

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**ANÁLISE QUANTITATIVA DAS REALAÇÕES ENTRE
USUÁRIOS NO *TWITTER***

ROBERTO FELÍCIO DE OLIVEIRA

JOÃO PESSOA-PB
Abril-2013

ROBERTO FELÍCIO DE OLIVEIRA

**ANÁLISE QUANTITATIVA DAS RELAÇÕES ENTRE
USUÁRIOS NO *TWITTER***

DISSERTAÇÃO APRESENTADA AO CENTRO DE INFORMÁTICA DA
UNIVERSIDADE FEDERAL DA PARAÍBA, COMO REQUISITO PARCIAL
PARA OBTENÇÃO DO TÍTULO DE MESTRE EM INFORMÁTICA
(SISTEMAS DE COMPUTAÇÃO).

Orientador: Prof. Dr. Alisson Vasconcelos de Brito

JOÃO PESSOA-PB
Abril-2013

O48a Oliveira, Roberto Felício de.

Análise Quantitativa das Relações entre Usuários no *Twitter*/

Roberto Felício de Oliveira. - - João Pessoa: [s.n.], 2013.

104 f. : il.

Orientador: Alisson Vasconcelos de Brito.

Dissertação (Mestrado) – UFPB/CCEN.

1. Informática, 2. Redes Sociais – análise. 3. Twitter. 4. Análise de Crédito. 5. Informática na educação.

UFPB/BC

CDU: 004(043)

(Ata de Sessão Pública)

Dedico à minha família material, em especial todos, pela ajuda e incentivo que me deram, para que eu pudesse transpor as distâncias geográficas existentes entre meu estudo e trabalho, pelas palavras de ânimo para transpor as dificuldades do dia a dia e, à minha família espiritual por estar ao meu lado todos os instantes, me amparando e ao mesmo tempo fortalecendo-me de modo a acreditar neste sonho de se tornar mestre e que, hoje torna-se uma realidade.

Agradecimentos

Agradeço à Deus em primeiro lugar, por ter me dado a oportunidade de concluir mais esta fase em minha vida.

Ao Prof. Dr. Alisson Vasconcelos Brito, orientador desta dissertação, por toda compreensão, acima de tudo, pela sua competência e exigência que me proporcionou estímulos a fim de que eu pudesse concluir este trabalho adquirindo assim novos conhecimentos.

Aos meus amigos, Jefferson Araújo e Danylo Wagner Albuquerque, pelos quais tenho hoje, uma grande consideração pela colaboração durante a elaboração desta dissertação.

Ao grupo de pesquisa referente à Redes Sociais, pertencente ao Programa de Pós Graduação em Informática - PPGI que, de uma maneira direta ou indireta, contribuíram para o término deste estudo.

Aos coordenadores do Programa de Pós Graduação em Informática - PPGI inicialmente à Prof^a Dra. Tatiana Aires Tavares e atualmente o Prof. Dr. Alisson Vasconcelos de Brito, pela oportunidade de crescimento, aprendizado, realização profissional e pessoal e pela confiança em mim depositada.

Resumo

Sob o aspecto de seres sociais, os seres humanos procuram incessantemente interagir uns com os outros, nas mais variadas ocasiões. Realidade esta, expandida ao mundo virtual, ao qual permite uma comunicação sem fronteiras através da Internet entre as pessoas de qualquer parte do mundo. Desta maneira, a presente dissertação visa analisar quantitativamente os relacionamentos entre usuários do *Twitter*, através da utilização de ferramentas de monitoramento de Redes Sociais aliados à técnicas de Análise de Rede Sociais, objetivando definir o perfil de relacionamento dos indivíduos dentro do cenário ao qual o mesmo encontra-se inserido quer na área de educação ou crédito. Os estudos realizados revelam que através da aplicação das métricas grau de centralidade, centralidade de intermediação e centralidade de proximidade é possível definir o perfil colaborativo das pessoas dentro de um determinado grupo e eventuais falhas na disseminação de informações no mesmo. Salienta-se também, que a referida pesquisa apresenta uma forma de definir a participação e a afinidade de pessoas em um determinado contexto. Os resultados obtidos mostram que é possível correlacionar os dados dos usuários no *Twitter*, com seus índices referente a aplicação das métricas de análise de redes sociais e com dados de empresas de crédito, a fim de analisar quantitativamente o relacionamento destes usuários em diversos aspectos votados à restrição de crédito.

Palavras-chave: Análise de Redes Sociais. *Twitter*. Análise de Crédito. Rede Sociais. Informática na Educação

Abstract

Under the aspect of social beings, humans seek relentlessly interact with each other, in various occasions. This reality, expanded to the virtual world, which allows communication without borders through the Internet between people anywhere in the world. Thus, this dissertation aims to analyze quantitatively the relationships among Twitter users, through the use of Social Media monitoring tools combined with the techniques of Social Network Analysis, aiming to define the profile relationship of individuals within the scenario of which the same is inserted either in education or credit. Studies show that by applying the metrics degree of centrality, brokerage and centrality centrality proximity is possible to define the profile of collaborative people in a particular group, and any flaws in the dissemination of information on the same. It should be noted also, that this research presents a way to define participation and affinity of people in a given context. The results show that it is possible to correlate the data of users on Twitter, with their indices regarding the application of the metrics data with social network analysis and credit companies in order to quantitatively analyze the relationship of these users on various aspects rated the credit crunch.

Keywords — *Social Network Analysis. Twitter. Credit Analysis. Social Networking. Computers in Education*

Lista de Ilustrações

Figura 1 - (a) Estrutura básica de um arquivo de rede para o <i>Pajek</i> , e em (b) a sua representação gráfica	23
Figura 2 - (a) Estrutura básica de um arquivo de rede tipo matriz de adjacência para o <i>Pajek</i> , e (b) sua representação gráfica	23
Figura 3 - (a) Trechos da estrutura básica de um arquivo de rede no <i>Prefuse</i> , e (b) sua representação gráfica	24
Figura 4 - (a) Trechos da estrutura básica de um arquivo de rede no <i>JUNG</i> , e (b) sua representação gráfica	26
Figura 5 - (a) trecho da estrutura básica de um arquivo de rede no Gephi, e (b) sua representação gráfica	27
Figura 6 - Grafo de Interação.....	31
Figura 7 - Cálculo do coeficiente de agrupamento <i>de</i> um nó em dois cenários	40
Figura 8 - Valores referentes à centralidade de informação	42
Figura 9 - Atividades de ARS.....	43
Figura 10 - Passos Metodológicos	45
Figura 11 - Interface da ferramenta <i>Twitter Acadêmico</i>	46
Figura 12 - Arquitetura da AppSNA.....	47
Figura 13 - Grafo de Interação entre os alunos.....	58
Figura 14 - Exclusão de um nó central da Rede: Formação de uma rede duomodal	59
Figura 15 - Desestruturação da rede: Exclusão de um nó de alto grau de centralidade	59
Figura 16 - Grafo de afinidade entre os alunos pesquisados.	64
Figura 17 - Grafo de Interação a_i para a_j	66
Figura 18 - Grafo da rede contendo somente usuários de João Pessoa-PB	73
Figura 19 - Representação do valor <i>in-degree</i> através de um grafo	77
Figura 20 - Porcentagem de pessoas Jurídicas e Físicas baseadas em uma variação <i>in-degree</i>	78
Figura 21 - Porcentagem das pessoas comuns e que são destaques na Grande João Pessoa-PB baseadas em uma variação <i>in-degree</i>	79
Figura 22 - Grafo representando os valores <i>out-deegree</i> de cada pessoa.....	80
Figura 23 - Porcentagem das pessoas que enviaram mensagens direcionadas.....	81
Figura 24 - Porcentagem de Pessoas Físicas e Jurídicas que interagiram diretamente na rede	82
Figura 25 - Comparativo de Proximidade.....	84
Figura 26 - Porcentagem de proximidade das pessoas físicas e jurídicas baseadas nos resultados do grau de proximidade	85
Figura 27 - Representação dos usuários identificados e não, pela Serasa Experian	88
Figura 28 - Representação da quantidade de pessoas identificadas em relação às restrições de crédito que cada nó possui	89
Figura 29 - Representação da quantidade de certidões negativas dos usuários.....	90

Figura 30 – Representação da quantidade de certidões negativas da vizinhança de cada um dos 26 usuários que são a base da rede.....	91
Figura 31 – Porcentagem comparativa da inadimplência.....	93

Lista de Tabelas

Tabela 1- Representação das pessoas como vértices.....	29
Tabela 2 - Interação entre os usuários	30
Tabela 3 – Matriz de Interação	30
Tabela 4 – Valor <i>In-Degree Centrality</i>	32
Tabela 5 – Valor <i>Out-Degree Centrality</i>	32
Tabela 6 – Valor da Centralidade de Proximidade.....	34
Tabela 7 – Valor da Centralidade de Intermediação	36
Tabela 8 – Valores <i>Page Rank</i>	38
Tabela 9 – Valor do coeficiente de agrupamento	40
Tabela 10 - Tecnologias utilizadas na construção da arquitetura AppSNA	47
Tabela 11 - Valores <i>In-Degree</i> e <i>OutDegree</i>	58
Tabela 12 – Grau de Participação dos acadêmicos sobre os assuntos: “Requisitos Funcionais” e “Não funcionais”	61
Tabela 13 – Grau de Participação Normalizado Referente ao tema Requisitos Funcionais e Não funcionais.....	62
Tabela 14 – Total de Mensagens de P_i e total de mensagens diretas de P_i para P_j	63
Tabela 15 – Menções de a_i para a_j	65
Tabela 16 - Grau centralidade <i>In-Degree</i> e <i>Out-Degree</i>	67
Tabela 17 - Valores medidos para o grau de proximidade	68
Tabela 18 – Valores do grau de intermediação	68
Tabela 19 – Valor <i>Page Rank</i>	74
Tabela 20- Valor <i>In-Degree Centrality</i>	75
Tabela 21 - Valor <i>out-degree centrality</i>	79
Tabela 22 – Grau de Intermediação.....	82
Tabela 23 – Valores referentes ao grau de proximidade das pessoas.....	83
Tabela 24 – Resultados da análise quantitativa	85
Tabela 25 – Índice de Inadimplência.....	92

Lista de Equações

Equação 1 - Grau de centralidade.....	31
Equação 2 – Equação Simplificada do Grau de Centralidade	32
Equação 3 – Centralidade de proximidade	33
Equação 4 – Centralidade de proximidade normalizada	34
Equação 5 – Centralidade da intermediação.....	34
Equação 6 – Grau de Intermediação de um nó.....	35
Equação 7 – Centralidade da intermediação normalizado.....	36
Equação 8 – Formula do <i>Page Rank</i>	37
Equação 9 – Coeficiente de agrupamento local.....	39
Equação 10 – Contabilização Global do Coeficiente de Agrupamento	40
Equação 11 - Centralidade da Informação	41
Equação 12 – Cálculo da Centralidade de Informação de uma matriz de adjacência	42
Equação 13 – Formula da inadimplência individual	91
Equação 14 – Formula da inadimplência da vizinhança	92

Lista de Abreviações

API – *Application Programming Interface*

ARS - *Análise de Redes Sociais*

CC - *Clustering Coefficient*

CSV - *Comma-separated Values*

MD - *Mensagens Diretas*

MVC - *Model View Controller*

RDF - *Resource Description Framework*

RT - *Retweet*

TIC - *Tecnologia da Informação e Comunicação*

TXT - *Text File*

XML - *Extensible Markup Language*

Sumário

1. Introdução.....	16
1.1. Motivação	17
1.2. Objetivo Geral.....	19
1.3. Objetivos Específicos	19
1.4. Metodologia	19
2. Fundamentação Teórica.....	21
2.1. Microblogging	21
2.2. Softwares para visualização de grafo.....	22
2.2.1 Pajek	22
2.2.2 Prefuse	23
2.2.3 JUNG.....	25
2.2.4 Gephi	26
2.3. Métricas de Análise de Redes Sociais	28
2.3.1 Cenário	29
2.3.1.1 Grau de Centralidade.....	31
2.3.1.2 Centralidade de Proximidade	33
2.3.1.3 Centralidade de Intermediação.....	34
2.3.1.4 Page Rank.....	36
2.3.2 Coeficiente de Agrupamento	38
2.3.3 Centralidade de Informação.....	40
3. Processo de Análise de Redes Sociais.....	43
3.1. Coleta de Dados	44
4. Trabalhos Correlatos.....	48
4.1. Monitoramento das redes sociais como forma de relacionamento com o consumidor – o que as empresas estão fazendo?.....	48
4.2. Detecção de conteúdo relevante e usuários influentes no <i>Twitter</i>	48
4.3. Mineração de dados e análise de opinião em redes sociais – Um estudo de caso sobre as Olimpíadas 2012 utilizando o <i>Twitter</i>	49
4.4. PANAS-t: Uma escala psicométrica para medição de sentimentos no <i>Twitter</i>	50
4.5. Sigam-me os bons! Transformando robôs em pessoas influentes no <i>Twitter</i> (influência).....	51
4.6. Implementing semantic queries in online social networks	52
4.7. On the predictability of elections using social media data	54

4.8.	Explorando redes sociais online: da coleta e análise de grandes bases de dados às aplicações.....	55
4.9.	Análise de Redes Sociais – aplicações nos estudos de transferência da informação	56
5.	Estudos de casos	57
5.1.	Estudos de casos na área de educação	57
5.1.1	Analysis of Students Interaction Profile in Online Social Networks	57
5.1.2	Estratégia Para Avaliação do Grau de Participação e Afinidade de Alunos Através de <i>Microbloggin</i>	60
5.1.3	Monitoramento das interações dos aprendizes na rede social <i>Twitter</i> como apoio ao processo de mediação docente.	64
5.2.	Conclusão dos estudos de casos na área acadêmica	69
5.3.	Estudos de casos de grupos específicos	70
5.3.1	Cenário utilizado nos experimentos dos grupos específicos	70
5.3.2	Classificação de usuários do <i>Twitter</i> pela análise estrutural e utilização de palavras-chave	71
5.3.2.1	Análise estrutural da rede	74
5.3.2.2	Análise quantitativa.....	85
5.3.2.3	Conclusão	86
5.4.	Análise quantitativa das relações dos usuários de uma rede social virtual frente à uma base de crédito	87
5.4.1	Conclusão do estudo de caso	94
6.	Considerações Finais	95
7.	REFERÊNCIAS	97
8.	ANEXOS.....	99

1. Introdução

A atualidade, em sua diversidade social e complexa formação leva consigo novas formas de relações entre indivíduos e grupos de indivíduos.

Os remotos conceitos de relacionamento, trocas e interações de informação, tão evidentes quanto à própria vida humana, adotaram novas variações, canais e interfaces, não com o surgimento da Internet, mas, com sua popularização e abrangência global. Sobrevém por decorrência, a evolução dos relacionamentos na era da Internet e surgem as redes sociais online enquanto forma mais viável de aproximar pessoas, trocar experiências, compartilhar interesses, além das mesmas proporcionarem uma infinidade de opções relacionais, comerciais, financeiras e afetivas.

As redes sociais online não atraem e conquistam apenas os jovens, mas também os adultos. Portanto, as redes sociais se prestam à múltiplas funcionalidades e são tidas como assunto dos mais interessantes para estudiosos/pesquisadores nos ramos do comportamento, finanças, economia, psicologia e demais áreas acadêmicas atualmente.

Diante de tal visão, constata-se que a temática "redes sociais online", constitui extraordinário assunto para aprofundamento em pesquisas de cunho acadêmico, podendo invariavelmente, produzir interessantes e diferentes tipos de produções textuais com resultados relevantes à atualidade.

Neste sentido, a presente pesquisa visa inquirir acerca da viabilidade de utilização do *Twitter* e das métricas de análise de redes sociais, como meios viáveis de caracterizar o perfil de relacionamento e interação entre os usuários nos diferentes grupos sociais.

Para o desenvolvimento deste processo de investigação, recorreu-se a uma bibliografia significativa, constituída por autores renomados e destacados na comunidade científica nacional e internacional, que auxiliaram a fundamentar o trabalho teórico.

Sendo assim, esta dissertação apresenta-se dividida em cinco capítulos: o primeiro capítulo corresponde à introdução onde se contextualiza brevemente o tema Redes Sociais, e posteriormente, são expostos a motivação do estudo e os objetivos do trabalho.

O segundo capítulo abrange uma perspectiva geral da técnica de Análise de Redes Sociais, proporcionando uma visão básica do seu histórico e aplicações existentes. O mesmo discorre ainda, sobre as principais Métricas de Análise de Redes Sociais.

No terceiro capítulo, a metodologia e a arquitetura do sistema para os experimentos práticos é apresentada de um modo abstrato e algumas características do *Twitter*, pois o mesmo é a base de nossa pesquisa. O capítulo quatro descreve alguns trabalhos correlatos que enfatizam as diversas áreas e campos que envolvem redes sociais, e que ainda poderão ser explorados. Seguido pelo quinto capítulo, que aborda os resultados da pesquisa através dos experimentos realizados mediante estudos de casos na área da educação e crédito.

1.1. **Motivação**

O presente trabalho visa investigar a viabilidade de utilizar-se do *Twitter* e das técnicas de análise de redes sociais como meios de analisar comparativamente o perfil de relacionamento dos usuários pertencentes a determinados grupos sociais online.

É notório que, desde a origem da humanidade as pessoas buscam se reunir em grupos. Sendo estes baseados em costumes, crenças, conceitos filosóficos ou até mesmo religiosos. Desta maneira, podemos definir uma Rede Social como sendo uma forma de representar os relacionamentos afetivos ou profissionais dos seres entre si ou seus agrupamentos de interesses mútuos. Por outro lado, sabe-se que, a qualidade e intensidade destes relacionamentos são os principais fatores de conexão entre cada pessoa e a rede como um todo; tendo como intuito, alcançar os objetivos tanto individuais quanto do grupo.

Segundo Berkowitz (1982), uma Rede Social é uma estrutura social composta por nós, (que são geralmente indivíduos ou organizações) que estão ligadas por um, ou mais, tipos específicos de interdependência, como: valores, visões, ideias, trocas financeiras, amigos, parentesco, antipatia, conflito, comércio, relações sexuais, dispersão de doenças (epidemiologia), rotas aéreas e outros.

Logo, através das Redes Sociais os indivíduos que possuem os mesmos interesses e/ou objetivos poderão partilhar ideias, informações entre outros aspectos. “As

interconexões entre indivíduos, seus agrupamentos e comunicações merecem a devida ênfase, para compreenderem a inserção social e a distribuição da informação – ao atuar nesta trama, cada indivíduo estabelece vínculos com outros, e a intersecção de seus interesses é amálgama suficiente para identificá-los não apenas como grupos, mas sim como uma rede social” - [Castells, 1999].

Baseado neste pensamento verifica-se a existência de uma interdependência social, onde cada parte desempenha um poder relativo que pressiona e influencia o todo, quer em maior, ou menor grau, ao mesmo tempo em que o todo limita a autonomia de cada uma das partes.

Segundo Capra (2002), o padrão de rede (*network pattern*), especificamente, é um dos padrões de organização mais básicos de todos os sistemas vivos. Em todos os níveis de vida – desde as redes metabólicas das células até as teias alimentares dos ecossistemas –, os componentes e os processos dos sistemas vivos se interligam em forma de rede. A compreensão sistêmica da vida ao domínio social permite identificar a aplicação de nossos conhecimentos dos padrões e princípios básicos de organização da vida – e, em específico, da nossa compreensão das redes vivas – à realidade social. Neste sentido, pode-se afirmar que as redes representam um modelo básico de relacionamento para todo e qualquer sistema vivente.

Castells (2003), destaca ainda que as Redes Sociais vêm incorporando as comunidades virtuais e na Internet caracterizam-se como uma “nova” forma de organizar a interação. Salienta-se que a formação de comunidades virtuais é uma prática cada vez mais utilizada em diversas áreas, como; educacionais, empresariais etc. Sendo que, diversas ferramentas síncronas e assíncronas estão presentes em ambientes que apoiam a colaboração entre pessoas, dentre elas destacam-se os fóruns, chats, e-mails e mais recentemente as redes sociais *onlines*.

Por conseguinte, a utilização das redes sociais mediadas pela Internet trouxe este tipo de estrutura social para um novo contexto, com uma abrangência social, econômica e geográfica antes impraticável.

Para Alencar (2007), tanto em redes sociais na Internet quanto em comunidade, as pessoas aprendem umas com as outras, compartilhando informações, desafios, conquistas e descobertas. Salienta-se que as redes sociais na internet são páginas da web que facilitam a interação entre os membros em diversos locais. Elas existem para proporcionar meios diferentes e interessantes de interação.

Partindo-se deste princípio, percebe-se que estes agrupamentos são fonte, para a atualização profissional, qualificação da prática e novas descobertas, sendo este ultimo o foco deste projeto.

1.2. Objetivo Geral

O propósito deste trabalho é investigar a viabilidade de utilização da rede social *Twitter* e das métricas de análise de redes sociais, como caminhos viáveis de caracterizar o perfil de relacionamento e interação entre os usuários em diferentes grupos sociais.

1.3. Objetivos Específicos

Os objetivos específicos almejados nesta pesquisa são as soluções para os itens abaixo descritos:

- Identificar métricas capazes de definirem o grau de afinidade entre as pessoas e o grau de influência que cada uma destas pessoas exerce sobre as outras em cada grupo a ser analisado;
- Definir caminhos que possam identificar o comportamento das pessoas pertencentes a uma rede social online, através de métricas de análise de rede social;
- Buscar estratégias para correlacionar o comportamento das pessoas de determinado grupo social, com a sua conduta dentro das redes sociais virtuais;

1.4. Metodologia

Para a realização da presente pesquisa, iniciou-se através de um estudo bibliográfico, onde se constatou a existência de algumas métricas divididas por categorias. Dentre as métricas estudadas optou-se por trabalhar o Grau de Centralidade, Centralidade de Proximidade, Centralidade de Intermediação e Centralidade de Informação,

pertencentes à categoria de Centralidade e Prestígio. Em relação à classe Grupos Estruturais, optou-se pela métrica Coeficiente de Agrupamento.

Esta opção foi fundamentada nos estudos dos casos correlatos, onde se verificou que referidas métricas possuem um elevado grau de eficiência para a solução de diversos problemas em cenários totalmente diferentes, razão pela qual, almejamos que elas dêem as respostas para os problemas apresentados no cenário ora em estudo.

Partindo desse pressuposto, experimentos foram realizados, mediante estudos de casos na área da educação e na área de crédito, almejando alcançar os objetivos específicos em grupos sociais distintos, além de verificar a possibilidade de elencar os pontos comuns entre estes.

Sendo assim, a realização deste projeto percorrerá as seguintes etapas:

- Adaptar a ferramenta de coleta de dados do *Twitter*, de maneira que a mesma possa coletar dados pertinentes, através de palavras-chave, referentes a cada grupo social a ser estudado;
- Gerar grafos que possam refletir a mudança do individualismo de cada indivíduo dentro do grupo em que se encontra inserido, de acordo com as técnicas de análise de redes Sociais;
- Analisar os grafos utilizando as técnicas e ferramentas disponíveis;
- Buscar relações entre o comportamento dos indivíduos no mundo real e suas relações no *Twitter*;

2. Fundamentação Teórica.

Este capítulo apresenta uma breve introdução referente ao microblog *Twitter*, posteriormente é exposto um estudo conciso sobre o software Pajek e as API's Prefuse e Jung, na sequência são abordados os aspectos da Análise de Redes Sociais (ARS), mais especificamente, as diversas técnicas existentes a fim de identificar dentro de um contexto, os membros que representam pontos estratégicos dentro da rede social.

2.1. Microblogging

Segundo Costa (2003) a cultura digital que nada mais é que a interação humana com a Tecnologia da Informação e Comunicação (TIC) no seu cotidiano, tem influenciado continuamente a relação/comunicação dos indivíduos em sociedade. Partindo deste pressuposto, verifica-se que atualmente a estrutura das redes sociais foca a interação entre pessoas de forma metafísica, ou seja, atingindo tanto o âmbito do mundo concreto quanto o do mundo virtual.

Baseado neste contexto, o *Twitter* (lançado em 2006) é considerado como uma forma de comunicação virtual entre indivíduos, além de representar uma espécie de serviço de microblogging que permite aos seus usuários publicarem pequenas menções (de até 144 caracteres), que são vistas por quem o segue (seguidores).

No *Twitter* a comunicação direcionada/referenciada entre pessoas é feita através da sintaxe “@” (menção). Além desta, utiliza-se também a convenção RT (*Retweet*) para o reenvio de mensagens de terceiros, para compartilhar descobertas e ampliar notícias na rede social. Estruturalmente um *Retweet* equivale ao encaminhamento de um e-mail, onde pessoas reenviam mensagens de outras para outros usuários. E finalmente, as mensagens privadas conhecidas como MD (Mensagens Diretas), que são visualizadas somente pelo remetente e o destinatário.

