

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

DANYLLO WAGNER ALBUQUERQUE

**UMA ABORDAGEM PARA ANÁLISE DA DISCUSSÃO POLÍTICA
UTILIZANDO DADOS DE USUÁRIOS DO TWITTER**

**JOÃO PESSOA - PB
JULHO - 2013**

DANYLLO WAGNER ALBUQUERQUE

**UMA ABORDAGEM PARA ANÁLISE DA DISCUSSÃO POLÍTICA
UTILIZANDO DADOS DE USUÁRIOS DO TWITTER**

Dissertação apresentada a banca examinadora do Programa de Pós-Graduação em Informática do Centro de informática da Universidade Federal da Paraíba como requisito para obtenção do título de mestre em informática na linha de pesquisa de computação distribuída.

Orientador: Prof. Dr. Alisson Vasconcelos de Brito.

**JOÃO PESSOA - PB
JULHO - 2013**

A345u Albuquerque, Danyllo Wagner.
Uma abordagem para análise da discussão política
utilizando dados de usuários do Twitter / Danyllo Wagner
Albuquerque.-- João Pessoa, 2013.
147f. : il.
Orientador: Alisson Vasconcelos de Brito
Dissertação (Mestrado) – UFPB/CI
1. Informática. 2. Redes sociais. 3. Mineração de dados.
4.Redes sociais – análise. 5. Redes complexas – teoria.
6.Mídias e interação.

UFPB/BC

CDU: 004(043)

Ata de defesa

“A imaginação é mais importante do que a ciência porque a ciência é limitada ao passo que a imaginação abrange o mundo inteiro”

(Albert Einstein)

Dedico este trabalho a todos aqueles que, direta ou indiretamente, contribuíram para o cumprimento de mais uma etapa de minha vida pessoal e profissional.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela minha existência, pelas bênçãos que me tem concedido ao longo de toda minha vida, por guiar-me e acompanhar-me diariamente e, por incentivar-me na conclusão de mais uma etapa.

Aos meus Pais Dorgival Albuquerque e Josélia Albuquerque, pelos exemplos de dignidade e bravura dados a cada dia, sem vocês, nada disso estaria acontecendo. Amo vocês!

Aos meus avós José Muniz e Antonieta Albuquerque (*In Memoriam*), não foram em vão todos os ensinamentos! Para sempre estarão guardados em meu coração. Eternamente grato!

Aos meus tios Josemar Albuquerque e Josenaldo Albuquerque (*In Memoriam*), Josefa Albuquerque e José Muniz Filho, aos meus irmãos Dorgival Junior e Maria Virgínia, pelos momentos de reflexão, pelos bons conselhos cedidos e por sempre estarem torcendo por mim. Meu sucesso também é de vocês!

A minha turma que, apesar dos obstáculos surgidos no caminho, nos mantivemos entusiasmados até este momento. Em especial aos companheiros Larissa Barbosa, Uirá Paiva, Filipe Moura e Diego Morais.

A todos integrantes do Grupo de Trabalho *LabSNA* os quais me acompanharam durante grande parte do meu trabalho de mestrado. Em especial agradeço ao professor Alexandre Duarte e aos alunos Moacir Lopes, Roberto Felício e Jansepetrus Brasileiro pelos momentos de estudo, pela compreensão e por todo apoio e ajuda nesta jornada.

Aos professores, Andrei Formiga, Álvaro Medeiros, Liliane dos Santos, Ronei Morais, Lucídio Formiga e Iguatemi Fonseca pela paciência e competência na transmissão dos conhecimentos.

Ao meu orientador e amigo Alisson Vasconcelos de Brito, pelo exemplo de profissionalismo, pela paciência, apoio e contribuição para realização deste trabalho, Muito Obrigado!

RESUMO

Redes sociais têm se tornado extremamente populares levando ao surgimento e a popularização de uma crescente onda de aplicações baseadas na web. Prioritariamente estes ambientes são utilizados como meio para partilhar ideias e trocar informações entre indivíduos que possuem interesses e/ou objetivos em comum de modo que a presença deste tipo de mídia fornece um meio especial de debate que as formas convencionais não conseguem desempenhar. Neste sentido o presente trabalho de pesquisa tem como objetivo investigar se as redes sociais são de fato ambientes utilizados para deliberação política e se através disto podemos obter informações semelhantes daquelas obtidas através de pesquisas de opinião eleitoral. Para isto utilizamos o contexto das eleições municipais da cidade de João Pessoa realizadas durante o ano de 2012. Como fonte de estudo nossa ferramenta de software conseguiu coletar aproximadamente 90 mil mensagens distintas contendo algum tipo de referência a um partido político ou mesmo a um político em específico.

Palavras-chave: *Redes Sociais, Eleições, Mídia, Análise.*

ABSTRACT

Social networks have become extremely popular leading to the emergence and popularization of a growing wave of web-based applications. These environments are primarily used as a means to share ideas and exchange information between individuals who have interests and / or common goals so that the presence of such media provides a means of special discussion that the conventional ways can not perform. In this sense, the present research aims to investigate whether social networks are actually used for political deliberation environments and through this we can obtain similar information obtained from those polls election. For this we use the context of municipal elections in the city of João Pessoa held during the year 2012. As our study source software tool able to collect about 90 000 distinct messages containing some kind of reference to a political party or even a politician in particular..

Keywords: *Social Network, Elections, Media, Analysis.*

LISTA DE ABREVIATURAS E SIGLAS

ABEP – *Associação Brasileira de Empresas de Pesquisa*
API – *Application Programming Interface*
ARS – *Análise de Redes Sociais*
BRASNAM – *Brazilian Workshop on Social Network Analysis and Mining*
CETIC – *Centro de estudos sobre Tecnologias de Informação e Comunicação*
CSBC - *Congresso da Sociedade Brasileira de Computação*
CSV – *Comma-Separated Values*
HITS - *Hyperlink-Induced Topic Search*
IBGE – *Instituto Brasileiro de Geografia e Estatística*
IBOPE - *Instituto Brasileiro de Opinião Pública e Estatística.*
IC – *Intervalo de Confiança*
IDP – *Instituto Data Popular*
IM – *Instant Message*
JSON – *JavaScript Object Notation*
JUNG – *Java Universal Network Graph*
LABSNA – *Laboratory of Social Network Analysis*
MVC – *Model, View, Controller*
PDCA - *Plan, Do, Check e Act*
PPGI – *Programa de Pós-Graduação em Informática*
SE – *Seção Eleitoral*
SMS - *Short Message Service*
SNA – *Social Network Analysis*
TIC – *Tecnologia da Informação e Comunicação*
TRE – *Tribunal Regional Eleitoral*
TSE – *Tribunal Superior Eleitoral*
UFPB – *Universidade Federal da Paraíba*
URL – *Uniform Resource Locator*
XML – *Extensible Markup Language*
ZE – *Zona Eleitoral*

LISTA DE FIGURAS

FIGURA 1 - Divisao dos bairros por zona.	36
FIGURA 2 - Intensidade de população nos bairros.	36
FIGURA 3 - Proporção de domicílios (em %).	45
FIGURA 4 - Proporção de domicílios (em %) por região.	46
FIGURA 5 - Percentual sobre toda a população.	46
FIGURA 6 - Proporção de indivíduos que nunca usaram a internet.	47
FIGURA 7 - Proporção de indivíduos que utilizam redes sociais.	48
FIGURA 8 - Proporção de indivíduos que usam o Twitter.	53
FIGURA 9 - Problema da ponte de Königsberg representado em grafo.	58
FIGURA 10 - Dígrafo Simples.	58
FIGURA 11 - Dois grafos. (a) Grafo simples. (b) Grafo completo.	59
FIGURA 12 - As redes representadas por matrizes de adjacência.	59
FIGURA 13 - Rede complexa aleatória.	61
FIGURA 14 - Rede complexa de pequeno-mundo.	61
FIGURA 15 - Rede complexa livre de escala.	63
FIGURA 16 - Exemplo de uma rede com estrutura modular.	64
FIGURA 17 - Nós com maior grau de intermediação.	69
FIGURA 18 - Hub e Autoridade.	70
FIGURA 19 - Rede com três subgrupos (cliques).	71
FIGURA 20 - Crescimento dos usuários do Twitter (em milhões).	83
FIGURA 22 - Proposta inicial da arquitetura.	88
FIGURA 23 - Fluxo de dados da <i>AppSNAElection</i>	90
FIGURA 24 - Metodologia para a construção da rede.	92
FIGURA 25 - Número de <i>Tweets</i> por semana.	95
FIGURA 26 - Número de Pessoas distintas por semana.	95

FIGURA 27 - Média de <i>Tweets</i> diários por semanas.	96
FIGURA 28 - Evolução das pesquisas eleitorais.....	96
FIGURA 29 - Número de <i>Tweets</i> diários.....	97
FIGURA 30 - Comunidades encontradas na rede.	98
FIGURA 31 - Partição do grafo de acordo com o grau ponderado.	99
FIGURA 32 - Partição do grafo de acordo com o grau.	100
FIGURA 33 - Número de <i>Tweets</i> por semana.	108
FIGURA 34 - Número de pessoas distintas por semana.	108
FIGURA 35 - Média de <i>Tweets</i> diários por semanas.	108
FIGURA 36 - Evolução das pesquisas eleitorais.....	109
FIGURA 37 - Número de <i>Tweets</i> diários.....	109
FIGURA 38 - Comunidades encontradas na rede.	111
FIGURA 39 - Partição do grafo de acordo com o grau ponderado.	112
FIGURA 40 - Partição do grafo de acordo com o grau.	112

LISTA DE TABELAS

TABELA 1 - Classificação etária.	34
TABELA 2 - Classificação por estado civil.	35
TABELA 3 - Classificação por escolaridade.	35
TABELA 4 - Distribuição da População de acordo com as regiões da cidade.	37
TABELA 5 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Norte.	38
TABELA 6 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Leste.	39
TABELA 7 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Oeste.	40
TABELA 8 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Sul.	40
TABELA 9 - Bairros com maior poder aquisitivo.	42
TABELA 10 - Bairros de população elevada.	43
TABELA 11 - Principais Redes Sociais.	57
TABELA 12 - Tecnologias utilizadas na construção da arquitetura.	89
TABELA 14 - Resultado do pleito eleitoral do 1º turno.	97
TABELA 15 - Total de votos nos bairros de maior poder aquisitivo.	97
TABELA 16 - Comparação dos resultados com a métrica Modularidade.	99
TABELA 17 - Comparação dos resultados com as métricas de grau.	101
TABELA 18 - Comparação entre as demais métricas ARS.	101
TABELA 19 - Menções (<i>Tweets</i>) com dados de geolocalização.	102
TABELA 20 - Distribuição das Menções (<i>Tweets</i>) por bairro (1).	102
TABELA 21 - Distribuição das Menções (<i>Tweets</i>) por bairro (2).	102
TABELA 22 - Dados para o cálculo do Intervalo de confiança.	103
TABELA 23 - Análise dos dados do Intervalo de confiança.	103
TABELA 24 - Novos dados para o cálculo do Intervalo de confiança.	104

TABELA 25 - Análise dos novos dados do Intervalo de confiança.....	104
TABELA 26 - Comparação geral dos resultados.....	105
TABELA 28 - Resultado do pleito eleitoral do 2º turno.....	110
TABELA 29 - Total de votos nos bairros de maior poder aquisitivo.	110
TABELA 30 - Comparação dos resultados com a métrica modularidade.	111
TABELA 31 - Comparação dos resultados com as métricas de grau.	113
TABELA 32 - Comparação entre as demais métricas ARS.....	113
TABELA 33 - Menções (<i>Tweets</i>) com dados de geolocalização.	114
TABELA 34 - Distribuição das Menções (<i>Tweets</i>) por bairro (1).	114
TABELA 35 - Distribuição das Menções (<i>Tweets</i>) por bairro (2).	114
TABELA 36 - Dados para o cálculo do Intervalo de confiança.	115
TABELA 37 - Análise dos dados do Intervalo de confiança.	115
TABELA 38 - Comparação geral dos resultados.....	116

LISTA DE EQUAÇÕES

EQUAÇÃO 1 - Equação do grau de um vértice.	60
EQUAÇÃO 2 - Equação do comprimento do caminho mínimo médio.	62
EQUAÇÃO 3 - Equação para o cálculo da modularidade.	65
EQUAÇÃO 4 - Equação para o cálculo do Pagerank.	66
EQUAÇÃO 5 - Equação para o grau de centralidade.	67
EQUAÇÃO 6 - Equação normalizada para o grau de centralidade.	67
EQUAÇÃO 7 - Cálculo do grau de saída (I) e grau de entrada (II).	68
EQUAÇÃO 8 - Equação para centralidade de intermediação.	68
EQUAÇÃO 9 - Equação normalizada para centralidade de intermediação.	68
EQUAÇÃO 10 - Equação para centralidade de proximidade.	69
EQUAÇÃO 11 - Equação normalizada para centralidade de proximidade.	69

