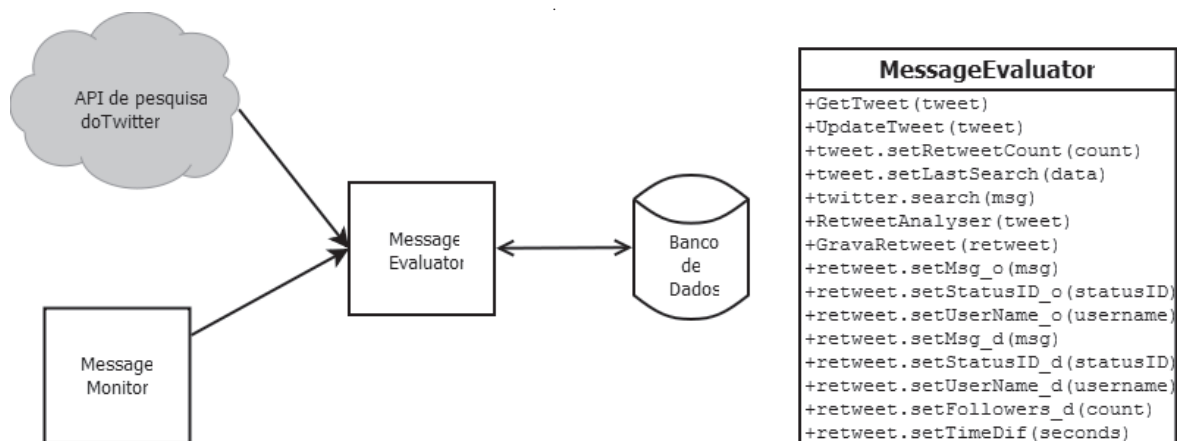
Figura 26: MessageMonitorTwitter



Fonte: elaborado pelo autor

no Twitter é muito restrita. Uma solução alternativa foi encontrada usando o comando *"search"*, da API do Twitter, que devolve as informações de forma mais concisa e que possui um limite de uso maior no Twitter. Para cada *tweet* recebido como resposta na pesquisa, este agente analisa o conteúdo para verificar se é um *retweet* ou não. Os *tweets* aceitos são gravados numa estrutura chamada *retweet* para posterior contabilização pelo agente DataAnalyzer. Os *tweets* rejeitados são gravados em uma estrutura do banco de dados chamada *retweet_garbage* para uma eventual conferência manual.

Figura 27: MessageEvaluator

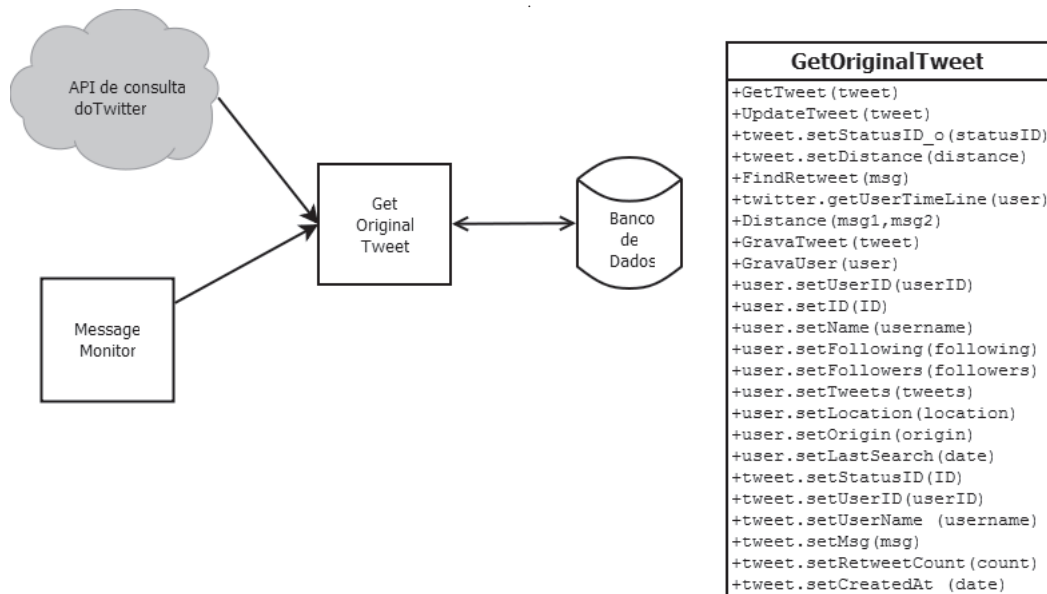


Fonte: elaborado pelo autor

- **GetOriginalTweet:** Quando uma mensagem enviada por um dos usuários monitorados é um *retweet*, a melhor forma de calcular a relevância da mensagem é através do cálculo da relevância do *tweet* original. O agente GetOriginalTweet é o agente responsável por buscar o *tweet* original destes *retweets*. Diferentemente do agente MessageEvaluator, isto é feito analisando a linha do tempo do usuário original que tipicamente consta na men-

sagem do *retweet* depois da expressão "RT @"ou "Via @". Frequentemente a mensagem que teve *retweet* sofre alguma alteração, assim sendo, se uma comparação simples não permitir identificar a mensagem que teve *retweet*, optamos por usar o algoritmo da Distância de Levenshtein ¹ sobre todas as mensagens da linha do tempo, a fim identificar qual a mensagem com menor diferença, sendo esta considerada a *tweet* original.

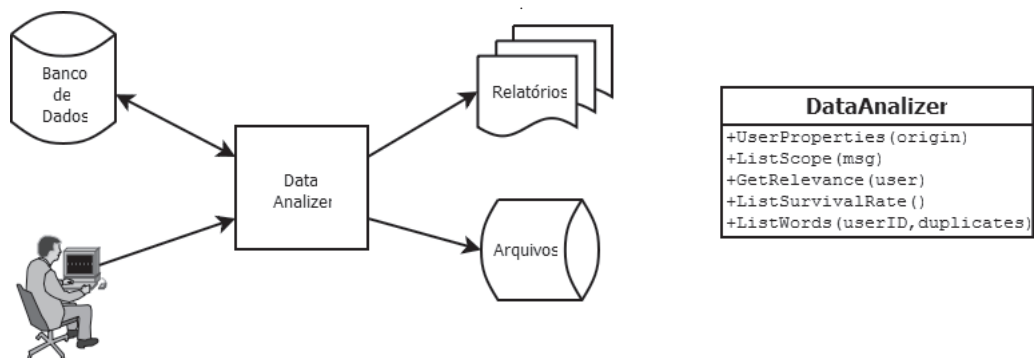
Figura 28: GetOriginalTweet



Fonte: elaborado pelo autor

- **DataAnalyzer:** O agente DataAnalyser, conforme figura 29, é o agente responsável por analisar, consolidar e reportar os dados do banco de dados gerando as informações sobre a difusão de mensagens, sua relevância, abrangência, velocidade de propagação e qualquer outra que se mostrar necessária ou oportuna.