2.2. Softwares para visualização de grafo

A respectiva seção visa apresentar alguns dos softwares existentes à geração de grafos referentes a um conjunto diverso de situações, tais como: processos industriais (circuitos impressos), redes sociais e mapas geográficos (rotas de transito).

Vale ressaltar que os grafos são uma das técnicas de visualização mais difundidas para representar os relacionamentos entre objetos, pessoas, estruturas hierárquicas entre outros. Os mesmos são modelos matemáticos constituídos de vértices que comumente representam objetos concretos ou abstratos e, um conjunto de arestas que indicam os relacionamentos entre estes objetos.

2.2.1 Pajek

O software *Pajek* tem como uma de suas funcionalidades a visualização e análise de amplas redes, sendo estas, sociais de coautoria e etc. Observa-se ainda que, o mesmo vem sendo utilizado pelos professores em suas aulas de grafos, visando tornar o processo ensino aprendizagem mais prazeroso e de fácil assimilação.

A geração de uma rede no *Pajek* pode ser feita através de uma representação dos vértices, *edges* e *arcs* por meio de um arquivo texto, com a extensão “.net”. É notório que para outros aplicativos esse tipo de arquivo não é comumente utilizado.

Na Figura 1 (a) pode-se ver a estrutura básica de uma rede descrita no arquivo “.net”. Inicialmente é representado a quantidade de vértices através da sintaxe **Vertices 6*, portanto pode-se constatar que o referido grafo a ser gerado terá 6 vértices. Nota-se ainda que, esta quantidade de vértices deverá ser sempre a primeira declaração na definição de uma rede no *Pajek*.

Na sequencia há representação de arestas direcionadas, denotadas pela sintaxe **Arcs*. Sendo assim, em cada linha visualiza-se a declaração das arestas direcionadas e o peso das mesmas para cada par de vértices. Posteriormente, a última sintaxe apresentada é **Edges* que representa arestas não direcionadas. Na Figura 1 (b) temos a representação gráfica da rede.

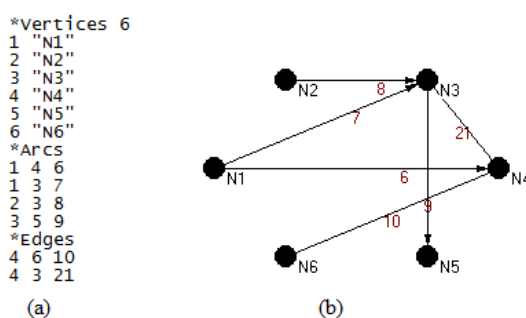


Figura 1 - (a) Estrutura básica de um arquivo de rede para o *Pajek*, e em (b) a sua representação gráfica

Percebe-se que outro meio de representar um grafo direcionado no *Pajek*, é através de uma matriz de adjacência. Figura 2 – (a) Estrutura básica de um arquivo de rede tipo matriz de adjacência para o *Pajek*, e (b) sua representação gráfica

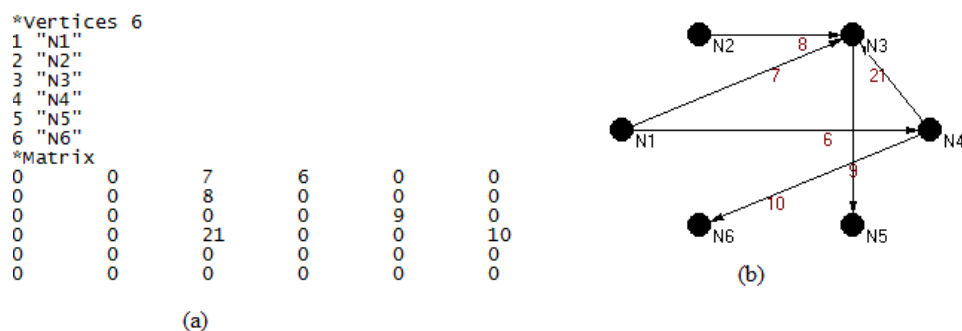


Figura 2 – (a) Estrutura básica de um arquivo de rede tipo matriz de adjacência para o *Pajek*, e (b) sua representação gráfica

Embora esta ferramenta permita uma ótima visualização dos dados através de um grafo e, possua algumas métricas de ARS implementadas, a mesma torna-se inviável à utilização permanente na pesquisa, pois sabe-se que o *Pajek* é um software de código fechado, ou seja, ele não possui uma API (*Interface de Programação de Aplicativos*) para permitir que a ferramenta de coleta de dados do *Twitter* possa fazer uma chamada de sistema ao referido software, com o intuito de executar as métricas existentes no mesmo.

2.2.2 Prefuse

O *Prefuse* é uma biblioteca para modelagem, visualização e interação de grafos, tabelas e árvores em Java. A partir de uma base de dados em arquivos no formato TXT ou XML, o referido framework permite gerenciar layouts de visualização, técnicas de codificação visual, animação, pesquisas integradas e conexão com banco de dados.

Destaca-se ainda neste framework, a base arquitetural com seus componentes distribuídos em um padrão MVC (*Model View Controller*), ou seja, um padrão de projeto de software que separa a “lógica” da aplicação da interface do usuário (manipulação dos dados), possibilitando assim, desenvolver e testar separadamente cada uma destas unidades.

Uma das grandes vantagens do *Prefuse* é o fato dele poder ser utilizado tanto para fins comerciais quanto não comerciais, pois o mesmo trata-se de uma API de código aberto.

Nota-se que o *Prefuse* utiliza-se de uma estrutura própria para armazenamento de dados. Desta maneira, o construtor da classe *Graph* é constituído por cinco atributos como parâmetro: especificação dos dados em cada nó, especificação dos dados em cada aresta, definição se o grafo é direcional ou não e os dois atributos em cada aresta que possuirão os identificadores dos nós adjacentes.

Na Figura 3 (a) verifica-se a definição dos parâmetros relacionados ao tipo de grafo, neste contexto não direcional; a sintaxe para construção dos vértices e seus respectivos rótulos além das arestas e seus respectivos pesos. A Figura 3 (b) temos sua representação gráfica.

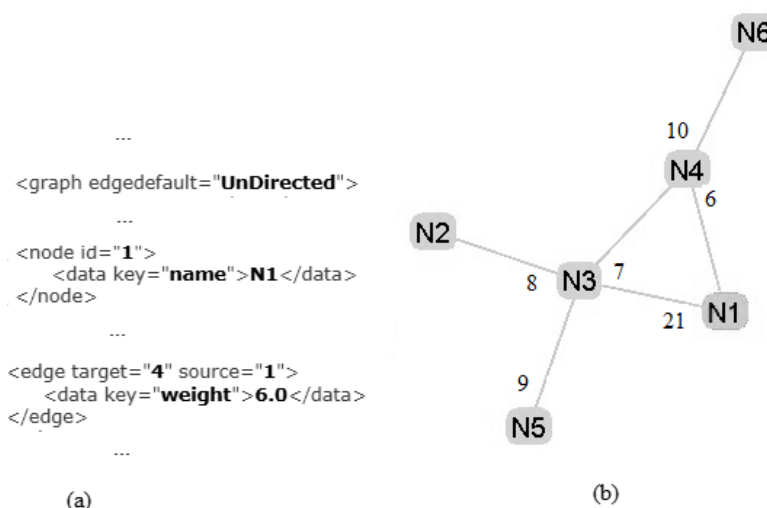


Figura 3 - (a) Trechos da estrutura básica de um arquivo de rede no *Prefuse*, e (b) sua representação gráfica

No entanto, a referida API apresenta como pontos negativos, a existência de uma documentação incipiente e apenas a inclusão de técnicas para análise de grafos e, não de métricas para análise de redes sociais. Por conseguinte, esta API torna-se inviável como base única para este projeto.

2.2.3 JUNG

O *JUNG* (*Java Universal Network / Graph Framework*) destaca-se pelo fato do mesmo ser uma biblioteca de software que proporciona uma linguagem comum e extensível para modelagem, análise e visualização de dados que possam ser representados como um grafo ou uma rede.

Sua arquitetura foi projetada para suportar uma gama de representações gráficas das relações existentes entre elementos de dados, tais como: grafos direcionais e não direcionais, multi-modal, grafos com bordas paralelas, hipergrafos entre outros. O *JUNG* inclui também, implementações de alguns algoritmos de teoria dos grafos, mineração de dados e análise de redes sociais, incluindo agrupamento, filtragem, geração de gráficos aleatórios, cálculo das distâncias das redes e fluxos, além de uma grande variedade de métricas tais como: *PageRank*, intermediação, proximidade entre outros.

Vale ressaltar que, o mesmo fornece um quadro de visualização que facilita a construção de ferramentas à exploração interativa de dados da rede. Deste modo, os usuários poderão usar um dos algoritmos de layout fornecido, ou usar a estrutura para desenvolverem seus próprios layouts.

Como uma biblioteca de código aberto, *JUNG* estabelece um quadro comum para geração de grafos e, visualização e análise de rede. Neste aspecto, ele torna-se um mecanismo facilitador para trabalhar com dados de grafo e de rede.

A Figura 4 (a) percebe-se a forma de definir um nó na rede e as duas maneiras de representação de arestas, uma através da sintaxe *EdgeType.DIRECTED*, indicando arestas direcionais e, a ausência desta sintaxe correspondendo a arestas não direcionais. Figura 4 (b) temos sua representação gráfica.

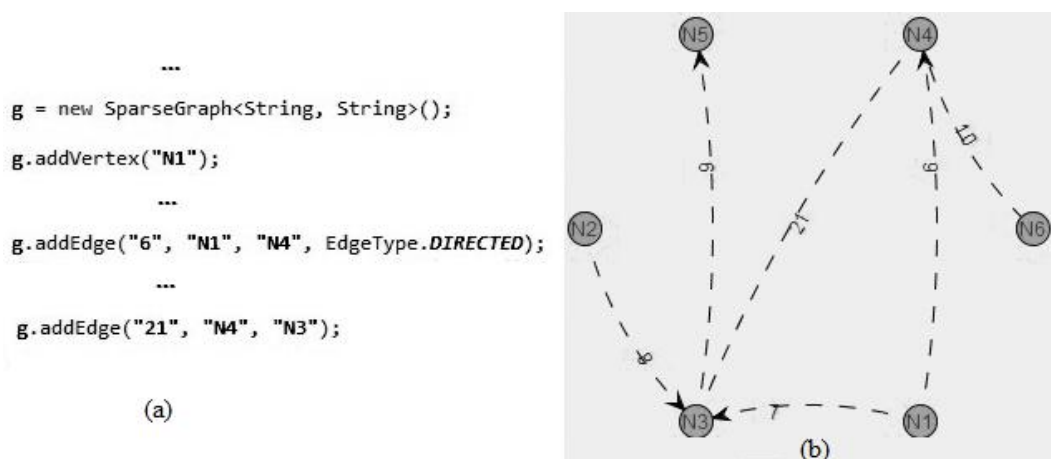


Figura 4 - (a) Trechos da estrutura básica de um arquivo de rede no *JUNG*, e (b) sua representação gráfica

2.2.4 Gephi

O Gephi é uma plataforma interativa que visa permitir a manipulação, exploração e compreensão de dados, através da representação dos mesmos por meio de grafos dinâmicos e/ou hierárquicos e, da aplicação de métricas de análise de redes sociais, tais como: intermediação, proximidade, coeficiente de agrupamento, *PageRank* entre outras. Por intermédio dele, o usuário pode interagir com as estruturas de dados através dos diversos *layouts* existentes. Desta maneira, através dos algoritmos de *layout* os usuários poderão alterar as configurações do grafo durante a execução do mesmo, e, portanto, aumentam consideravelmente o *feedback* do usuário frente às informações apresentadas, fazendo com que as propriedades mais ocultas passem a ser visualizadas e compreendidas.

Um dos objetivos do Gephi é auxiliar os analistas de dados de modo que, os mesmos a partir de hipóteses, descubram intuitivamente, padrões, singularidades e isolamentos estruturais das informações representadas por intermédio de um grafo.

É notório que este software poderá ser utilizado como ferramenta complementar para estudos estatísticos dentro de um determinado contexto, através da correlação do cenário analisado e, os resultados gerados pela ferramenta por meio da aplicação de métricas de análise de redes sociais.

Salienta-se que, este software foi desenvolvido visando análise de dados de maneira exploratória, identificando relacionamentos sistemáticos entre variáveis quando não existem expectativas *a priori* acerca da natureza destes relacionamentos ou estas são incompletas, então, percebe-se uma grande preocupação por parte de seus idealizadores no que diz respeito ao campo de visualização de dados e, deste modo, visando permitir uma fácil assimilação das informações geradas, pois ele tornou-se o mais interativo possível.

Dentre as suas características, destacam-se a possibilidade de geração de redes de até 50.000 nós e 1.000.000 de arestas e, a possibilidade de percorrerem dinamicamente a visualização dos dados dentre outras.

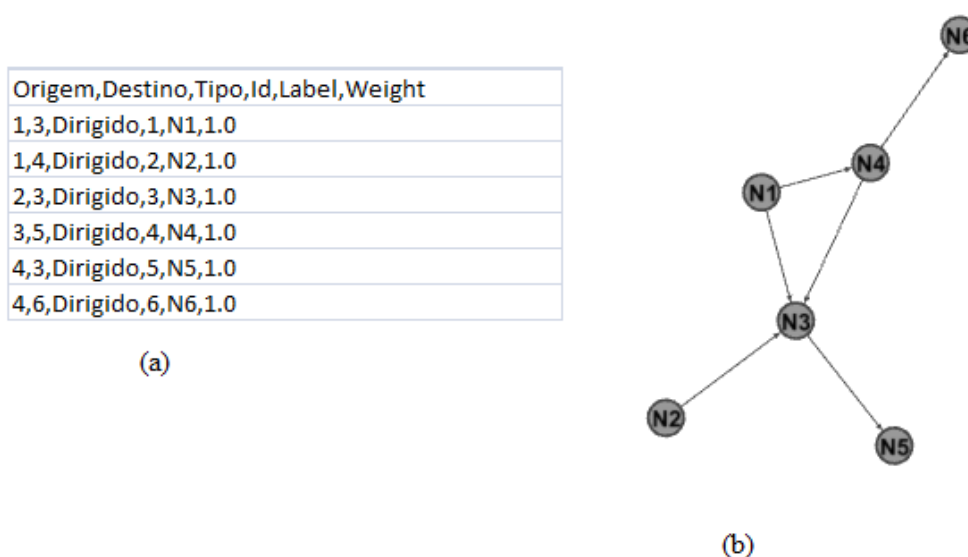


Figura 5 – (a) trecho da estrutura básica de um arquivo de rede no Gephi, e (b) sua representação gráfica

A Figura 5 (a) apresenta-se a estrutura de um arquivo CSV que permite a geração de uma rede no Gephi, deste modo, verifica-se de qual nó sai a aresta e, para qual nó ela se direciona, se o grafo é dirigido, ou não, campo este representado pelo tipo. A identificação do nó, nome do nó no grafo e por fim, a espessura das arestas. Figura 5 (b) verifica-se sua representação gráfica.

Em vista de todas estas informações, o referido software foi escolhido para ser a base para geração e análise de redes sociais do *Twitter*.

2.3. Métricas de Análise de Redes Sociais

O conceito de Análise de Redes Sociais surgiu a partir de estudos realizados pela Sociologia e Antropologia Social, cita-se como um dos marcos iniciais os experimentos realizados em 1954 pelo antropólogo J. A. Barnes, que passou a mapear no sudoeste da Noruega, as relações familiares de indivíduos pertencentes a uma vila de pescadores (Bianco, 1987, p.130). Segundo Barnes, "toda a vida social" poderia ser vista como "um conjunto de pontos", alguns dos quais são unidos por linhas para formarem uma "rede total" das relações.

Atualmente, este experimento poderá ser estendido a todo problema da vida real que possa ser mapeado e representado através de um grafo. Desta maneira, diferentes situações poderão ser representadas dentro das redes sociais como um conjunto de pontos (vértices) que representam indivíduos/grupos, ligados por arcos (arestas) que descrevem a relação/interação entre as pessoas e/ou organizações.

Conforme Wasserman e Faust (1994), a noção de rede social e métodos de análise de redes sociais têm atraído consideravelmente a comunidade das ciências sociais e comportamentais a fim de analisar os relacionamentos entre entidades sociais assim como os padrões e implicações destes relacionamentos.

A Análise de Redes Sociais estabelece um novo paradigma na pesquisa sobre a estrutura social. Para estudar como os comportamentos ou as opiniões dos indivíduos dependem das estruturas, nas quais eles se inserem nas unidades de análise, não somente os atributos individuais (classe, sexo, idade e gênero) mas também, o conjunto de relações que os indivíduos estabelecem através de suas interações uns com os outros [Marteleto, 2001].

Através da ARS, é possível compreender e acompanhar de forma mais eficaz a disseminação de informações e a interação entre as pessoas que compõem a rede. Tal acompanhamento pode auxiliar na identificação de conectores de redes isoladas, melhorar a atuação de atores críticos, identificar a fragilidade da rede em relação à comunicação dos membros dentre outras.

Portanto, na próxima seção será utilizado um cenário hipotético com o intuito de elucidar as métricas Grau de Centralidade, Centralidade de Proximidade, Centralidade de Intermediação e *PageRank*, e, na sequência, serão analisadas teoricamente as métricas Centralidade de Informação e Coeficiente de Agrupamento.

2.3.1 Cenário

O presente estudo realizar-se-á por intermédio de um cenário hipotético, sobre a aplicação de técnicas de Análise de Redes Sociais, como instrumento para analisar o comportamento dos indivíduos com base na comunicação e compartilhamento de informações. através das redes sociais online, em especial, o *Twitter*.

Entre várias ferramentas para geração, manipulação e visualização de grafos disponíveis, destaca-se o Gephi, que foi utilizado nesta pesquisa para geração e plotagem de grafos e aplicação de métricas de análise de redes sociais. Salienta-se que o mesmo foi escolhido, devido a gama de métricas normalizadas existentes na referida biblioteca, o auto desempenho no processamento de grande quantidade de dados e o fato da mesma ser de caráter livre.

Utilizando-se de uma hipótese de relacionamento envolvendo quatro pessoas, pode-se associar cada membro a um vértice, e o respectivo nome deste, dentro do grafo, como visto na Tabela 1.

Tabela 1- Representação das pessoas como vértices

USUÁRIOS	VÉRTICE	NOME DO VÉRTICE NO GRAFO
Usuário 1	1	"N1"
Usuário 2	2	"N2"
Usuário 3	3	"N3"
Usuário 4	4	"N4"

A relevância da Tabela 1, deve-se ao fato da mesma possibilitar que a ferramenta de coleta de dados, possa classificar cada usuário como um vértice do grafo e, posteriormente, coletar as mensagens enviadas por cada um.

Na sequência, é necessária a construção da tabela de interação entre os usuários, contendo informações pertinentes ao emitente das mensagens e a quantidade de *mentions* correspondendo as mensagens do remetente, para o destinatário específico da mesma linha a que ele (remetente) pertence.

A coluna correspondente à quantidade total de mensagens dissipadas na rede refere-se tanto às mensagens direcionadas quanto às não direcionadas a um destinatário específico. Inclui-se também nesta coluna, a quantidade de mensagens direcionadas (*mentions*), a quantidade de *retweets*, isto é, mensagens recebidas pelo destinatário e reenviadas a terceiros.

Tabela 2 - Interação entre os usuários

INTERAÇÃO				
REMETENTE	QTDE. DE MENTIONS	QTDE. DE MENSAGENS ENVIADAS	QTDE. DE RETWEETS	DESTINATÁRIO
Usuário 1	2	23	0	Usuário 2
Usuário 1	3	23	0	Usuário 4
Usuário 2	1	46	0	Usuário 1
Usuário 2	4	21	0	Usuário 4
Usuário 3	1	15	0	Usuário 4
Usuário 4	1	40	0	Usuário 1
Usuário 4	1	40	0	Usuário 2

Mediante os dados preliminares que resultaram na Tabela 2, realiza-se a somatória dos campos valorados (*mentions* e *retweets*) de um usuário para outro e, o resultado obtido será utilizado para elaboração da matriz de interação conforme Tabela 3 – Matriz de Interação.

Tabela 3 – Matriz de Interação

De/Para	1	2	3	4
1	0	2	0	3
2	1	0	0	4
3	0	0	0	1
4	1	1	0	0

Tendo por base os dados da Tabela 3, gerou-se o grafo constante na Figura 6, onde cada nó corresponde a um dos usuários de 1 à 4. O relacionamento entre eles é gerado através das arestas direcionais que contêm a quantidade de mensagens de um usuário para outro. Sendo assim, percebe-se na Tabela 3 que o Usuário 1, enviou duas mensagens destinadas ao Usuário 2 e, três mensagens ao Usuário 4.

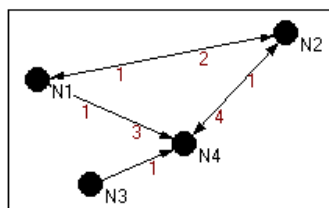


Figura 6 – Grafo de Interação

2.3.1.1 Grau de Centralidade

Em uma Rede Social, é de suma importância o conhecimento do grau de centralidade individual de um ator. Vale ressaltar que um ator nada mais é que, uma unidade discreta, ou seja, pode ser uma pessoa, um grupo ou instituições e etc.

Segundo Marteleto (2001), a centralidade de um ator significa a identificação da posição em que o mesmo encontra-se em relação às trocas e a comunicação na rede. Neste aspecto, o grau de centralidade de um nó (*Degree Centrality*), define que o ator (nó), com o maior número de enlaces (ligações), possui o maior grau de centralidade, em relação à topologia na qual se encontra. Isto ocorre, o nó com o maior número de enlaces conectados a ele, possui conexão direta a um número maior de nós dentro da topologia da rede.

Em relação a um grafo não direcional, o grau de um nó é simbolizado por $C_D(n_i) = d(n_i)$, onde $d(n_i)$, representa o número de arestas incidentes, ou ainda, de forma equivalente, o número de nós adjacentes a ele. Vale ressaltar que o grau de um nó pode variar de 0, caso este nó seja isolado, até $g-1$, onde g , é a quantidade de nós na rede, caso os mesmos estejam conectados com os demais nós do grafo.

No caso de um grafo direcionado, considera-se o grau de entrada (*In-Degree*), $d_I(n_i)$, e o grau de saída (*Out-degree*), $d_O(n_i)$, de acordo com a direção das arestas que chegam ou partem do nó.

Assim, sugestões feitas por Proctor e Loomis (1951) e Shaw (1954), além de outros pesquisadores, definem $C_D(n_i)$, como o índice referente a centralidade de um nó. A Equação 1 - Grau de centralidade define o grau de centralidade.

$$C_D(n_i) = d(n_i) = \sum_{i=1}^g x_{ij} = \sum_{j=1}^g x_{ji}$$

Equação 1 - Grau de centralidade

Na Equação 1, poderão ser utilizados os valores das linhas i de uma matriz de conectividade, ou das colunas j da mesma matriz, para o cálculo do grau de centralidade de um nó. O valor x_{ij} ou x_{ji} , indica posição de linha/coluna (ij) ou coluna/linha(ji) de uma matriz de conectividade. O valor x igual a 1, indica que existe um enlace entre os nós i e j . A soma dos valores de x_{ij} ou x_{ji} indica o valor do grau do nó. O valor g , representa o número total de nós ou de linhas/colunas da matriz de adjacências.

Segundo Donner (1986), a Equação 2, baseada nos modelos gráficos probabilísticos de Erdos e Reyni (1960), poderá ser utilizada para determinar a Centralidade de um nó,

$$C'_D(n_i) = \frac{d(n_i)}{g - 1}$$

Equação 2 – Equação Simplificada do Grau de Centralidade

Onde:

- $d(n_i)$ é o grau do nó a ser calculado
- g é o total de nós do grafo.

A partir de então passou-se a aplicar as métricas *In-Degree* e *Out-degree*

Centrality no grafo de interação referente a Figura 6 – Grafo de Interação.

A Tabela 4, apresenta a quantidade de arestas que chegam a cada vértice. Sendo assim, verifica-se que o vértice $N4$ é o que possui o maior número, totalizando 3 (três). Para melhor compreensão, verificam-se os valores normalizados de cada vértice.

Tabela 4 – Valor *In-Degree Centrality*

Vértice	<i>In-Degree</i>	<i>Normalizado</i>
N1	2	$2/(4-1) = 0,667$
N2	2	... = 0,667
N3	0	... = 0,000
N4	3	... = 1,000

Em relação à quantidade de arestas que saem de um nó, constata-se através da Tabela 5 que o vértice $N4$ é o que tem mais possibilidade de disseminar mensagens por caminhos distintos dentro da rede.

Tabela 5 – Valor *Out-Degree Centrality*

Vértice	Out-Degree	Normalizado
N1	2	$2/(4-1) = 0,667$
N2	2	$2/(4-1) = 0,667$
N3	1	$1/(4-1) = 0,333$
N4	2	$2/(4-1) = 0,667$

2.3.1.2 Centralidade de Proximidade

É conhecido que em um grafo, existe uma distância natural entre dois pares de nós, distância esta, representada pelo menor caminho entre eles.