SUMÁRIO

Resumo

Abstract

Lista de Siglas e Abreviaturas

Lista de Figuras

Lista de Tabelas

Lista de Equações

1. Introdução	19
1.1 Motivação	20
1.2.21 Objetivos	21
1.3 Contribuição	22
1.4 Relevância	23
1.5 Estrutura do Trabalho	24
2. Trabalhos correlatos	25
2.1 Trabalhos desenvolvidos pelo LabSNA	25
2.2 Trabalhos em outras áreas	26
2.3 Trabalhos relacionados com eleições	27
2.4 Considerações.....	29
3. Embasamento teórico	32
3.1 Conceitos sobre a cidade de João Pessoa	33
3.1.1 Perfil Eleitoral	33
3.1.2 Perfil demográfico.....	35
3.1.3 Perfil econômico	37
3.1.4 Considerações	41
3.2 Acesso a informação e debate político	44
3.2.1 Demografia do uso das mídias.....	44
3.2.2 O uso das Mídias para o debate político	48
3.2.3 Conceitos sobre pesquisas eleitorais	51
3.2.4 Considerações	53
3.3 Redes Sociais.....	55
3.3.1 Conceitos Básicos.....	55

3.3.2 Teoria dos grafos	57
3.3.3 Modelos de Redes	60
3.3.4 Métricas para o Estudo de Redes Complexas	63
3.3.5 Ferramentas para Análise de Redes Sociais.....	72
3.3.5 Considerações	75
3.4 Rede Social Twitter.....	77
3.4.1 Conceitos Básicos.....	77
3.4.2 Twitter no debate político	79
3.4.3 Acesso aos dados do Twitter	81
3.4.4 Considerações	83
4. Metodologia	85
4.1 Metodologia de Análise.....	85
4.2 Arquitetura da <i>AppSNAElection</i>	87
4.3 Funcionamento da <i>AppSNAElection</i>	90
4.4 Modelagem da Rede	91
4.5 Considerações.....	93
5. Análise dos resultados	94
5.1 Resultados do 1º Turno	94
5.1.1 Estatísticas do Twitter e pesquisas eleitorais	94
5.1.2 Resultado final da eleição	97
5.1.3 Aplicação de métricas de ARS	98
5.1.4 Aspectos sobre Geolocalização	101
5.1.5 Utilização do Intervalo de Confiança	103
5.1.6 Considerações sobre o 1º Turno	105
5.2 Resultados do 2º Turno	107
5.2.1 Estatísticas do Twitter e pesquisas eleitorais	107
5.2.2 Resultado final da eleição	110
5.2.3 Aplicação de métricas de ARS	110
5.2.4 Aspectos sobre Geolocalização	113
5.2.5 Utilização do intervalo de Confiança	115
5.2.6 Considerações sobre o 2º Turno	116
5.3 Publicações.....	117
6. Considerações	119
6.1 Limitações do trabalho	119

6.2 Discussão	120
6.3 Trabalhos Futuros	122
6.4 Considerações Finais	122
Referências Bibliográficas	123
Anexos	130
Anexo A – Seções eleitorais e eleitores por bairro (TRE-PB, 2012).	131
Anexo B – Divisão dos bairros de João Pessoa por região (TSJP, 2009).	132
Anexo C – Intensidade de população dos bairros (TSJP, 2009).....	133
Anexo D - Mapa da exclusão/inclusão social (TSJP, 2009).....	134
Anexo E – Incidência % de responsáveis por domicílio (TSJP,2009)	135
Anexo F - Programas ARS (Adaptada de Huisman e Van Duijn, 2004).	136
Anexo G – API's para ARS (Adaptada de Huisman e Van Duijn, 2004).....	138
Anexo H – Rede com partição Modularidade (1º Turno) – (1)	139
Anexo I – Rede com partição Modularidade (1º Turno) – (2).....	140
Anexo J – Rede com partição Modularidade e Grau ponderado (1º Turno)	141
Anexo K – Rede com partição Modularidade e Grau (1º Turno)	142
Anexo L – Rede com partição Modularidade (2º Turno) – (1)	143
Anexo M – Rede com partição Modularidade (2º Turno) – (2)	144
Anexo N – Rede com partição Modularidade e Grau ponderado (2º Turno)	145
Anexo O – Rede com partição Modularidade e Grau (2º Turno).....	146

1. Introdução

Com a crescente utilização da internet e, em particular dos serviços fornecidos pela camada web, temos uma mudança no paradigma da interação entre os grupos sociais onde esta atividade, que outrora era realizada de modo presencial, hoje acontece naturalmente também no meio virtual. A cada dia surgem novas e diferentes formas de interação entre pessoas no meio virtual sendo as mais comuns às chamadas Redes Sociais *Online* tais como o *Facebook*, *Twitter* e *Orkut*.

De acordo com Stanley e Faust (1994), as redes sociais podem ser utilizadas como meio para compartilhamento de ideias e troca de informações entre indivíduos que possuem interesses ou objetivos em comum. Por tratar-se de redes de interação, muitos conceitos já consolidados em outras áreas como a antropologia, sociologia e psicologia podem ser usados para analisar o comportamento de cada indivíduo bem como da rede de relacionamento como um todo, mesmo sendo estas redes formadas a partir de meios virtuais. Tais estudos indicam que o comportamento dos indivíduos nas redes sociais *online* possuem características semelhantes ao comportamento dos indivíduos nas redes sociais convencionais (Simões *et al*, 2011).

A popularização dessas novas formas de disseminação de conteúdo através das redes sociais tem chamado à atenção de diversos segmentos da sociedade. No caso específico do ambiente eleitoral essa nova forma de fazer política tem como princípio o envolvimento entre políticos profissionais e a população, eleitora ou não. E esse envolvimento supõe a superação do modelo de mão única, em que fala o político e o povo apenas escuta (como visto no horário eleitoral ou mesmo comícios); e propõe uma via de mão dupla, tal como o “corpo a corpo” que se faz comumente nas ruas, mas com a possibilidade de alcançar e garantir voz a mais pessoas de maneira direta, podendo dispensar a intermediação da mídia tradicional (Silva, 2012b).

Em diversos países do mundo os principais atores envolvidos em um pleito eleitoral tem utilizados as redes sociais para diversos fins, como o caso emblemático das eleições presidenciais americanas no ano de 2008. No caso do Brasil percebemos que o uso massivo deste tipo de mídia ocorreu apenas a

partir das eleições presidenciais do ano de 2010. Desde então os candidatos aos cargos públicos têm recorrido com maior frequência às redes sociais como forma de interação direta com seus eleitores e também como plataforma para angariar apoio a votos dos usuários dessas redes. Em contrapartida os eleitores também utilizam as redes sociais para promover seus candidatos favoritos e criticar seus adversários através da publicação de conteúdo politicamente relevante através destes ambientes.

1.1 Motivação

A pesquisa de intenção de voto realizada pelas diversas instituições e empresas da área tem como principal finalidade projetar um cenário quantitativo de um determinado público ou região. Nos dias atuais podemos afirmar que este artefato figura como o meio mais utilizado por políticos para medir seu grau de aceitação/rejeição e também para os eleitores analisarem a probabilidade do candidato de sua preferência vencer o pleito eleitoral em questão.

Este tipo de ferramenta considera os aspectos metodológicos e os processos contemporâneos de formação das opiniões políticas e suas influências, objetivando, em síntese, analisar o desempenho e a aceitação dos candidatos ou partidos, traçando um perfil, do momento, sob a ótica do eleitorado, elementos fundamentais para futuras e imediatas posturas dos candidatos, em suas campanhas, de forma a adaptarem suas propostas às expectativas dos eleitores.

Contudo, diante do último pleito eleitoral para presidência da república federativa do Brasil realizado no ano de 2010, podemos afirmar que ocorreram algumas disparidades entre as projeções das pesquisas previamente realizadas e o resultado das eleições. Isto foi confirmado através da análise dos resultados finais das eleições onde algumas pesquisas não conseguiram, mesmo com a margem de erro imposta, apontar com razoável fidelidade o resultado do final do pleito. A matéria intitulada “Muito além da margem de erro” veiculada na Revista Época¹ reafirma a ideia anterior conforme trechos abaixo:

¹ Revista Época - Exemplar de 11 de outubro 2010 - Nº 647, p. 68.

“No primeiro turno da eleição, os institutos que fazem pesquisa de intenção de voto travaram uma disputa paralela à eleitoral”. “A disparidade de previsões de algumas pesquisas colocou a credibilidade dos resultados em xeque. (...) Com a apuração das urnas, o coro de suspeitas ganhou eco por causa da diferença entre a votação final e a expectativa criada pelas pesquisas. (...)”

“Se os institutos erraram até na pesquisa que entrevistou quem acabou de votar, será que não estavam errando o tempo todo?, diz o cientista político Jairo Nicolau, uma das vozes mais críticas aos institutos. As pesquisas não estão conseguindo captar a intenção dos eleitores, isso é grave. (...)”

É dentro deste contexto perante os meios convencionais de pesquisas eleitorais que este trabalho surge para analisar as informações provenientes de usuários das redes sociais com vistas a subsidiar este meio de pesquisa.

1.2 Objetivos

O presente trabalho tem como objetivo principal demonstrar que através de uma análise minuciosa da interação e relacionamento entre os usuários de uma rede social durante o período eleitoral, podemos obter informações semelhantes daquelas obtidas através de pesquisas de opinião eleitoral, com vistas a prever o resultado final do pleito.

Os objetivos específicos a seguir são de fundamental importância para o seu desenvolvimento do presente projeto de pesquisa:

- Realizar uma revisão bibliográfica abrangente sobre redes sociais, formas de representação e medidas para análise visando à definição de uma metodologia básica para realizar a análise de dados.
- Pesquisar as principais tecnologias envolvidas para criação do software para coleta de dados da rede social Twitter, escolhida como meio de estudo do presente trabalho.
- Identificar palavras ou termos relevantes que possam ser associados a um dos candidatos ao cargo de prefeito da cidade de João Pessoa para que o software possa filtrar as mensagens relevantes.

- Durante o processo de coleta das mensagens do Twitter, analisar quais delas contém informações de geolocalização no intuito de realizar uma comparação com os dados provenientes destes locais e os dados obtidos na votação cujas seções estejam situadas nestes locais.
- Identificar métricas capazes de definir 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;
- Buscar estratégias para correlacionar o perfil e o comportamento das pessoas de determinado grupo social, com a sua conduta dentro das redes sociais virtuais.

1.3 Contribuição

Este trabalho traz como contribuição uma metodologia utilizada para realizar uma análise quantitativa da interação entre os usuários de uma rede social. A partir deste trabalho pretendemos avançar o estado da arte relativo ao tema de pesquisa no sentido de obtermos futuramente novos métodos preditivos para vários contextos em apoio aos métodos utilizados convencionalmente.

A segunda contribuição deste trabalho é um modelo de processo para o desenvolvimento de aplicações para análise de redes sociais, que tem como objetivo analisar principalmente os aspectos relacionados à quantidade, procurando identificar problemas, falhas e melhorias que possam ser feitas durante todo processo de desenvolvimento.

Por fim, a discussão da aplicação de software, do seu processo de desenvolvimento, das tecnologias e ferramentas envolvidas bem como da metodologia utilizada para obtenção de dados podem estimular pesquisas científicas para o desenvolvimento de novos meios e métodos que possam ser utilizados para este fim, bem como podendo ser apontadas melhorias nos que foram utilizados durante este trabalho.

1.4 Relevância

Para os candidatos a cargos eleitorais pensar em uma candidatura ou mesmo em administrar uma cidade, estado ou país nos dias de hoje sem o auxílio de informações estratégicas pode ser uma tarefa impraticável. No caso dos eleitores, estes necessitam receber informações qualitativas e quantitativas dos diversos candidatos em questão para facilitar o processo decisório do voto. Neste contexto podemos afirmar que o principal meio utilizado para obtenção destas informações são as pesquisas eleitorais. São inúmeras as informações factíveis de serem extraídas desse tipo de pesquisas tais como: Qual o potencial de um candidato? Quais são os interesses da população? Onde há mais chances votos? Qual o índice de rejeição?