Figura 29: DataAnalyzer

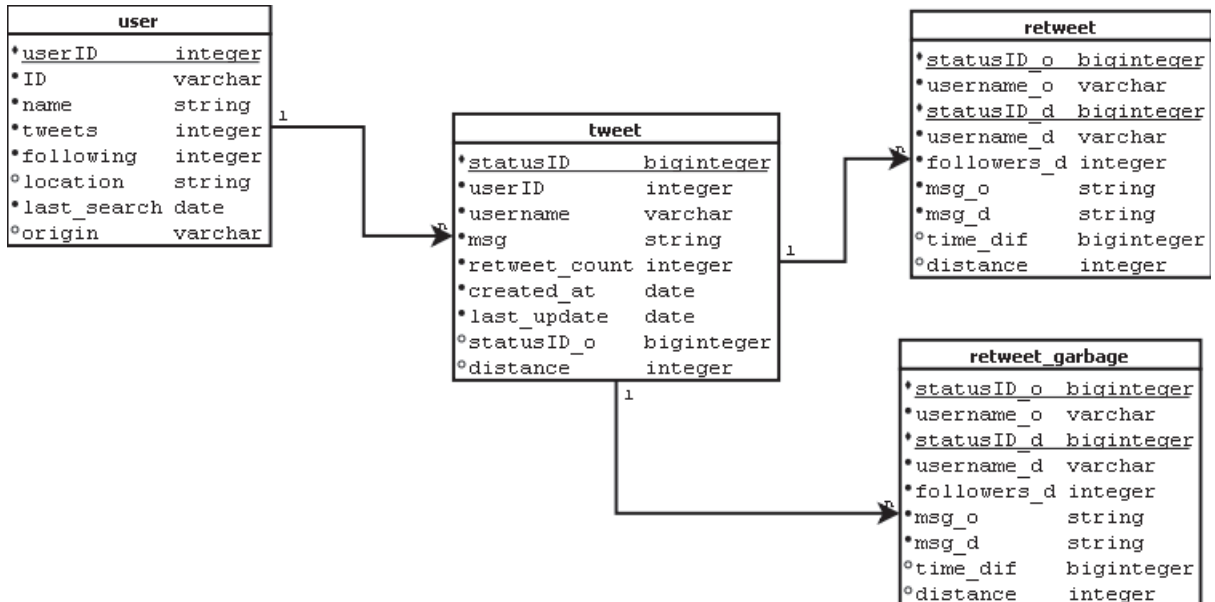


Fonte: elaborado pelo autor

¹http://pt.wikipedia.org/wiki/Distância_Levenshtein

Estes agentes usam a API do Twitter através da biblioteca JAVA Twitter4J e guardam informações em um banco de dados MySQL, cujo Diagrama ER (Entidade Relacionamento) pode ser visto na figura 30.

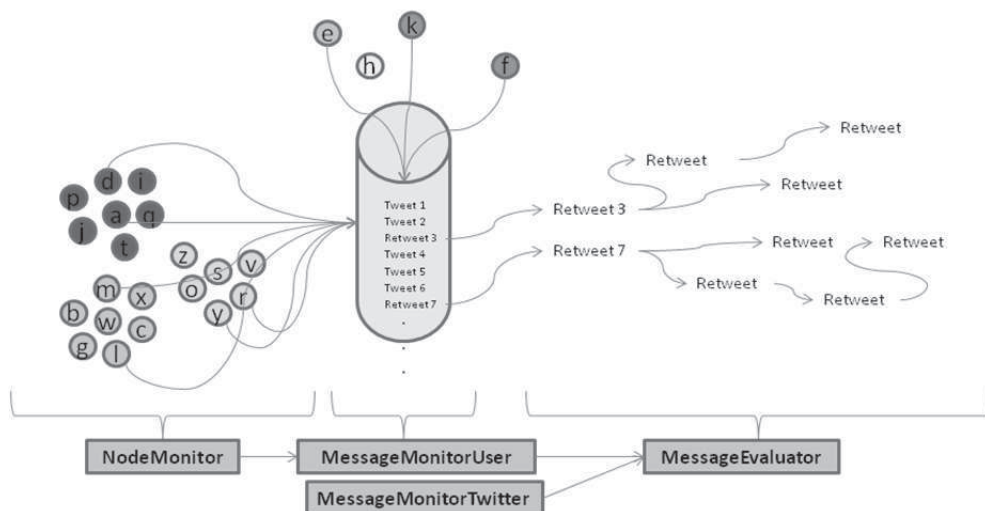
Figura 30: Diagrama de Entidade Relacionamento



Fonte: elaborado pelo autor

O funcionamento colaborativo dos principais agentes do sistema (NodeMonitor, MessageMonitorUser, MessageMonitorTwitter e MessageEvaluator) pode ser visto na figura 31.

Figura 31: Principais Agentes do Sistema



Fonte: elaborado pelo autor

Tudo inicia com o NodeMonitor analisando se houve alguma mudança nos dados dos usuários monitorados e cadastrando algum novo usuário caso solicitado. Em seguida o agente MessageMonitorUser analisa se houve alguma mensagem nova nos usuários monitorados. Esta análise é feita olhando a linha do tempo do usuário. Caso seja encontrado algum *tweet* novo este é

armazenado em banco de dados para posterior processamento pelo agente MessageEvaluator.

Simultaneamente o agente MessageMonitorTwitter monitora o Twitter para verificar a recepção de algum *tweet* que satisfaça os seus critérios de pesquisa. Da mesma forma que o agente MessageMonitorUser, ao encontrar, este novo *tweet* também é armazenado em banco de dados para posterior processamento pelo agente MessageEvaluator.

A partir deste momento, o principal agente entra em execução: o MessageEvaluator. Ele analisa cada um dos *tweets* armazenados no banco de dados, verificando se houveram *retweets*. A ideia inicial deste trabalho era analisar a linha do tempo de cada um dos seguidores do usuário e verificar se o *tweet* havia sofrido *retweet*. Logo percebemos que seria impossível fazer esta tarefa desta forma pelas limitações impostas pela API do Twitter (CENTER, 2011), que para determinados tipos de consultas são limitadas em 350 requisições por hora, e pela quantidade de seguidores de determinados usuários serem na ordem de milhões. Partimos então para outra solução, usando a API de pesquisa do Twitter (DEVELOPERS, 2011), que atende as necessidades exceto pelo fato de mostrar tão somente os *tweets* de 6 a 9 dias atrás. Esta limitação impede que se analise *tweets* mais antigos, cujas informações não tenham sido previamente armazenadas.

Para finalizar, quando identificamos que uma mensagem analisada é um *retweet*, o agente GetOriginalTweet é acionado para identificar o *tweet* original, e quando queremos alguma análise sobre as informações do sistema, acionamos o agente DataAnalyzer.

4.3 Tecnologias e Ferramentas Adotadas

Para o desenvolvimento da ferramenta proposta, foram utilizadas as tecnologias e softwares listados a seguir:

- Java² JDK 1.6: é uma linguagem de programação orientada a objeto, compilada para um bytecode que é executado por uma máquina virtual;
- Eclipse³ SDK 3.5.2: ambiente de desenvolvimento integrado (IDEs) com forte orientação ao desenvolvimento baseado em plug-ins;
- MySQL⁴ 5.5: é um SGBD (Sistema de Gerenciamento de Banco de Dados) livre, muito popular por causa de seu alto desempenho, alta confiabilidade e facilidade de uso, conforme (WWW.MYSQL.COM, 2011). Utiliza a linguagem SQL (Structured Query Language) como interface;
- MySQL Front⁵ 5.1: software livre de administração do banco de dados MySQL;

²<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

³<http://www.eclipse.org/>

⁴<http://www.mysql.com>

⁵<http://www.mysqlfront.de/wp/>

- Cytoscape ⁶ versão 2.8.2: Plataforma de software livre para análise e visualização de redes complexas.