Conforme Marteleto (2001), a centralidade de proximidade (*Closenes Centrality*) permite medir a independência de um ator em relação aos outros e, ele é tão mais central, quanto menor o caminho a percorrer para atingir os outros elos da rede.

Sendo assim, compreende-se que através desta métrica, é possível verificar o quão rápido um nó pode interagir com os demais, pois, quanto menor distância total referente à soma dos menores caminhos de um nó para com os demais, maior é sua aproximação perante todos. Logo, se um nó possuir menos nós intermediários para chegar a todos os outros, ele estará mais próximo de todos e, consequentemente mais central.

A Centralidade de Proximidade poderá ser definida através da Equação 3.

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

Equação 3 – Centralidade de proximidade

Neste caso $d(n_i, n_j)$ representa a distância geodésica – menor caminho entre dois pontos - entre n_i e n_j . A proximidade do centro da topologia será o inverso da soma das distâncias de n_i , para todos os outros nós, sendo $i \neq j$. Esta propriedade indica quais nós são mais importantes na troca de informações entre os nós da topologia. O nó com a maior Centralidade de Proximidade da topologia indica o “atalho” para os demais nós.

A normalização do índice de centralidade de proximidade, poderá ser feita multiplicando-se, $C_C(n_i)$ por $g - 1$, obtendo-se a Equação 4.

$$\begin{aligned}
 C'_C(n_i) &= \frac{g-1}{\left[\sum_{j=1}^g d(n_i, n_j)\right]} \\
 &= (g-1)C_C(n_i).
 \end{aligned}$$

Equação 4 – Centralidade de proximidade normalizada

Com as informações contidas na Figura 6, aplicou-se a métrica centralidade de proximidade, em relação aos enlaces que saem de cada vértice, obtendo como resultado de maior grau de proximidade, os vértices *N1*, *N2* e *N4*; (vide Tabela 6).

Tabela 6 – Valor da Centralidade de Proximidade

Vértice	Centralidade de Proximidade	Normalizada
N1	4	$(4-1)/4 = 3/4 = 0,750$
N2	4	$\dots = 3/4 = 0,750$
N3	5	$\dots = 3/5 = 0,600$
N4	4	$\dots = 3/3 = 7,500$

2.3.1.3 Centralidade de Intermediação

De acordo com Newman e Girvan (2004), a centralidade de intermediação $B(e)$ de uma aresta “e”, é definida como o número de caminhos mínimos entre todos os pares de nós em um grafo que passam por “e”. Que poderá ser representado da seguinte forma:

$$B(e) = \sum_{u \in V, v \in V} \frac{\sigma_e(u, v)}{\sigma(u, v)}$$

Equação 5 – Centralidade da intermediação

Onde: $\sigma(u, v)$ representa o número de caminhos mínimos entre **u** e **v**, e $\sigma_e(u, v)$ representa o número de caminhos mínimos entre **u** e **v** que incluem **e**.

Segundo Benevenuto et. al (2011), a centralidade de intermediação de uma aresta indica sua relevância para o grafo em termos de sua localização.

De modo similar, a centralidade de intermediação poderá ser aplicada para um nó ao invés de uma aresta. Neste caso, esta centralidade medirá o número de caminhos mínimos que passarem pelo nó dado. Nós que possuem muitos caminhos mínimos que

passem por eles, possuem maior intermediação indicando sua importância à estrutura da rede.

A centralidade de intermediação é o potencial daqueles que servem de intermediários. Calcula quanto um ator atua como “ponte”, facilitando o fluxo de informação em uma determinada rede [Marteleto, 2001].

Em muitos casos, um indivíduo pode não possuir um grande número de contatos, ou até mesmo não estabelecer elos muito fortes dentro da rede social, no entanto, o mesmo poderá ser de suma importância à mediação de informações entre atores não adjacentes.

Conforme Wasserman e Faust (1994), a intermediação de um ator em presença de dois outros se define por sua faculdade de situar sobre o caminho ou os caminhos geodésicos, (isto é, de menor distância), ligando estes dois pontos. De tal forma que, para ter uma grande centralidade de intermediação, um ator deverá estar entre muitos atores, através dos seus caminhos geodésicos.

Várias pesquisas sobre centralidade vêm reconhecendo a importância estratégica de atores através dos caminhos geodésicos. Bavelas (1948) e Shaw (1954), sugeriu que os atores localizados em muitos caminhos geodésicos, são de fato centrais à rede; enquanto Shimbel (1953) e Cohn e Marriott (1958), observou que tais atores centrais poderiam desempenhar papéis importantes na rede. A seguir é apresentada a fórmula para definição do grau de intermediação de um nó, onde g_{jk} é o número de caminhos geodésicos (mais curtos e do mesmo tamanho), que ligam os nós j e k , e $g_{jk}(n_i)$, o número de tais caminhos no total de g_{jk} , que passa pelo nó n_i .

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

Equação 6 – Grau de Intermediação de um nó

A normalização deste índice poderá ser feita através da divisão do mesmo pelo seu máximo possível, que é o número de pares de nós no grafo, que não incluem n_i , ou seja: $(g-1)(g-2)/2$, chegando-se ao seguinte índice de Centralidade de Intermediação normalizado por ator, apresentado na Equação 7.

$$C'_B(n_i) = \frac{C_B(n_i)}{(g-1)(g-2)/2}$$

Equação 7 – Centralidade da intermediação normalizado

Conforme o cenário hipotético em estudo percebe-se que o vértice *N4*, possui o maior índice de intermediação da rede. Tabela 7.

Tabela 7 – Valor da Centralidade de Intermediação

Vértice	Centralidade de Intermediação
N1	0,000
N2	0,000
N3	0,000
N4	0,333

2.3.1.4 Page Rank

O *Page Rank* é um algoritmo de análise de rede que dá pesos numéricos a cada elemento de uma coleção de documentos hiperligados, como as páginas da WEB, com o propósito de determinar a sua importância nesse grupo por meio de um método de busca. Sabe-se que este tipo de algoritmo pode ser aplicado a qualquer coleção de objetos com ligações recíprocas, ou não e, referências (*links*).

O mesmo leva em consideração a quantidade de ligações (*links*) que este elemento recebe em toda a rede. E assim, o mesmo busca organizar os resultados obtidos através de um ranking, onde apresenta a importância de cada elemento baseado nos valores de seus índices.

O sistema *Page Rank* é usado pelo motor de busca *Google* para ajudar a determinar a relevância ou importância de uma página. O mesmo foi desenvolvido pelos fundadores do *Google*, Larry Page e Sergey Brin enquanto cursavam a Universidade de Stanford em 1998, por isso, a patente do processo de *Page Rank* pertence a universidade e não ao *Google*.

Chen, Gan e Suel, (2002), o *Page Rank* trata-se de um sistema para dar notas às páginas da Web, no qual usa-se de uma estrutura gráfica de *links* da Web como uma ferramenta organizacional e, como um indicador do valor de uma página individual, visando.

Este algoritmo leva em consideração não só a existência do link de uma página X para uma página Y, mas também, a importância da página que citou Y, neste caso a página X. Desta maneira, os links dados por uma página importante, pesam mais e ajudam a tornar as páginas citadas (através de links) mais importantes.

Logo, a nota calculada pelo *Page Rank* é independente dos termos da consulta, pois, para cada documento gera-se um valor considerando a importância das páginas que apontam para uma outra página. Vale ressaltar, que este valor é recursivo, ou seja, a importância da página depende e, influencia a importância de outras páginas.

Por isso, o *Page Rank* é computado pesando-se cada *in-link* (elo para dentro) para uma página proporcionalmente com a qualidade da página enviando o *in-link*. A qualidade destas páginas referentes também é determinada pelo *Page Rank* (Brin & Page, 1998)

Segundo Fortunato et al. (2006), independente de sua origem, duas páginas possuem importância diferente e todos os *links* para uma página não são considerados iguais, deste modo, o *Page Rank* examina uma página junto com seus vizinhos predecessores de maior ou menor nota, representando uma mudança de medida local para uma global, em que páginas contribuem para definir a importância de outras páginas.

Sendo assim, este índice poderá ser calculado, conforme a seguinte equação:

$$PR(P_i) = (1 - d) + d * ((PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

Equação 8 – Formula do *Page Rank*

Onde:

- d é um parâmetro compreendido entre 0 e 1 (habitualmente $d = 0,85$)
- $PR(P_i)$ designa o valor do *Page Rank* da página P_i
- $PR(T_1), \dots, PR(T_n)$ são os índices das páginas que possuem um link para a página P_i
- $C(T_1), \dots, C(T_n)$ designa o número de links destas páginas

Deste modo, o valor mínimo do *Page Rank* atribuído a cada página, dado por $1 - d$, junta-se com as parcelas que resultam do modo como a página P_i está relacionada na rede. Neste aspecto, o *Page Rank* de cada página é transmitido equitativamente por todos os seus links, sendo este valor ponderado pelo parâmetro d .

Verifica-se que o *Page Rank* da página P_i depende de um cálculo iterativo, portanto a condição inicial é $r_i(P_i) = 1/n$, sendo que r_i corresponde a uma iteração, e n o número total

de páginas. Logo, baseado na Figura 6, a computação *Page Rank* para o referido grafo, começar-se-ia com $r_0(P_i) = 1/4$, uma vez que há 4 elementos rotulados por $N_1, \dots, N_4$. Após uma iteração obter-se-iam os valores seguintes:

$${}_mPR(P_i) = (1 - d) + d * ((PR(T_1)/C(T_1) + PR(T_n)/C(T_n)) =$$

$$r_1PR(N_1) = 0,15 + 0,85 * (0,25/2) = 0,256$$

$$r_1PR(N_2) = 0,15 + 0,85 * ((0,25/1) + (0,25/3)) = 0,208$$

$$r_1PR(N_3) = 0,15 + 0,85 * 0 = 0,15$$

Tabela 8 – Valores *Page Rank*

Iteração	N1	N2	N3	N4
r0	0,25	0,25	0,25	0,25
r1	0,256	0,208	0,15	0,439
r2	0,238	0,476	0,15	0,554
r4	0,352	0,508	0,15	0,665
r5	0,365	0,648	0,15	0,735
r6	0,425	0,719	0,15	0,816
r7	0,455	0,767	0,15	0,862
r8	0,475	0,797	0,15	0,892
r9	0,488	0,817	0,15	0,912
r10	0,497	0,830	0,15	0,925
r11	0,502	0,838	0,15	0,932

Vê-se assim que, após algumas iterações, o nó N_4 apresenta valor de *Page Rank* mais elevado, e o nó N_3 valor mais baixo. Percebe-se ainda que, após algumas iterações o nó N_1 que possuía o maior índice passa a ser o 3º maior.

Verifica-se através deste exemplo, que um link de uma página X para uma página Y poderá ser interpretado como um “voto” da página X à página Y. Assim, para que uma página tenha um valor elevado de *Page Rank*, não interessa apenas que receba um grande número de ligações de outras páginas, mas também que, estas páginas tenham um *Page Rank* elevado.

2.3.2 Coeficiente de Agrupamento

Segundo Watts e Strogatz (1998), Coeficiente de Agrupamento (*Clustering Coefficient*) é o grau em que os laços adjacentes de um dado nó também são adjacentes a

este mesmo nó, ou seja, a presente métrica estrutural descreve com propriedade uma densidade de conexões que abrange um agrupamento de vértices. Desta maneira, o $cc(i)$ de um nó é o resultado da equação que envolve o número de arestas existentes entre os vizinhos de i e o número máximo de arestas possíveis entre estes vizinhos.

Salienta-se que a densidade das relações apresenta a aproximação existente entre dois ou mais indivíduos. De tal modo que, a mesma permite a formação de grupos, levando a um comportamento em termos de troca de recursos e informação [Wasserman e Faust, 1994].

O Coeficiente de Agrupamento localmente é definido através da contabilização do sub-grafo do nó i , isto é, dos valores referentes ao subconjunto formado por este nó, seus primeiros vizinhos e os laços entre eles, (vide Equação 9).

$$C_i \equiv \frac{2y_i}{k_i(k_i - 1)}$$

Equação 9 – Coeficiente de agrupamento local

Onde: y_i é o número de conexões existentes entre os vizinhos mais próximos do nó i e, k_i é o número de vizinhos mais próximos deste nó, que coincidem com a conectividade do nó i . Concluindo, pode-se definir tal densidade por intermédio do resultado da equação que abrange y_i e o número de conexões possíveis no subgrafo do nó i , $k_i(k_i-1)/2$.

Salienta-se ainda, que a referida métrica mede a densidade local da rede, cujo resultado varia de 0 a 1. Destaque-se que quanto maior o valor do coeficiente de agrupamento, maior é a densidade local.

A Figura 7 apresenta o valor do coeficiente de agrupamento do nó preenchido em dois casos distintos. No primeiro (a), todos os vizinhos do nó em destaque estão ligados entre si e, conseqüentemente, o cc (*Clustering Coefficient*) do nó é 1. No outro caso (b), existem apenas duas arestas entre os vizinhos do nó dentre as três possíveis. Seguindo a Equação 9, o cálculo seria $cc = (2*2)/(3*2)$, ou seja, 0,667.

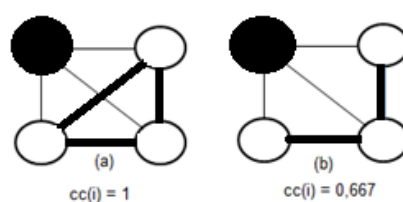


Figura 7 – Cálculo do coeficiente de agrupamento *de* um nó em dois cenários

Nota-se que o Coeficiente de Agrupamento representa uma medida da densidade de arestas estabelecidas entre os vizinhos de um nó. A seguir podemos visualizar o cálculo de todos os vértices. Tabela 9.

Tabela 9 – Valor do coeficiente de agrupamento

Vértice	Coeficiente De Agrupamento
N1	$(2*2)/(3*2) = 0,667$
N2	... = 1,000
N3	... = 0,667
N4	... = 1,000

Globalmente, o coeficiente é obtido através da contabilização total do grafo:

$$C \equiv \frac{1}{N} \sum_{i=1}^N \frac{2y_i}{k_i(k_i - 1)}$$

Equação 10 – Contabilização Global do Coeficiente de Agrupamento

Baseado na Equação 10 pode-se alcançar o coeficiente de agrupamento de uma rede, de maneira que o CC é calculado como a média dos coeficientes de agrupamento de todos os seus nós. Para isso, utilizando a Equação 10 na Figura 7 obtêm-se o resultado $CC=0,833$.

2.3.3 Centralidade de Informação

Conforme Ruhnau (2000), a Centralidade de Informação tem como princípio fundamental, a noção de que o número de relações diretas estabelecidas entre um vértice com os demais, é um aspecto importante que determina sua posição estrutural e está relacionado ao número de interações ou conexões deste vértice em uma rede. Desta forma, sua posição na rede está associada ao número de elementos que interage com ele.

Salienta-se que o conceito de informação é bastante antigo, e tem uma rica aplicação na estatística (Shannon e Weaver 1949; Khinchin 1957; Kullback 1959; Gokhale e Kullback 1978; assim como Coleman 1964; Theil 1967; e Allison 1978), para aplicações em economia e sociologia. Porém, o mesmo também vem sendo estendido de maneira a ser aplicado na teoria de estimativa e análise de dados categóricos.

Marteletto (1994), salienta que a informação “é praticada em um contexto sociocultural de produção de discursos, representações e valores que informem cada existência, fornecendo a cada sujeito um modelo de competência [...] para dirigir suas vidas, relacionar-se com outros e com a sociedade”. Neste aspecto, pode-se afirmar que a informação está marcada por duas características: o tempo para a comunicação e, sua explosão quantitativa. De tal modo que, esta ocorre graças ao advento da escrita através de vários meios, quer eletrônico ou não, e aquela na quebra da distância entre as pessoas.

Deste modo, o índice de centralidade desenvolvido por Stephenson e Zelen (1989), considera o caminho combinado de um vértice para outro tomando todos os caminhos, incluindo os geodésicos (mais curtos), e atribuindo-lhes peso. Portanto, na referida métrica descarta-se quem transmite a informação a quem, mas leva em consideração, se existe ou não, um caminho em que a informação possa fluir. Logo, seu valor é calculado pelo grau do vértice em questão, conforme a seguinte equação:

$$C_D(v_i) = d(v_i), 1 \leq i \leq n.$$

Equação 11 - Centralidade da Informação

Onde:

- $d(v_i)$ é o grau do nó a ser calculado

A Figura 8 apresenta o valor da centralidade de cada vértice do grafo. Observa-se que no referido exemplo a média de centralidade de informação na rede é de 2, por conseguinte, verifica-se a existência de um número significativo de canais possíveis para o fluxo da informação. Visto que, para cada vértice, existem apenas três vizinhos.

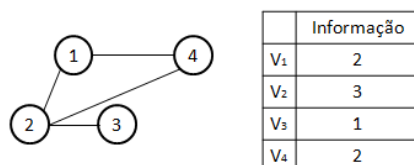


Figura 8 – Valores referentes à centralidade de informação

Verifica-se que, $C_D(v_i)$, também poderá ser obtido através da soma dos elementos i – ésima linha da matriz de adjacência A_G , que representa uma referida rede, conforme a Equação 12.

$$C_D(v_i) = \sum_{j=1}^n a_{ij}$$

Equação 12 – Cálculo da Centralidade de Informação de uma matriz de adjacência

3. Processo de Análise de Redes Sociais

A investigação proposta para análise de redes sociais, em especial, o *Twitter*, é baseada num conjunto de atividades que objetivam mapear a relação entre os participantes.

Deste modo, a presente seção visa apresentar de maneira abstrata todo o processo sistematizado adotado nesta pesquisa, a partir da realização das seguintes atividades: planejamento, monitoramento, grafo, métricas e análises constantes na Figura 9.



Figura 9 – Atividades de ARS

Os cinco estágios da demonstram as atividades fundamentais para obtenção de resultados em experimentos distintos, baseados em nossa arquitetura de coleta e análise de dados de redes sociais:

- **Planejamento:-** neste estágio são traçados os objetivos, ou seja, as indagações referentes a um determinado ponto de vista, que buscam respostas, quer estas positivas ou negativas. É nesta etapa que se definem alguns atributos, os quais serão utilizados no monitoramento, são eles: os grupos a serem pesquisados, ou seja, se os mesmos serão alunos, pessoas com restrições de crédito ou grupos aleatórios gerados pela ferramenta de coleta de dados; as palavras-chave que serão utilizadas para distinguirem as mensagens relevantes das que não são, para serem coletadas; dentre outros.
- **Monitoramento:-** Fase destinada à coleta de dados que serão a base de realização dos experimentos. Sendo assim, os atributos definidos como parâmetros no planejamento são inseridos na ferramenta de coleta de dados, de maneira que a mesma retorne um arquivo XML ou CSV contendo as informações solicitadas.

- **Grafo:-** O ponto inicial ao estudo e análise dos dados coletados no monitoramento. A partir do arquivo XML ou CSV, gerado na tarefa anterior, é construído um grafo que representa a relação e interação dos indivíduos monitorados, deste modo, pode-se obter grafos direcionados, ou não e, ao mesmo tempo contendo, ou não, os pesos de suas arestas.
- **Métricas de ARS:-** A partir do grafo, aplicam-se as métricas de análise de redes sociais, visando buscar um padrão comum, resultados individuais e gerais entre os membros da rede, através de relatórios que apresentem os índices em relação às métricas aplicadas.
- **Análise:-** Os resultados gerados são estudados minuciosamente, gerando assim, um documento contendo quais dos objetivos declarados no planejamento foram obtidos, ou não. Em caso de insucesso, inicia-se novamente este ciclo através do planejamento do novo experimento, porém, corrigindo as falhas.

3.1. Coleta de Dados

A coleta de dados para o presente trabalho foi realizada por intermédio de duas ferramentas desenvolvidas à esta finalidade: Twitter Acadêmico e AppSNA. Ambas utilizaram-se da API Twitter4J, que é uma biblioteca Java não oficial para ter acesso ao conjunto de dados e serviços oferecidos pelo *Twitter*.

A API do *Twitter*, ora utilizada, é dividida em quatro ferramentas:

- para *WebSites* - permite um conjunto de produtos para integração ao ambiente web do *Twitter*, como botões de compartilhamento, exibição e escrita da *timeline* de usuários, entre outros,
- *Search API* – torna possível consultar o conteúdo do *Twitter*, isto inclui encontrar um conjunto de *tweets* com palavras-chave específicas, que referenciam algum usuário, ou até mesmo se originou de algum usuário.
- *Stream API* - utilizada por desenvolvedores com necessidades de dados intensivos, obtidos através de mineração de dados ou análise de pesquisas entre outros. Também permite um grande número de palavras-chave para serem especificadas e rastreadas.

- *Rest API* - possui um conjunto de ferramentas que dá acesso à informações dos usuários do *Twitter*, como seguidores, seguidos, quantidade de amigos, *timeline* entre outros.

De acordo com os conceitos acima, foi eleito como ferramenta à aplicação a *Rest API*, uma vez que, a referida aplicação necessita realizar filtros nas mensagens postadas (em tempo real) no *microblogging*, de acordo com as palavras-chave de interesse a um determinado contexto

Cabe a ressalva de que todas as ferramentas de *API* descritas dependem da utilização do protocolo de autenticação do próprio *Twitter*, o *OAuth*, pois permite que os usuários compartilhem os recursos privados da aplicação acessando todos os dados disponíveis no *Twitter*.

O Twitter Acadêmico, utilizado nas pesquisas da área acadêmica e AppSNA, aplicada nos estudos de grupos específicos, tem como objetivo capturar dados do *Twitter*, possibilitando que os mesmos sejam analisados utilizando técnicas de ARS. Os passos metodológicos do sistema são apresentados na Figura 10 – . Inicialmente, o filtro de busca deverá ser configurado segundo parâmetros de acordo com o objetivo do monitoramento. Estes parâmetros dizem respeito aos indivíduos que serão monitorados, as mensagens capturadas de acordo com as palavras-chave e o intervalo de duração do monitoramento. A partir daí, a ferramenta de coleta de dados inicia o monitoramento e, todas as mensagens enquadradas nos critérios de busca serão armazenadas num Banco de Dados.

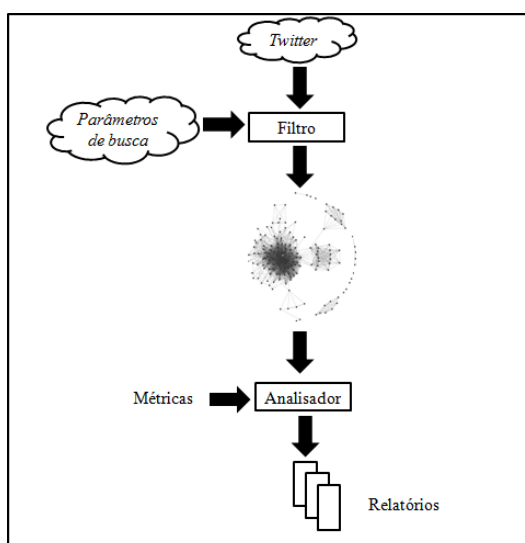


Figura 10 – Passos Metodológicos

Estas informações geraram um arquivo que posteriormente foi aplicado a uma ferramenta de ARS que possibilitou a representação e interação entre os usuários monitorados. Várias regras poderão ser seguidas na elaboração do grafo, tais como: grafos direcionados, onde cada vértice representa um indivíduo e, cada aresta uma citação de uma pessoa à outra.

O peso da aresta pode ser, por exemplo, a quantidade de vezes que um indivíduo citou o outro. Neste caso, quanto mais vezes uma pessoa A citar um colega B mais forte será a ligação de A para B. Talvez, o mesmo não ocorra na relação inversa de B para A, havendo assim, uma relação não recíproca entre ambos.

Uma vez que o grafo de interação é gerado através de uma dos softwares de visualização de grafos, ele poderá ser analisado segundo métricas de ARS. Cada métrica aplicada resultará na geração de um relatório com os resultados encontrados.

O Twitter Acadêmico, cuja interface está representada na Figura 11, exibe as mensagens monitoradas, apresentando a data, hora de suas postagens, o conteúdo e o remetente; assim como duas opções de ação: analisar e excluir. Na opção Excluir, o administrador poderá remover a mensagem da lista de mensagens a serem analisadas, caso a mesma fuja dos objetivos do monitoramento. Já na opção Analisar, é possível atribuir um conceito à mensagem de 1 à 5, de acordo com o grau de relevância da mensagem dentro do contexto analisado. Ao término, também será possível a geração de um relatório com a pontuação geral que cada um recebeu.