Por outro lado podemos concluir também que o uso das pesquisas eleitorais apesar de fornecerem boas informações a candidatos e eleitores, possuem além da sua margem de erro, certo grau de atraso em relação ao período em que estas são captadas até o momento em que são processadas e entregues as partes interessadas. Ainda, devido a quantidade de pessoas envolvidas na tarefa de captar as informações junto a população e de todo apoio técnico para dar suporte a esta atividade, a realização deste tipo de pesquisa possuem um custo relativamente elevado.

No caso das redes sociais, estas têm força de manipulação, velocidade e fragmentação do que acontece o tempo todo. Hoje, no Brasil, mais de 80 milhões de pessoas têm acesso à internet, sendo que 90% desse universo faz parte de alguma mídia social (Recuero, 2009). Com o grande número de pessoas conectadas, a comunicação na rede pode ser considerada uma ferramenta tão relevante para as campanhas que, em termos de importância no ranking dos meios de comunicação, ficaria atrás apenas dos meios televisivos.

Deste modo a partir da definição de uma metodologia para coleta dos dados gerados a partir destes ambientes e de sua posterior análise através dos princípios da teoria dos grafos e redes complexas, poderíamos compor uma ferramenta para medir a discussão política de modo a fornecer diversas informações aos principais atores de eleição.

Ainda, este tipo de aplicação poderia ser utilizado em apoio às pesquisas eleitorais convencionais, cabendo à ressalva de que os custos envolvidos em sua realização bem como o tempo de resposta são inferiores se comparado a modalidade supracitada.

1.5 Estrutura do Trabalho

O presente trabalho encontra-se dividido em seis capítulos. O [Capítulo 1](#) apresenta o trabalho de maneira geral descrevendo sua motivação, objetivos, contribuição, relevância e estrutura da dissertação. No [Capítulo 2](#) encontram-se descritos os trabalhos correlatos ao nosso tema de pesquisa. No [Capítulo 3](#) encontra-se descrito o embasamento teórico do trabalho com os conceitos e definições relacionados aos processo eleitoral brasileiro, o perfil demográfico e econômico do município de João Pessoa; os principais conceitos relativos a redes sociais e a metodologia para análise destas. No [Capítulo 4](#) é apresentado o desenvolvimento do trabalho propriamente dito com a metodologia utilizada no estudo, arquitetura da aplicação de software utilizada e as etapas necessárias para extração de dados na rede social *twitter*. No [Capítulo 5](#) estão sumarizados os resultados obtidos nos experimentos realizados a partir dos dados obtidos durante o 1º e 2º turno das eleições municipais de João Pessoa. Ainda, estão discriminados os artigos científicos submetidos e publicados a partir da metodologia proposta no estudo. Por fim, no [Capítulo 6](#) temos as limitações do trabalho, a discussão a respeito dos resultados obtidos, as sugestões de trabalhos futuros e considerações finais do presente trabalho de mestrado. As informações complementares do projeto são apresentadas nos diversos [Anexos](#) deste trabalho.

2. Trabalhos correlatos

Desde 2006, *microblogging* tornou-se uma ferramenta cada vez mais utilizada para comunicação na Internet. Recentemente podemos verificar a existência de diversos trabalhos sugerindo que os dados produzidos no Twitter podem possuir um impressionante poder preditivo. Verificamos trabalhos relacionados aos mais diversos temas, onde todos estes contextos estão passíveis de previsões a partir dos dados de usuários desta rede social.

O objetivo deste capítulo é apresentar os principais trabalhos relacionados direta ou indiretamente ao tema de pesquisa no intuito de verificar de que modo o twitter está sendo utilizado com vistas a fornecer meios de previsão de diversos acontecimentos do mundo real. Com isto este capítulo será dividido em quatro seções: Trabalhos correlatos encontrados desenvolvidos pelo LabSNA ([Seção 2.1](#)); Trabalhos correlatos em outras áreas ([Seção 2.2](#)); Trabalhos que utilizaram as redes sociais, em particular o twitter, no contexto de eleições ([Seção 2.3](#)) e; por fim considerações ([Seção 2.4](#)) onde iremos discutir brevemente o estado da arte da utilização de redes sociais com vistas a realizar previsões e apontar os principais diferenciais da presente pesquisa frente as técnicas utilizadas até o momento da realização da pesquisa.

2.1 Trabalhos desenvolvidos pelo LabSNA

O programa de pós-graduação em informática (PPGI) da Universidade federal da Paraíba (UFPB) dispõe de um Grupo de Trabalho para pesquisas de temas relevantes na área de redes sociais denominado LabSNA. Dentre os trabalhos desenvolvidos recentemente citamos **“Análise Quantitativa das Relações entre Usuários no Twitter”** (Oliveira, 2013). Este trabalho teve como propósito investigar a viabilidade de utilização da rede social Twitter e das métricas de análise de redes sociais, como caminhos para caracterizar o perfil de relacionamento e interação entre os usuários em diferentes grupos sociais. Para isto foi desenvolvida uma metodologia de análise de redes sociais apoiada por um conjunto de ferramenta de coleta e análise de dados, sendo realizados alguns estudos de caso na área de educação e de concessão de crédito financeiro.

2.2 Trabalhos em outras áreas

Cheng *et al* (2011) investigaram o Twitter como uma ferramenta para monitorar e capturar as tendências e padrões emergentes sobre conhecimento de tempo crítico. Eles avaliaram extensivamente um quadro de recomendação baseado em difusão e propuseram um algoritmo que utilizava dados do Twitter recolhidos durante o surto inicial da gripe H1N1. Lamos e Cristianini (2010) mostraram um método que se baseia na análise de centenas de milhares de tweets por dia, em busca de declarações de sintomas relacionados a alguma doença. Este estudo foi realizado no Reino Unido por um período de 24 semanas durante o período da pandemia de gripe H1N1.

Estudos de Bruns e Liang (2012) e Cheong e Cheong (2011) analisaram a comunicação no Twitter baseado no contexto de desastres naturais, centrando-se sobre as inundações ocorridas na Austrália durante o ano de 2011. Através de métodos de análise de redes sociais, eles descobriram que diversos grupos de atores, incluindo os moradores dos locais afetados, os serviços de emergência e os componentes da imprensa em geral, desempenharam papéis importantes no intercâmbio de informações sobre o desastre.

Outros estudos (Vis, 2012; Lotan *et al*, 2011) examinaram o uso do Twitter durante a grande agitação civil no contexto dos motins realizados no Reino Unido e nos levantes da Primavera Árabe em uma série de países do Norte Africano e do Oriente Médio. Eles identificaram uma ampla gama de uso das mídias sociais para a divulgação de informações relativa a cada um dos incidentes. Analisaram também o uso de outros canais de mídia e informação “boca-a-boca”, apontando diferenças significativas entre as informações obtidas pelos observadores locais e pelos mais distantes destes eventos.

Na área comercial foi analisado o trabalho de Stieglitz e Krüger (2011) que investigaram a crise envolvendo a fabricante de automóvel Toyota no ano de 2010 e mostrou que, medido pelo número de *tweets* publicados, as discussões sobre a crise são caracterizadas por períodos de pico e períodos de silêncio na comunicação de questões relacionadas à empresa. Além disso, eles descobriram que os usuários indiretamente envolvidos no debate pelo Twitter desempenham um papel importante na discussão (por exemplo, através da

publicação de uma quantidade significativamente elevada de *tweets* pode ser gerada uma grande quantidade de retweets).

Park *et al* (2009) investigaram a crise envolvendo a marca Domino's Pizza através da análise da difusão de más notícias através do Twitter. Eles classificaram separadamente sentimentos nos *tweets* gerados pelos clientes e os gerados pela própria empresa. Baseado em seu estudo provam que a difusão das notícias ruins é mais rápida do que a difusão de outros tipos de conteúdo (por exemplo, pedidos de desculpas).

2.3 Trabalhos relacionados com eleições

Muito embora a ferramenta Twitter exista desde o ano de 2006, podemos verificar a publicação de diversas pesquisas a partir do ano de 2008 acompanhando a discussão política em diversas partes do mundo como as eleições americanas (Diakopoulos e Shamma, 2010), na Nigéria (Ifukor, 2010), na Suécia (Larsson, 2011), em Singapura (Skoric *et al*, 2012), no Reino Unido (Anstead, 2012) e diversos outros países, inclusive no Brasil. Este tema ganhou notável discussão principalmente após as publicações de Tumasjan *et al* que concluíram positivamente a respeito da possibilidade de prever resultados eleitorais utilizando dados provenientes desta rede social.

O primeiro trabalho "***Predicting elections with twitter: What 140 characters reveal about political sentiment***" (Tumasjan *et al*, 2010) tem como finalidade verificar se as mensagens do *microblog* refletiram o sentimento político no contexto das eleições federais da Alemanha no ano de 2009. Para este estudo foi realizada uma análise de sentimentos de mais de 100.000 mensagens que continham uma referência a um político ou mesmo ao partido ao qual este pertence.

No segundo estudo "***Elétron Forecasts With Twitter: How 140 Characters Reflect the Political Landscape***" (Tumasjan *et al*, 2011) são expostos resultados que demonstram que o Twitter é realmente utilizado de modo extensivo para a deliberação política na Alemanha. É sugerido que o simples número de mensagens (*tweets*) que mencionam um partido/candidato pode refletir o resultado das eleições. Ainda, foi relatada uma análise de

sentimento dos *tweets* em grau mais elevado que o trabalho anterior, indicando que o conteúdo de mensagens do Twitter pode refletir plausivelmente o cenário político de uma eleição.

Após estas duas últimas pesquisas, diversos outros trabalhos surgiram em resposta criticando a metodologia empregada e as conclusões obtidas no contexto das eleições da Alemanha no ano de 2009. Citamos o trabalho de Jungher *et al* (2011) “***Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. “Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment”*** que é uma resposta direta ao trabalho de Tumasjan *et al* (2011). Como foi explicado a cerca deste trabalho, onde é alegado que o simples número de tweets é um bom indicador para o resultado das eleições e que a distribuição de tweets acompanha de perto a distribuição de votos para as diferentes partes, Jungher *et al* (2011) apontou que o método do trabalho de Tumasjan *et al* (2011) baseou em escolhas arbitrárias e, além disso, os seus resultados são variados, dependendo da janela de tempo utilizado para realização dos cálculos. Com base nestas conclusões Jungher *et al* (2011) refutou a alegação de Tumasjan *et al* (2011).

Deve-se notar que posteriormente, este trabalho fez com que Tumasjan *et al* (2011) publicassem “***Where There is a Sea There are Pirates: Response to Jungherr, Jurgens, and Schoen***” em resposta as críticas impostas no artigo anteriormente citado. Neste trabalho os autores tentaram suavizar suas conclusões anteriores, afirmando que os dados provenientes do Twitter não possuem capacidade de substituir as pesquisas convencionais, mas para simplesmente complementá-los. Ainda, declararam que o método de predição não foi a principal conclusão de seu artigo original.

O trabalho “***Limits of Electoral Predictions using Twitter***” (Gayo-Avello, 2011b) é um dos poucos trabalhos que constestam os supostos poderes de previsão de dados do Twitter sobre as eleições e sugerindo a realização de diversas atividades como linha de base para pesquisas deste tipo. Ao analisar

os resultados de um número de diferentes eleições concluiu que os dados Twitter não possuem um poder convincente de previsão de eleições.

Além da discussão internacional a respeito do assunto podemos encontrar algumas publicações relevantes no Brasil. O trabalho denominado **“O uso do twitter nas eleições de 2010: o microblog nas campanhas dos principais candidatos ao governo do Paraná”** (Cervi *et al*, 2011) reflete os resultados de uma pesquisa que analisa como os dois principais candidatos ao governo do estado do Paraná utilizaram o twitter como ferramenta de campanha durante o período eleitoral de 2010.

Por fim, temos o trabalho de título: **“Observatório da web: Uma plataforma de monitoração, síntese e visualização de eventos massivos em tempo real”** (Filho *et al*, 2010), esta pesquisa teve como principal propósito monitorar, em tempo real, o que estava sendo veiculado sobre as eleições de 2010 nas várias mídias e pelos vários usuários da Web. O seu objetivo era ajudar a traçar um panorama do cenário eleitoral do ponto de vista das informações e das opiniões que circulavam na Web e nas redes sociais online, incluindo jornais, revistas, portais e o Twitter. Foi construído um portal utilizando dezenas de ferramentas softwares inéditas de captura e análise de dados baseadas em código livre ou aberto.