4.4 Análise dos Dados Obtidos

Foram analisados 4343 *tweets* enviados do dia 1/12/2011 até o dia 15/02/2012. Estes *tweets* tiveram por base 10 usuários conforme a tabela 6:

Tabela 6: Algumas Propriedades dos Usuários Objetos do Estudo

| | Total de Tweets | Seguindo | Seguidores | Tweets de 1/12/2011 a 15/02/2012 |
|------------|-----------------|----------|------------|----------------------------------|
| Usuário 1 | 19701 | 92605 | 165874 | 2051 |
| Usuário 2 | 2796 | 53 | 35814 | 225 |
| Usuário 3 | 5672 | 3056 | 626542 | 817 |
| Usuário 4 | 438 | 22 | 96 | 37 |
| Usuário 5 | 11060 | 391 | 30115 | 732 |
| Usuário 6 | 525 | 103 | 120 | 33 |
| Usuário 7 | 3579 | 498 | 282 | 334 |
| Usuário 8 | 83 | 50 | 21 | 2 |
| Usuário 9 | 439 | 38 | 30 | 69 |
| Usuário 10 | 205 | 7 | 80 | 23 |

Fonte: elaborado pelo autor

- Usuário 1: é um *twitter* de um jornal diário;
- Usuário 2: é um *twitter* de um site de tecnologia;
- Usuário 3: é um *twitter* de uma revista mensal;
- Usuário 5: é um *twitter* de um consultor de empresas e conferencista;
- Demais usuários: são *twitters* de pessoas comuns.

Com base nestes *tweets* obtivemos informações sobre: abrangência, relevância, sobrevida, órbita, centralidades, assunto e algumas apreciações adicionais.

4.4.1 Abrangência

O modelo proposto permite a análise de *tweets* e *retweets* com as mais diversas características de abrangência. Uma vez havendo um *retweet* de uma mensagem, podemos saber para quem ela foi, pois para cada mensagem sabemos que usuário enviou o *retweet* através do agente

⁶<http://www.cytoscape.org/>

MessageEvaluator. Por exemplo, a mensagem "**Usuários do Twitter estão ficando mais inteligentes - e mal-humorados <http://t.co/iCOOBCKL>**" sofreu 221 *retweets* e podemos fornecer todos os usuários que o fizeram.

Também identificamos quando uma mensagem analisada é um *retweet* e buscamos o *tweet* original através do agente de software chamado GetOriginalTweet. Por exemplo, quando o usuário 5 fez *retweet* com a mensagem "**RT @carolinagontijo: o Brasil é o país onde tudo é urgente, mas nada é prioridade**", o agente GetOriginalTweet buscou o *tweet* original da usuária @carolinagontijo, disponibilizando o mesmo para análise pelo agente MessageEvaluator. Isto permite que, além dos *retweets* da mensagem com *retweet* do usuário 5, no caso 44 *retweets*, também sejam contabilizados os *retweets* sobre a mensagem original, no caso 4 *retweets*, totalizando 48 *retweets*.

Adicionalmente, além das 4343 mensagens do grupo de 10 usuários anteriormente mencionados, obtidas através do agente MessageMonitorUser, também obtivemos alguns *tweets* adicionais através do agente MessageMonitorTwitter. Este agente é flexível o suficiente para ficar monitorando o Twitter e obter *tweets* por algum critério específico. Nos testes feitos, verificamos em tempo real mensagens com algum assunto específico, monitoramos mensagens com alta taxa de *retweets* em um curto espaço de tempo e combinações destas duas condições. Em um dos testes, monitoramos por 6 horas o Twitter, buscando por *tweets* aonde a API informava que havia mais de 20000 *retweets*. Obtivemos 16 mensagens diferentes. Destas 16 mensagens, analisando visualmente, somente três mensagens possuíam conteúdo "interessante", as mensagens 8, 10 e 16, classificadas pelo número de *retweets* e que seguem:

- Mensagem 8 (57342 *retweets* em 2 dias) - RT @WizKhalllifa: *#myweakness is always forgiving people who don't deserve it, just because I miss having them in my life.*
- Mensagem 10 (36260 *retweets* em 1 dia) - RT @KevinHart2Real: *Age 5 : "#iLove Mommy"Age 15: "#iHate HER!"Age 20: "Mom was right..."Age 57: "#iWish my Mom was still here."R ...*
- Mensagem 16 (21185 *retweets* em 5 dias) - RT @ddlovato: *I don't care what my haters say, cause they're still talking about me ;)*

Em outro teste, monitoramos por um dia mensagens com mais de nove *retweets*, contendo a palavra *education*. Obtivemos 279 mensagens, seguem as seis primeiras, por ordem de número de *retweets*:

- Mensagem 1 (7486 *retweets* em 50 dias) - RT @DalaiLama: *My hope and wish is that one day, formal education will pay attention to what I call "education of the heart."*
- Mensagem 2 (6730 *retweets* em 55 dias) - RT @kanyewest: *Help education. School systems were designed to turn people into factory workers.*

- Mensagem 3 (4230 *retweets* em 3 dias) - *#SomeWhereInTheGhetto SOMEONE on welfare with NO job, education, future, food, grape drink, just woke up and tweeted "Rise and grind"*
- Mensagem 4 (4081 *retweets* em 40 dias) - *RT @TheIlluminati: Education is not preparation for life; education is life itself.*
- Mensagem 5 (3938 *retweets* em 62 dias) - *Put a girl on the path to education & she will lead the way for the next generation. Read these stories to be inspired! <http://t.co/v9B9yN1B>*
- Mensagem 6 (3635 *retweets* em 5 dias) - *#doNOTbother with boys who claim to have "swag". Look for a man with ambition and an education. In ten years, "swag" won't pay the bills.*

Nesta segunda relação, observa-se uma nítida melhora na qualidade do conteúdo dos *retweets*.

4.4.2 Relevância

Podemos saber que impacto a mensagem causou, pois para cada mensagem sabemos quantas vezes ela teve *retweet*. Além disto, para cada mensagem que teve *retweet*, temos o usuário que fez o *retweet* e temos a quantidade de seguidores que cada usuário possui. Com isto, podemos medir quantos usuários podem ter lido a mensagem e consequentemente mensurar o seu impacto. Por exemplo, a mensagem, "**#Frase da semana: "A vida é o que acontece enquanto você está ocupado fazendo outros planos"**<http://t.co/xKVMSx04>" foi a segunda mensagem com mais *retweets* entre as 4343, com 260 *retweets*. Estes 260 *retweets* foram feitos por usuários cuja soma de seguidores é 62420. Levando em consideração os 626542 seguidores do usuário responsável pela envio do *tweet* original temos um total de 688962 pessoas que podem ter lido a mensagem. A este escore total é que atribuímos neste trabalho o conceito de relevância. Abaixo segue a figura 7, com 12 mensagens com mais de 200 *retweets*, identificados pelo sistema proposto, e seus respectivos escores.

- Mensagem 1 - *Homophobia: the fear that gay men will treat you the way you treat women;*
- Mensagem 2 - *#Frase da semana: "A vida é o que acontece enquanto você está ocupado fazendo outros planos"*<http://t.co/xKVMSx04>;
- Mensagem 3 - *Não sou burro, sou um depósito de informações inúteis;*
- Mensagem 4 - *Entenda o que é SOPA e por que a internet está protestando contra isso* <http://t.co/JrvFlaYy>;

Tabela 7: Escores das mensagens com mais retweet

| Mensagem | Seguidores do Usuário da Mensagem | Retweet | Soma de Seguidores dos Retweet | Escore Total |
|----------|-----------------------------------|---------|--------------------------------|--------------|
| 1 | 1660 | 1243 | 515246 | 516906 |
| 2 | 626542 | 260 | 62420 | 688962 |
| 3 | 92723 | 239 | 53981 | 146704 |
| 4 | 626542 | 230 | 275378 | 901920 |
| 5 | 1396574 | 229 | 146530 | 1543104 |
| 6 | 1319794 | 226 | 226103 | 1545897 |
| 7 | 626542 | 221 | 55227 | 681769 |
| 8 | 626542 | 221 | 97806 | 724348 |
| 9 | 626542 | 219 | 71716 | 698258 |
| 10 | 626542 | 216 | 73534 | 700076 |
| 11 | 626542 | 208 | 86977 | 713519 |
| 12 | 626542 | 204 | 42259 | 668801 |