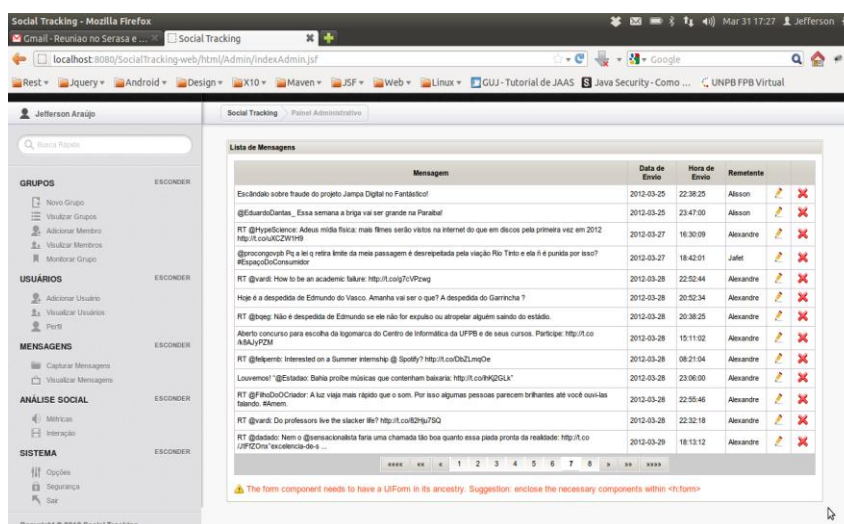


Figura 11 – Interface da ferramenta *Twitter Acadêmico*

Na construção do sistema de software AppSNA foi necessário um “*Tool Chain*” (Cadeia de Ferramentas, em tradução livre). A partir deste conceito as ferramentas poderão ser utilizadas conforme cadeia, de modo que, a saída de cada uma das ferramentas tornar-se-á entrada para a outra, mas o termo é utilizado amplamente para se referir a qualquer conjunto de ferramentas interligadas, visando o desenvolvimento de uma ferramenta de propósito único. Desta maneira, na construção da ferramenta foram utilizadas as seguintes tecnologias listadas na Tabela 10

Tabela 10 - Tecnologias utilizadas na construção da arquitetura AppSNA

Descrição	Tecnologia	Ferramenta
Camada de Persistência	Banco de Dados	MySQL
	Armazenamento	Redis
Camada de Integração	API Twitter	Twitter4J
	Métricas em Grafos	JUNG
Camada de Visualização	Visualização de Grafos	Prefuse e JUNG
Ferramentas de Autoria	Visualização e Métricas em grafos	Pajek, Gephi

A Figura 12 demonstra de modo gráfico o fluxo de dados da AppSNA, que possibilitam os usuários realizarem comunicação no *Twitter* através das mensagens, que são verificadas de acordo com o parâmetro de filtragem (palavras-chave).

API *Twitter4J* verifica existência de palavras-chave dentro de cada *tweets* dos usuários que estão sendo monitorados. Após localizar estas palavras no contexto da mensagem, as mesmas serão coletadas juntamente com as informações de seu remetente, tais como nome, biografia, localização e etc, de modo a serem armazenadas e manipuladas pela AppSNA

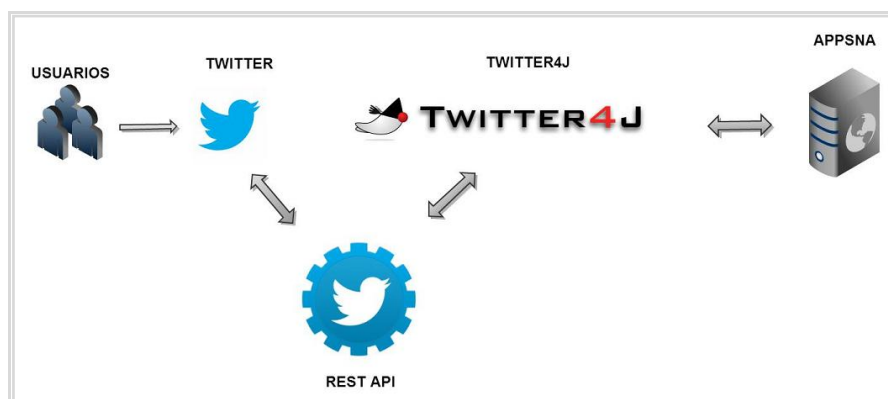


Figura 12 – Arquitetura da AppSNA

4. Trabalhos Correlatos

Esta seção tem por objetivo apresentar alguns trabalhos relacionados à Análise de Redes Sociais (ARS) online, enfatizando a existência de várias linhas de pesquisa na mesma área. Baseado nos artigos a seguir; percebe-se que a formação de redes de interação vem atingindo as mais diversas esferas e campos do conhecimento. Por conseguinte, a comunicação em rede tem sido explorada como instrumento de ativação de movimentos sociais, culturais e científicos, sendo este último, o foco desta dissertação.

4.1. Monitoramento das redes sociais como forma de relacionamento com o consumidor – o que as empresas estão fazendo?

Este estudo foi desenvolvido pelo Departamento de Gestão Contemporânea da FAPA, por [Filho, N. R. S., 2011], cujo intuito foi a realização de uma pesquisa descritiva objetiva, visando conhecer e interpretar através de questionários voltados às empresas de diversos segmentos e análise bibliográfica, como as empresas comportam-se mediante nova realidade de interação virtual e, se estão monitorando as redes sociais, bem como quais as formas de interação das organizações nas mesmas.

Conclui-se que, através do estudo bibliográfico realizado foi possível analisar como as organizações lidam com estes espaços virtuais, de forma a atingir seus objetivos e estratégias, com o objetivo de tornarem-se sempre competitivas.

Segundo Filho (2011), o principal resultado desta pesquisa foi à descoberta das organizações estarem monitorando as redes sociais e, através de informações obtidas nelas, promoverem alterações nas suas estratégias, porém, a grande maioria das empresas apenas monitora, sem saber ao certo como proceder com este novo agente em suas estratégias competitivas.

4.2. Detecção de conteúdo relevante e usuários influentes no *Twitter*

Trabalho este foi desenvolvido pelo Departamento de Ciência da Computação da UFMG, por [Valiati et al, 2012], cujo plano, foi a aplicação de uma técnica intuitiva e

circular; de relevância e influência, para determinar tanto os usuários influentes, quanto os conteúdos importantes dissipados por um grupo de usuários do *Twitter*.

De acordo com o artigo, existem alguns pontos desafiadores a fim de serem levados em consideração, no intuito de atingir o objetivo principal do referido trabalho que são:

- à rede de disseminação de informação é pouco observável, dificultando a detecção exata do processo de disseminação e também da importância do conteúdo e influência dos usuários;
- quando é possível observar esta difusão e identificar os principais fatores que a explicam;
- o comportamento dos usuários muda ao longo do tempo, e, mais uma vez de forma pouco previsível e observável.

A técnica intuitiva e circular utilizada para ordenar usuários e conteúdos de acordo com sua influência e prestígio, foi modelada como uma extensão do problema de *PageRank* [Brin and Page, 1998], e implementa eficientemente, uma vez que não requer a obtenção de rede de seguidores e, determina as ordenações através de uma estratégia randomizada. Posteriormente, esta técnica foi avaliada em dois cenários amplamente discutidos na rede social *Twitter*: Política e automóveis.

Sendo assim, utilizando-se de equações matemáticas, os usuários influentes e conteúdos relevantes são identificados através de uma associação circular entre ambos e, como resultado, foi constatado que tal técnica apresenta ganhos de 37% quando comparada a um método de filtragem colaborativa, ao mesmo tempo que, tanto usuários influentes, quanto conteúdo relevante se mostram superiores qualitativamente.

4.3. Mineração de dados e análise de opinião em redes sociais – Um estudo de caso sobre as Olimpíadas 2012 utilizando o *Twitter*

Tal projeto vem sendo desenvolvido no Departamento de Sistemas de Informação da UFV do Campus Rio Paranaíba, por [Oliveira, Campos e Silva, 2012], sendo sua proposta fazer uso dos dados produzidos no *microblog Twitter* para obter informações relacionadas às olimpíadas 2012.

O trabalho em si, apresenta as possibilidades de utilizarem o *Twitter*, como meio de classificação de usuários dentro de grupos de interesses, análise de opiniões e aceitação de diversos assuntos.

Outro aspecto importante apresentado pelos autores é a metodologia de desenvolvimento, que esta dividida em 5 etapas:

- Coleta de dados;
- Pré-processamento;
- Mineração;
- Pós-processamento;
- Informações;

Através destas atividades, este projeto busca responder as seguintes perguntas:

- Qual atleta atraiu maior atenção do público em determinado esporte?
- Qual esporte foi mais comentado na rede social?
- Existem relações entre as pessoas que comentam sobre dois esportes distintos?

Salienta-se que a coleta de dados iniciar-se-á no dia 14 de julho, duas semanas antes do início das Olimpíadas, e terminará em 19 de agosto, uma semana após o término da mesma. Por conseguinte, o referente artigo ainda não possui resultados concretos.

4.4. PANAS-t: Uma escala psicométrica para medição de sentimentos no *Twitter*

Trabalho este, desenvolvido no Departamento de Computação da UFOP, por [Gonçalves, Dores e Benevenuto, 2012], nele foi adaptada uma escala psicométrica conhecida como PANAS-x (*Positive Affect Negative e Affect Scale*), comumente aplicada em forma de questionário, para medir os sentimentos dos usuários do *Twitter*, sobre uma série de eventos, tema social, político e esportivo.

O PANAX-t consiste em sentimentos positivos e negativos, sendo adequado para medição de qualquer evento aleatório no *Twitter*. Para tal procedimento, utilizou-se dados empíricos a partir de um conjunto único de dados coletados contendo cerca de 1,8 milhões de *tweets*. Dados estes que foram utilizados para computar valores de normalização para cada sentimento, de modo que o crescimento ou decrescimento de cada um poderia ser

medido na base de dados. Logo após, a validação da metodologia era baseada em uma análise sobre variações no humor dos usuários, em relação a eventos que abrangessem alguma tragédia, política, saúde bem como eventos esportivos.

O processo de medição de sentimentos com o PANAS-x é simples, a princípio todos os usuários recebem um questionário contendo adjetivos que descrevem cada sentimento. Cada palavra deste questionário, é então classificada ao seu sentimento correspondente em uma escala de 1 à 5; de acordo com o sentimento do entrevistado. O PANAS-x mede 11 afetos específicos, tais como: medo, tristeza e etc.

Já o PANAS-t, que é uma variação proposta pelos autores, utiliza-se de sete eventos que foram amplamente populares no *Twitter*, eles abrangem tópicos como: tragédias, lançamentos e etc. Para extrair apenas *tweets* relevantes de cada evento, utilizou-se palavras-chave, que os descrevem em consultas a websites, blogs e etc. Como forma de medir as variações de sentimentos, o PANAS-t usou também da representação *Kiviat* para geração de grafos.

O primeiro tópico para análise, foi a partir das declarações da Organização Mundial da Saúde referente a gripe suína. Deste ponto, iniciou-se uma coleta de dados a fim de definir o sentimento correspondente a tal evento. Como resultado deste estudo de caso, constatou-se que após referidas declarações, o estado de humor dos usuários modificou-se, aumentando os sentimentos de atenção e medo nos dias posteriores ao anúncio.

É notório que outros estudos de casos foram aplicados, tais como; à queda de um avião em 1º de julho, as eleições americanas entre outros. Destaca-se ainda, a comparação através de testes do PANAS-t com o tradicional PANAS-x.

4.5. Sigam-me os bons! Transformando robôs em pessoas influentes no *Twitter* (influência)

Este estudo foi desenvolvido no Departamento de Computação da UFOP, por [Messias et al, 2012], o qual apresenta alguns sistemas que classificam usuários conforme sua influência nas redes sociais, em especial, o *Twitter*. Consideramos esta mediação algo complexo e subjetivo, os mesmos buscaram caminhos para demonstrarem a vulnerabilidade e facilidade de manipulação em tais sistemas.

Então, realizaram-se experimentos e análises em dois sistemas de classificação de influência: Klout e Twitalyzer. Para tanto, foram desenvolvidos simples algoritmos robôs usando API *Python* do *Twitter* capazes de interagirem através de contas no *Twitter*, como se fossem usuários da rede, trocando informações, seguindo e conquistando novos seguidores, por um período de 90 dias.

Para seguir os usuários, os autores abordaram um método de busca em largura, ou seja, a partir de um usuário inicial aleatório, o algoritmo o inclui em uma lista, seguindo-o e selecionando aleatoriamente 30 usuários de sua lista de seguidores, que também serão inclusos na lista.

Para interagir através de troca de mensagens, uma sequência de algoritmos lê um dicionário de palavras relacionadas ao tema, em seguida, capta quatro dos *tweets* retornados na pesquisa e os posta em um intervalo de tempo aleatório.

Os resultados das análises obtidas, mostram que as ferramentas Klout e Twitalyzer, não apresentam métricas ideais para classificarem a influência de um indivíduo na rede. Fato este, comprovado através de *spammers* espalhados na rede que, puderam facilmente passarem-se por celebridades e pessoas populares apenas utilizando contas automatizadas.

4.6. Implementing semantic queries in online social networks

Este projeto foi desenvolvido no Instituto de Informática da UFGO, por [Alarcón, Xavier e Carvalho, 2011], visando a implementação de uma estrutura que se utiliza da Web Semântica para fazer relacionamentos em redes, e assim, gerar consultas semânticas em redes sociais online.

Deste modo, foi desenvolvido um *framework*, denominado Airetama, o qual baseia-se em sistemas multi-agente ao desenvolvimento de redes sociais online com os princípios da Web Semântica, com o objetivo de gerar consultas mais expressivas em relação as redes sociais convencionais.

Salienta-se que, a Web Semântica é uma extensão da Web atual, que permite interação “homem-máquina” (pessoas e computadores), pois, a mesma interliga significados de palavras e, neste contexto, tem como finalidade conseguir atribuir um sentido aos conteúdos publicados na *Internet*, de maneira que seja perceptível tanto pelo usuário, quanto pelo computador.

Portanto, a *Web Semântica* não se trata de uma nova rede de informações/relacionamentos, mas sim de um projeto para aplicar conceitos inteligentes na internet atual, ou seja, uma interação inteligente do usuário com os conteúdos organizados de maneira eficiente na rede.

Baseado neste contexto o *framework* Airetama é composto por três componentes:

- Portal:- ponto de entrada para solicitações de serviços feitos pelos usuários através de um navegador *Web* a um servidor *Web*. Cada solicitação do usuário é encaminhada ao servidor *Web* que a redireciona ao sistema multi-agente, onde são processados.
- Sistema Multi-agente: - nesta camada temos o processamento das solicitações dos usuários pelos agentes. Sobressai-se que, o uso de agentes permite que uma solicitação do usuário, possa permanecer funcionando de forma autônoma, mesmo se ele, usuário, estiver inativo.
- Repositório:- as ferramentas ligadas ao *framework* deverão ser tão independentes quanto possível, pois, as mesmas compartilharão o mesmo banco de dados. E, o repositório semântico é o local onde os dados são armazenados.

A base de dados do Airetama é coletada da rede social online *Facebook*. Vale ressaltar que, o *Facebook* é um site e serviço de rede social que, permite criar um perfil pessoal, adicionar outros usuários como amigos e trocar mensagens; incluindo notificações automáticas quando der atualização dos perfis. Além disso, os usuários poderão participar de grupos de interesse comum de outros utilizadores.

Logo, o *framework* Airetama também chamado pelos idealizadores de “*Facebook Intepretador*”, busca o perfil e os interesses do usuário conectado e, o perfil e interesses dos seus amigos através da API do *Facebook*. De imediato, a ferramenta gera arquivos RDF (*Resource Description Framework*) contendo estes dados coletados que serão então salvos no repositório. Salienta-se que, RDF nada mais é que, uma linguagem para representar informações na Internet, quer dizer, modelos ou fonte de dados conhecidos como *metadata*, tendo como principais objetivos criar um modelo simples de dados, com uma semântica formal.

A partir dos arquivos RDF, iniciou-se os experimentos através da utilização da linguagem SPARQL para consulta à banco de dados. Destaca-se que estas consultas visam os interesses comuns de todos os usuários que foram importados do *Facebook*. Dentre os resultados obtidos, foi possível identificar que nos casos em que os usuários possuíam mais de 50 interesses comuns, muitos deles não eram amigos uns dos outros, mas tinham um amigo em comum.

Como contribuições deste trabalho, coloca-se em evidencia a capacidade de realizar consultas semânticas em dados pré-existentes de usuário da rede social e, conversão de dados de usuários do *Facebook* em arquivos RDF.

4.7. On the predictability of elections using social media data

Esta pesquisa foi desenvolvido no Departamento de Ciência da Computação do *Wellesley College*, por [Metaxas, 2011], com o intuito de inquirir sobre a possibilidade da utilização do *Twitter* e outras redes sociais online, como mecanismo para prever consistentemente os resultados das eleições do Congresso dos EUA.

Segundo Metaxas, a capacidade de se fazer previsões com base de dados disponíveis publicamente, pode levar a inúmeros benefícios dentre diversas áreas, tais como; na saúde através das previsões de epidemias, na economia, através das previsões sobre as tendências do mercado de ações, de habitação entre outras áreas.

Mediante a proposta deste trabalho, a referida pesquisa apresentou em grau de porcentagem, os seguintes resultados obtidos pela análise feita em relação aos conteúdos postados na rede social online *Facebook* logo após as eleições: Previsões em relação aos vencedores para a Câmara atingiu a margem de 74% de acertos; no caso das previsões relacionadas aos candidatos ao Senado, a precisão de acertos chegou a 81%.

Quanto ao *Twitter*, Metaxas (2011) levou em consideração um estudo diferenciado realizado por O' Connor et AL (2010), no qual analisaram a maneira em que métodos de análise simples de sentimento, poderiam ser aplicados a *tweets* como uma ferramenta automática para determinar opiniões. Tais pesquisadores correlacionaram os resultados da ferramenta com a evolução temporal de diferentes índices de sentimento dos eleitores, tais como, o índice de aprovação do trabalho presidencial e vários índices pré eleitorais. Segundo os resultados desta pesquisa, os índices obtidos e a correlação com os sentimentos

não foram significativos às pesquisas pré eleitorais, no entanto, eles apontam que, a análise de dados do *Twitter* ainda é um campo promissor às pesquisas com estas finalidades.

Partindo deste pressuposto, Metaxas (2011) utilizou-se de dois conjuntos de dados relacionados às eleições. Logo, as previsões foram calculadas com base no volume de *tweets* e *retweets*, e, em seguida, com base em análise de sentimento destas mensagens. Previsões estas que foram comparadas aos resultados reais das eleições a Senador, e, o total de acerto referente aos candidatos ganhadores, foi mais de 50%.

Verifica-se que, as pesquisas acadêmicas nesta área e desta maneira, tendem a apoiar um parecer positivo sobre o poder preditivo das mídias sociais, como uma promissora linha de pesquisa, embora ela não seja ainda tão aceitável como as pesquisas e relatórios feitos presencialmente.

4.8. Explorando redes sociais online: da coleta e análise de grandes bases de dados às aplicações.

Este trabalho foi desenvolvido por (Benevenuto, Almeida e Silva, 2011), com a finalidade de apresentar a origem das redes sociais até sua expansão no campo virtual.

O mesmo oferece uma introdução aos pesquisadores desta área, através da apresentação das principais características e mecanismos de interação, entre os usuários das redes sociais virtuais mais populares: como; *Twitter*, *Facebook* e outros. Por intermédio destas características torna-se possível compreender as definições referentes à algumas funcionalidades/características das redes sociais online, assim como, perfis dos usuários, atualizações de conteúdos, compartilhamento e avaliação de conteúdo/comentários, lista de favoritos, lista mais populares e metadados.

Na sequência, são expostas as principais métricas e análise para estudo de grafos que formem a topologia das redes sociais. Nas quais encontram-se, *Betweenness*, Coeficiente de Agrupamento, *PageRank* etc.

Por fim, realiza-se uma sumarização das principais abordagens para coleta e tratamento de dados de redes sociais virtuais e, o uso das principais métricas de ARS.

4.9. Análise de Redes Sociais – aplicações nos estudos de transferência da informação

Esta pesquisa foi desenvolvida por (Marteleto, 2001), onde reflete à aplicação da metodologia de análise de redes sociais nos estudos do fluxo e transferência da informação, junto aos movimentos sociais.

Por conseguinte, o respectivo trabalho procurou perceber como as informações circulam e são empregadas nas lutas por melhores condições de vida dentro de uma rede social. Tal forma que, realizou-se uma pesquisa junto a uma comunidade no subúrbio do Rio de Janeiro.

A partir dos elos entre os diversos movimentos sociais, aplicou-se algumas métricas de ARS, combinadas com instrumentos relacionados à avaliação qualitativa. A seguir, concluiu-se que, tal como apontou a análise de redes sociais, algumas personagens são responsáveis pela dinamização das redes, ocupando posições estratégicas que lhe garantam inclusive, o reconhecimento dos demais agentes. Além disso, percebeu-se que, no interior destas redes, sempre existirá uma constante tensão entre a diversidade de pensamentos que ela comporta, formando assim uma “rede de conhecimentos” que, alimentam e redefinem constantemente as ações de seus membros.

5. Estudos de casos

Nesta seção são apresentados quatro casos de análise de redes sociais, sendo que, três abordam uma investigação do perfil na área acadêmica e um, no âmbito de crédito e fraudes dos usuários do *Twitter*.

5.1. Estudos de casos na área de educação

5.1.1 Analysis of Students Interaction Profile in Online Social Networks

O referido estudo é fruto de uma pesquisa publicada no congresso internacional *IADIS*, deste modo o mesmo apresenta uma análise de uma rede social virtual acadêmica a partir de um cenário hipotético, onde foram aplicadas as seguintes métricas de análise de redes sociais: grau de centralidade e centralidade de intermediação. (OLIVEIRA, R. F., DUARTE, A. N. and BRITTO, A. V., 2011).

Para tal, simulou-se uma possível interação entre alunos por intermédio do *Twitter*, visando comparar a participação e colaboração de cada um deles, com intuito de apresentar através desta pesquisa, meios para que em cenários reais os professores possam de fato, aplicar os mesmos mecanismos de análise e, assim, identificar o perfil colaborativo de cada aluno através da comparação de sua participação e interação nas redes sociais presenciais (sala de aula) e virtuais (*Twitter*).

Deste modo, criou-se um cenário hipotético com 25 alunos de Engenharia de Software, que criaram seus perfis no *Twitter*, sob a orientação do professor da disciplina, com o intuito de, compartilhar as informações relacionadas com o conteúdo ministrado presencialmente. Salienta-se também que estas informações disponibilizadas na rede passariam a ser cobradas nas avaliações da aprendizagem.

Posteriormente, referidos alunos foram mapeados e identificados individualmente como um nó (numerados de 1 à 25), a partir daí, passaram a ser representados em um grafo com seus devidos relacionamentos de interação, de um aluno para com o outro, conforme Figura 13.

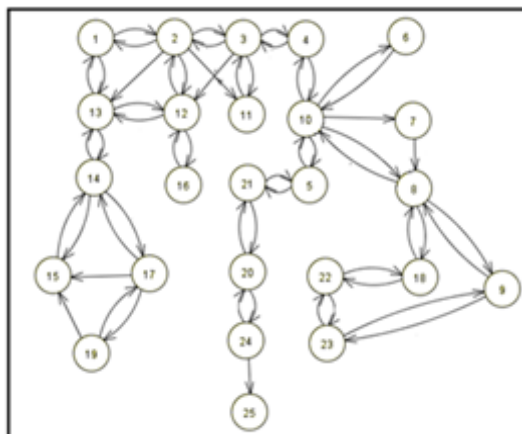


Figura 13 – Grafo de Interação entre os alunos

Vale ressaltar que, na construção deste grafo estrutural com seus respectivos fluxos unidirecionais e bidirecionais (relacionamento de interação), foram levados em consideração as pessoas que são seguidas, ou que seguem um respectivo aluno no *Twitter*, conforme apresentado na Tabela 11. Deste modo, todos os acadêmicos foram mapeados, a fim de analisar o perfil de cada um na rede.

Tabela 11 - Valores *In-Degree* e *OutDegree*

NÓ	IN-DEGREE CENTRALITY	OUT-DEGREE CENTRALITY
1	0,083	0,083
2	0,125	0,167
3	0,083	0,167
4	0,083	0,083
5	0,042	0,083
6	0,042	0,042
7	0,042	0,042
8	0,167	0,125
9	0,083	0,083
10	0,167	0,208
11	0,083	0,042
12	0,167	0,125
13	0,167	0,125
14	0,125	0,125
15	0,083	0,042
16	0,042	0,042
17	0,083	0,125
18	0,083	0,083
19	0,042	0,083
20	0,083	0,083
21	0,083	0,083
22	0,083	0,083
23	0,083	0,083
24	0,042	0,083
25	0,042	0,000

A partir da métrica *In-Degree Centrality* constatou-se que, os alunos representados pelos nós 8, 10, 12 e 13 do grafo possuem maior grau de centralidade local, conforme a Tabela 11. Sendo assim, todos são peças chaves dentro da rede, pois, são estes os nós que mais recebem informações de diferentes pessoas. A fim de confirmar este fato, foi realizado o seguinte experimento: eliminou-se o nó 8 e, automaticamente a rede *monomodal* deixou de existir, dando lugar a uma rede *duomodal* conforme visto na Figura 14.