2.4 Considerações

Gayo-Avello (2012) afirma existir muitas falhas e distorções nos dados que são utilizados em pesquisas. A maior parte desta problemática é causada pela metodologia empregada na coleta, tratamento e análise dos dados. Em virtude disto a maioria dos trabalhos que utilizam dados provenientes das redes sociais com vistas a prever os resultados, sejam estes eleitorais ou não, não conseguem chegar a alguma conclusão contundente.

A partir da revisão da literatura anterior, todo trabalho que se propõe a desenvolver algum tipo de pesquisa relacionada a eleições deverá ter ciência das seguintes premissas:

1) A população votante não se encontra fielmente representada nas mídias sociais, desta forma temos que considerar que as mídias sociais não

constituem uma amostra imparcial da população votante. Vamos mostrar em nossa revisão que alguns segmentos estão sub-representadas enquanto outros estão sobrerrepresentados no Twitter.

2) Com relação aos usuários do Twitter, podemos afirmar que apenas parte deste universo utiliza este meio para “falar sobre política”. Apenas uma parcela dos usuários pode ser responsável pela maior parte da conversa política e, portanto, as suas opiniões podem conduzir o que pode ser previsto a partir de mídias sociais.

3) Sabemos que existe contas no Twitter que não representam pessoas físicas votantes ou ainda, são contas “falsas” utilizadas para fins próprios. Isto implica que uma substancial quantidade de dados não é de confiança e, portanto, deverá ser descartada.

4) Métodos de análise de sentimento simplistas devem ser evitados. O discurso político é atormentado com humor e sarcasmo, o que torna a determinação da preferência política dos usuários difícil de inferir.

De acordo com os pontos elencados anteriormente, identificamos alguns pontos críticos em estudos desta natureza. Desta forma, a partir da análise dos principais estudos relacionados diretamente ao nosso tema e, visando avançar o estado da arte nesse tipo de pesquisa, nosso trabalho irá apresentar os seguintes diferenciais:

1) Enquanto a maioria dos trabalhos se limitaram a coletar os dados do Twitter apenas em uma ou duas semanas (no máximo) antes das eleições, nosso trabalho se propõe a realizar uma pesquisa mais abrangente durante um período de (no mínimo) três semanas antes da data prevista para o pleito eleitoral.

2) A forma de coleta de dados será através de termos ou palavras que façam menção a algum candidato. Serão evitados termos genéricos de partidos ou agremiações políticas uma vez que o mesmo partido pode ter representação em outras cidades o que diminuiria a confiabilidade dos dados. Durante o período de coleta serão analisadas novas palavras ou termos utilizados na campanha visando um aumento na precisão da coleta dos dados.

3) Trabalharemos com conceitos de teoria dos grafos e da teoria das redes complexas para inferirmos sobre o resultado eleitoral. Desta forma nosso trabalho será um dos pioneiros a realizar o emprego dos conceitos de redes sociais (consequentemente os conceitos de teoria dos grafos e redes complexas) nesta área de estudo. Verificamos que a maioria dos trabalhos utilizou a contagem de tweets, contagem de usuários ou mesmo a análise de sentimento para este fim.

4) Visando simplificar nosso trabalho, não utilizaremos o contexto de análise de sentimentos. Uma vez que na maioria das vezes, os classificadores baseados sentimentos são apenas ligeiramente melhor do que classificadores aleatórios. No trabalho de Gayo-Avello (2011a) demonstra que a margem de erro desta técnica está em torno de 8%, um percentual muito acima dos oferecidos pelas pesquisas eleitorais convencionais.

5) Na constituição de nossas hipóteses, iremos priorizar métodos que não valorizem a quantidade de menções realizadas por cada conta. Sabemos que podem existir contas “falsas”, contas “robôs” ou contas oficiais de campanha que realizam inúmeras menções a um determinado candidato e isto poderia influenciar negativamente os resultados obtidos.

6) Com relação à demografia, iremos propor uma análise comparativa dos nossos resultados com os resultados obtidos nos bairros de maior poder aquisitivo uma vez que partimos do pressuposto de que os maiores usuários de redes sociais são os elementos que possuem maior poder aquisitivo. Nenhum dos trabalhos anteriores partiu deste pressuposto. Isto se deve a características sociais próprias do Brasil que serão devidamente expostas durante a nossa revisão bibliográfica.

7) Por fim, utilizaremos os dados de geolocalização (caso existam) das mensagens do Twitter. Com isso poderemos identificar o local de origem das mensagens e comparar estes números com os resultados obtidos nos locais de votação correspondentes. Este tipo de análise é favorecido pelo processo de alistamento eleitoral brasileiro que implica que um cidadão irá votar no seu bairro de residência ou, na ausência de um local de votação no seu bairro, seria remanejado para um bairro próximo.

3. Embasamento Teórico

Este capítulo pretende modelar o referencial teórico que embasou a presente pesquisa. Por ser um assunto amplo e de domínio de varias áreas de conhecimento, procuramos deixar explicito apenas os principais conceitos utilizados. Iniciamos esta parte apresentando os principais conceitos sobre a cidade de João pessoa ([Seção 3.1](#)) no que tange ao perfil do eleitor, o perfil demográfico e perfil econômico. Com esta seção pretendemos compreender quem é o eleitor da capital paraibana, qual o perfil do morador desta cidade e qual classe social ele está inserido, com vistas a obtermos um referencial comparativo para ser utilizado em nossas hipóteses.

Na seção que trata sobre as questões de acesso a informação e meios para o debate político ([Seção 3.2](#)) serão expostos as principais condicionantes para o acesso a informação do ponto de vista socioeconômico utilizando como referencial dados de pesquisas realizadas no país. Isto servirá de subsidio para realizarmos a comparação dos votos de determinado segmento social. Ainda, serão discutidas as principais formas de coleta e transmissão de dados em uma eleição através das pesquisas convencionais de intenção de voto e o modo como elas são utilizadas pelos principais atores de uma eleição.

Na seção Redes Sociais ([Seção 3.3](#)) abordaremos os conceitos de redes, suas formas de representação, citaremos de modo sucinto a teoria dos grafos que servirá de suporte para uso da teoria das redes complexas e seus principais modelos. Ainda, demonstraremos as principais medidas para estudo das redes complexas (compatíveis para análise de redes sociais), falaremos das principais ferramentas utilizadas para ARS, focando principalmente naquelas utilizadas em nosso estudo.

Por fim, na seção sobre o Twitter ([Seção 3.4](#)) iremos apresentar os principais conceitos sobre a rede social utilizada para estudo. Serão expostas ainda as formas de utilização desta rede para o debate político (tanto por parte dos políticos como, principalmente, por parte dos eleitores) e uma das formas de coleta de dados dos usuários desta rede através de uma API que servirá de fonte de dados para modelagem da nossa rede utilizada na pesquisa.

3.1 Conceitos sobre a cidade de João Pessoa

O objetivo desta seção é apresentar os principais conceitos sobre a cidade de João Pessoa no sentido de apontar as principais características da população e, em particular, do eleitorado da capital paraibana. Com isto esta seção encontra-se dividida em quatro subseções: Em perfil eleitoral ([Seção 3.1.1](#)) iremos discorrer sobre os principais aspectos envolvidos em um processo eleitoral e procuraremos apontar um perfil do eleitorado da cidade; Em perfil Demográfico ([Seção 3.1.2](#)) apresentamos algumas características relativas à população, divisão do espaço urbano da cidade e distribuição dos domicílios ao longo dessa área; Em perfil econômico ([Seção 3.1.3](#)) serão demonstrados com base nos dados do IBGE (2000, 2010) detalhes sobre a distribuição de renda com vistas a determinar um conjunto de bairros de poder aquisitivo relevante para utilização como grupo de estudo; Finalizando, na seção considerações ([Seção 3.1.4](#)) iremos expor quais serão os candidatos que serão monitorados no nosso estudo, bem como mostraremos como foi selecionado o grupo de bairros para compor nossa hipótese de votação nos bairros de maior poder aquisitivo.

3.1.1 Perfil Eleitoral

De acordo com o exposto no Glossário Eleitoral Brasileiro (2013) do Tribunal Superior Eleitoral (TSE), toda cidade sujeita as atividades de um processo eleitoral deve organizar seus locais de votação de acordo com Zonas Eleitorais. Estas zonas são regiões geograficamente delimitadas, gerenciadas por um cartório eleitoral, e que podem ser compostas por mais de um município ou apenas parte dele, como ocorre nas cidades mais populosas.

Cada Zona Eleitoral é subdividida em Seções Eleitorais. Este é o local onde fica instalado a urna eletrônica - equipamento responsável pelo registro dos votos dos eleitores. Um dos critérios para a associação de um eleitor a uma Seção Eleitoral é a proximidade do local de residência deste em relação ao local de votação. A título de exemplificação, eleitores que moram em um determinado bairro muito provavelmente votarão neste bairro, ou mesmo em alguma seção alocada em outro bairro próximo de sua residência.

Um problema que ocorre devido à legislação brasileira é que essa associação é realizada de modo permanente, fazendo com que o eleitor, embora se mude para outro bairro de uma mesma cidade, continue votando no mesmo lugar, salvo quando transfere seu título de eleitor para outra cidade, estado ou país. Isto acontece em virtude de não existir a possibilidade de troca entre seções eleitorais.

De acordo com o exposto no item anterior, podemos afirmar que uma cidade pode ser constituída por mais de uma zona eleitoral. Desta maneira o município de João Pessoa é subdividido em cinco zonas eleitorais exclusivas: 01, 64, 70, 76 e 77. Cada Zona eleitoral possui um conjunto de seções eleitorais sendo que o município de João Pessoa possui 1.283 destas seções distribuídas ao longo de 202 Locais de Votação, abrangendo um eleitorado de aproximadamente 480.237 pessoas. No [Anexo A](#) encontra-se descritas algumas dessas estatísticas onde podemos verificar a distribuição do eleitorado e locais de votação de acordo com os bairros da capital paraibana.

Buscando identificar padrões comuns entre os eleitores da cidade de João Pessoa, de posse de alguns dados obtidos junto ao TSE e ao TRE-PB, classificamos o eleitorado de acordo com alguns critérios. O primeiro critério descrito na **Tabela 1** diz respeito à divisão da faixa etária por sexo. De acordo com os dados podemos concluir que 55,3% do eleitorado são do sexo feminino e que 53,4% do eleitorado encontram-se na faixa etária entre 17 e 40 anos.

TABELA 1 - Classificação etária.

Fonte: TRE-PB (2012).

Sexo	Até 17	18 a 40	41 a 55	56 a 69	70 a 79	79	Total
Fem	2.300	133.550	70.562	37.266	13.534	7.187	264.399
Masc	2.170	117.391	55.276	26.646	8.196	4.093	213.772
NI	0	0	21	27	10	12	70

* NI – Não informado.

O segundo critério de classificação descrito na **Tabela 2**, apresenta informações a respeito da divisão do eleitorado de acordo com o estado civil. A respeito deste dado concluímos que 66,55 % do eleitorado possui estado civil solteiro. Ainda percebemos que a quantidade de mulheres separadas ou divorciadas supera em três vezes o número de homens com estas situações.

TABELA 2 - Classificação por estado civil.
Fonte: TRE-PB (2012).

Sexo	Solteiro	Casado	Viúvo	Separado	Divorciado	NI	Total
Fem	169.862	78.628	6.645	2.493	6.666	105	264.399
Masc	148.394	59.731	923	1.647	2.995	82	213.772
NI	20	28	3	2	0	17	70

* NI – Não informado.

Finalizando, o terceiro critério de classificação diz respeito à divisão do eleitorado de acordo com o grau de escolaridade. Com os dados expostos na **Tabela 3** chega-se a conclusão de que 32,7% deste universo possuem grau de escolaridade de 2º grau completo ou superior e que 46,65% possuem 1º grau completo ou inferior.

TABELA 3 - Classificação por escolaridade.
Fonte: TRE-PB (2012).

Sexo	A	LE	1º I	1º C	2º I	2º C	3º I	3º C	Total
Fem	9.869	23.800	69.651	15.612	53.097	51.363	14.742	26.214	264.348
Masc	7.199	19.653	64.465	12.818	45.649	34.609	13.053	16.273	213.719
NI	13	19	11	10	2	13	0	1	69

* NI – Não informado, A – Analfabeto, LE – Lê e escreve, I – Incompleto, C – Completo.

3.1.2 Perfil demográfico

De acordo com dados do IBGE contidos no censo demográfico realizado no ano de 2010, João Pessoa distribui seus 597.934 mil habitantes em pouco mais de 155 mil domicílios, ou seja, quase 4 habitantes por moradia. Certamente existe uma variação deste números em cada uma das 65 unidades territoriais com as quais se identifica a população local (59 bairros propriamente ditos e seis regiões especiais de baixa incidência populacional). No [Anexo B](#) deste trabalho pode ser verificado (em detalhes) o modo de divisão geográfica dos bairros de João Pessoa.