Fonte: elaborado pelo autor

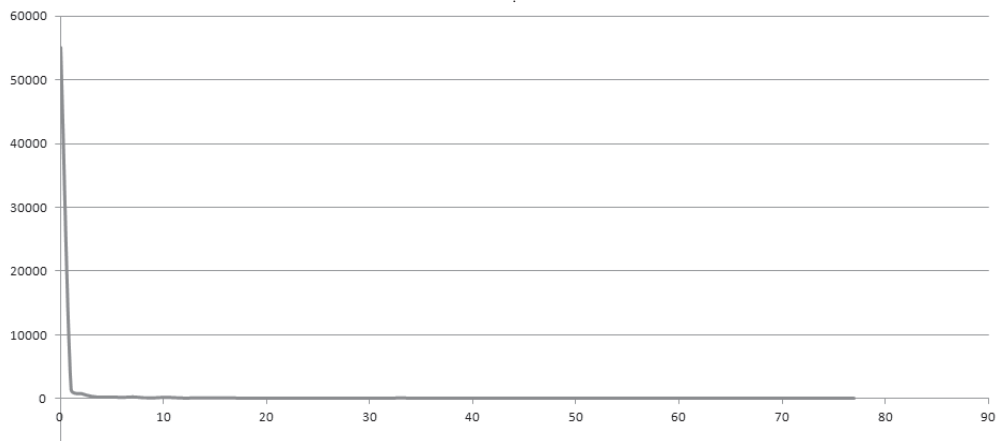
- Mensagem 5 - *Breaking: The U.S. economy added 120,000 jobs in November. The unemployment rate fell to 8.6%: <http://t.co/aSwhokfY>;*
- Mensagem 6 - *Cientistas criam composto capaz de matar o vírus da aids <http://t.co/KNsHlmU4>;*
- Mensagem 7 - *Casais que bebem juntos têm menos problemas <http://t.co/Oo2EY1fv>;*
- Mensagem 8 - *Usuários do Twitter estão ficando mais inteligentes - e mal-humorados <http://t.co/iCOOBckL>;*
- Mensagem 9 - *EUA vão arquivar todas as mensagens publicadas no Twitter (inclusive as suas) <http://t.co/lxsOVqoP>;*
- Mensagem 10 - *Pessoas que bebem ganham mais <http://t.co/MqAN1FnE>;*
- Mensagem 11 - *Brasil tem o iPad mais caro do mundo <http://t.co/h2sIktwG>;*
- Mensagem 12 - *#Comunicado: circula um e-mail falso "Parabéns, você ganhou um iPad2"em nome da Ed. Abril. Não respondam nem passem qualquer dado pessoal.*

4.4.3 Sobrevida

Também podemos identificar quanto tempo uma mensagem fica circulando na *Web*, quanto tempo ela "sobrevive", levando em conta o primeiro *tweet* e o último *retweet* que o sistema identificar. Pegamos como exemplo o *tweet* "**A repetição deixa sua marca até nas pedras.**" **Provérbio Árabe**, de 02/12/2011, que teve até o momento da redação deste trabalho 110 *retweets*, tendo o primeiro ocorrido 29 segundos após o *tweet* original e o último *retweet* tendo ocorrido 25 dias após. Particularmente com relação a este *tweet*, 99 *retweets* ocorreram

dentro das primeiras 24 horas, 6 *retweets* ocorreram entre 24 e 48 horas, um ocorreu dentro de 5 dias, outro dentro de 8 dias, outros dois ocorreram no vigésimo terceiro dia e o último no vigésimo quinto dia. Analisando 60.241 *retweets* do sistema observamos que 55.044 *retweets* ocorreram dentro das primeiras 24 horas, o que corresponde a 91,37% do total. A figura 32 mostra esta evolução.

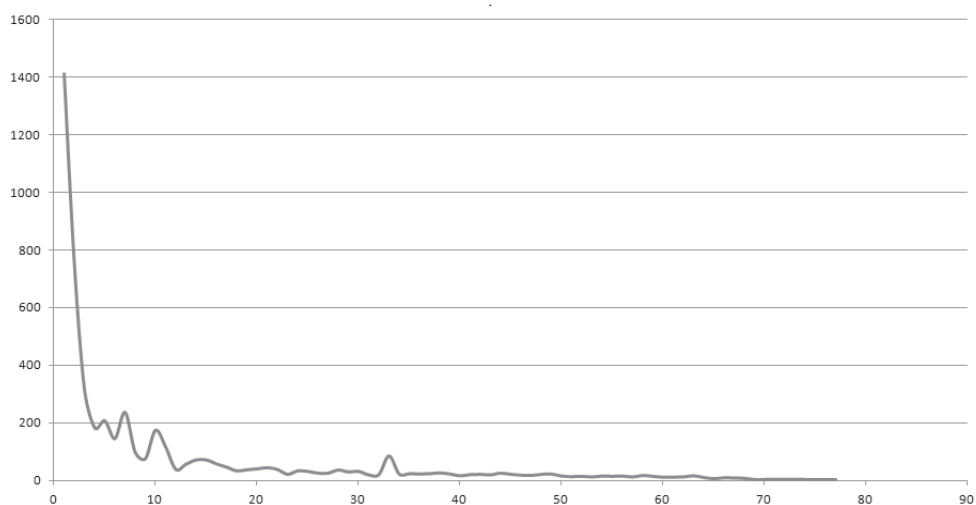
Figura 32: Quantidades de retweets por dia



Fonte: elaborado pelo autor

No gráfico da figura 33 eliminamos as informações do primeiro dia, para podermos ter uma ideia melhor do que acontece nos dias subsequentes. Com isto observamos que nos dez dias seguintes, ainda existe uma quantidade de *retweets* bem elevada. Depois disto ainda ocorre *retweets* com alguma frequência até próximo do trigésimo quinto dia.