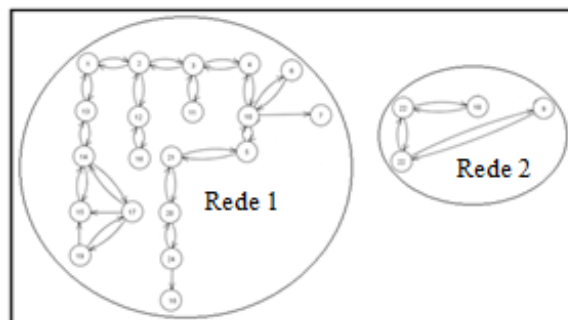


Figura 14 - Exclusão de um nó central da Rede: Formação de uma rede duomodal

Diante destas descobertas, o professor, ao analisar a rede monomodal deverá dar atenção especial aos nós mais receptores e transmissores de informações, pois são eles o centro do fluxo de informações e motivação da rede, acarretando uma dependência da rede em relação aos mesmos.

Segundo a métrica *out-degree centrality*, observou-se que, o nó 10 é o mais significativo, pois, através dele, um número maior de pessoas recebem determinada informação postada na rede, ver Tabela 11. A exclusão deste indivíduo conforme apresentado na Figura 15, gera um grande particionamento da rede, levando ao isolamento de alguns membros.

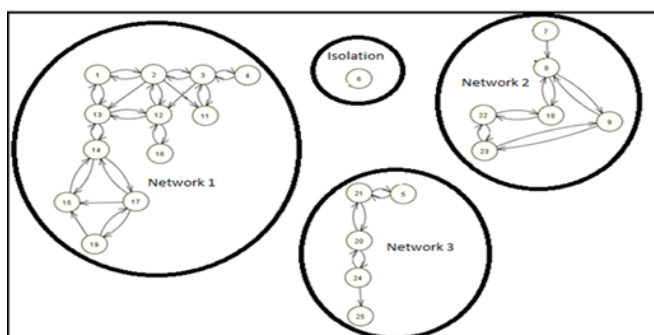


Figura 15 – Desestruturação da rede: Exclusão de um nó de alto grau de centralidade

Baseados nestas duas métricas, dois aspectos poderão ser analisados pelos professores: os alunos com o maior e, menor grau de centralidade local. No primeiro caso, nota-se, neste experimento, que os alunos mais comunicativos em sala são

também os que mais produzem e/ou consomem informações na rede. Este fato foi constatado através da observação da participação, interação, troca de informação e comunicabilidade entre os alunos em sala de aula e, a quantidade de mensagens que eles dissipavam na rede social virtual.. No segundo aspecto, pode-se abordar os casos dos alunos apenas receptivos, como exemplo, o nó 25 representado na Figura 13. Neste item o docente poderá analisar se existe alguma semelhança deste isolamento virtual com o cenário presencial em sala, observando seu relacionamento frente aos demais colegas nos dois âmbitos e, em caso afirmativo, encontrar caminhos que possam integrá-los cada vez mais, por exemplo, utilizando-se de dinâmicas de grupos dentre outros mecanismos.

Aplicando a métrica centralidade de intermediação, analisa-se o grau de intermediação de um nó, entre outros dois pontos da rede. Para exemplificar, calculou-se o valor da centralidade de intermediação do nó 13, para os nós 1 e 12 tendo como resultado, o valor normalizado de 0,667.

Ao fazer uso desta métrica, o professor poderá avaliar o grau de aproximação e/ou intermediação de determinados alunos. Se o resultado não for satisfatório em relação à aproximação, o docente poderá buscar meios para reverter esta situação, a fim de, propiciar um maior relacionamento entre os alunos.

5.1.2 Estratégia Para Avaliação do Grau de Participação e Afinidade de Alunos Através de *Microbloggin*

O referente estudo expõe a utilização do *Microblogging* mais popular, denominado *Twitter* a fim de definir o Grau de Afinidade entre alunos de uma mesma classe, assim como seus conhecimentos individuais relacionados a determinadas matérias.

Desta forma, o referido experimento baseia-se no monitoramento dos acadêmicos do curso de Sistemas de Informação da Universidade Estadual de Goiás, em especial, os matriculados no 3º ano na disciplina de Engenharia de Software, totalizando 14 alunos.

O processo teve início com o monitoramento das mensagens enviadas por este grupo de alunos através do *Twitter*. De tal forma a ferramenta específica denominada *Twitter Acadêmico* foi desenvolvida, bem como, uma *hashtag* exclusiva para obter, acompanhar e avaliar apenas os dados necessários ao experimento. A utilização da

hashtag #esuegposse passou a ser condição essencial aos alunos para postarem suas mensagens, a fim de que, o monitoramento e a coleta dos dados pudessem ser realizados com eficácia.

A *hashtag* criada foi divulgada aos alunos que, passaram a utilizá-la juntamente com suas postagens no *Twitter*. Deste princípio, deu-se início o monitoramento através do *Twitter Acadêmico*, gerando assim os dados necessários para aplicação das fórmulas matemáticas, referentes ao Grau de Afinidade entre os alunos e, suas respectivas habilidades em determinados assuntos de Engenharia de Software.

Para obter a quantização da participação dos acadêmicos, foi necessário a utilização de duas regras de contagens, baseadas nas menções e *retweets* efetuados por cada aluno sobre um determinado assunto:

O aluno P_i , que fizer uma menção sobre um determinado assunto que fora proposto pelo professor, receberá automaticamente através da ferramenta *Twitter Acadêmico*, um incremento em sua pontuação sobre aquele tema.

O aluno P_j , que reenviar a mensagem do aluno P_i , incrementará uma unidade em sua pontuação e na pontuação do aluno P_i .

Tabela 12 – Grau de Participação dos acadêmicos sobre os assuntos: “Requisitos Funcionais” e “Não funcionais”

ID	REQUISITOS FUNCIONAIS E NÃO FUNCIONAIS
P_1	14
P_2	20
P_3	0
P_4	1
P_5	20
P_6	5
P_7	8
P_8	20
P_9	20
P_{10}	20
P_{11}	12
P_{12}	20
P_{13}	20
P_{14}	20

O resultado final é a soma destas duas regras. Presume-se que, podemos então verificar através da Tabela 12 o grau de participação dos acadêmicos sobre os assuntos: “Requisitos Funcionais” e “Não funcionais”; calculados através das informações extraídas da ferramenta *Twitter Acadêmico*.

Já, com o grau de participação dos alunos aferidos, percebeu-se a necessidade de determinar uma relação em porcentagem destes valores, sobre todo o conhecimento do assunto dispersado em nossa Rede Social Virtual. Então foi realizada uma normalização de 0 à 1 para os valores, dividindo o grau de habilidade do aluno P_i , pela soma total dos Graus de Participação do grupo analisado.

Tabela 13 – Grau de Participação Normalizado Referente ao tema Requisitos Funcionais e Não funcionais

ID	REQUISITOS FUNCIONAIS E NÃO FUNCIONAIS
P ₁	0,07
P ₂	0,1
P ₃	0
P ₄	0,005
P ₅	0,1
P ₆	0,025
P ₇	0,04
P ₈	0,1
P ₉	0,1
P ₁₀	0,1
P ₁₁	0,06
P ₁₂	0,1
P ₁₃	0,1
P ₁₄	0,1

A equação resultante dos cálculos apresentados é dada por:

$$Gh_{ij} = (Qt_{ij} + Qrt_{ij} \text{ (efetuados)} + Qrt_{ij} \text{ (recebidos)}) / N_j$$

Onde Gh_{ij} , é o Grau de Participação da pessoa “i”, para o requisito “j”. Qt_{ij} , é a quantidade de mensagens que a pessoa “i”, realizou sobre o requisito “j”. Qrt_{ij} , é a quantidade de *retweets* que a pessoa “i” efetuou, e também recebeu sobre o requisito “j”. Por fim, temos o N_j , o total de todas as menções e *retweets* efetuados e recebidos pelos alunos sobre o requisito “j”. O resultado do cálculo do Grau de Participação é mostrado na Tabela 13.

Posteriormente, foi desenvolvida uma métrica para identificação do Grau de Afinidade de um aluno “i”, para um aluno “j”. No entanto, foi realizada uma pesquisa observando o relacionamento dos acadêmicos dentro da Rede Social (*Twitter*), baseada no fluxo da troca de informações na mesma.

Verificou-se também que, a afinidade pode não ser recíproca, ou seja, a afinidade do aluno “i” para o aluno “j”, pode não ser a mesma do aluno “j” para o aluno “i”. Esta última, em alguns casos até não existiu.

A Tabela 14 apresenta a quantidade total de mensagens postadas por cada aluno e, quantas vezes cada um mencionou, ou reenviou (*retwitte*) mensagem de outro aluno. Tais dados são a base para o cálculo do peso da afinidade entre eles.

Tabela 14 – Total de Mensagens de P_i e total de mensagens diretas de P_i para P_j

ID	QTDE DE MENSAGENS	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁	P ₁₂	P ₁₃	P ₁₄
P ₁	14	-	0	0	0	0	0	0	0	0	0	0	0	0	0
P ₂	20	0	-	0	0	0	0	0	0	0	1	0	0	0	0
P ₃	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
P ₄	1	0	0	0	-	0	0	0	0	0	0	0	0	0	0
P ₅	20	0	0	0	0	-	0	0	7	0	0	0	0	0	0
P ₆	5	0	0	0	0	0	-	0	0	0	0	0	0	0	0
P ₇	8	0	1	0	0	0	0	-	0	0	0	0	0	0	0
P ₈	20	0	0	0	0	0	0	0	-	0	0	0	0	0	0
P ₉	20	0	0	0	0	0	0	0	0	-	0	0	0	0	0
P ₁₀	20	0	0	0	0	0	0	0	0	0	-	0	0	0	0
P ₁₁	12	0	0	0	0	0	0	0	0	0	0	-	0	0	0
P ₁₂	20	0	0	0	0	0	0	0	0	0	0	0	-	0	4
P ₁₃	20	0	0	0	0	0	0	0	0	0	0	0	0	-	1
P ₁₄	20	0	1	0	0	0	0	0	0	0	1	0	0	0	-

O cálculo da afinidade é baseado na seguinte fórmula:

$$W_{ij} = \frac{(MD_{ij} + RT_{ij})}{M_i}$$

Onde:

W_{ij} é o peso do nó i , para o nó j ;

MD_{ij} é a quantidade de menções diretas enviadas de i , para j ;

RT_{ij} é a quantidade de *Retweet* feitos por i , dos *Tweets* (menção) de j ;

M_i é a quantidade de mensagens enviadas por i , na rede.

Após a aplicação dos dados coletados na fórmula do peso, foi possível a representação dos valores obtidos através do grafo direcionado. Consequentemente, no estudo ora analisado notou-se a existência de afinidades unidirecionais, conforme apresentado na Figura 16.

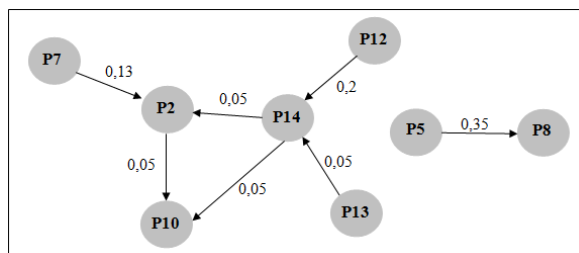


Figura 16 – Grafo de afinidade entre os alunos pesquisados.

Com base na Figura 16 – Grafo de afinidade entre os alunos pesquisados., é verídico, que através da utilização de uma rede social virtual, nem sempre os que mais colaboram, têm um olhar de prestígio pelos demais membros. Fato este, confirmado através dos alunos P₅ e P₈ que, mesmo fazendo parte dos acadêmicos mais participativos, continuaram isolados dos demais. Deste jeito, vislumbramos a necessidade de intervenção por parte do professor, através de mecanismos que possam integrá-los com o restante do grupo.

Conforme a Tabela 14, é possível notar que os alunos P₆, P₈, P₉, P₁₀ e P₁₁, embora sejam grandes colaboradores da rede, jamais mencionaram os demais colegas. Logo, deve-se analisar se estes alunos trazem consigo individualismo, ou seja, esperam que todos os assuntos na Rede Social girem em torno de suas mensagens postadas.

Outro fator a ser investigado são os conteúdos postados, pois às vezes os mesmos podem não ser tão relevantes ou, serem bastante técnicos para o grupo. Por outro lado, o aluno colaborador jamais será mencionado. Nesta situação, podemos citar o P₅, P₆, P₇, P₉, P₁₁, P₁₂ e P₁₃ da Tabela 14.

Desta forma, percebe-se que a interação entre os alunos através do *Twitter* não ocorre de maneira recíproca, existindo assim, interrupções na disseminação das informações na rede social virtual.

5.1.3 Monitoramento das interações dos aprendizes na rede social *Twitter* como apoio ao processo de mediação docente.

Neste estudo de caso, aplicaram-se as seguintes métricas de análise de redes sociais: grau de centralidade, centralidade de proximidade e centralidade de intermediação, dentro do cenário do estudo do caso anterior (Estratégia Para Avaliação do Grau de Participação e Afinidade de Alunos Através de *Microblogging*), com a finalidade de definir o perfil dos alunos comparando a atuação deles em dois âmbitos:

rede social presencial (sala de aula) e virtual (*Twitter*). (OLIVEIRA, R. F., ARAÚJO, J., PRETÔNIO, F. e BRITTO, A. V., 2012).

Os resultados obtidos foram comparados pelo professor, com a conduta presencial dos alunos, ajudando a definir o perfil de cada um de seus discentes, em relação à interatividade e colaboração e possibilitando a realização de um trabalho que estimule a participação em grupo.

O primeiro passo na modelagem de interação, é a construção da matriz relacional $M \times N$ que, irá representar os laços de comunicação entre os acadêmicos.

De acordo com a Tabela 15¹, verifica-se que, existe ligação de um aluno para outro, somente quando $a[i,j] = 1$. Sendo assim:

$a[i,j] = 1$ se e somente se, existir uma menção de a_i para a_j .

$a[i,j] = 0$ se e somente se, não existir nenhuma menção de a_i para a_j .

Tabela 15 – Menções de a_i para a_j .

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	1	0	0	0	0	1	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	1	0	0	0	0	1	0	0	0
6	0	0	0	0	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0	1	0	0	0
10	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11	0	0	0	0	1	1	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	1	0	0	0	0	0	1	0	0	0

Após a construção da matriz de interação $M \times N$, foram representados os referidos dados através de um grafo direcionado, conforme a Figura 17, onde os nós representam os alunos de 1 à 14 e, as linhas as comunicações $a[i,j]$.

¹ Por motivos de privacidade, os *logins* dos alunos foram omitidos.

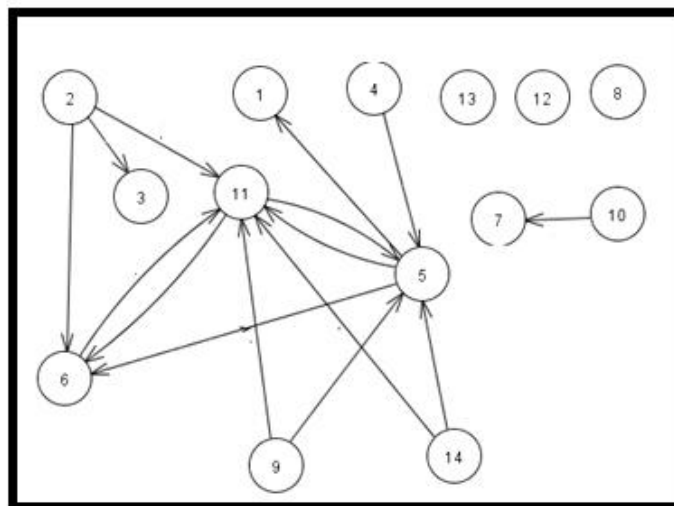


Figura 17 – Grafo de Interação a_i para a_j

Uma vez que, o estudo de caso é baseado em um grafo direcionado, foi aplicada as métricas de centralidade *In-Degree* e o *Out-degree*, conforme apresentado na Tabela 16 - Grau centralidade *In-Degree* e *Out-Degree*, onde:

- ID é a identificação do aluno
- Total de mensagens refere-se à quantidade de mensagens enviadas por cada aluno na rede virtual
- *In-Degree* o grau de entrada
- *Out-degree* o grau de saída

De posse dos resultados, o perfil dos acadêmicos no espaço virtual (*Twitter*) foi comparado com seu comportamento em sala de aula. De maneira que, constatou-se que os alunos representados pelos nós 2 e 5, possuíam os maiores índices *Out-degree*. Coincidentemente, eles foram os mesmos que mais contribuíram tanto na sala de aula, quanto na rede virtual. Por outro lado, foi observado os altos índices *In-Degree* dos alunos indicados pelos nós 5 e 11, destacando o aluno 5, como sendo o aluno central do grupo, pois o mesmo participa contribuindo com seus colegas (*Out-degree*), e foi também citado por eles (*In-Degree*).

Cumprе destacar também que, o aluno representado pelos nós 7 e 10 obtiveram um alto índice *in-degree* e *out-degree*, respectivamente pelo fato do grafo ser disjunto, ou seja, para o cálculo desta centralidade, leva-se em consideração apenas os nós que interagem entre si, neste caso, apenas um nó interage com cada um.

Tabela 16 - Grau centralidade *In-Degree* e *Out-Degree*

ID	MENSAGENS	<i>In-Degree</i>	<i>Out-Degree</i>	8	23	0,000	0,000
1	05	0,125	0,0	9	46	0,000	0,250
2	40	0,000	0,375	10	24	0,000	1,000
3	10	0,125	0,0	11	59	0,625	0,250
4	21	0,000	0,125	12	07	0,000	0,000
5	59	0,500	0,375	13	13	0,000	0,000
6	40	0,375	0,125	14	43	0,000	0,250
7	20	1,000	0,0				

O professor deverá ficar atento às altas discrepâncias entre os Graus de Centralidades de seus alunos. Isto pode significar que, a interação está centralizada em poucos alunos. Como é o caso dos alunos 2, 11 e, principalmente do aluno 5. Após esta análise, o professor deverá verificar se este comportamento de isolamento também está ocorrendo fora da Rede Social Online. Se isto acontecer, o professor deverá buscar intervir de alguma forma, como por exemplo, através de dinâmicas de grupos. Este foi o caso dos alunos 8, 12 e 13.

Os índices *out-degree* dos alunos 7 e 10, surpreenderam negativamente visto que, estes alunos, foram muito dedicados em sala de aula e, grandes colaboradores no cenário virtual; fato este, comprovado pela quantidade de mensagens postadas, no entanto, os mesmos não citaram, nem foram citados suficientemente pelos seus colegas, significando um certo isolamento.

Conclui-se que, através desta métrica, é possível encontrar os canais mais importantes por onde circulam as informações, pois, um nó central passa a estar em contato direto ou indireto com muitos outros nós. Deste modo, este ator em destaque deverá ser reconhecido na rede, a fim de ser peça fundamental na dissipação das informações.

Apresenta-se na Tabela 17, o grau de proximidade dos alunos monitorados. Neste caso, o aluno 2 possui maior índice, seguido pelos nós 5 e 9; por conseguinte, eles possuem um contato mais próximo perante os demais atores, deste modo, acabam tendo acesso às informações mais confiáveis. Eles também recebem as mensagens praticamente do ator que originou as mesmas, ou seja, em muitos casos, não existe um intermediário entre eles.

Tabela 17 - Valores medidos para o grau de proximidade

ID	Closeness Centrality	8	0,000
1	0,000	9	0,000
2	0,294	10	0,076
3	0,00	11	0,192
4	0,153	12	0,000
5	0,230	13	0,000
6	0,141	14	0,230
7	0,000		

O *Betweenness* é uma métrica que quantifica o grau de intermediação de um ator dentro da rede. O resultado para esta métrica é apresentado na Tabela 18.

Considerando estes fatores pode-se afirmar que, o grau de intermediação está concentrado em três alunos. Verifica-se a proeminência e o prestígio do nó 5, pois, de acordo com o valor de seu *betweenness*, o mesmo é altamente relevante. Sua ausência irá isolar a comunicação entre os outros. Deste modo, ele será considerado um ator com alto grau de influência sobre os demais.

Com base neste resultado, fica comprovado a necessidade do professor identificar os motivos pelos quais, levaram a maioria dos alunos a não obterem prestígio na rede. Neste contexto, o professor deverá analisar a disponibilidade de tempo do aluno para ter acesso à Rede Social Online, e se existe alguma antipatia na sala de aula, que possa comprovar esta desunião entre eles.

Tabela 18 – Valores do grau de intermediação

ID	Beteweenness Centrality	8	0,000
1	0,000	9	0,000
2	0,000	10	0,000
3	0,000	11	0,597
4	0,000	12	0,000
5	0,958	13	0,000
6	0,305	14	0,000
7	0,000		

Com a utilização das métricas de análise de redes sociais foi possível avaliar todos os alunos participantes e, dentre as conclusões constatou-se a existência de eventuais falhas de interação entre os mesmos, na rede social virtual. Fato este,

comprovado através dos resultados obtidos pela aplicação da métrica *out-degree*, onde os alunos representados pelos nós 7 e 10 tiveram um índice baixíssimo, pois, ambos, de acordo com os dados, foram alunos que contribuíram com inúmeras mensagens, no entanto, não obtiveram nenhum retorno por parte dos demais alunos.

Através da métrica *in-degree*, constata-se que grande parte dos alunos tiveram um papel de meros receptores na rede social, pois, a nulidade de seus índices indicam que, os mesmos não contribuíram de maneira ativa na inserção de conteúdos relevantes para à “comunidade virtual”.

A escolha pela métrica de proximidade está intrinsecamente ligada à busca pelos alunos que estão mais próximos dos demais. Com isto, o professor poderá atuar de maneira a iniciar questionamentos/respostas utilizando-se destes alunos, a fim de que, as informações atinjam rapidamente todos os membros da rede e, conseqüentemente, solucione uma dúvida ou gere uma nova discussão o mais rápido possível entre os participantes.

Observa-se por fim que, quanto ao grau de intermediação, o professor terá a possibilidade de identificar quais os membros da rede, cuja ausência poderão quebrar o ciclo de transmissão das informações, uma vez que, o afastamento deles poderá dividir uma rede em várias. Logo, caso o aluno representado pelo nó 5 que, possui maior índice de intermediação se ausentar da rede, eliminará a comunicação dos alunos 1 e 4 com os demais membros da rede, ou seja, deixa de existir um caminho para eles.

5.2. Conclusão dos estudos de casos na área acadêmica

As aplicações das métricas grau de centralidade, centralidade de intermediação e centralidade de proximidade, definem o perfil colaborativo das pessoas dentro de um determinado grupo e, eventuais falhas na disseminação de informações no mesmo.

Nas pesquisas realizadas nos estudos de caso das seções: *Analysis of Students Interaction Profile in Online Social Networks* e Monitoramento das interações dos aprendizes na rede social *Twitter* como apoio ao processo de mediação docente., foram aplicadas as métricas acima citadas e, o docente definiu o perfil colaborativo de seus alunos e, identificou os fatores que eventualmente obstruam as trocas de informações.

Pela fórmula do grau de participação e a do cálculo da afinidade define-se a afinidade das pessoas frente a um contexto.

Neste estudo da seção Estratégia Para Avaliação do Grau de Participação e Afinidade de Alunos Através de *Microbloggin*, utilizando-se das fórmulas retrocitadas,

o professor identificou a participação dos alunos em relação a um determinado conteúdo ministrado em sala de aula e a afinidade destes acadêmicos no cenário virtual (*Twitter*).

Para trabalhos futuros, devem ser realizados experimentos que identifiquem procedimentos necessários para modificar a interação entre os alunos, e, se existe algum mecanismo ou forma padrão de interferência no relacionamento dos acadêmicos. Pleiteia-se ainda, a verificação da possibilidade de aplicação da métrica de Clusterização, pois através dela é possível realizar a formação de grupos. O intuito deste agrupamento é viabilizar ao aluno possibilidades de construir através de pequenos grupos, o conhecimento; exercitando ao mesmo tempo inúmeras habilidades, tais como: capacidade de escolha, avaliação e decisão de quais informações ele disponibilizará na rede. Simultaneamente, possibilita o professor a monitorar as informações transmitidas pelos grupos, impedindo que os mesmos percam o foco pedagógico.