Mangabeira é considerado o bairro mais populoso da capital paraibana com mais de 70mil habitantes em forte contraste com a região especial denominada Mussurê que registra 18 habitantes que vivem em apenas 7 domicílios distintos. Desta forma, podemos perceber que existe uma forte discrepância na concentração de população e domicílios de João Pessoa. A divisão por bairros e regiões especiais da cidade de João Pessoa pode ser visualizada na **Figura 1** a seguir:

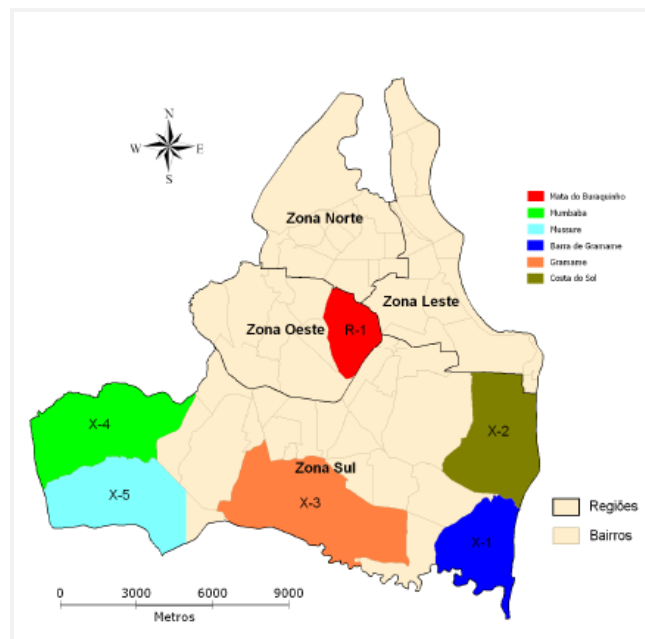


FIGURA 1 - Divisão dos bairros por zona.
Fonte: Topografia Social de João Pessoa (2009).

A distribuição populacional mostra que 40% dos bairros de João Pessoa possuem até 5.000 habitantes. Portanto, pouco mais do que 1.000 famílias. Considerando os agregados até 10.000 habitantes, verifica-se que na organização intraurbana 72% de João Pessoa têm bairros de pequeno porte. Na **Figura 2** podemos acompanhar a distribuição da população por bairro:

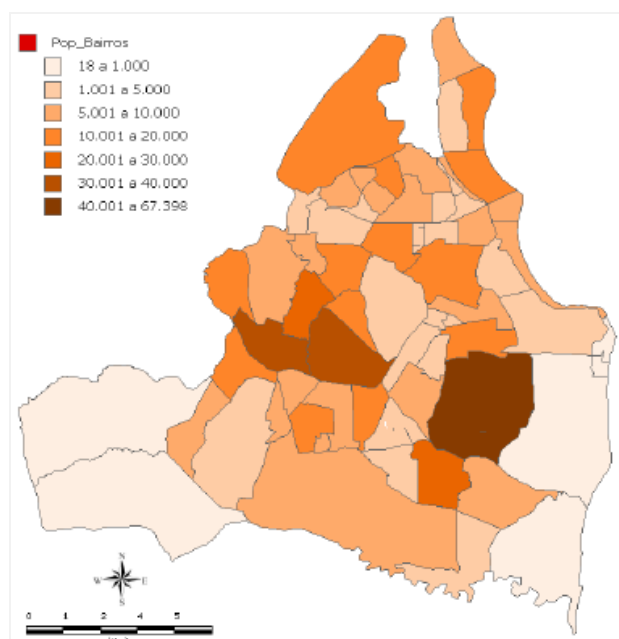


FIGURA 2 - Intensidade de População nos bairros.
Fonte: Topografia Social de João Pessoa (2009).

Com base nos dados expostos na figura acima (Veja em detalhes no [Anexo C](#)), podemos tabular estas informações com a quantidade de habitantes por zona da cidade e sua representatividade na população total do município de João Pessoa.

TABELA 4 - Distribuição da População de acordo com as regiões da cidade.

Fonte: Censo IBGE (2010)*.

Regiões	População	
	Quantidade de Habitantes	% da População Total
Zona Norte	107.849	14,91
Zona Leste	138.593	19,16
Zona Oeste	155.098	21,44
Zona Sul	317.079	43,82
Total	718.619 (*)	99,33 (*)

(*) A Tabela 4 mostra uma diferença entre a população total registrada pelo Censo do IBGE e a soma do número de habitantes por bairro, pois algumas zonas da cidade não tiveram seus dados registrados completamente, uma vez que algumas residências não puderam ter seus dados coletados ou então tratam-se de Zonas Especiais (com baixa, ou nenhuma, incidência populacional).

De acordo com a **Tabela 4**, observamos que a zona sul é mais populosa do município, abrangendo quase que metade da quantidade total de habitantes. Isso se deve ao fato de que essa zona compreende quatro dos oito bairros com população acima dos 20 mil habitantes, incluindo Mangabeira que é o bairro mais populoso, alcançando quase um quarto da população total do município. Outra zona bastante populosa é a chamada zona oeste que, embora apresente um número reduzido de bairros em relação às outras zonas, alcança o segundo lugar em número de habitantes.

3.1.3 Perfil econômico

De acordo com dados da Associação Brasileira de Empresas de Pesquisa (ABEP, 2010) uma pessoa pode ser considerada como parte componente da chamada Classe Média (ou Classe C), se a Renda Média da família estiver situada na faixa entre R\$ 1.540 e R\$ 2.313, de acordo com dados salariais relativos ao ano de 2010.

Para fins de análise, são computados apenas os valores por habitante acima de 10 anos e que possuem algum tipo de rendimento, ou seja, pessoas sem nenhum rendimento formal e abaixo dos 10 anos serão desconsideradas. Serão usados dois critérios de amostragem: o **Rendimento Nominal Médio**

Mensal, que representa uma média dos rendimentos totais recebidos pela amostra da população; e o **Rendimento Nominal Mediano Mensal**, que é basicamente o valor que divide uma amostra exatamente ao meio (50% dos elementos terão valor maior ou igual enquanto 50% terão valor menor ou igual). Todos os dados foram retirados do Censo de 2010 (IBGE, 2010).

De modo análogo ao perfil demográfico descrito na [seção 3.1.2](#), os bairros serão dispostos em grupos de acordo com a zona à qual pertencem. Com isso, podemos identificar quais os bairros que concentram a maior parte do poder aquisitivo do município, em média, identificando os que concentrem maior parte da população de Classe C ou superior. Vale salientar que os dados do IBGE (2010) não consideram a renda por domicílio, ou seja, a classificação usada no presente estudo pode não refletir a situação real das famílias de alguns bairros.

TABELA 5 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Norte.

Fonte: Censo IBGE (2010).

Bairros da Zona Norte	População com 10 anos ou mais, e com rendimentos	Rendimento Nominal Médio Mensal	Rendimento Nominal Mediano Mensal
Tambauzinho	3.074	R\$ 3.208,54	R\$ 1.500,00
B. dos Estados	4.727	R\$ 3.101,19	R\$ 2.000,00
Pedro Gondim	2.077	R\$ 2.847,59	R\$ 1.700,00
Expedicionários	2.156	R\$ 2.129,34	R\$ 1.020,00
Centro	2.255	R\$ 1.547,26	R\$ 800,00
Treze de Maio	4.511	R\$ 1.543,55	R\$ 800,00
Torre	9.412	R\$ 1.432,41	R\$ 760,00
Tambiá	1.483	R\$ 1.309,28	R\$ 779,00
Bairro dos Ipês	4.896	R\$ 1.276,62	R\$ 600,00
Roger	5.336	R\$ 787,41	R\$ 510,00
Mandacarú	6.618	R\$ 759,75	R\$ 510,00
Varadouro	1.979	R\$ 721,49	R\$ 510,00
Alto do Céu	7.110	R\$ 574,03	R\$ 510,00
Padre Zé	3.654	R\$ 539,65	R\$ 510,00
Total	59.288	R\$ 1.555,58	R\$ 893,50

Na Zona Norte, observamos os seguintes bairros com Rendimento Nominal Médio Mensal acima dos 1.540 reais: Tambauzinho e Bairro dos Estados (ambos superando a faixa dos três mil reais), Pedro Gondim, Expedicionários, Centro e Treze de Maio. O bairro da Torre, apesar de não possuir uma Renda alta (7º em sua Zona), fazendo com que fique fora da Classe C, possui uma

população elevada em relação aos demais (mais que a soma dos dois mais ricos) podendo ser utilizado para análise em algum estudo futuro.

TABELA 6 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Leste.

Fonte: Censo IBGE (2010).

Bairros da Zona Leste	População com 10 anos ou mais, e com rendimentos	Rendimento Nominal Médio Mensal	Rendimento Nominal Mediano Mensal
Cabo Branco	5.161	R\$ 4.510,36	R\$ 3.000,00
Tambaú	6.443	R\$ 4.266,78	R\$ 2.700,00
Manaíra	16.687	R\$ 3.941,03	R\$ 2.500,00
Jardim Oceania	9.594	R\$ 3.691,68	R\$ 2.500,00
Brisamar	2.496	R\$ 3.677,97	R\$ 2.000,00
Miramar	5.400	R\$ 3.482,74	R\$ 2.000,00
Ponta dos Seixas	301	R\$ 3.025,81	R\$ 1.600,00
Aeroclube	5.975	R\$ 2.927,76	R\$ 1.800,00
Portal do Sol	2.112	R\$ 2.844,59	R\$ 1.000,00
João Agripino	736	R\$ 2.599,28	R\$ 1.020,00
Bessa	7.772	R\$ 2.590,06	R\$ 1.500,00
Altiplano Cabo Branco	2.996	R\$ 2.442,13	R\$ 1.000,00
Bancários	7.038	R\$ 1.704,67	R\$ 1.000,00
Castelo Branco	6.595	R\$ 1.022,78	R\$ 550,00
Penha	376	R\$ 783,56	R\$ 510,00
São José	3.002	R\$ 469,63	R\$ 510,00
Total	82.684	R\$ 2.748,80	R\$ 1.574,38

A Zona Leste representa a área de maior concentração de renda da capital paraibana e, em virtude disso, uma das áreas de melhor qualidade de vida (Veja [anexo D](#)) e uma das áreas que concentra a população de maior nível de instrução (Veja [anexo E](#)). De acordo com a **Tabela 6** apenas três bairros não podem ser considerados Classe C ou superior, de acordo com os dados do nosso estudo, uma vez que o Castelo Branco, Penha e São José não conseguiram atingir a Renda Média Mensal mínima para esta classe.

Podemos destacar os bairros de Cabo Branco e Tambaú que detém uma Renda Média Mensal acima dos R\$ 4.000. O bairro de Manaíra consegue, mesmo sendo mais populoso que os dois juntos, chegar próximo aos 4 mil reais em rendimentos mensais. Outro bairro com população considerável e que também merece destaque é o Jardim Oceania com seus quase 10 mil habitantes e com uma renda média acima dos R\$ 3.500.

TABELA 7 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Oeste.

Fonte: Censo IBGE (2010).

Bairros da Zona Oeste	População com 10 anos ou mais, e com rendimentos	Rendimento Nominal Médio Mensal	Rendimento Nominal Mediano Mensal
Jaguaribe	8.875	R\$ 1.378,78	R\$ 800,00
Cristo Redentor	19.967	R\$ 1.029,53	R\$ 510,00
Trincheiras	3.421	R\$ 784,77	R\$ 510,00
Cruz das Armas	13.052	R\$ 725,98	R\$ 510,00
Alto do Mateus	7.785	R\$ 678,16	R\$ 510,00
Varjão	8.390	R\$ 668,19	R\$ 510,00
Oitizeiro	14.476	R\$ 638,03	R\$ 510,00
Ilha do Bispo	3.028	R\$ 540,66	R\$ 510,00
Total	78.994	R\$ 805,51	R\$ 546,25

Na Zona Oeste de João pessoa percebemos que, de acordo com os dados da **Tabela 7**, não existe nenhum bairro que esteja de acordo com os critérios para Classe C. Cabe comentário de que o bairro de Jaguaribe é o mais próximo destes critérios de classificação. Ainda, há mais três bairros que merecem destaque por serem bastante populosos: Cristo Redentor, Oitizeiro e Cruz das Armas. Cabe a ressalva de que estes bairros também poderão servir como base para algum estudo futuro.