Figura 33: Quantidades de retweets após o primeiro dia

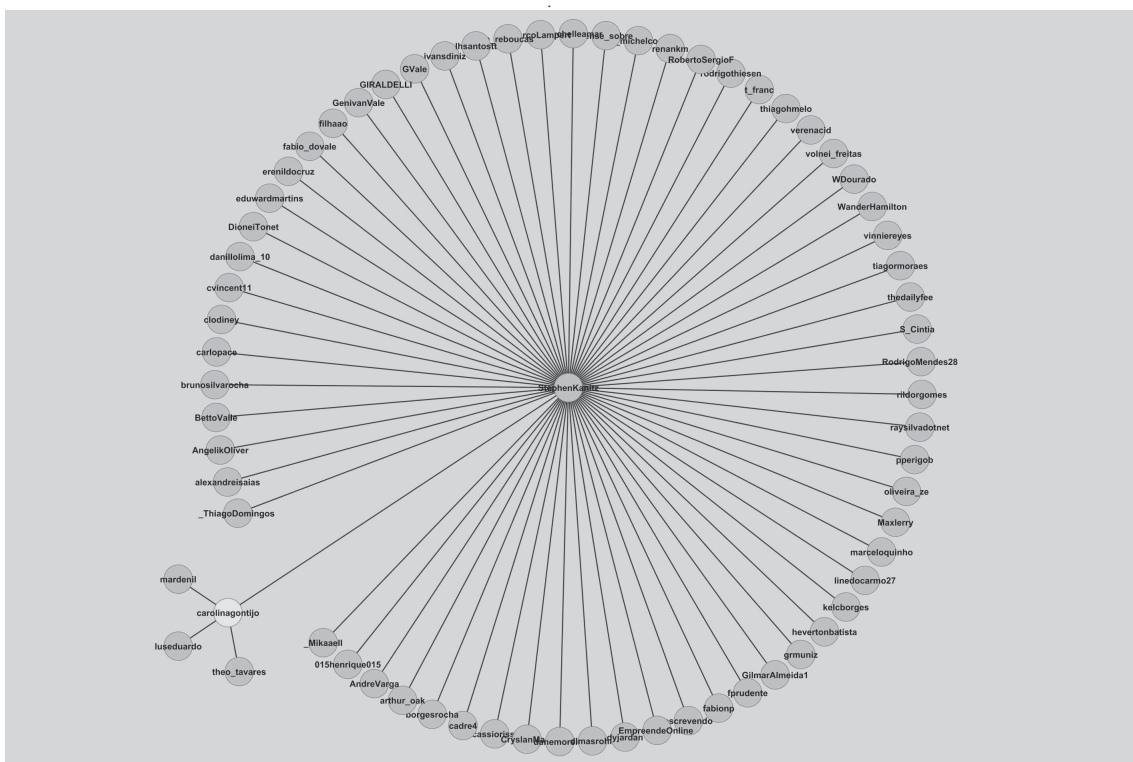


Fonte: elaborado pelo autor

4.4.4 Órbita

Podemos saber que usuários orbitaram em torno de uma mensagem. A figura 34, gerada pelo software Cytoscape, a partir dos *retweet* ocorridos de uma mensagem selecionada, nos fornece uma percepção visual do que aconteceu com a mensagem e quem orbitou ao seu redor. A mensagem, ou *tweet*, original foi criada pelo nó mais claro. Este *tweet* original sofreu quatro *retweets*, mas um destes *retweets*, feito por um usuário com muitos seguidores, gerou mais de 60 *retweets*.

Figura 34: Usuários orbitando em torno de uma mensagem



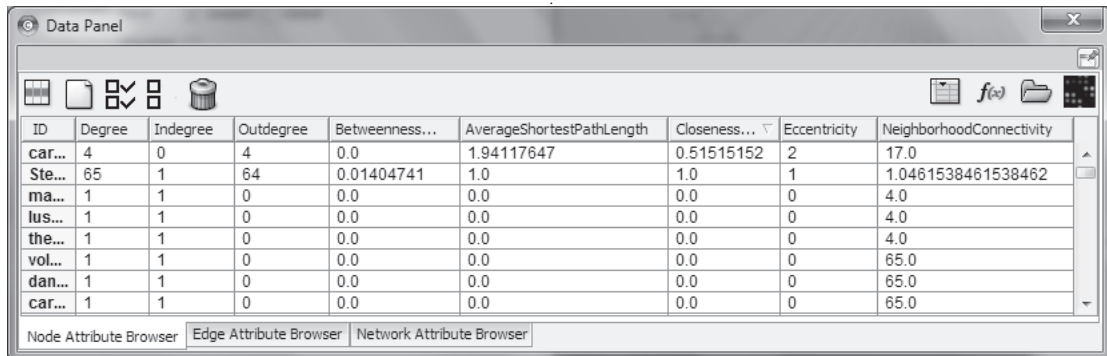
Fonte: elaborado pelo autor

4.4.5 Centralidades

Podemos também calcular as medidas de centralidade da rede gerada pelos usuários que orbitaram em torno de uma mensagem. A figura 35, também gerada pelo software Cytoscape, nos mostra as principais medidas de centralidade da rede da figura 34.

- **Grau (*Degree*):** O número de outros atores (vértices) na rede que um determinado ator está conectado, ou, de forma equivalente, o número de nós adjacentes a ele;
- **Grau de entrada (*Indegree*):** O número de outros atores (vértices) na rede que um determinado ator segue (recebe informações);

Figura 35: Medidas de Centralidade



| ID | Degree | Indegree | Outdegree | Betweenness... | AverageShortestPathLength | Closeness... | Eccentricity | NeighborhoodConnectivity |
|--------|--------|----------|-----------|----------------|---------------------------|--------------|--------------|--------------------------|
| car... | 4 | 0 | 4 | 0.0 | 1.94117647 | 0.51515152 | 2 | 17.0 |
| Ste... | 65 | 1 | 64 | 0.01404741 | 1.0 | 1.0 | 1 | 1.0461538461538462 |
| ma... | 1 | 1 | 0 | 0.0 | 0.0 | 0.0 | 0 | 4.0 |
| lus... | 1 | 1 | 0 | 0.0 | 0.0 | 0.0 | 0 | 4.0 |
| the... | 1 | 1 | 0 | 0.0 | 0.0 | 0.0 | 0 | 4.0 |
| vol... | 1 | 1 | 0 | 0.0 | 0.0 | 0.0 | 0 | 65.0 |
| dan... | 1 | 1 | 0 | 0.0 | 0.0 | 0.0 | 0 | 65.0 |
| car... | 1 | 1 | 0 | 0.0 | 0.0 | 0.0 | 0 | 65.0 |

Fonte: elaborado pelo autor

- **Grau de Saída (*Outdegree*):** O número de outros atores na rede que seguem um determinado ator (envio de informações);
- **Intermediação (*Betweenness*):** Indica quantas vezes um vértice cai no caminho mais curto entre dois outros vértices (FREEMAN, 1979);
- **Média dos caminhos mais curtos (*Average Shortest Path Length*):** É a média do tamanho de todos os caminhos mais curtos de um vértice (WATTS; STROGATZ, 1998);
- **Proximidade (*Closeness*):** Medida para indicar a distância total de um vértice para os outros vértices na rede (FREEMAN, 1979);
- **Excentricidade (*Eccentricity*):** É a maior distância geodésica de um vértice em relação a todos os outros vértices (NEUSTDADTL, 2006);
- **Conectividade da vizinhança (*Neighborhood Connectivity*):** A conectividade da vizinhança de um autor (vértice) é definida como a conectividade média de todos os vizinhos do autor (MASLOV; SNEPPEN, 2002).

4.4.7 Apreciações Adicionais

Para fins de comparação com outros trabalhos acrescentamos algumas apreciações adicionais obtidas nas análises feitas.

Verificando as URLs dos *tweets* e *retweets* do banco de dados, obtivemos que 74,5% dos *tweets* e 84,2% dos *retweets* tem pelo menos uma URL no seu texto. Estes resultados estão muito acima dos estudos de (ZARRELLA, 2009a) e do trabalho da seção 3.2.2.1 e deve-se ao fato de que os principais usuários que geraram os estudos deste trabalho postam quase todas as suas mensagens com URLs.

O que também chamou a atenção, na análise das URLs, é que 98,9% e 99,4% dos *tweets* e *retweets* com URLs, respectivamente, usam o *link* <http://t.co>. O domínio <http://t.co> é um serviço de encurtamento do Twitter, também usado para proteger os usuários contra sites mal-intencionados que se dedicam a espalhar *malware*, ataques de *phishing* e outras atividades nocivas (CENTER, 2012a).

Com relação ao uso de *hashtags*, os estudos deste trabalho revelaram que 7,9% de *tweets* possuem pelo menos um *hashtag* em seu texto, enquanto 14,8% dos *retweets* contém *hashtags*. Estes valores são condizentes com outros estudos em particular com o trabalho da seção 3.2.2.1. A tabela 8 mostra os *hashtags* mais populares obtidos neste estudo.