5.3. Estudos de casos de grupos específicos

5.3.1 Cenário utilizado nos experimentos dos grupos específicos

Visando alcançar os objetivos dos estudos de casos de grupos específicos foi realizado uma parceria com a empresa Serasa Experian, a qual forneceu informações de seu banco de dados a fim de que os mesmos fossem correlacionados com os dados de usuários do *Twitter*.

Inicialmente, realizou-se um experimento por intermédio de uma coleta aleatória de pessoas que se utilizam do *Twitter*. Neste caso, a ferramenta de coleta de dados AppSNA, desenvolvida pelo grupo de pesquisa, focou o campo localização existente nos perfis dos usuários, e fez uso das seguintes palavras-chave: “JP” e João Pessoa, como parâmetro de seleção e coleta de usuários no *Twitter*, pois a finalidade do mesmo foi analisar apenas as pessoas de João Pessoa- Paraíba.

Sendo assim, no mês de maio de 2012, a ferramenta de coleta de dados identificou maior número de pessoas que se enquadravam nos parâmetros da pesquisa, pois, foi gerado um arquivo CSV contendo o nome, login, cidade, de 3000 mil usuários. Quando disponível, foram também armazenados a latitude e longitude da pessoa que enviou a mensagem

De posse destes dados, realizou-se um correlacionamento entre os nomes dos perfis dos 3000 mil usuários, com nomes existentes na base de dado da Serasa Experian.

Processo este, cujo intuito foi identificar o nome completo e CPF destes usuários outrora coletados, a fim de conhecer a situação deles em relação a restrição de crédito.

Após uma análise minuciosa, constatou-se que a maioria dos usuários colocam em seus perfis codinomes ou abreviações de seus nomes e, esta ausência de informações concretas, tornou-se um obstáculo para identificação destes usuários. Deste então, foi possível classificar 26 pessoas, ou seja, 0,87% da lista inicial.

Com estes 26 usuários definidos, iniciou-se a coleta de dados dos seus seguidores para obter uma rede completa e, gerou-se um arquivo CSV contendo algumas informações do perfil dos mesmos, tais como: id no *Twitter*, nome no perfil, *screen name*, biografia, localização total de seguidores e seguidos, total de *tweets*, site, fuso horário, idioma, data de criação do perfil e foto, obtendo um arquivo completo com informações de 6518 usuários.

Este cenário foi a base para realização dos estudos de caso que serão expostos a seguir.

5.3.2 Classificação de usuários do *Twitter* pela análise estrutural e utilização de palavras-chave

Este estudo tem como finalidade classificar cada um dos indivíduos que compõem a rede, através da aplicação de métricas de análise de redes sociais e, da quantidade de mensagens disseminadas por elas baseadas em palavras-chave, tais como: crédito, empréstimo, dívidas, financiamentos, SPC, Soluções Financeiras, orçamento doméstico, juros etc

Optou-se por esta pesquisa para delimitar o universo do grupo pesquisado, permitindo que empresas identifiquem usuários que possam ser úteis em seus interesses, tornando-se assim, seus parceiros diretos ou indiretos. Estes usuários são selecionados mediante a análise dos seus índices estruturais, gerados a partir das métricas de ARS e da análise qualitativa do conteúdo das mensagens que possuem palavras-chave de interesse das empresas.

O presente trabalho restringe-se porém à primeira etapa, ou seja, à identificação dos indivíduos com potencialidade útil às empresas através das métricas e das palavras-chave.

Inicialmente utilizou-se dos experimentos realizados na seção 5.3.1, e a partir daí, com base nos 26 usuários identificados e no arquivo contendo todos seus respectivos seguidores, realizou-se uma classificação para selecionar apenas os seguidores destes usuários e que pertencentes a cidade de João Pessoa e, obteve-se então um arquivo contendo 451 usuários da respectiva cidade que, ao mesmo tempo são seguidores de um ou mais dos 26 usuários.

Com estas informações, iniciou-se um monitoramento das mensagens postadas por eles, visando coletar e identificar as mensagens direcionadas de um para outro. O intuito desta identificação é modelar a interação entre as 477 pessoas, ou seja, constrói-se uma matriz relacional, na qual, representar-se-ão os laços de comunicação entre estas pessoas, onde, só existirá um elo nos casos em que houver pelo menos uma mensagem direcionada (*mentions*) a outra.

A Figura 18, gerada através da ferramenta Gephi, representa o relacionamento entre os 477 usuários de João Pessoa baseados no envio e recebimento de mensagens direcionadas (*mentions*), deste modo, constatou-se que na maioria dos casos estes relacionamentos são unidirecionais, como ocorre com o usuário 1668, que enviou mensagens direcionadas (*mentions*) para os usuários 1681, 1672, 1680, 1675, 1683, e 1682 não obtendo reciprocidade destes. Verificou-se também, a existência de usuários isolados tais como: 1496 e 1456 que não mantêm relacionamento de mensagens direcionadas com quaisquer outros usuários. Há também os casos de usuários semi-isolados, formando redes “secundárias”, que são: o usuário 579 que enviou *mentions* apenas para os usuários 581 e 582. Informações estas marcadas através de um círculo vermelho. Através deste grafo, verificou-se uma intensidade na disseminação de mensagens direcionadas por um pequeno grupo de usuários, no entanto, percebe-se que, embora vários usuários tenham recebidos estas mensagens, o conteúdo de alguma forma não lhes interessou ou as respostas não foram direcionadas aos respectivos emissores e sim, exposta de maneira geral para todos os usuários da rede, não permitindo desta forma, a verificação da reciprocidade de relacionamento entre os emissores e receptores dos *mentions*.

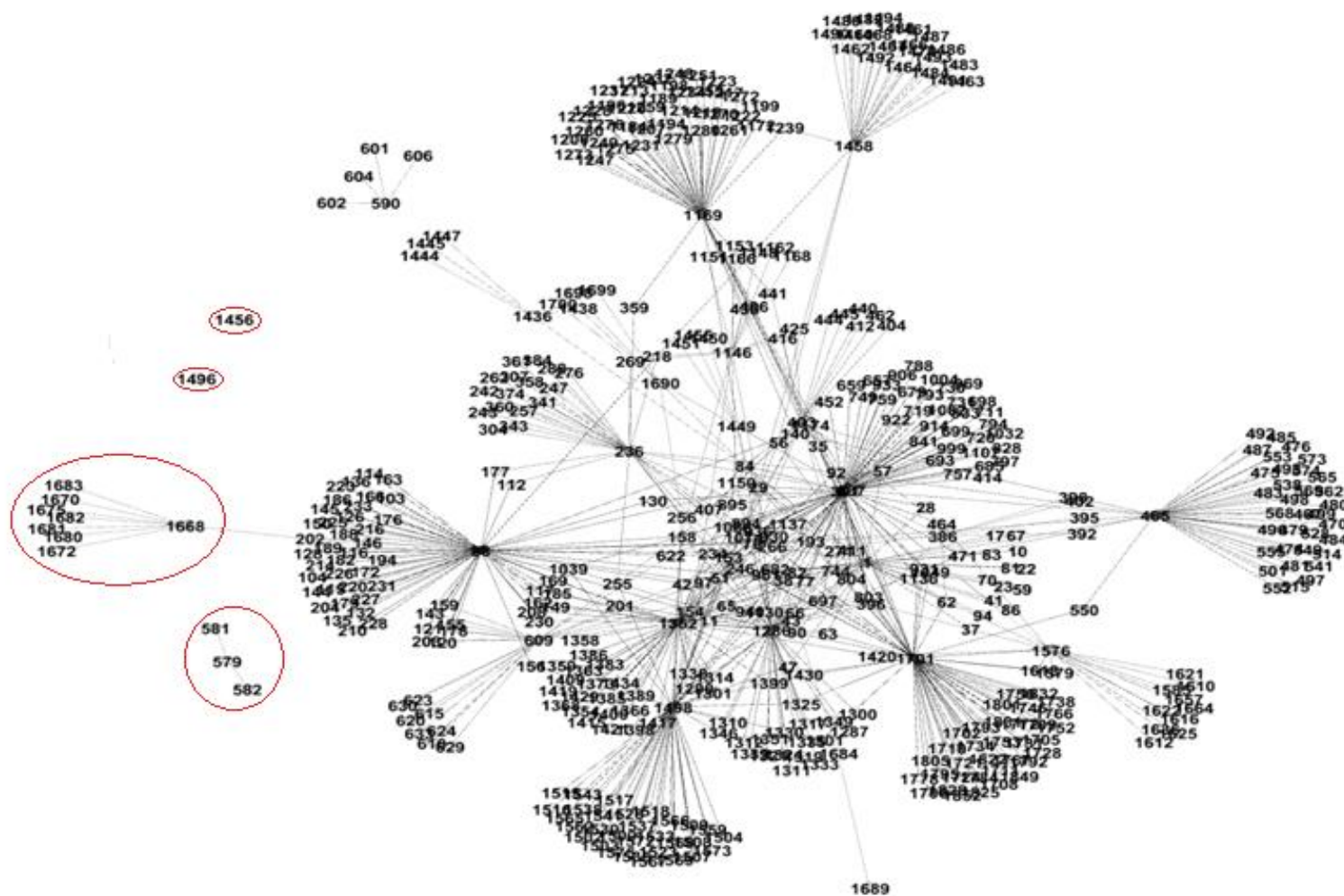


Figura 18 – Grafo da rede contendo somente usuários de João Pessoa-PB

5.3.2.1 Análise estrutural da rede

Conforme já citado, o presente trabalho visa identificar quais as pessoas que possuem maior grau de centralidade enquanto emissor e receptor de informações, qual a sua natureza jurídica, os relacionamentos que elas possuem, eventuais reciprocidades destes relacionamentos e o grau de importância de cada um na rede em decorrência da aplicação das métricas de análise de redes sociais.

Nesta pesquisa realizou-se experimentos com as métricas: grau de centralidade, de intermediação, de proximidade e *Page Rank*, através da ferramenta Gephi.

Deste modo, ao aplicar a métrica *Page Rank*, os resultados obtidos apresentaram um valor uniforme entre a maioria das pessoas, fato este constatado pelo número similar de suas ligações, o que gerou um índice de baixa relevância entre as mesmas.

Verificou-se também que, um pequeno grupo de pessoas, ou seja, apenas 2,09%, se destacaram das demais, apenas com uma diferença pouco significativa em relação as outras, conforme apresentado na Tabela 19.

Tabela 19 – Valor *Page Rank*

Nó	<i>Page Rank</i>	Nó	<i>Page Rank</i>
38	0.004	29	0.002
387	0.003	158	0.002
84	0.003	895	0.002
581	0.003	1447	0.002
582	0.003	1445	0.002
1689	0.003	1444	0.002
51	0.003	...	...
301	0.003	609	0.002
1438	0.003	1137	0.002
76	0.003	1352	0.002
601	0.002	1436	0.002
602	0.002	1449	0.002
604	0.002	1456	0.002
606	0.002	1496	0.002
396	0.002	1684	0.002
153	0.002	1690	0.002
256	0.002	1701	0.002

Optamos por utilizar a métrica *in-degree centrality* para definir o grau de centralidade dos receptores de informações dentro da rede, pois o método de identificação desta métrica é através da quantidade de arestas recebidas por cada um dos

nós, obtendo já de imediato um resultado satisfatório e capaz de responder os questionamentos acima citados.

Ao aplicarmos esta métrica um dos resultados obtidos foi a identificação do valor *in-deegree* de cada um dos 477 usuários. Mediante esta informação, foi elaborada a Tabela 20, onde identificamos o campo “valor *in-deegree*” em colorações diferentes que, posteriormente possibilitará a visualização de cada indivíduo no grafo. Destaca-se que quanto mais escura a coloração, maior é o grau de centralidade do indivíduo. O outro campo apresentado corresponde ao dos usuários que, permite identificar a qual indivíduo pertence cada valor.

Tabela 20- Valor *In-Degree Centrality*

VALOR IN-DEGREE	USUÁRIO	VALOR IN-DEGREE	USUÁRIO	VALOR IN-DEGREE	USUÁRIO	VALOR IN-DEGREE	USUÁRIO
12	38	3	464	2	255	1	1
11	387	3	435	2	359	1	41
9	84	3	471	2	269	1	17
9	76	3	682	2	266	1	67
8	29	3	940	2	425	1	70
7	92	3	895	2	416	1	83
7	158	3	803	2	452	1	22
7	153	3	961	2	441	1	81
6	65	3	1290	2	550	1	10
6	396	3	1301	2	622	1	23
5	42	3	1325	2	1130	1	94
5	51	3	1399	2	744	1	86
5	97	2	90	2	1049	1	37
5	234	2	28	2	1074	1	59
5	271	2	11	2	930	1	98
4	66	2	43	2	1136	...	...
4	35	2	57	2	931	1	1709
4	82	2	47	2	936	1	1706
4	77	2	156	2	804	1	1778
4	193	2	154	2	894	1	1832
4	218	2	185	2	1086	1	1795
4	256	2	130	2	1150	1	1711
4	411	2	149	2	1174	1	1731
4	697	2	169	2	1239	0	403
4	1039	2	120	2	1310	0	579
3	62	2	203	2	1314	0	590
3	63	2	208	2	1346	0	609
3	56	2	162	2	1338	0	1137
3	201	2	117	2	1420	0	1352
3	230	2	121	2	1417	0	1436
3	140	2	155	2	1430	0	1449
3	246	2	177	2	1438	0	1456
3	301	2	143	2	1501	0	1496

3	414	2	112	2	1576	0	1684
3	407	2	178	2	1618	0	1690
3	406	2	159	2	1579	0	1701

A fim de visualizar os referidos resultados através de um grafo direcionado, gerou-se a Figura 19, onde os usuários com maior grau de centralidade são representados pelas colorações mais escuras.

A visualização do grafo permite identificar de imediato os dois indivíduos com o maior grau de centralidade que estão representados com o círculo vermelho. Possibilitando verificar através da coloração que os indivíduos agrupados nas extremidades da rede, ou seja, as folhas; são os que possuem menores valores *in-degree*.

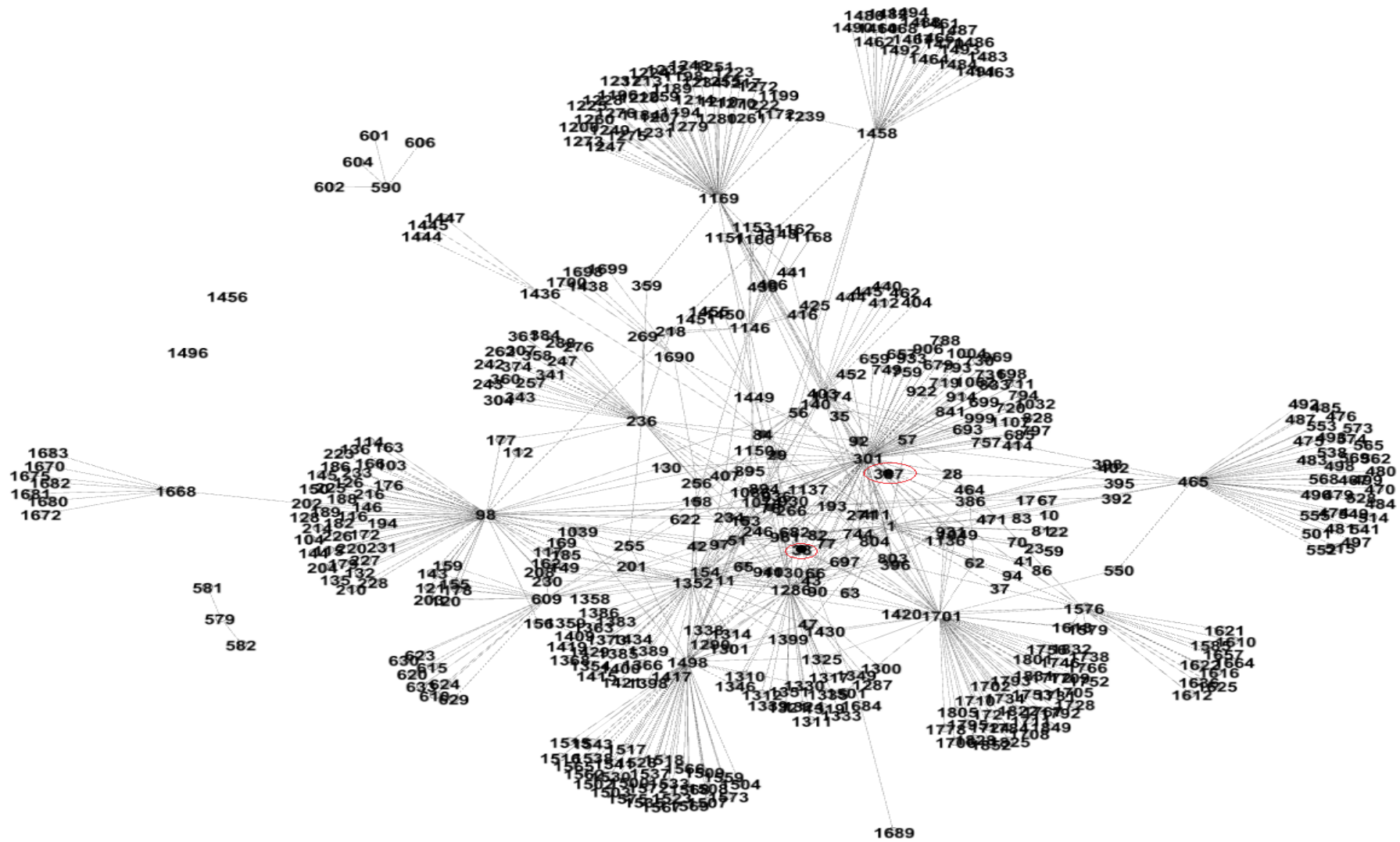


Figura 19 – Representação do valor *in-degree* através de um grafo

Analisando os resultados foi possível perceber que dentre os valores resultantes variando de 3 à 11, 78% são perfis de pessoas físicas e, 22% perfis de pessoas Jurídicas. Conforme Figura 20.

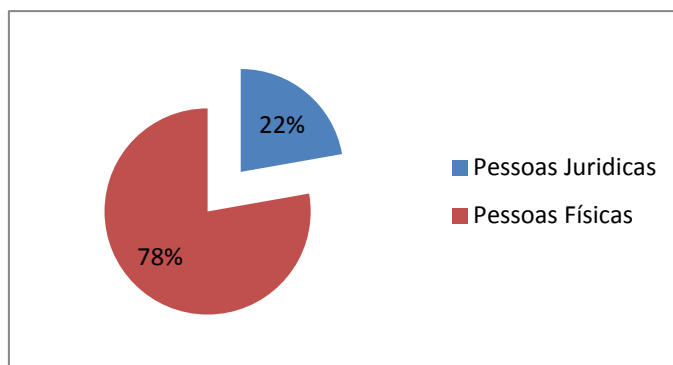


Figura 20 – Porcentagem de pessoas Jurídicas e Físicas baseadas em uma variação *in-degree*

Com o resultado *in-degree* verificou-se também que o usuário de maior índice recebe mensagens diretas de apenas 12 pessoas, ou seja, dentre todas as pessoas existentes na rede, apenas 2,53% interagem com ele diretamente através de *mentions*.

Outro aspecto observado neste usuário é que, ele corresponde a uma pessoa física e não jurídica. O que vem corroborar o fato de que a maioria dos usuários das redes sociais ainda são pessoas físicas, pois elas veem inicialmente as redes sociais como um instrumento de comunicação entre amigos e familiares. Todavia, este cenário vem se modificando e, os usuários das redes sociais já as identificam também como um instrumento de relacionamento com as pessoas jurídicas, seja para esclarecimento de dúvidas ou reclamações.

Realizou-se também dentro desta variação, um estudo com a finalidade de identificar a quantidade de pessoas que são “celebridades”, ou seja, destaques na cidade de João Pessoa-PB e, obteve-se um percentual de 21% para pessoas tidas como destaques na cidade de João Pessoa e, 79% referente às pessoas anônimas. Figura 21.

Vale ressaltar que correlacionando estas informações com os resultados da Tabela 20, constatou-se que os dois maiores índices são pessoas identificadas como “celebridades”, sendo o primeiro, referente a uma pessoa ligada a política paraibana e, o segundo à área jornalística da Paraíba.

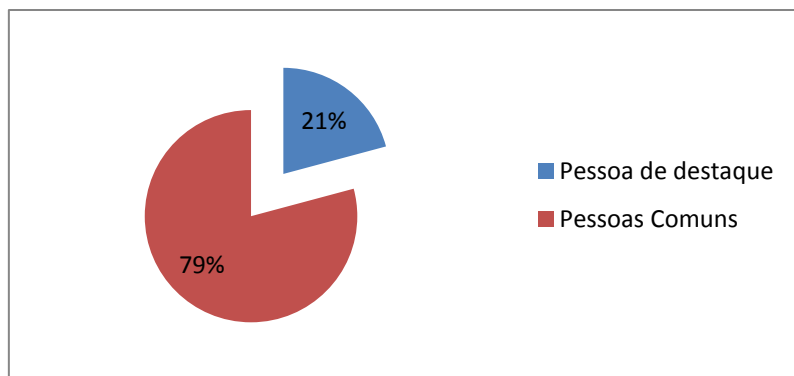


Figura 21 – Porcentagem das pessoas comuns e que são destaques na Grande João Pessoa-PB baseadas em uma variação *in-degree*

Utilizando-se da métrica *out-degree centrality*, que identifica o grau de centralidade de cada emissor de informações, gerou-se a Tabela 21, onde se observa uma disseminação maior entre os membros da rede, todavia, a um grupo pequeno de usuários. A coluna valorada *out-degree* indica a quantidade de pessoas a quem cada usuário enviou uma ou mais mensagens direcionadas.

Tabela 21 - Valor *out-degree centrality*

OUT-DEGREE	USUÁRIO	OUT-DEGREE	USUÁRIO
77	301	9	1668
69	98	7	1690
64	1701	6	1449
59	1352	5	1436
57	1498	4	590
50	1169	3	1137
48	1286	2	579
43	465	2	1684
36	1	0	92
32	236	0	42
26	1458	0	62
25	403	...	...
23	609	0	1832
18	1576	0	1795
17	1146	0	1711
10	386	0	1731

Vale ressaltar que as informações constantes na Tabela 21, também podem ser demonstradas através de um grafo (Figura 22), onde o grau de centralidade *out-degree* é representado por colorações, sendo que, as cores mais escuras representam maiores índices de centralidade da rede. Percebe-se ainda que, os usuários circulados em vermelho são os que possuem maior grau de centralidade *out-degree*.

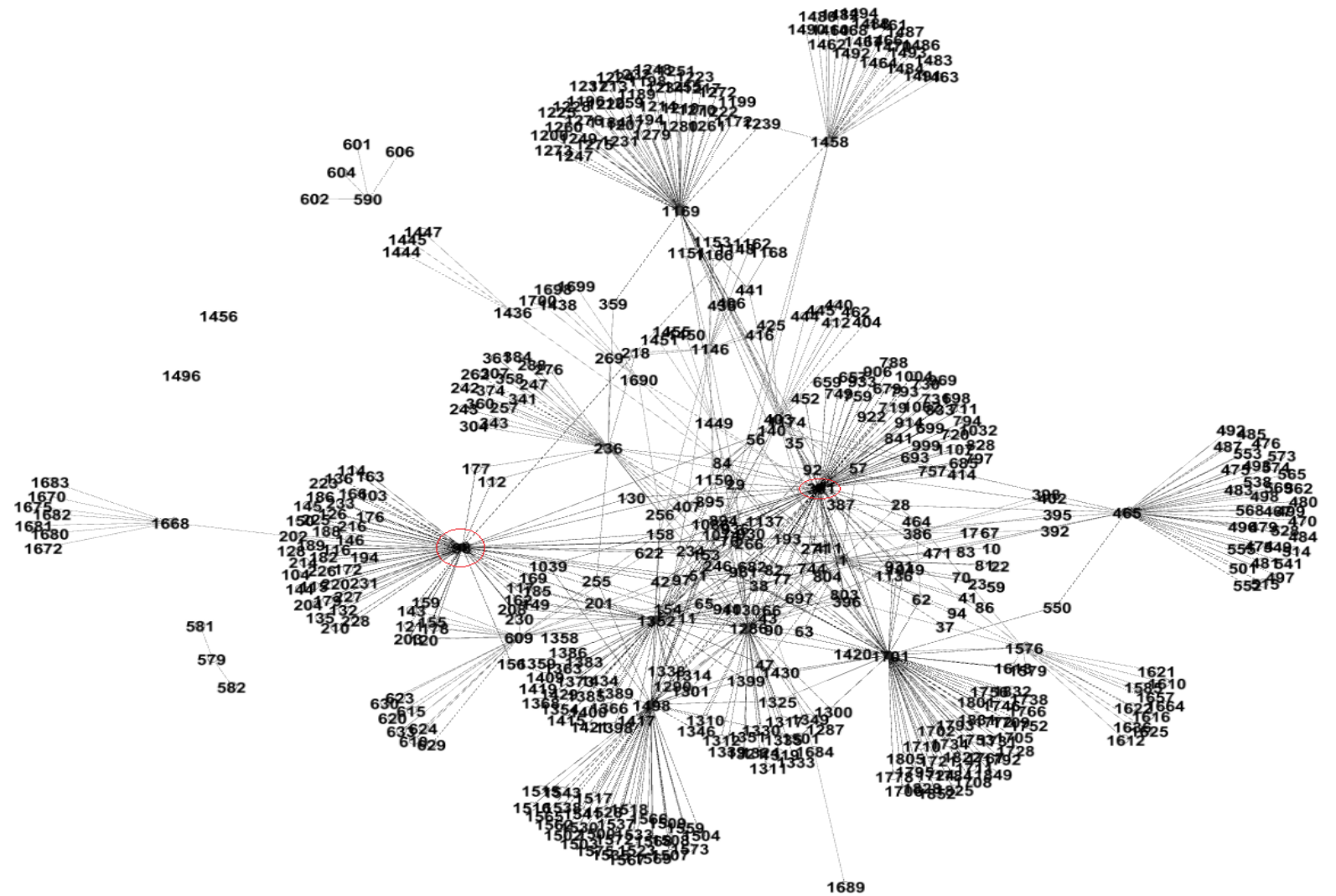


Figura 22 – Grafo representando os valores *out-deegre* de cada pessoa

Tabulados os dados resultantes da métrica *out-degree*, percebe-se uma grande discrepância entre a quantidade de pessoas existentes na rede e das que disseminam as informações, por exemplo, maior índice *out-degree* enviou mensagens direcionadas a 77 usuários enquanto, 95% dos membros da rede, não enviaram nenhuma mensagem direcionada. Figura 23.