TABELA 8 - Quantidade de habitantes acima dos 10 anos e seus rendimentos Nominal Médio Mensal e Mediano Mensais por bairros da Zona Sul.

Fonte: Censo IBGE (2010).

Bairros da Zona Sul	População com 10 anos ou mais, e com rendimentos	Rendimento Nominal Médio Mensal	Rendimento Nominal Mediano Mensal
Anatólia	761	R\$ 2.036,68	R\$ 1.400,00
Jardim C. Universitária	12.764	R\$ 1.854,20	R\$ 1.200,00
Jardim São Paulo	2.915	R\$ 1.744,96	R\$ 1.200,00
Água Fria	3.708	R\$ 1.577,96	R\$ 1.100,00
Cuiá	3.876	R\$ 1.229,27	R\$ 800,00
Ernesto Geisel	8.219	R\$ 1.115,34	R\$ 780,00
José Américo	8.598	R\$ 1.055,06	R\$ 600,00
Cidade dos Colibris	2.085	R\$ 965,27	R\$ 586,00
Valentina	12.371	R\$ 897,46	R\$ 600,00
João Paulo II	7.547	R\$ 874,16	R\$ 510,00
Mangabeira	41.731	R\$ 846,91	R\$ 550,00
Funcionários	8.271	R\$ 837,10	R\$ 516,00
Boa Esperança	3.178	R\$ 791,33	R\$ 521,50
Ernani Sátiro	4.575	R\$ 760,47	R\$ 510,00
Barra de Gramame	174	R\$ 668,90	R\$ 510,00
Bairro das Indústrias	4.208	R\$ 658,59	R\$ 510,00
Costa e Silva	4.104	R\$ 656,14	R\$ 510,00

Costa do Sol	3.877	R\$ 641,71	R\$ 510,00
Paratibe	6.136	R\$ 610,85	R\$ 510,00
Jardim Veneza	5.560	R\$ 608,40	R\$ 510,00
Muçumago	2.957	R\$ 606,90	R\$ 510,00
Gramame	11.706	R\$ 586,52	R\$ 510,00
Mumbaba	3.587	R\$ 585,06	R\$ 510,00
Distrito Industrial	841	R\$ 554,98	R\$ 510,00
Grotão	3.129	R\$ 525,21	R\$ 510,00
Total	82.684	R\$ 2.748,80	R\$ 1.574,38

Por fim, na Zona Sul, temos quatro bairros que se encaixaram na Classe C: Anatólia, Jardim Cidade Universitária, Jardim São Paulo e Água Fria. Nesta zona destacamos o bairro de Mangabeira com seus quase 42 mil habitantes. Apesar de possuir mais da metade do total da Zona Sul, podemos classificá-lo como um bairro de poder aquisitivo relativamente baixo.

Uma consideração a se fazer é que essa classificação dos bairros em Classe C é reflexo dos dados coletados. Contudo esta classificação não exclui a possibilidade da existência de moradores que pudessem ser classificados em outras classes (até mesmo classes superiores). Ou seja, pode ser que no bairro São José (obteve a menor Renda Mensal) existam moradores com poder aquisitivo bem superior aos aproximadamente R\$ 470,00 da média.

3.1.4 Considerações

Para nosso estudo teremos como referência o processo eleitoral do município de João Pessoa, na Paraíba, constituindo-se por sete postulantes ao cargo de prefeito no ano de 2012, sendo eles: Antonio Ferreira Lima Neto, Cícero de Lucena Filho, Estelizabeth Bezerra de Souza, José Targino Maranhão, Luciano Cartaxo Pires de Sá, Maria de Lourdes Sarmiento, Renan Palmeira Costa, de acordo com dados do TSE (2012).

Tendo como premissa os dados obtidos nas pesquisas de intenção de voto bem como com os próprios dados obtidos através da ferramenta utilizada no estudo, nossa pesquisa limitou-se a analisar o comportamento dos eleitores em relação aos candidatos José Targino Maranhão, Luciano Cartaxo Pires de Sá, Cícero de Lucena Filho e Estelizabeth Bezerra de Souza, considerados os principais candidatos na disputa eleitoral de João Pessoa.

Nosso estudo também procurou analisar a interação entre os eleitores e os candidatos ao cargo de vereador. Para o pleito em questão verificamos o registro de 551 candidatos concorrendo diretamente as 27 vagas previstas para a câmara municipal. O elevado número de candidatos conduz a uma enorme polaridade na discussão política. Isto se tornou um obstáculo para o estudo realizado aliado ao fato de que a propaganda eleitoral dos candidatos ao cargo de vereador não ocorre de acordo com critérios bem definidos como é o caso dos candidatos ao cargo de prefeito.

Com relação ao perfil socioeconômico nos interessa os dados da população que possui maior renda. Este dado será importante para nosso estudo no sentido de que poderemos utilizar os dados de votação dos bairros de maior poder aquisitivo (uma vez que, de acordo com a descrição da [seção 3.1.1](#), a inscrição do título eleitoral está vinculada diretamente ao bairro de residência do eleitor) para compará-los com os dados obtidos junto à ferramenta de busca e coleta de dados (partindo do pressuposto de que alguns dados poderão vir com o referencial de geolocalização) e, com isso, buscarmos alguma semelhança ou padrão.

Utilizando o critério de rendimento nominal médio e rendimento nominal mediano da população dos bairros de João Pessoa, podemos classificar 7 bairros de interesse para o nosso estudo: Cabo Branco, Tambaú, Manaíra, Miramar, Tambauzinho, Bessa e Altiplano. Embora existam outros bairros com esta classificação, procuramos priorizar aqueles componentes (ou próximos) da zona leste da capital paraibana.

TABELA 9 - Bairros com maior poder aquisitivo.

Fonte: Censo IBGE (2010), TRE-PB (2012).

Bairros	Rendimento Nominal Médio	Rendimento Nominal Mediano	Seções eleitorais	Eleitorado
Cabo Branco	R\$ 4.510,36	R\$ 3.000,00	4	1.682
Tambaú	R\$ 4.266,78	R\$ 2.700,00	39	16.176
Manaíra	R\$ 3.941,03	R\$ 2.500,00	27	10.949
Miramar	R\$ 3.482,74	R\$ 2.000,00	6	1.942
Tambauzinho	R\$ 3.208,54	R\$ 1.500,00	17	6.190
Bessa	R\$ 2.590,06	R\$ 1.500,00	44	19.413
Altiplano	R\$ 2.442,13	R\$ 1.000,00	17	6.730

Os bairros elencados na **Tabela 9** acima somam mais de 63 mil eleitores (13,2% do total de eleitores) e uma população estimada de mais de 120 mil habitantes (17,6% do total de habitantes). Estes bairros pertencem à região leste de acordo com a divisão proposta na [seção 3.1.2](#) anteriormente.

Podemos identificar alguns bairros que, embora não tenham sido escolhidos para compor os bairros de maior poder aquisitivo, se destacam por possuir (entre outras características) população elevada. São eles os bairros do Cristo Redentor, Mangabeira, Jardim cidade universitária e Bairro dos Ipês.

TABELA 10 - Bairros de população elevada.
Fonte: Censo IBGE (2010), TRE-PB (2012).

Bairros	Rendimento Nominal Médio	Rendimento Nominal Mediano	Seções eleitorais	Eleitorado
Cristo	R\$ 1.029,53	R\$ 510,00	69	30.983
Mangabeira	R\$ 846,91	R\$ 550,00	149	50.128
C. Universitária	R\$ 1.854,20	R\$ 1.200,00	24	9.760
Bairro dos Ipês	R\$ 1.276,62	R\$ 600,00	35	14.224

3.2 Acesso a informação e debate político

O objetivo desta seção é demonstrar que a utilização dos meios de Tecnologia da Informação e Comunicação (TIC) se dá de forma bastante segmentada na população brasileira. Ainda, buscaremos demonstrar que o uso deste tipo de meio podem auxiliar na expansão do debate político. Assim, a presente seção encontra-se dividida em quatro subseções: Em Demografia do uso das mídias ([Seção 3.2.1](#)) iremos traçar um perfil dos usuários deste tipo de meio e, em particular, dos usuários das redes sociais; Em uso das mídias para o debate político ([Seção 3.2.2](#)) ampliaremos a discussão sobre os modos de utilização destes meios na disseminação de informações de cunho político no contexto de uma eleição; Em conceitos sobre as pesquisas eleitorais ([Seção 3.2.3](#)) iremos expor os principais conceitos de uma das principais fontes da discussão política no nosso país; Por fim em considerações ([Seção 3.2.4](#)) iremos analisar o perfil do “internauta” brasileiro que faz uso das redes sociais (em particular os usuários do Twitter) comparando este com o perfil do eleitor da capital paraibana.

3.2.1 Demografia do uso das mídias

A popularização do acesso à internet a partir da melhoria da infraestrutura de telecomunicações do país, criação de diversos programas para a inclusão digital e, principalmente, através da melhoria das condições financeiras dos brasileiros nos últimos anos, tem gerado um aumento substancial na utilização dos ambientes digitais como meio para realização de intercâmbio de informações entre as pessoas.

Estudos recentes apontam que cada vez mais pessoas passam do que é considerado Classe Baixa para a Classe Média. Dados do IDP (Instituto Data Popular) indicam que no período compreendido entre os anos de 2002 e 2012 houve uma redução de 48% para 28% da população classificada como Classe Baixa, ao mesmo tempo em que houve um incremento da população de classe média de 38% para 52%, ou seja, traduzindo estes índices em números, temos aproximadamente 104 milhões de pessoas.

Conforme já citamos na [seção 3.1.3](#), segundo a ABEP (2010) uma pessoa faz parte da chamada Classe Média (ou Classe C) se a Renda Média da família estiver entre R\$ 1.540 e R\$ 2.313, baseado com os dados salariais do ano de 2010. Dados de pesquisas realizadas pelo IDP (2012) apresentados no Fórum Novo Brasil¹, indicaram que no ano de 2012, cerca de 60% das pessoas que utilizam as Redes Sociais eram componentes da Classe Média, contra 23% da Alta e 20% da Baixa.

Outro aspecto saliente no novo cenário nacional é a queda da taxa de analfabetismo. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2010), a taxa de analfabetismo da população de 15 anos ou mais caiu de 13,63%, em 2000, para 9,6% em 2010. Junto à diminuição do número de analfabetos, o Brasil acompanha também o aumento do número pessoas com ensino superior completo, que passou de 5,5 milhões, em 2000, para 25,5 milhões em 2010.

A seguir, serão apresentados alguns dados da pesquisa TIC Domicílios realizada pelo Centro de estudos sobre a Tecnologia da Informação e Comunicação (Cetic, 2012) que tem por objetivo geral medir o acesso e os usos da população brasileira em relação às Tecnologias de Informação e Comunicação. Os primeiros índices dizem respeito à proporção de domicílios (em %) que possuem computador em casa e acesso a internet.

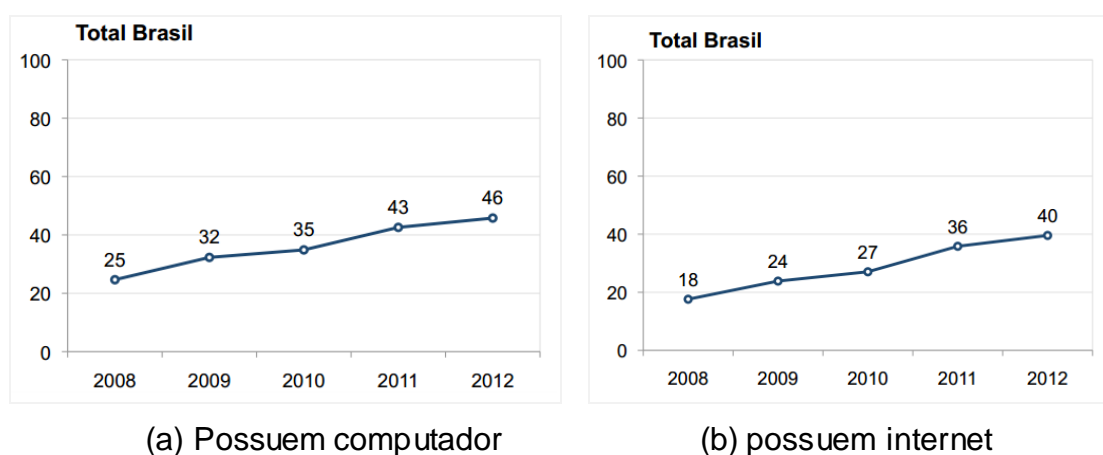


FIGURA 3 - Proporção de domicílios (em %).
Fonte: Cetic (2012).