Tabela 8: Hashtags mais populares nos dados obtidos

| Hashtag | Em Tweets | Em Retweets | Taxa de Retweets |
|---------------------|-----------|-------------|------------------|
| #CiênciaMaluca | 1844 | 29 | 0,42 |
| #Superlink | 959 | 29 | 0,81 |
| #Superdestaques2011 | 886 | 30 | 0,91 |
| #teste | 668 | 20 | 0,80 |
| #Frase | 405 | 2 | 0,13 |
| #Supermanual | 393 | 12 | 0,82 |
| #natal | 291 | 4 | 0,37 |
| #Cultura | 283 | 5 | 0,47 |
| #EnxergueSUPER | 277 | 2 | 0,19 |
| #ADM | 270 | 11 | 1,09 |
| #islam | 236 | 2 | 0,23 |
| #Apple | 224 | 4 | 0,48 |
| #pelasruas | 215 | 30 | 3,75 |
| #Comunicado | 204 | 1 | 0,13 |
| #saude | 145 | 1 | 0,19 |
| #SOPA | 144 | 1 | 0,19 |
| #frasedasemana | 130 | 2 | 0,41 |
| #DiaDoLivro | 118 | 1 | 0,23 |
| #Desafio | 105 | 1 | 0,26 |

Fonte: elaborado pelo autor

5 ANÁLISE COMPARATIVA ENTRE OS TRABALHOS

Pela modernidade da presente proposta, considerando que o Twitter possui pouco mais de cinco anos de existência, os trabalhos encontrados e avaliados, com exceção do WhoReTweetedMe, da seção 3.2.3, não tratavam exatamente das mesmas questões aqui abordadas. Ainda assim, por tratarem de questões afins ao tema em questão, foram aqui elencados.

O trabalho apresentado na seção 3.1.1, *Mapping Communities In Large Virtual Social Networks: Using Twitter Data To Find The Indie Mac Community*, isola sub-comunidades significativas, que partilham interesses comuns dentro de uma rede maior, através de um rápido algoritmo guloso, tendo como critério de análise nuvens de palavras e comparação das centralidades. A base do estudo são alguns poucos usuários iniciais sobre os quais todo o estudo se desenvolve.

O sistema aqui proposto parte de uma perspectiva diferente, o foco é na mensagem e a partir das mensagens é que se geram as redes, através do estudo dos *retweets*. Apesar disto, a análise de nuvens de palavras e medidas de centralidades também podem ser obtidas conforme visto na seção 4.4.6, figuras 36 e 37, e seção 4.4.4, figura 35, respectivamente.

O trabalho apresentado na seção 3.2.1, *Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter*, descreve a prática de enviar *retweets*. Os dados do estudo são obtidos via a API do Twitter e através de uma pesquisa qualitativa com poucas centenas de respostas. O estudo é importante porque descreve as diversas variações na prática de enviar *retweets* de mensagens no Twitter e as maneiras pelas quais estilos variados procuram manter a autoria e fidelidade na conversação, no formato restrito de 140 caracteres. Explica porque as pessoas enviam *retweets*, o que as pessoas enviam neles e para quem os enviam.

O trabalho apresentado na seção 3.2.2, *Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network* examina, em *tweets* e *retweets*, as ocorrências de URLs e *hashtags*, a influência do número de seguidores, de pessoas seguidas (amigos) e da idade da conta no Twitter. Os resultados sugerem que a taxa de *retweets* possuem uma relação muito estreita com o contexto de rede social dos autores e a informação de conteúdo e valor contido nos *tweets*.

O trabalho apresentado na seção 3.2.3, WhoReTweetedMe, é o que mais se aproxima do objetivo principal desta proposta: a análise da relevância de *tweets*. Em uma página da internet, ao se digitar uma mensagem, que obrigatoriamente deve ser um *link* de um site, como resposta obtemos a quantidade de *retweets*, a quantidade média de seguidores do usuário que enviaram *retweets* da mensagem, uma informação de potencial alcance da mensagem, um gráfico com a linha do tempo dos *retweets* e uma lista dos 20 usuários mais influentes que enviaram *retweets* do *link*. Este trabalho possui sérias limitações, vejamos:

- restrição de ter uma URL - a consulta no site só trabalha com *tweets* que contenham *links* de sites na internet, abandonando toda e qualquer pesquisa e consulta em um *tweet* que não possua URL;

- restrição de tempo - a consulta no site só funciona para *tweets* que ocorreram no período de um dia até duas semanas da consulta;
- restrição quanto ao número de consultas - muitas vezes, durante as consultas no site, recebemos mensagem de erro avisando que o limite de consultas previstos por aplicação/usuário havia sido atingido;
- restrição de quantidade de usuários - a consulta no site mostra somente os vinte usuários mais relevantes.

Apesar das diferentes abordagens, a tabela 9 sistematiza de alguma forma os trabalhos relacionados, resumindo e comparando os estudos feitos.

Tabela 9: Tabela comparativa entre os principais trabalhos

| | trabalho 1 | trabalho 2 | trabalho 3 | trabalho 4 | trabalho proposto |
|-------------------------|------------|------------|------------|------------|-------------------|
| foco no(a) | usuário | mensagem | mensagem | mensagem | mensagem |
| analisa a abrangência | sim | não | não | parc. | sim |
| analisa a relevância | usuário | não | não | mensagem | mensagem |
| analisa a sobrevida | n.a. | não | não | sim | sim |
| analisa a órbita | sim | não | não | não | sim |
| analisa a centralidade | sim | não | não | não | sim |
| analisa assuntos | sim | sim | sim | não | sim |
| analisa URLs | n.a. | parc. | sim | sim | sim |
| analisa <i>hashtags</i> | n.a. | não | sim | não | sim |
| analisa seguidores | sim | não | sim | sim | sim |
| analisa seguidos | sim | não | sim | não | sim |

Fonte: elaborado pelo autor

- trabalho 1 - Mapping Communities In Large Virtual Social Networks: Using Twitter Data To Find The Indie Mac Community;
- trabalho 2 - Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter;
- trabalho 3 - Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network;
- trabalho 4 - WhoReTweetedMe;
- trabalho proposto - Análise da Relevância de Mensagens no Twitter através de um Sistema Multi-Agente.

6 CONCLUSÕES

Redes sociais são estruturas que exercem papel fundamental na vida social moderna. O microblog Twitter é uma destas redes sociais atuais em amplo desenvolvimento e crescimento. A modelagem, interpretação e conhecimento da dinâmica da troca de informações nesta rede permite que exploremos e analisemos as rápidas modificações que ocorrem na sociedade atual.

Neste sentido, a maior contribuição deste trabalho está em desenvolver um sistema que nos permite analisar a relevância de mensagens no Twitter, de uma forma não encontrada nos trabalhos avaliados. Adicionalmente objetivos importantes também foram atingidos. Vejamos:

- **Abrangência:** na seção 4.4.1 identificamos como abranger todos usuários envolvidos no envio de um *tweet* e de um *retweet*, como abranger assuntos e como abranger mensagens com rápida evolução;
- **Relevância:** na seção 4.4.2 calculamos o escore possível de pessoas a serem sensibilizadas pela mensagem;
- **Sobrevida:** na seção 4.4.3 mostramos qual a expectativa de tempo que uma mensagem circula na rede;
- **Órbita:** na seção 4.4.4 explicitamos visualmente o espaço demográfico, a rede, dos usuários que orbitam em torno de uma mensagem;
- **Centralidades:** na seção 4.4.5 calculamos os indicadores de centralidade dos principais nós (usuários) da rede;
- **Assuntos:** na seção 4.4.6 mostramos uma possibilidade de analisar os principais assuntos abordados por um usuário;
- **Apreciações Adicionais:** na seção 4.4.7 analisamos os percentuais de uso de URLs e hashtags em *tweets* e *retweets*.

Finalizando, com os objetivos acima expostos atingidos, podemos concluir que verificamos e validamos o modelo proposto em uma rede social real, no caso o Twitter .

6.1 Limitações da Análise

Embora a análise acima ajude na obtenção de informações que não está prontamente disponível fora do mundo do Twitter, há limitações a esta inteligência. São elas:

- **Representatividade:** A maior parte das análises fornecidas começaram a partir de um conjunto de usuários iniciais que enviaram *tweets* ou de algumas mensagens selecionadas segundo critérios específicos definidos. Estes usuários, critérios e mensagens muito provavelmente não são representativos de todo o Twitter.