Figura 23 – Porcentagem das pessoas que enviaram mensagens direcionadas

Outro aspecto observado é a ausência de afinidade entre as pessoas, visto que, em muitos casos, os usuários que mais disseminaram mensagens direcionadas na rede, não obtiveram uma reciprocidade de respostas ou comentários de maneira direcionada. Por exemplo, o usuário 301, e que possui maior grau *out-degree* com 77 mensagens direcionadas enviadas na rede, recebeu apenas 3 mensagens direcionadas, conforme seu valor *in-deegree* - Tabela 4 - dentre os 476 usuários existentes. Em outros, como por exemplo o usuário 1701 que enviou 64 mensagens direcionadas, e não recebeu nem uma mensagem direcionada.

Por fim, esta métrica permitiu verificar que dentre as pessoas que mais interagiram na rede diretamente com outros usuários, foram em sua maioria pessoas físicas e tidas como comuns perante a sociedade. Figura 24.

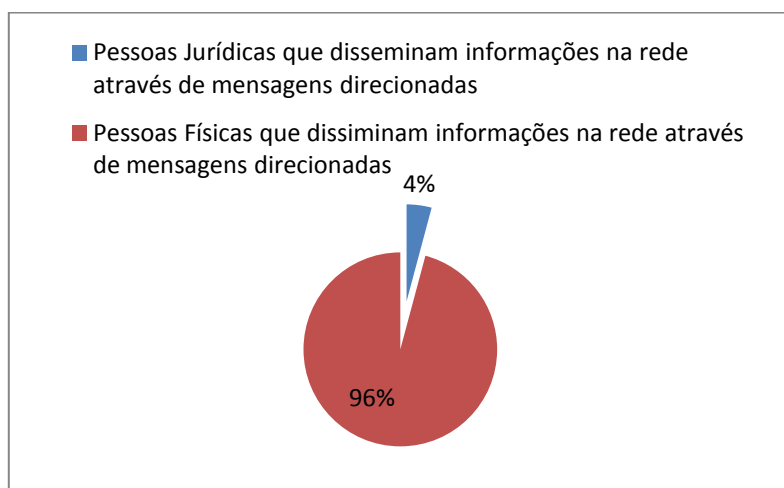


Figura 24 – Porcentagem de Pessoas Físicas e Jurídicas que interagiram diretamente na rede

Dentre as métricas estudadas optou-se pela utilização da métrica de centralidade de intermediação, visando assim, analisar o grau de intermediação de um nó, entre outros dois pontos da rede. Com estes resultados obtidos pela aplicação desta métrica, foi possível verificar uma independência de comunicação de quase todas as pessoas, onde, apenas dois usuários, que são pessoas físicas e comuns na sociedade, possuem um valor acima de 0. Tabela 22.

Constatou-se também que, a ausência de reciprocidade na comunicação direta entre as pessoas, é um dos fatores que contribuem para não haver esta intermediação, pois pelo fato de ser um grafo direcionado, em muitos casos acaba não existindo caminhos de um nó para outro, resultando assim, na maioria das vezes um índice nulo em relação ao grau de intermediação.

Tabela 22 – Grau de Intermediação

Grau de Intermediação	Usuário
0,001	301
0,0001	1576
0	1
0	92
0	42
...	...
0	1795
0	1711
0	1731

Embora existam apenas dois usuários com valores relevantes de intermediação, a presença deles é importante na rede, pois a ausência dos mesmos implicará no isolamento de outras pessoas, ou seja, dos usuários que eles estão intermediando a comunicação.

Com base na aplicação desta métrica, constata-se que mais uma vez o usuário 301, qual corresponde a uma pessoa física e não tido como “celebridade” pela sociedade, se sobressai dentre os demais membros da rede.

Dando sequencia aos experimentos da presente pesquisa, passou-se a analisar o grau de proximidade das pessoas monitoradas, utilizando-se assim a métrica centralidade de proximidade que, ao ser aplicada gerou o resultado constate na Tabela 23.

Salienta-se que as pessoas detentoras dos valores iguais ou bem próximo de 1 (um), são as que possuem maior contato com os demais membros da rede. Neste aspecto, as mesmas passam a ter acesso a mensagens mais confiáveis, ou seja, sem grandes alterações por parte de outras pessoas da rede.

Tabela 23 – Valores referentes ao grau de proximidade das pessoas

Grau de Proximidade	Usuário
1.0	1
1.0	98
1.0	301
1.0	386
1.0	403
1.0	465
1.0	579
1.0	590
1.0	609
1.0	1146
1.0	1169
1.0	1286
1.0	1352
1.0	1436
1.0	1449
1.0	1458
1.0	1498
1.0	1576
1.0	1668
1.0	1684

1.0	1690
0.864	1701
0.589	236
0.510	1137
0.0	92
0.0	42
0.0	62
...	...
0.0	1795
0.0	1711
0.0	1731

Analisando os dados da Tabela 23, verificou-se que 95% (Figura 25) das pessoas que compõem a rede estão distantes dos demais membros. Isto demonstra que existe uma ausência de interação entre a maioria dos usuários. Um dos aspectos que pode estar gerando este fato é que a maioria dos membros da rede não é seguida e, nem seguidora das demais pessoas, de tal forma que, a ausência destas características poderá comprometer os valores de proximidade.

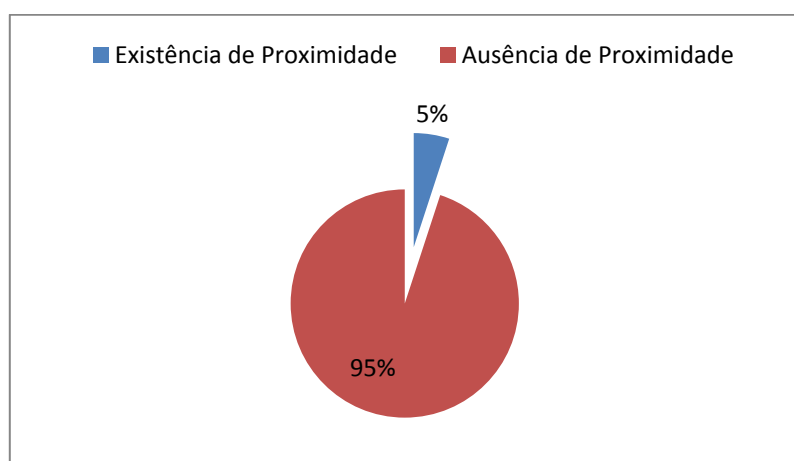


Figura 25 – Comparativo de Proximidade

Outra característica verificada é que, dentre os índices de proximidade diferentes de nulo, 4% são de pessoas jurídicas conforme Figura 26. Nota-se com este resultado, que a maioria das empresas existentes nesta rede social não possui uma comunicação mais constante com seus clientes através da rede social *Twitter*; pois, dos 22% dos membros da rede que se enquadram neste perfil - Figura 20 -, apenas 4% possuem um índice de proximidade aceitável. Fato este que torna-se evidente uma distância na troca de informações entre estas empresas e seus clientes no mundo virtual.

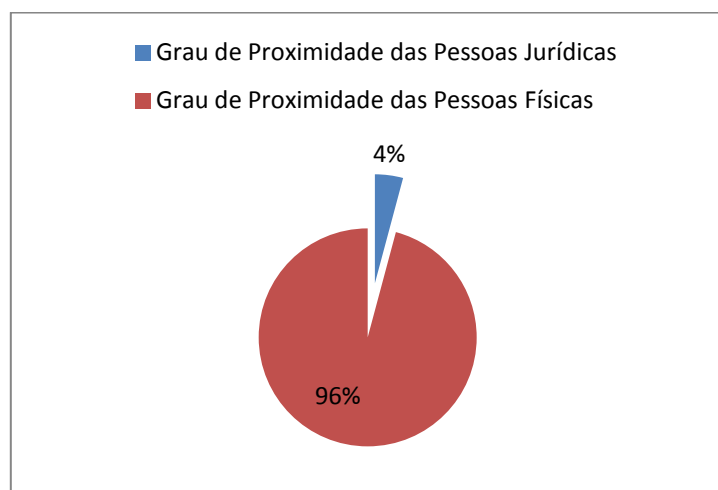


Figura 26 – Porcentagem de proximidade das pessoas físicas e jurídicas baseadas nos resultados do grau de proximidade

5.3.2.2 Análise quantitativa

Visando classificar quantitativamente os usuários durante um período de três meses, a ferramenta de coleta de dados monitorou os 477 usuários, objetivando coletar todos os *tweets* postados por eles, gerando uma base de dados. A partir deste banco de dados, consultas *mysql* foram realizadas com a finalidade de identificar pessoas que possuem interesse em assuntos relacionados a economia e/ou orçamento familiar, para tal, foram utilizadas as seguintes palavras-chave: empréstimo, taxas, juros e financiamento, que resultou na Tabela 24.

Tabela 24 – Resultados da análise quantitativa

Usuário	Quant. de <i>tweets</i> com a palavra chave “Empréstimo”
301	5
98	4
609	3
1286	2
1701	1
Usuário	Quant. de <i>tweets</i> com a palavra chave “Financiamento”
1668	1
1352	1
Usuário	Quant. de <i>tweets</i> com a palavra chave “taxas”
301	4

609	2
1286	1
1576	1
Usuário	Quant. de <i>tweets</i> com a palavra chave “juros”
301	13

A análise quantitativa apresentada na tabela acima identifica dos 477 usuários, apenas aqueles que emitiram *tweets* direcionados contento as palavras-chave utilizadas na pesquisa.

Em destaque, aparece o usuário 301 que mencionou em seus *tweets*, em número elevado, 3 das 4 palavras pesquisadas. Analisando este usuário comparativamente com o resultado da análise estrutural, constatou-se que o mesmo está em primeiro lugar na intermediação de comunicação (Tabela 22) e envio de mensagens (Tabela 21); em terceiro lugar em relação a proximidade com os demais usuários (Tabela 23) e, 26º em relação aos que mais receberam mensagens (Tabela 20).

Desta forma, o usuário 301 por ter interesse em assuntos relacionados às palavras-chave e ter uma boa posição na análise estrutural, tem o perfil que interessa às empresas ou pessoas físicas ligadas à economia e ou orçamento doméstico, podendo até mesmo ser seu eventual parceiro.

5.3.2.3 Conclusão

Após a realização desta pesquisa, os 477 usuários de João-Pessoa-PB que se utilizam do *Twitter* foram classificados estruturalmente e quantitativamente. Referidos resultados foram cruzados e constatou-se que, para os interesses de empresas ou demais clientes a classificação estrutural dos usuários tornar-se-á mais completa quando utilizada com a classificação quantitativa, pois poderá ser analisado dentro de um mesmo contexto não apenas sua relevância estrutural como também ser identificado com interesse em determinado assunto, no caso as palavras-chave escolhidas.

Futuramente é interessante a realização de uma análise qualitativa das mensagens postadas, pois com isto, poderá ser observado o real sentido em que as palavras-chave estarão sendo utilizadas no trabalho.

5.4. Análise quantitativa das relações dos usuários de uma rede social virtual frente à uma base de crédito

Esta pesquisa tem por objetivo representar através de um grafo, a correlação de um grupo de usuários do *Twitter* com uma base de dados fornecida por empresas de crédito, com o intuito de analisar quantitativamente este correlacionamento em diversos aspectos voltados a restrição de crédito.

Para tanto, utilizou-se o cenário apresentado na seção 5.3.1, e a partir do arquivo contendo os 6518 usuários da rede, que representam as 26 pessoas identificadas e seus respectivos seguidores, foi realizado um correlacionamento entre os dados existentes nos perfis destes usuários, com informações existentes na base de dado da referida empresa. Processo este, com o objetivo de identificar o nome completo, CPF e dados de crédito dos mesmos, ou seja, se possuem alguma restrição de crédito ou não. Deste modo, após uma análise minuciosa, foi possível classificar 504 pessoas, correspondentes à 7,74% da lista inicial.

As informações correlacionadas permitiram a criação de um arquivo que foi analisado através da ferramenta Gephi.

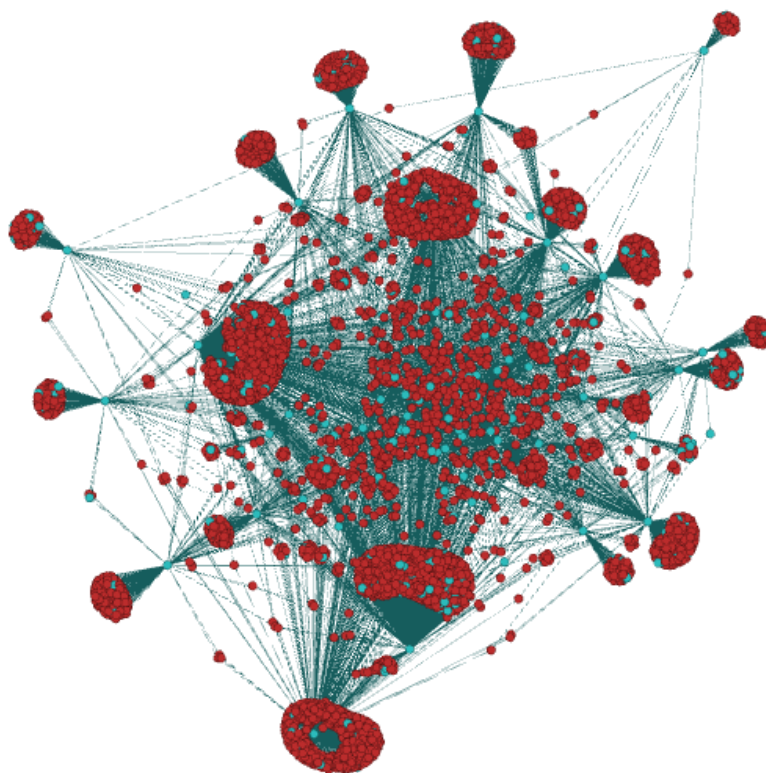


Figura 27 - Representação dos usuários identificados e não, pela Serasa Experian

Neste grafo (Figura 27), verifica-se em cada clusters em destaque, isto é, cada grupo existente a partir de uma das 26 pessoas e de seus respectivos seguidores, os que mais possuem indivíduos identificados ou não em relação às restrições de créditos. Esta classificação torna-se importante pois, por intermédio dela, permite-se realizar uma análise de crédito em cada grupo a partir dos seus membros.

Salienta-se que os nós em azul representam as 504 pessoas classificadas. Os demais usuários representados pela cor vermelha (6014), não foram identificados em relação as restrições de crédito, pois seus dados no *Twitter* não foram completamente cadastrados, alguns possuem em seus perfis apenas codinomes ou abreviações de seus nomes e, esta ausência de informações concretas, tornou-se um obstáculo na identificação precisa destes usuários.

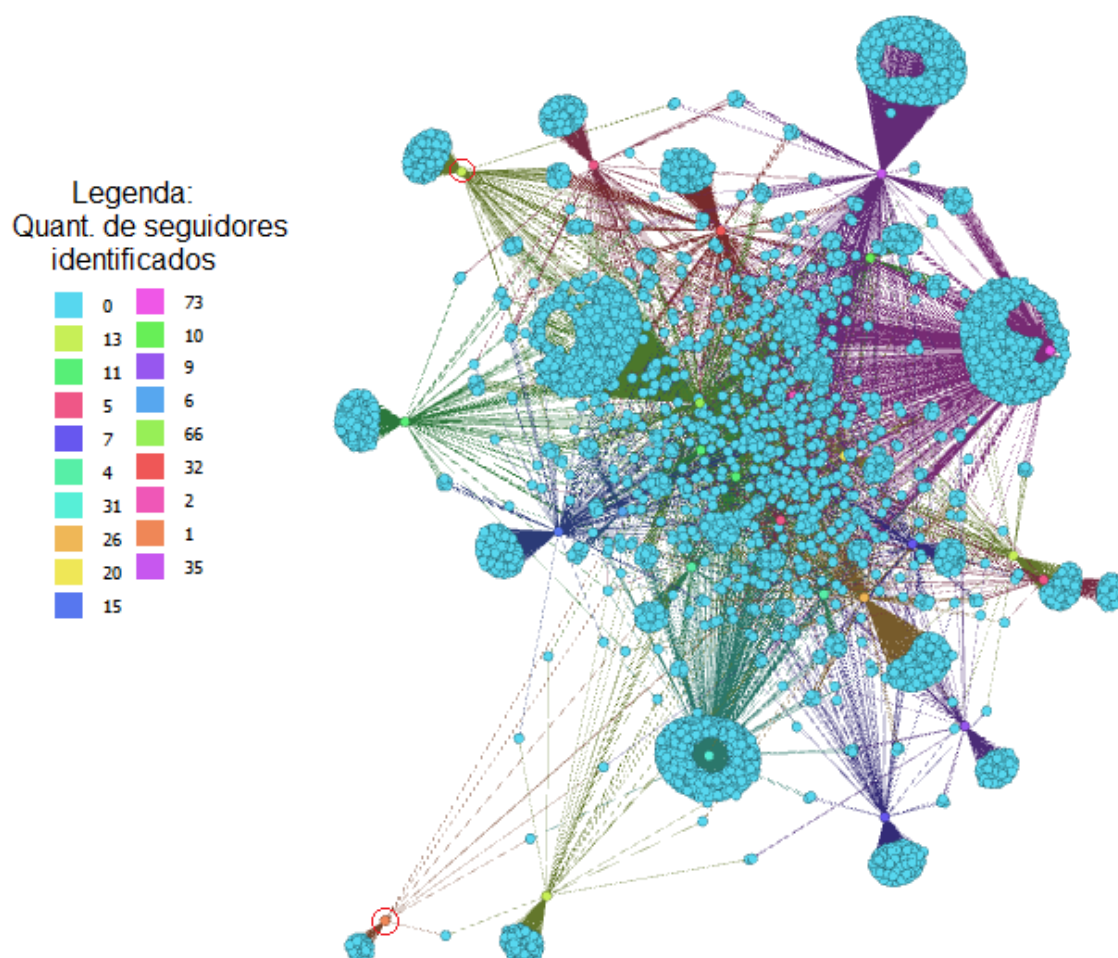


Figura 28 – Representação da quantidade de pessoas identificadas em relação às restrições de crédito que cada nó possui

O grafo abaixo (Figura 28), representa através das cores a quantidade de seguidores identificados em relação às restrições de crédito que cada um dos 6518 usuários da rede possui. Com estas informações é possível analisar cada pessoa e/ou grupo a partir dos dados correlacionados. Deste modo percebe-se que, o nó (3246), circulado em preto na parte superior possui 13 seguidores identificados em relação às restrições de crédito, e, já o nó (4412), circulado em vermelho na parte inferior, possui apenas uma pessoa.

Percebe-se também que, as folhas deste grafo são azuis e os nós que as representam possuem um índice 0 (zero) de seguidores classificados, fato este, causado pela ausência de ligações saindo deles para outros usuários da rede.

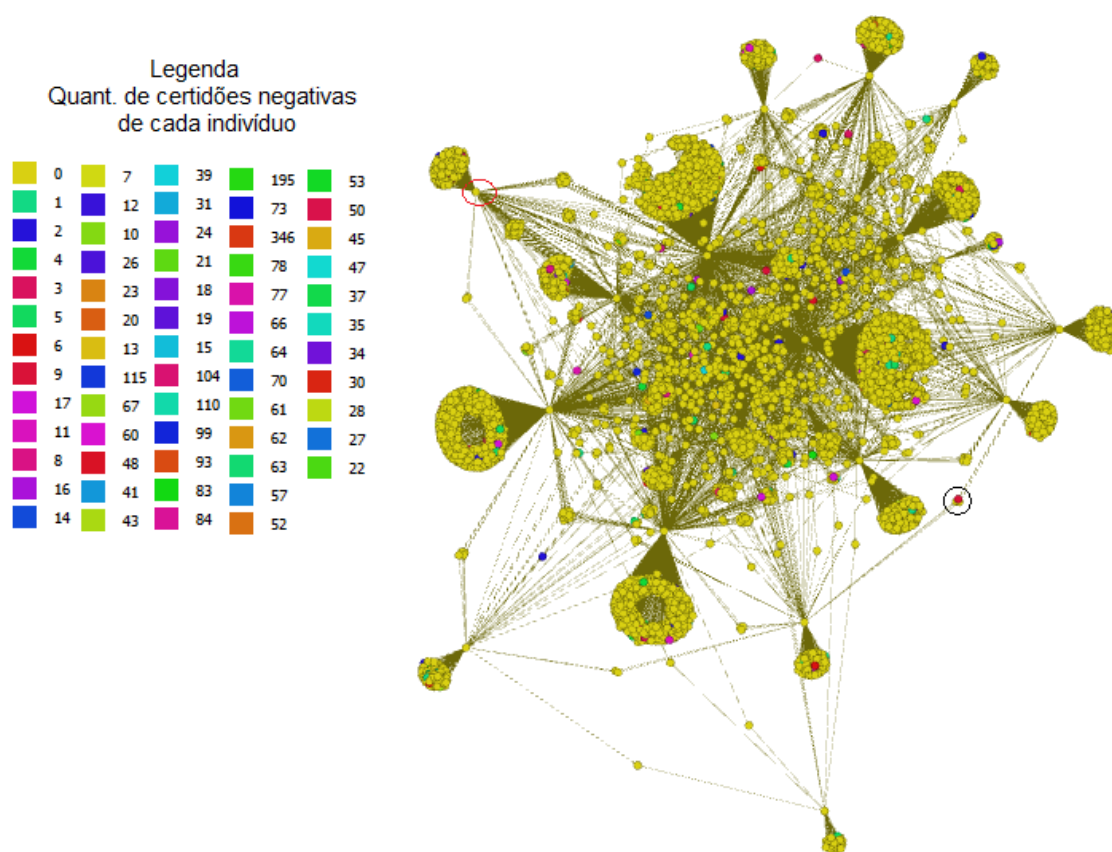


Figura 29 – Representação da quantidade de certidões negativas dos usuários

A Figura 29, os usuários são classificados de acordo com a quantidade de certidões negativas existentes na base de dados da Serasa Experian. Assim, o nó (2504), circulado em vermelho na parte superior, não possui nenhuma certidão negativa, no entanto, 9 de seus 13 seguidores classificados, ou seja, 69%, têm certidões negativas, logo, possuem restrições de crédito junto à Serasa Experian.

Em relação ao nó (5004), circulado em preto na parte inferior, que é seguidor de mais de um dos 26 usuários iniciais da rede, possui 9 certidões negativas.