¹ <http://www.ecofinancas.com/noticias/redes-sociais-sao-usadas-pela-classe-media>

Nos dados descritos na **Figura 3** (a) concluímos que 46% dos domicílios (equivalente a 28,1 milhões em número de domicílios) possuem computador. Na **Figura 3** (b) concluímos que cerca de 40% dos domicílios (equivalente a 24,3 milhões de domicílios) possuem acesso regular a internet.

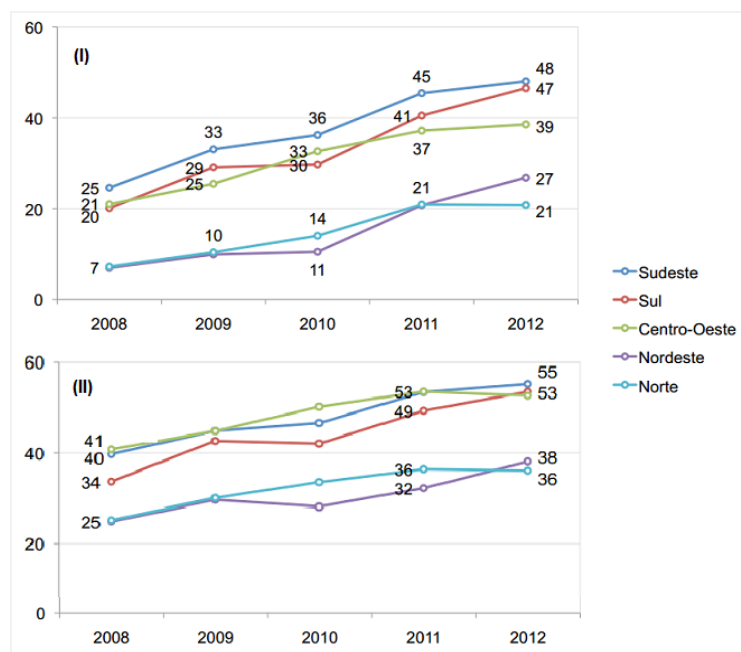


FIGURA 4 - Proporção de domicílios (em %) por região.
Fonte: Adaptado do Cetic (2012).

Na **Figura 4** (I) temos o percentual sob o total de domicílios de cada região que possuem acesso regular a internet. Na **Figura 4** (II) temos o percentual sob o total de domicílios de cada região que possuem computador. Concluímos que a região nordeste é a que apresentou o maior crescimento em termos percentuais em comparação com as demais regiões, tanto nos domicílios com acesso a internet quanto nos domicílios com computador.

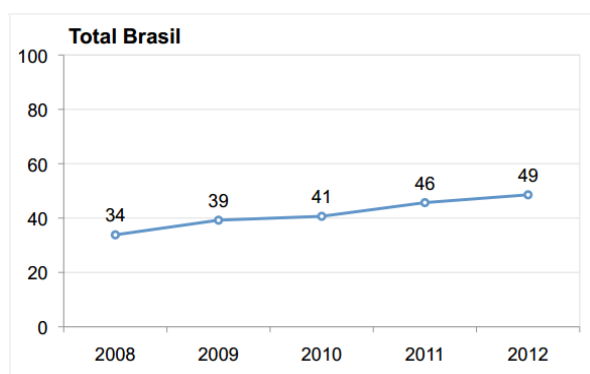


FIGURA 5 - Percentual sobre toda a população.
Fonte: Cetic (2012).

Na **Figura 5** podemos analisar o percentual sobre o total da população que possuem acesso a internet. Desta forma concluímos que no ano de 2012 49% da população possui acesso regular a internet o que representa mais de 80 milhões de habitantes se traduzido em números.

De acordo com os dados demonstrados, independentemente das discrepâncias existentes nos números de pesquisas de outras instituições, até aqui podemos certamente afirmar que o número de internautas vem crescendo a cada ano e que cada vez mais a população brasileira tem recorrido à internet como fonte de informação através de blogs ou sites informativos, para compartilhar interesses através de redes sociais, para realização de compras *online*, para divertimento e lazer entre diversos outros fins.

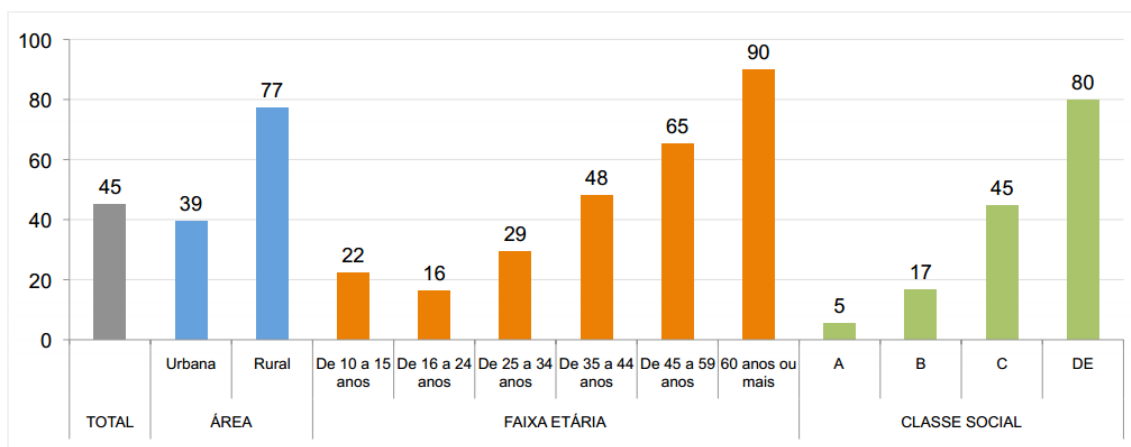


FIGURA 6 - Proporção de indivíduos que nunca usaram a internet.
Fonte: Cetic (2012).

Diante disto podemos identificar o perfil dos indivíduos que não realizam uso regular da internet no Brasil. De acordo com os dados da **Figura 6**, afirmamos que os indivíduos que não fazem uso do computador e internet são, em sua maioria, moradores da zona rural, possuem 45 anos de idade ou mais (representando aproximadamente 43 milhões de habitantes) e são prioritariamente da classe D e E (representando aproximadamente 68 milhões de habitantes).

Em contrapartida, se analisarmos os dados referentes à proporção de indivíduos que utiliza as redes sociais descrito na **Figura 7** (Cabe comentário de que esta pesquisa levou em consideração os meses de dezembro de 2012 e janeiro e fevereiro de 2013), tais como facebook, Orkut e Badoo, podemos

constatar que em sua maioria são jovens (entre 18 e 24 anos) e que sua utilização encontra-se de certa forma distribuída “uniformemente” entre as classes sociais, contudo o nível de instrução é um fator que deve ser levado em consideração.

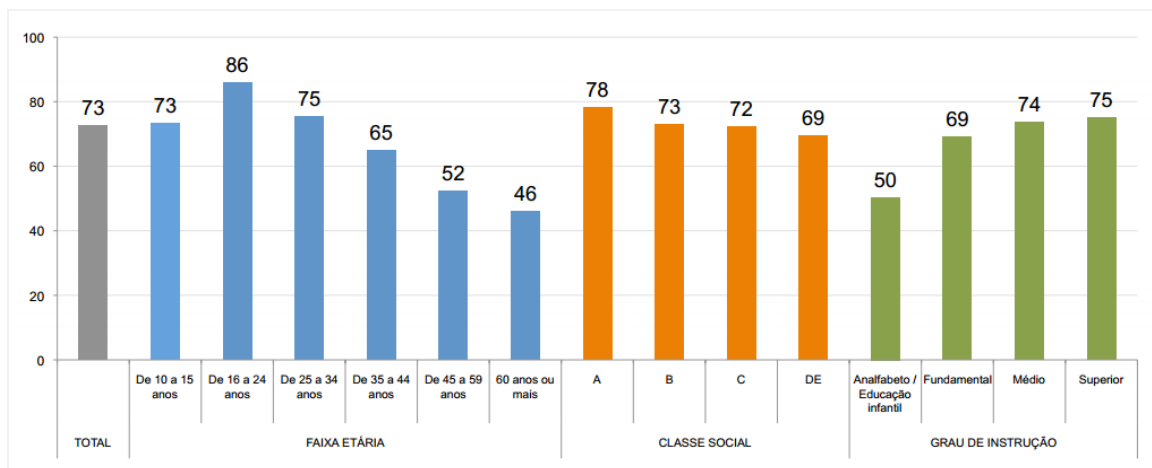


FIGURA 7 - Proporção de indivíduos que utilizam redes sociais.
Fonte: Cetic (2012).

Todas estas informações corroboram com a hipótese de que a classe média, além de outros fatores, por possuir maior poder aquisitivo e melhor nível de instrução, pode ser considerada responsável por grande parte do consumo de eletrônicos (computadores) e dos serviços de telefonia/internet, acarretando desta forma (muito provavelmente) em maior acesso as mídias sociais como meio para busca e compartilhamento de informações.

3.2.2 O uso das Mídias para o debate político

Nos anos 90 tiveram início as discussões relacionado à participação política nas novas tecnologias desenvolvidas a partir da internet (Silva, 2006), como uma possível forma de aproximar os representantes de seus representados. Além da presença nas eleições, Coleman (1999) relaciona as novas tecnologias com o processo de representatividade e democracia, pois tem sido usada também para melhorar a governança.

A campanha eleitoral de 2002 foi um dos primeiros eventos eleitorais em que a internet atuou de forma estratégica (Aldé, 2004). Com isso ela se converteu em um instrumento de marketing político, possibilitando dentre outras coisas, ampliar o público-alvo e reduzir os custos de campanha.

Contudo apenas nas eleições de 2010, devido ao crescimento considerável de candidatos que passaram a utilizar sites, blogs e redes sociais, tivemos a criação de uma Lei Federal¹ que normatizasse a utilização dos meios digitais no contexto eleitoral.

Apesar de não representar mudanças significativas nos resultados, o que é explicado ainda pelo alcance social de certa forma limitado no caso brasileiro (veja mais detalhes na [seção 3.2.1](#)), a internet ganha destaque durante corrida eleitoral do ano de 2012 onde, candidatos e suas respectivas assessorias passaram a dar maior atenção a esse mecanismo como forma de atingir a camada segmentada da população que utiliza a internet.

Mesmo não sendo possível identificar os impactos diretos para a decisão do voto, as redes sociais, por exemplo, tornaram-se espaço de debate, pois as discussões que ocorrem nesses meios alternativos ganham espaço na mídia tradicional, seja como fonte para os jornalistas ou para ampliar o debate a um número maior de eleitores de forma indireta ou indireta (Stanley e Weare, 2003). Isso significa que os “afetados” pela internet não foram apenas aqueles que acessam periodicamente as páginas eletrônicas ou seguem os candidatos através de redes sociais, mas também de forma indireta, pois os que recebem as informações as retransmitem em conversas “*off-line*” a outros eleitores.

Isso justifica, em parte, a produção dessa pesquisa, pois se trata de um debate que transcende o espaço digital proporcionado por uma rede social. Embora este tipo de meio possa excluir um número considerável de indivíduos, um grupo maior é afetado pela sua lógica das relações de interação que ocorrem nas redes globais (Castells, 1999).

O primeiro problema constatado não apenas no Brasil, mas em diversos países, é a diferença no acesso pelas pessoas, o que significa desigualdade quanto no acesso quanto ao recebimento de informações. Segundo Levy (1999), o que se vê neste momento é que existe uma distribuição desigual de competências técnicas, de recursos financeiros e de habilidades educacionais, e que desta forma, ao invés das novas mídias resolverem as desigualdades

¹ Lei Federal 9.504 – Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/L9504.htm>

anteriores, podem torná-las ainda mais relevantes quando as oportunidades digitais de participação política ficam fora do alcance de parcela da população.

Além das dificuldades (de caráter técnico) para o acesso à rede, outro problema apontado por Castells (2005) é a cultura da sociedade que ainda não considera as potencialidades da “ferramenta”. Segundo o autor, não é a tecnologia que determina a sociedade, mas é esta que dá forma à tecnologia de acordo com suas necessidades, valores e interesses. É preciso que o cidadão seja interativo para além das tecnologias interativas (Coleman, 1999). A noção de cultura cívica aponta para as características do mundo sociocultural que constituem condições prévias para a participação democrática (Dahlgren, 2000), sendo que sem elas há dificuldade de aumentar as capacidades de intervenção da rede digital nos debates de temas públicos.