- **A Qualidade dos Dados:** Algumas partes do Twitter possuem dados falsos. Biografias de usuários podem ser falsificadas. Os nomes podem ser deliberadamente alterados. Robôs de *spam* podem criar *tweets* em tópicos populares. Tudo isso corrompe os dados (MARASANAPALLE et al., 2010).
- **Automação:** Muitas vezes, ao se enviar um *retweet* de uma mensagem, alguma informação é alterada. A rotina de cálculo da distância de Levenshtein permite, com uma boa precisão, comparar quão parecida uma mensagem é com a outra. Dependendo da precisão que for adotada, podemos errar assumindo que uma mensagem é um envio de *retweet* de outra, quando não é, ou o contrário podemos errar assumindo que uma mensagem não é um envio de *retweet* de outra, quando na verdade é;
- **Sobreposição:** Usuários diferentes que enviaram um *retweet* da mesma mensagem podem possuir seguidores em comum. Estes usuários em comum, causam uma sobreposição e conseqüentemente aumento indevido no cálculo da relevância;
- **Limites do Twitter:** Com a intenção de aliviar um pouco a pressão sobre a infraestrutura do Twitter, reduzir o tempo ocioso, reduzir páginas de erro e aumentar a confiabilidade, a equipe de desenvolvimento do Twitter coloca inúmeros limites nas transações feitas (CENTER, 2012b). Estes limites impedem que muitas vezes seja usada a melhor consulta existente em detrimento de outra não tão precisa, mas com limite mais flexível na API do Twitter.

6.2 Trabalhos Futuros

O trabalho aqui proposto cumpriu as metas que foram traçadas em seu planejamento. Entretanto, no decorrer de seu desenvolvimento e de sua validação, percebeu-se que alguns novos recursos poderiam ser implementados, dentre eles, merecem destaque:

- Gravar em banco de dados todos os retornos de consulta feitas ao Twitter, com a ideia de reaproveitar as informações, evitando consultas duplicadas e minimizando as restrições às consultas da API do Twitter;
- Utilizar ontologias para descrever os conceitos, relações, tarefas e comportamentos;
- Utilizar técnicas de mineração de dados sobre as informações obtidas, a fim de se descobrir tendências e mudanças que possam ser utilizadas em futuros projetos;
- Melhorar o algoritmo para identificação de um *retweet* por meio da aplicação de técnicas específicas para o reconhecimento de expressões e processamento de linguagem natural;
- Criar novas categorias de agentes, voltados a monitorar outros aspectos do Twitter não previstos, como por exemplo a remoção de *tweets*;

- Analisar as diferenças e peculiaridades existentes no Twitter para usuários brasileiros, comparando-os com o resto do mundo.

Finalizando, possivelmente uma das melhores contribuições futura desta trabalho, que aproveitaria a sua característica interdisciplinar, seria a gravação sistemática em banco de dados de mensagens do Twitter, conforme parâmetros e configurações que poderiam ser definidas para um grupo de agentes. Estas mensagens, em anos vindouros, poderiam ser uma base inestimável para pesquisas nas mais diversas áreas, em especial a sociologia e psicologia.

REFERÊNCIAS

- ALANI, H.; DASMAHAPATRA, S.; O'HARA, K.; SHADBOLT, N. Identifying Communities of Practice through Ontology Network Analysis. **IEEE Intelligent Systems**, Piscataway, NJ, USA, v. 18, p. 18–25, March 2003.
- AVRAM, A. **Twitter, uma Arquitetura Evoluindo**. 2009.
- BACKSTROM, L.; BOLDI, P.; ROSA, M.; UGANDER, J.; VIGNA, S. **Four Degrees of Separation**. 2011.
- BAKSHSHANDEH, R.; SAMADI, M.; AZIMIFAR, Z.; SCHAEFFER, J. **Degrees of Separation in Social Networks**. 2011.
- BARABASI. **Rich Get Richer**.
<http://www.barabasilab.com/LinkedBook/chapters/7Ch-TheRichGetRicher.pdf>, acesso em 19/06/2011.
- BERNSTEIN, M. S.; SUH, B.; HONG, L.; CHEN, J.; KAIRAM, S.; CHI, E. H. Eddi : interactive topic-based browsing of social status streams. **Fortune**, [S.l.], p. 303–312, 2010.
- BLOOD, R. **Weblogs: a history and perspective**. 2000.
- BONACICH, P. **Power and centrality: a family of measures**. 1987.
- BORDINI, R. H. **Uma Introdução aos Sistemas Multiagentes**. 2001.
- BOURDIEU, P. **The Rules of Art**. 1996.
- BOYD, D.; GOLDBERGER, S.; LOTAN, G. Tweet, Tweet, Retweet: conversational aspects of retweeting on twitter. **Hawaii International Conference on System Sciences**, Los Alamitos, CA, USA, 2010. volume 0, páginas 1-10.
- BOYD, D. M. **Taken out of context: american teen sociality in networked publics**. UNIVERSITY OF CALIFORNIA, BERKELEY, 409 pages.
- BRESCIANI, P.; GIORGINI, P.; GIUNCHIGLIA, F.; MYLOPOULOS, J.; PERINI, A. **Tropos: an agent oriented software development methodology**. 2004.
- BRUNO MARCOS GOMES DA SILVA BRUNO; MONTEIRO, R. L. **Empresa deve ter regra interna sobre redes sociais**.
<http://www.conjur.com.br/2010-jun-26/empresa-regra-interna-redes-sociais-evitar-conflitos>, acesso em fevereiro de 2012.
- CAVES, R. E. **Creative Industries**. 2000.
- CENTER, T. H. **How Do I Get Whitelisted?** <https://support.twitter.com/articles/160385>, acesso em 21/11/2011.
- CENTER, T. H. **About Twitter's Link Service (<http://t.co>)**.
<https://support.twitter.com/entries/109623>, acesso em fevereiro de 2012.

CENTER, T. H. **About Twitter Limits (Update, API, DM, and Following)**.

<https://support.twitter.com/articles/15364>, acesso em 02/02/2012.

CHENG, A. **Six Degrees of Separation, Twitter Style**. 2010.

CONVERTINO, G.; KAIRAM, S.; HONG, L.; SUH, B.; CHI, E. H. Designing a cross-channel information management tool for workers in enterprise task forces. In: INTERNATIONAL CONFERENCE ON ADVANCED VISUAL INTERFACES, 2010, New York, NY, USA. **Proceedings...** ACM, 2010. p. 103–110. (AVI '10).

DELOACH, S. A.; WOOD, M. F. **An Overview of the Multiagent Systems Engineering Methodology**. 2001.

DEVELOPERS, T. **Using the Twitter Search API**. <https://dev.twitter.com/docs/using-search>, acesso em 20/10/2011.

FERREIRA, M. L. T. **Redes Sociais: alguns modelos e conceitos básicos - parte 1**. <http://www.oficinadanet.com.br/artigo/midias-sociais/redes-sociais-alguns-modelos-e-conceitos-basicos-parte-1>, acesso em maio de 2011.

FREEMAN, L. C. **Centrality in Social Networks: conceptual clarification**. 1979.

GIRVAN, M.; NEWMAN, M. E. J. **Community structure in social and biological networks**. 2002.

GRANOVETTER, M. S. **The Strength of Weak Ties**. American Journal of Sociology, 1360 - 1380.

HANNEMAN, R. A.; RIDDLE, M. **Introduction to Social Network Methods**. <http://faculty.ucr.edu/hanneman/>, acesso em 19/06/2011, Riverside: Univ. of California, Riverside.

HUBERMAN, B. A.; ROMERO, D. M.; WU, F. **Social Networks that Matter: twitter under the microscope**. First Monday 14(1), 2009.

INC, M. Social Networks vs. Management? Harness the Power of Social Media. **Fresh Perspectives. Manpower Inc.**, [S.l.], p. 1 – 8, 2010.

IZQUIERDO, L. R.; HANNEMAN, R. A. **Introduction to the formal analysis of social networks using Mathematica**. 2006.