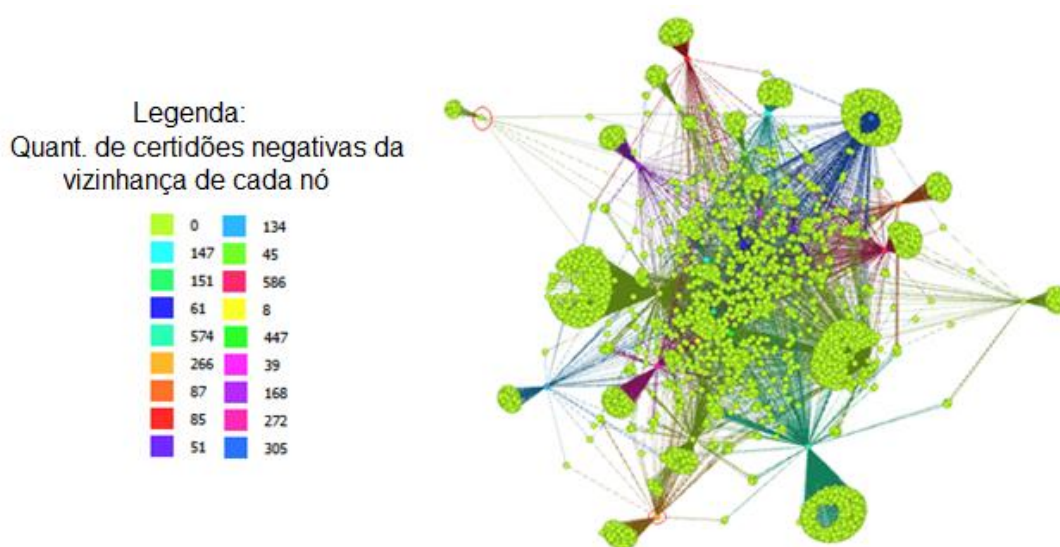


Figura 30 – Representação da quantidade de certidões negativas da vizinhança de cada um dos 26 usuários que são a base da rede

A Figura 30, apresenta quantidade de certidões negativas, da vizinhança de cada indivíduo existente, deste modo, verifica-se que a vizinhança do nó (5466), circulado em vermelho na parte superior, possui 45 certidões negativas, enquanto que, a vizinhança do nó (3246), circulado em vermelho na parte inferior, possui 266 certidões negativas.

Todos os resultados obtidos pela pesquisa acima realizada, ou seja, a identificação das pessoas com restrição de crédito e, a relação delas com os outros usuários do *Twitter*, foram utilizados para definirem o índice de inadimplência das 26 pessoas identificadas inicialmente pela Serasa Experian e de sua vizinhança através da seguinte fórmula:

Sendo assim, estes índices poderão ser calculados, conforme as fórmulas de inadimplência individual e do grupo, expostas a seguir:

$$GI_i = (Qn_i / Qng_i)$$

Equação 13 – Formula da inadimplência individual

Onde:

- GI_i designa o valor de inadimplência da pessoa i
- Qn_i a quantidade de negativas da pessoa i
- Qng_i é a quantidade de negativas do grupo ao qual a pessoa i encontra-se inserida

• i – Para melhor visualização dos grafos acima os mesmos encontram-se disponíveis em tamanhos maiores em anexos.

$$GIV_i = ((QtnV_i / QtpV_i)/100)$$

Equação 14 – Formula da inadimplência da vizinhança

Onde:

- GIV_i designa o valor de inadimplência da vizinhança referente a pessoa i
- $QtnV_i$ a quantidade de negativas da vizinhança referente a pessoa i
- $QtpV_i$ número de pessoas que compõem a vizinhança da pessoa i

Com os dados obtidos com a aplicação da fórmula de inadimplência foi possível elaborar a Tabela 25, na qual consta o grau de inadimplência normalizado de cada um dos 26 usuários do *Twitter* identificados e de sua vizinhança.

Tabela 25 – Índice de Inadimplência

vértice	Inadimplência do vértice	Inadimplência da vizinhança
916	0,194	0,174
923	1	0,121
1122	0,264	0,134
1126	0,003	0,164
1730	0,512	0,098
2453	0,229	0,137
2499	0,220	0,149
2504	0	0,000
2839	0,055	0,098
2905	0,110	0,224
3014	0	0,000
3246	0,567	0,205
3508	0,005	0,129
3519	0	0,000
3712	0,578	0,040
4253	0,001	0,225
4412	0	0,000
4662	0,035	0,181
4811	0,061	0,122
5466	0,676	0,064

5747	0	0,000
5810	0	0,000
6229	0,02	0,134
6296	0	0,000
6915	0	0,000
6960	0,163	0,085

Conclui-se que em alguns casos as pessoas representadas por vértices, são mais inadimplentes que toda sua vizinhança, por exemplo, o vértice 3712, e, na maioria dos casos, 78% ocorre o inverso, isto é, vizinhança é mais inadimplente. Figura 31.

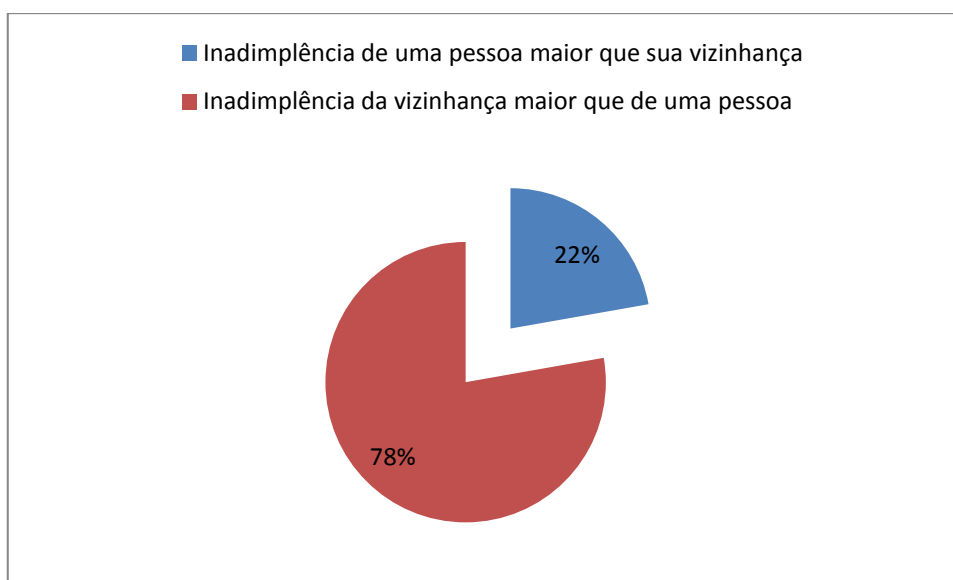


Figura 31 – Porcentagem comparativa da inadimplência

Esta análise permitiu também, verificar a existência da inadimplência ou a inexistência destas em grupos, por exemplo, o grupo formado pela vizinhança do vértice 5747 não possui inadimplência, enquanto que, o grupo formado pela vizinhança do vértice 4253 possui o maior grau. Conforme Tabela 25.

Os resultados obtidos através dos índices de inadimplência, possibilitam empresas, bem como, pessoas físicas ligadas ao ramo do comércio; a identificarem possíveis riscos de crédito, frente aos consumidores pertencentes à estes grupos de inadimplentes e, da mesma

forma permite identificar grupos que se demonstrem seguros, em relação ao crédito em decorrência de seu baixo, ou inexistente índice de inadimplência.

Percebe-se através desta pesquisa que, muitos fatores relacionados ao, crédito poderão ser estudados a fim de definirmos não só o perfil de crédito de cada pessoa, como também do grupo ao qual o mesmo se insere.

5.4.1 Conclusão do estudo de caso

Através desta pesquisa conclui-se que, a utilização da AppSNA na coleta de dados e o Gephi na geração de grafos e aplicação das métricas de análise de redes sociais foi eficaz e, ao mesmo tempo, permitiu alcançar objetivos almejados. Assim, foi possível correlacionar dados fornecidos pela Serasa Experian, com dados dos usuários selecionados do *Twitter*, visando à identificação das pessoas quanto à sua restrição de crédito, possibilitando a realização do trabalho em sequencia, ou seja, análise quantitativa dos relacionamentos existentes entre usuários, baseados nas restrições de crédito.

6. Considerações Finais

Através da realização deste trabalho, ficou evidente a utilização das redes sociais virtuais como novas formas de interação e comunicação entre as pessoas nos diversos campos da sociedade. Percebe-se que, cada vez mais, a partir de interesses comuns, grupos se formam não só no âmbito do mundo concreto, mas também no virtual, compartilhando uma vasta gama de conteúdos entre as pessoas. Destaca-se, que as redes sociais online e os conteúdos dissipados por intermédio delas, passaram a pouco tempo despertar interesse dos estudiosos que, iniciaram pesquisas aprofundadas em diversos âmbitos.

Verifica-se ainda que, estes ambientes virtuais são propícios à realização de novas descobertas em diversos temas, pois, estas informações poderão ser utilizadas de várias formas conforme o contexto a ser analisado.

Buscando aprofundar-se em alguns aspectos das redes sociais online, este trabalho realizou estudos através do emprego de métricas de análise de redes sociais, o que permitiu a identificação das pessoas centrais de determinados grupos monitorados, bem como, comparação da atuação de seus membros dentro do mundo virtual e do concreto.

Outro aspecto identificado foi que os indivíduos com maior grau de centralidade, foram os que responderam pela dinamização e mobilização da rede, isto é, ao aplicar cada métrica de ARS, foi possível verificar sua atuação, por exemplo: as pessoas com maiores índices *in-degree* consequentemente foram as que mais dissiparam informações.

Através de aplicação das métricas ARS, peculiaridades e características do mundo real, identificou-se dentro de um determinado contexto, o perfil de relacionamento virtual de cada indivíduo, ou grupo a ser monitorado.

A participação de cada indivíduo foi monitorada através da identificação das palavras-chave, do interesse de cada grupo. Este método possibilitou apenas uma análise automatizada quantitativamente, abrindo portanto, espaço à análise qualitativa para que a mesma fosse realizada manualmente, conforme efetivado no estudo de caso acadêmico.

Com relação afinidade das pessoas monitoradas, constatou-se que elas não ocorrem de forma objetiva, ou seja, o fato do indivíduo ser mais central conforme índices

resultantes das métricas de ARS, não interfere na afinidade com os demais membros da rede, deste modo, concluiu-se neste ponto que, afinidade ocorrerá por outro fator, por exemplo, o subjetivo.

Todo caminho percorrido ao desenvolvimento deste estudo, possibilitou verificar que as informações existentes nas redes sociais online, são fontes úteis na geração de conhecimentos a diversos contextos.

7. REFERÊNCIAS

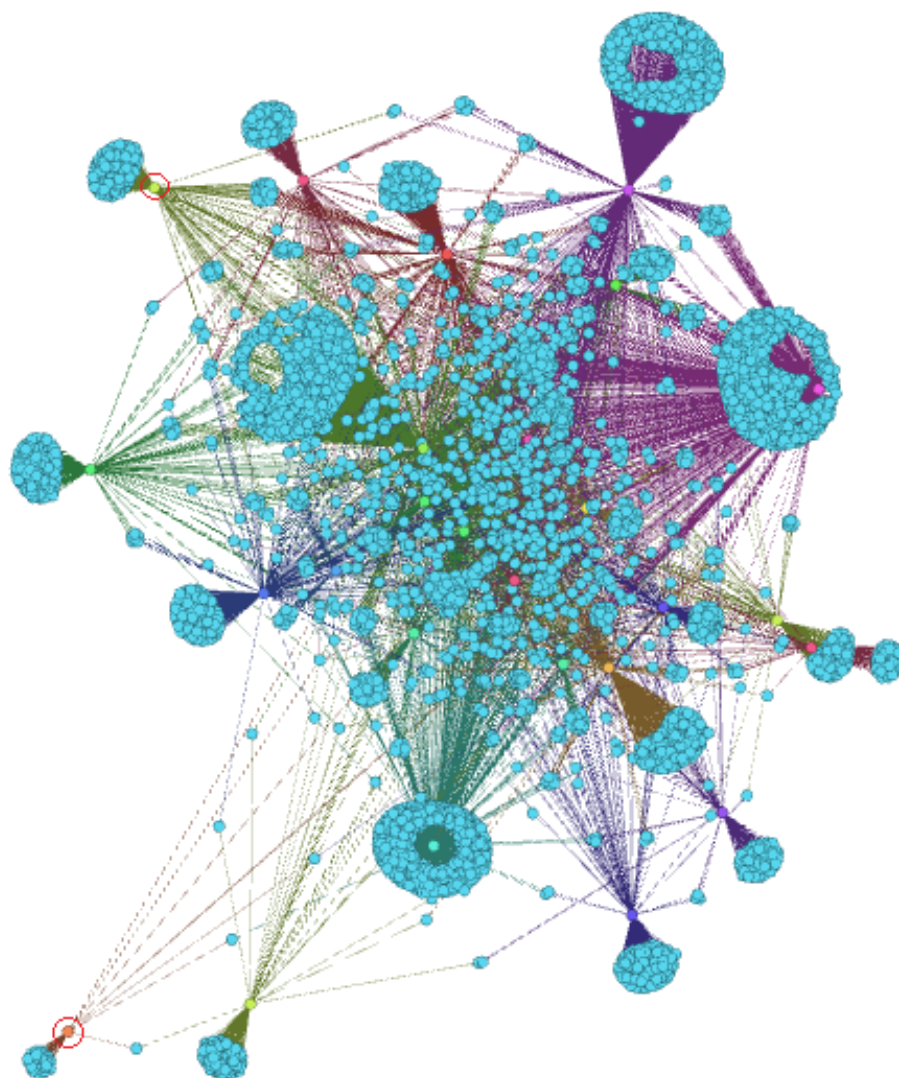
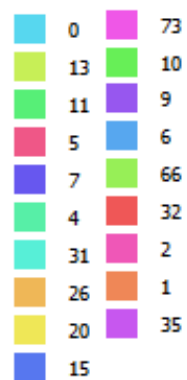
- ALARCON, J. A. B. L., XAVIER, O. C. e CARVALHO, C. L., 2011. *Implementing semantic queries in online social networks*. IADIS - International Conference Applied Computing. Rio de Janeiro, Brasil, pp 1-7.
- ALISSON, P. D., (1978). Measures of inequality. *American Sociological Review*.
- ALENCAR, A. F. (2007) “A pedagogia da migração do software proprietário para o livre: uma perspectiva” 246 p. Dissertação (Mestrado) – Universidade de São Paulo, Faculdade de Educação, São Paulo.
- BAVELAS, A., (1948). A mathematical model for group structure. *Human organizations*. 7, 16-30.
- BENEVENUTO, F., ALMEIDA, J., and SILVA, A. (2011). *Explorando redes sociais online: Da coleta e análise de grandes bases de dados às aplicações*. In Mini-cursos do Simpósio Brasileiro de Redes de Computadores (SBRC).
- BERKOWITZ, S. D., (1982). *An Introduction to Structural Analysis: The Network Approach to Social Research*. Toronto: Butterworth.
- BIANCO, B. F., (1987). *A Antropologia das sociedades contemporâneas*. Produção Gráfica: presser & bertelli Consultoria Editorial. São Paulo, Brasil.
- BRIN, S and PAGE, L., (1987). *The anatomy of a large-scale hypertextual web search engine*. *Comput Netw. ISDN Syst.* 30(1-7):107-117.
- CAPRA, F., (2002). *As conexões ocultas: ciência para uma vida sustentável*. 2. ed. São Paulo: Cultrix, p.93.
- CASTELLS, M., (1999). *A sociedade em rede*. São Paulo: Paz e Terra.
- COHN, B. S., and MARRIOTT, M., (1958). Networks and centres of integration in Indian civilization. *Journal of Social Research*. 1, 1-9.
- CASTELLS, M., (2003). *A galáxia da Internet*. Rio de Janeiro: Jorge Zahar.
- COLEMAN, J. S., (1964). *Introduction to Mathematical Sociology*. New York: Free Press.
- COSTA, R., (2003). *A Cultura Digital*. Publifolha. pp 1- 82.
- CHEN, Y-Y.; GAN, Q.; SUEL, T. I/O efficient techniques for computing pagerank, In: **Proceedings** of the CIKM'02, McLean, Virginia, USA, November, 2002.
- DONNINGER, C. (1986). The distribution of centrality in social networks. *Social Networks*, 8, 191-203.
- O'CONNOR, B.; Balasubramanyan, R.; Routledge, B.R.; Smith, N.A. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of ICWSM'2010*.
- ERDOS, P., and RENYI, A. (1960). On the evolution of random graphs. *Publications of the mathematical Institute of the Hungarian Academy of Sciences*. 5, 17-61.
- Filho, N. R. S., (2011). Monitoramento das redes sociais como forma de relacionamento com o consumidor. O que as empresas estão fazendo? *Gestão Contemporânea*, Porto Alegre, Brasil, ano 8, n 9, p.63-86.
- FORTUNATO, S. et al. How to make the top ten: approximating pagerank from indegree, In: **Proceedings** from WWW2006, Edinburgh, UK, May 22–26, 2006
- GEPHI. *Tutorial Visualization*. Disponível em: < <https://gephi.org/users/tutorial-visualization/>>. Acesso em: 1 set. 2012.
- GOKHALE, D. V., e KULLBACK, S., (1978). *The Information in Contingency Tables*. New York: Marcel Dekker.
- GONLAVES, P. , DORES, W. e BENEVENUTO, F., 2012. *PANAST-t Uma Escala Psicométrica para Medição de Sentimentos no Twitter*. BraSNAM – *Brazilian Workshop on Social Network Analysis and Mining*. Curitiba, Brasil, pp 1-12.
- KHINCHIN, A. I., (1957). *Mathematical Foundations of Information Theory*. New York: Dover.
- KULLBACK, S., (1959). *Information Theory and Statistics*. New York: John Wiley and Son.
- MARTELETO, R. M., (2001). Análise de Redes Sociais – aplicações nos estudos de transferência da informação. *Ci. Inf.* Brasília, Brasil, PP 1-10.
- MARTELETO, R. M., (1994). *Cultura da modernidade: discursos e práticas informacionais*. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v.23, n.2, p.115-137.
- MEXIAS, J. et al, 2012. *Sigam-me os bons! Transformando robôs em pessoas influentes no Twitter*. BraSNAM – *Brazilian Workshop on Social Network Analysis and Mining*. Curitiba, Brasil, pp 1-12.
- METAXAS, P. T., 2011. *On the predictability of elections using social media data*. IADIS - International Conference Applied Computing. Rio de Janeiro, Brasil, pp 1-4.
- NEWMAN, M. and GIRVAN, M., 2004. Coauthorship networks and patters of scientific collaboration..

- NEWMAN, M. and GIRVAN, M., 2004 Finding and evaluating community structure in networks. *Physical Review E*.
- NIEMINEN, J., (1974) *On centrality in a graph*. *Scandinavian Journal of Psychology*.
- NOOY, W., MRVAR, A. and BATAGELJ, V. (2005). *Exploratory Social Network Analysis With Pajek*. Cambridge University Press, New York, EUA.
- OLIVEIRA, R. F., DUARTE, A. N. and BRITTO, A. V., 2011. *Analysis of Students Interaction Profile in Online Social Networks*. IADIS - International Conference Applied Computing. Rio de Janeiro, Brasil, pp 1-4.
- OLIVEIRA, R. F., ARAÚJO, J., DUARTE, A. N. and Britto, A. V., 2011. Estratégia para Avaliação do Grau de Participação e Afinidade de Alunos Através de Microblogging. IV WAVALIA – Workshop sobre Avaliação e Acompanhamento da Aprendizagem em Ambientes Virtuais – 22º SBIE. Aracaju, Brasil, pp 1-10.
- OLIVEIRA, R. F., ARAÚJO, J., PRETÔNIO, F. e BRITTO, A. V., 2012. *Monitoramento das Interações dos Aprendizes na Rede Social Twitter como Apoio ao Processo de Mediação Docente*. BraSNAM – Brazilian Workshop on Social Network Analysis and Mining. Curitiba, Brasil, pp 1-12.
- OLIVEIRA, L. S., CAMPOS, G. O. e SILVA, R.S., 2012. *Mineração de dados e análise de opinião em redes sociais – Um estudo de caso sobre as Olimpíadas 2012 utilizando o Twitter*. BraSNAM – Brazilian Workshop on Social Network Analysis and Mining. Curitiba, Brasil, pp 1-12.
- PAJEK. Disponível em: < <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>>. Acesso em: 10 jun. 2012.
- PREFUSE. User's Manual. Disponível em: < <http://prefuse.org/doc/manual/>>. Acesso em: 04 jul. 2012.
- RUHNAU, B. (2000) *Eigenvector-centrality – a Node-centrality?*, *Social Networks* 22, 357-365, Dortmund, Germany.
- SANTOS, R. e GRECIO, A. Introdução à representação e análise de grafos com a API JUNG. 49. Ed. Revista MundoJ.
- SANTOS, R. e GRECIO, A. Introdução visualização de grafos com a API JUNG. 50. Ed. Revista MundoJ.
- SCHRICKEL, W. K. (1997). *Análise de crédito: concessão e gerência de empréstimos*. 3. Ed. São Paulo: Atlas, 1997.
- SHAW, M., (1954). Group structure and the behavior of individuals in small groups. *Journal of Psychology*. 38, 139-149.
- SHANNON, C. E., e WEAVER, W. W., (1949). *The mathematical theory of communication*. Champaign: The University of Illinois Press.
- SHIMBEL, A., (1953), Structural parameters of communication networks. *Bulletin of mathematical Biophysics*. 15, 501-507.
- STEPHENSON, K., ZELEN, M, (1989). *Rethinking centrality: methods and examples*. *Social Networks*, v.11, n.1, p.1-37.
- THEIL, H., (1967). *Economics and Information Theory*. Chicago: Rand McNally.
- VALIATI, H. et al, 2012. *Deteção de Conteúdo Relevante e Usuários influentes no twitter*. BraSNAM – Brazilian Workshop on Social Network Analysis and Mining. Curitiba, Brasil, pp 1-12.
- WATTS, D.J.; STROGATZ, S.H., (June 1998) *Collective Dynamics of “Small-World” Networks*, *Nature*, v.393, p. 440-442.
- WASSERMAN, S., FAUST, K., (1994). *Social Network Analysis: Methods and Applications*. 8. Ed. Cambridge University Press, New York, EUA.

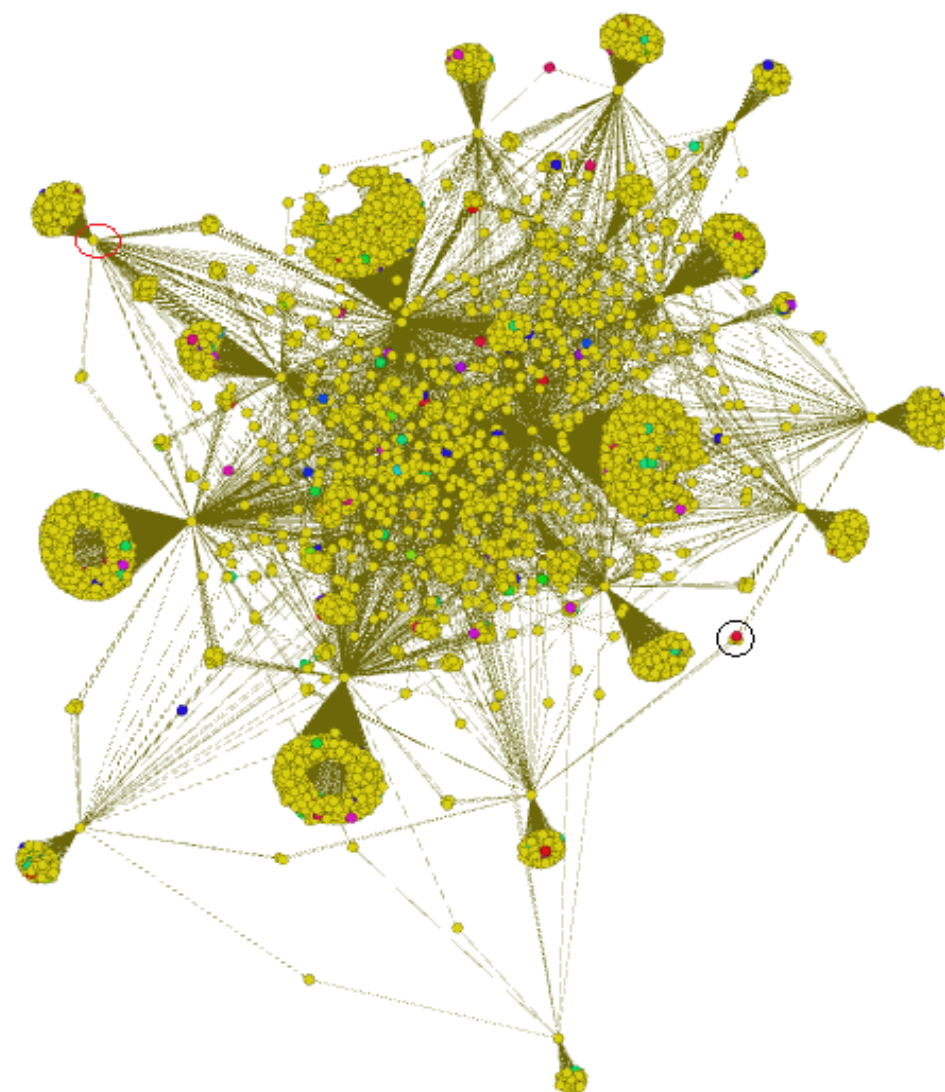
8. ANEXOS



Legenda:
Quant. de seguidores
identificados



Legenda
Quant. de certidões negativas
de cada indivíduo



Legenda:
Quant. de certidões negativas da
vizinhança de cada nó