JANSEN, B. J.; ZHANG, M.; SOBEL, K.; CHOWDURY, A. Twitter power: tweets as electronic word of mouth. **J. Am. Soc. Inf. Sci. Technol.**, New York, NY, USA, v. 60, p. 2169–2188, November 2009.

JAVA, A.; FININ, T. Why We Twitter: understanding microblogging usage and communities. **Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop**, [S.l.], 2008.

JAVA, A.; SONG, X.; FININ, T.; TSENG, B. Why We Twitter: understanding microblogging usage and communities. **Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007**, [S.l.], August 2007.

KAPLAN, A. M.; HAENLEIN, M. The early bird catches the news: nine things you should know about micro-blogging. **Business Horizons**, [S.l.], v. 54, n. 2, p. 105–113, March 2011.

KWAK, H.; LEE, C.; PARK, H.; MOON, S. What is Twitter, a social network or a news media? In: WWW '10: PROCEEDINGS OF THE 19TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 2010, New York, NY, USA. **Anais...** ACM, 2010. p. 591–600.

LINE, G. O. **Facebook ganha 45 milhões de usuários em três meses.**

<http://gazetaonline.globo.com/-conteudo/2012/02/a-gazeta/indice/vida/1106191-facebook-ganha-45-milhoes-de-usuarios-em-tres-meses.html>, acesso em fevereiro de 2012.

MARASANAPALLE, J.; VIGNESH, T. S.; SRINIVASAN, P. K.; SAHA, A. Business intelligence from Twitter for the television media: a case study. In: _____. **2010 IEEE International Workshop on Business Applications of Social Network Analysis BASNA**. [S.l.]: IEEE, 2010. p. 1–6.

MASLOV, S.; SNEPPEN, K. **Specificity and stability in topology of protein networks**. 2002.

MEETEREN, M. V. **An ethnography of Indie software developers.**

<http://www.humangeographer.com/pdf/ethnographyofIndiesWP.pdf>, acesso em 23/06/2011.

MEETEREN, M. V.; POORTHUIS, A.; DUGUNDJI, E. Mapping communities in large virtual social networks: using twitter data to find the indie mac community. In: BUSINESS APPLICATIONS OF SOCIAL NETWORK ANALYSIS (BASNA), 2010, IEEE, INTERNATIONAL WORKSHOP ON, 2010. **Anais...** [S.l.: s.n.], 2010. p. 1–8.

MILGRAM, S. **The small world problem**. 1967.

NAAMAN, M.; BOASE, J.; LAI, C.-H. Is it really about me?: message content in social awareness streams. **Computer Supported Cooperative Work**, [S.l.], p. 189–192, 2010.

NEUSTADTL, A. **Social Network Analysis**. 2006.

NEWMAN, M. E. J. **The structure of scientific collaboration networks**. 2001.

NOORDHUIS, P.; HEIJKOOP, M.; LAZOVIK, A. Mining Twitter in the Cloud: a case study. In: IEEE 3RD INTERNATIONAL CONFERENCE ON CLOUD COMPUTING, 2010., 2010, Washington, DC, USA. **Proceedings...** IEEE Computer Society, 2010. p. 107–114. (CLOUD '10).

NORVIG, S. R. P. **Inteligencia Artificial**. 2004.

ODELL, J.; PARUNAK, H. V. D.; BAUER, B. **Extending UML for Agents**. 2000.

OH, S.; YEOM, H. Y.; AHN, J. A method to extract a social network based on semantic association. In: INTERNATIONAL CONFERENCE ON UBIQUITOUS INFORMATION MANAGEMENT AND COMMUNICATION, 5., 2011, New York, NY, USA. **Proceedings...** ACM, 2011. p. 10:1–10:5. (ICUIMC '11).

- PASSANT, A.; HASTRUP, T.; BOJARS, U.; BRESLIN, J. Microblogging: a semantic and distributed approach. In: WORKSHOP ON SCRIPTING FOR THE SEMANTIC WEB, TENERIFE, SPAIN, JUNE 02, 2008, CEUR WORKSHOP PROCEEDINGS, 4., 2008. **Proceedings...** [S.l.: s.n.], 2008.
- PERERA, R. W. D.; ANAND, S.; SUBBALAKSHMI, K. P.; CHANDRAMOULI, R. Twitter analytics: architecture, tools and analysis. **MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010**, [S.l.], 2010.
- PHITHAKKITNUKON, S.; DANTU, R. Mobile social closeness and communication patterns. In: IEEE CONFERENCE ON CONSUMER COMMUNICATIONS AND NETWORKING CONFERENCE, 7., 2010, Piscataway, NJ, USA. **Proceedings...** IEEE Press, 2010. p. 319–323. (CCNC'10).
- PIETRI, W. **Best Java Twitter Library**.
<http://stackoverflow.com/questions/851767/best-java-twitter-library>, acesso em 19/06/2011.
- RAMAGE, D.; DUMAIS, S.; LIEBLING, D. Characterizing Microblogs with Topic Models. In: ICWSM, 2010. **Anais...** [S.l.: s.n.], 2010.
- ROSA, A. M. **Elementos para uma teoria geral de rede**.
<http://www.cecl.com.pt/redes/pdf/amr.pdf>, página 10, acesso em 19/06/2011.
- SATHIK M.M.; RASHEED, A. **Social networks of buying - Likely patterns**. 2009.
- SHENQING, X. H. C. D. K. **The research about virtual community impact knowledge exchange and diffusion based on the perspective of social network**. 2010.
- SUH, B.; HONG, L.; PIROLI, P.; CHI, E. H. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In: IEEE SECOND INTERNATIONAL CONFERENCE ON SOCIAL COMPUTING (SOCIALCOM), 2010. **Proceedings...** [S.l.: s.n.], 2010. p. 177–184.
- TAYLOR, C. **Twitter has 100 million active users**. Acessado em 11/02/2011.
- TOMAEI, M. I.; ALCARA, A. R.; CHIARA, I. G. D. **Das redes sociais à inovação**. 2005.
- TWITTERSEARCH. **The Engineering Behind Twitters New Search Experience**. 2011.
- UGANDER, J.; KARRER, B.; BACKSTROM, L.; MARLOW, C. **The Anatomy of the Facebook Social Graph**. 2011.
- WASSERMAN, S.; FAUST, K. **Social Network Analysis: methods and applications (structural analysis in the social sciences)**. 1. ed. [S.l.]: Cambridge University Press, 1994. (Structural analysis in the social sciences, 8).
- WATTS, D. J.; STROGATZ, S. H. **Collective dynamics of small world networks**. 1998.
- WIKIKOFF, L. P. M. **Prometheus: a practical agent-oriented methodology**. Chapter 5 in Agent-Methodologies, 2005.
- WOOLDRIDGE, M. **An Introduction to Multiagent Systems**. 2002.
- WRIGHT, A. **Mining the Web for Feelings, Not Facts**. New York Times, 2009-08-23.

WWW.MYSQL.COM. **Why MySQL**. 2011.

ZARRELLA, D. **Science of Retweets**.

<http://danzarella.com/the-science-of-retweets-report.html>, acesso em novembro de 2011.

ZARRELLA, D. **Dan Zarrella Bio**. <http://danzarella.com/bio>, acesso em novembro de 2011.

ZARRELLA, D. **Introducing a New Tool**: whoretweetedme.

<http://danzarella.com/introducing-a-new-tool-whoretweetedme.html>, acesso em novembro de 2011.

ZHAO, D.; ROSSON, M. How and why people Twitter: the role that micro-blogging plays in informal communication at work. In: ACM 2009 INTERNATIONAL CONFERENCE ON SUPPORTING GROUP WORK, 2009. **Proceedings...** [S.l.: s.n.], 2009. p. 243–252.