De acordo com o autor, as novas tecnologias da informação “manifestam-se de diversas formas, conforme a cultura, as instituições e a trajetória histórica de cada uma” (Castells, 2005, p. 18). Em decorrência dessas características intrínsecas da sociedade, tudo vai depender de onde, por quem e para quem são usadas as tecnologias de comunicação e informação. O fator que determina o grau de democratização que as TICs significam é a cultura política onde ela se desenvolve (Coleman, 1999).

Segundo Gomes (2005), a ausência de debate político nas novas mídias é, também, resultado da falta de outros requisitos da vida democrática. Para que os eleitores possam compreender a internet como um espaço de participação política, é necessária uma mudança na mentalidade, nos modos de organização e hábitos políticos (Levy, 1999). É preciso ainda levar em consideração, além da oferta de informação política, que há política, o que está relacionado com a cultura política. Castells (2005) afirma que se existe um real e significativo interesse da esfera civil em adquirir informações e participar da vida, logo existe um interesse significativo do usuário da internet em informação política, pois não basta apenas a existência das ferramentas para que haja aumento da participação. Isso explica o porquê de muitas pessoas terem acesso à rede e mesmo assim não a utilizarem para fins políticos.

3.2.3 Conceitos sobre pesquisas eleitorais

As pesquisas de intenção de voto são um dos tipos mais conhecidos de levantamentos estatísticos feitos com a população. Apenas as pesquisas do censo do IBGE podem se comparar as pesquisas eleitorais em termos de importância e nível de inserção na mídia. As pesquisas de opinião (eleitorais) são ferramentas importantes e eficazes para o conhecimento da opinião e do comportamento dos grupos, entre outros aspectos, e nos possibilita entender como se manifestam os aspectos individuais dentro do grupo social.

Basicamente existem dois tipos de pesquisa: **(1)** as pesquisas quantitativas e **(2)** as pesquisas qualitativas (Fernandes e Carneiro, 2013). **(1)** As quantitativas são pesquisas feitas entre a população em geral, com uma amostra selecionada, geralmente para saber a intenção de votos dos eleitores. Elas são as mais conhecidas e também as mais divulgadas pela mídia em geral. **(2)** As qualitativas servem para avaliar as diferenças entre as opiniões, através de discussões em grupo e entrevistas mais profundas. Elas não costumam ser divulgadas ao público e é um tipo mais utilizado por partidos e políticos em geral.

Normalmente a pesquisa quantitativa é feita utilizando-se de uma amostra. As pessoas podem ser abordadas em casa, pelo telefone ou mesmo na rua, dependendo das preferências e metodologias do instituto de pesquisa encarregado pela pesquisa. Cada amostra deve retratar de forma fiel o universo de eleitores. Por exemplo, se a população for formada por 50% de homens e 50% de mulheres, a amostra deverá conter também esta mesma proporção. O mesmo acontece para outras características da população como faixa etária, renda, classe social entre outros (Santos e Lopes, 2002).

A margem de erro é um componente essencial neste tipo de pesquisa. Ela existe porque não são entrevistados todos os eleitores e sim uma parcela deste universo. O erro amostral serve para estimar o resultado dos candidatos na pesquisa eleitoral e entender o intervalo de confiança onde eles podem estar oscilando (Santos e Lopes, 2002). Em estatística, um **intervalo de confiança** (IC) é um intervalo estimado de um parâmetro estatístico. Em vez de estimar o parâmetro por um único valor, é dado um intervalo de estimativas prováveis.

Para calcular este valor precisamos conhecer o nível de significância, Desvio padrão e do tamanho da amostra. Maiores detalhes sobre esta técnica pode ser encontrada no trabalho de Guimarães (2007).

Continuando o exemplo, se um candidato tem 27% das intenções de voto, considerando uma margem de erro em três pontos percentuais e um intervalo de confiança de 95%, este valor pode variar entre 24% e 30%. Isto quer dizer que de cada cem amostras, a pesquisa tem 95 resultados entre 24% e 30%.

Dentre outros fatores a data de uma pesquisa é uma informação extremamente importante (Beraldo, 2002). As pesquisas eleitorais são simples ferramentas científicas que se propõem a detectar e antecipar uma determinada tendência, a partir de levantamentos sucessivos que tenham características semelhantes de desenho amostral. Deve-se, então, buscar interpretá-la consoante a mensagem principal que transmite: *Snapshots* (fotografias) de um determinado momento. Ainda é interessante analisar o conteúdo dos questionários aplicados no intuito de confirmar a lisura do processo (Venturini, 1995).

Para o Ibope (2013), a pesquisa tem uma influência mais institucional do que no eleitor propriamente dito. Do ponto de vista do eleitor, ele pode tomar as mais variadas decisões a partir das informações de uma pesquisa. Pode por exemplo votar em quem está na frente (porque quer votar no vencedor), votar em quem está em segundo (porque não quer que quem esteja na frente ganhe), ou pode simplesmente votar em nulo ou branco. Os efeitos são os mais diversos possíveis de modo que não existe estudo que comprovem a influência direta das pesquisas nos votos dos eleitores.

Para o Datafolha *apud* Fernandes e Carneiro (2013) a influência das pesquisas sobre o eleitor existe e é válida. O eleitor leva em conta, na hora de decidir seu voto, o conjunto de informações que ele recebe, seja essas provenientes de campanhas, do noticiário, dos seus amigos e também dos resultados das pesquisas. Se um eleitor opta por votar naquele candidato que tem mais chances, é uma opção racional e um voto decidido de forma mais pensada. Nesse sentido é legítimo que a pesquisa tenha influência.

3.2.4 Considerações

Nesta parte do trabalho podemos confirmar que é crescente o número de pessoas que fazem uso da internet e, em particular das redes sociais. Verificamos ainda que os ambientes proporcionados pelas mídias digitais (em particular as redes sociais) ampliam sobremaneira o espaço para a deliberação política uma vez que, no contexto da discussão política, praticamente tudo que pode ser feito na vida real também pode ser feito no mundo virtual.

Ainda, analisamos os conceitos das pesquisas eleitorais e percebemos que existe um “*trade off*” de opiniões entre os principais institutos de pesquisas do nosso país quanto a real influência que as pesquisas eleitorais exercem sobre o voto do eleitor. Não cabe entrar nesse mérito, primeiro por motivos de simplicidade do trabalho e, segundo por falta de “*expertise*” no assunto por parte deste autor. Contudo uma boa análise sobre os impactos das apresentações das pesquisas no aumento/diminuição da discussão política nas redes sociais poderia ser analisada em maiores detalhes.

Sobre a rede social Twitter, objeto de estudo da presente pesquisa, ela apresenta um perfil de usuário diferenciado se comparado aos usuários das demais redes sociais (Figura 7). Diferentemente dos usuários de outras redes sociais como Facebook e Orkut, os usuários do Twitter tem perfil de usuário centrado nas classes sociais mais elevadas e nos níveis de escolaridades mais altos. Encontramos fortes semelhanças apenas na faixa etária.

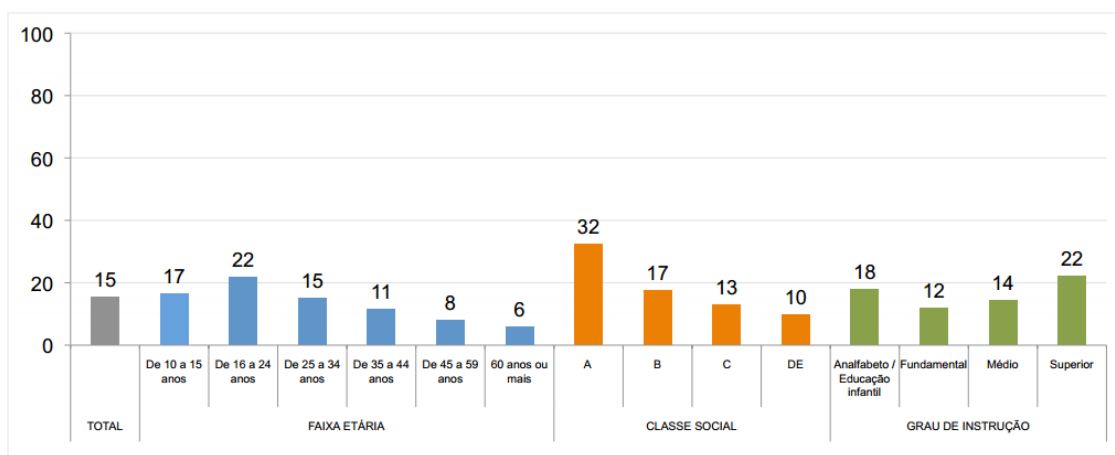


FIGURA 8 - Proporção de indivíduos que usam o Twitter.
Fonte: Cetic (2012).

Cumprir dizer que o perfil dos usuários do twitter, de acordo com os dados da pesquisa da Cetic (2012) se assemelha fortemente ao perfil dos eleitores da capital paraibana (em particular no que diz respeito à classe social e grau de instrução) que são moradores dos bairros selecionados de acordo com a **Tabela 9** descrita na seção [3.1.4](#) o presente trabalho.

Com relação ao debate político através da internet temos que levar em consideração que esta ainda não alcança todos aqueles que têm acesso à rede, pois, na maioria dos casos, não há interesse em interagir sobre esse assunto. Aqueles que utilizam a rede para mobilizações, interação e maior proximidade com os governantes muito provavelmente são os que já possuem interesse prévio pela política antes da utilização das novas mídias. Dessa forma, existe uma migração da militância política **off-line** para **online** e não uma ampliação a partir deste último. Os mesmos líderes de opinião que atuavam e militavam em favor de candidatos na esfera pública são aqueles que procuram interação com os candidatos pelas redes sociais, continuando seu papel de mediador de opinião que já ocupava na sociedade (Silva, 2006).

Ressalta-se ainda que em relação aos meios prioritários de campanha eleitoral (comícios, presença em eventos, passeatas e relação face a face com o eleitor), estes são os meios tradicionais e não há indícios de mudanças em um curto período de tempo. Segundo Dahlgren (2000), é preciso observar alguns aspectos da cultura cívica, centrando-se em como a rede é usada em conjunto com outras atividades *off-line* e não que ela substitua outros espaços da campanha massiva. A internet atua como ferramenta de campanha e é um espaço público específico que até o momento consegue atingir apenas um segmento da população.

3.3 Redes Sociais

O objetivo desta seção é expor os principais conceitos e teorias que fornecem embasamento para análise das redes sociais. Assim, a presente seção encontra-se dividida em seis subseções: Em Conceitos Básicos ([Seção 3.3.1](#)) serão expostos os principais tipos de redes, conceituando distintamente redes sociais e redes sociais *online*; Em Teoria dos Grafos ([Seção 3.3.2](#)) será discutido a possibilidade de representação de redes sociais através dos grafos e as implicações advindas desta possibilidade; Em Modelos de Redes ([Seção 3.3.3](#)) ampliaremos a discussão sobre as redes complexas e os seus principais modelos; Em métricas para estudo de redes complexas ([Seção 3.3.4](#)) serão apresentadas as principais métricas que podem ter aplicação em nosso estudo em virtude das características esperadas; Em Ferramentas para ARS ([Seção 3.3.5](#)) iremos expor as ferramentas de autoria e as API's que foram utilizadas no presente estudo; Por fim em considerações ([Seção 3.3.6](#)) iremos analisar de que maneira os conceitos de teoria dos grafos e das redes complexas podem beneficiar nosso estudo.

3.3.1 Conceitos Básicos

Existe uma abundância de redes pelo mundo, os exemplos incluem desde a Internet (rede mundial de computadores), redes sociais, conexões entre quaisquer indivíduos até redes de conexão comercial, redes metabólicas, redes de distribuição tais como energia elétrica, serviços telefônicos, serviços de correio, entre outros (Newman, 2003).

Uma rede, em sua forma mais simples, pode ser definida como um conjunto de pontos que às vezes chamaremos de nós, interligados por conexões denominadas arestas (Newman, 2003). Esta morfologia permite identificar as relações entre pares de nós que podem desde páginas web, pessoas, computadores, até mesmo cidades, países ou aeroportos.

Um tipo específico de rede é de extremo interesse ao presente trabalho: rede social que pode ser definida como qualquer conjunto limitado de unidades sociais interligadas (Berkowitz, 1982). Esta definição destaca três características importantes de redes sociais:

Primeiro, as redes têm limites, isto é, existe algum critério para determinar a associação na rede. O **segundo** elemento chave da definição é "conexão" em redes sociais. Para fazer parte de uma rede social, cada membro deve ter ligações reais ou potenciais para pelo menos outro membro da rede. Esses links podem ser diretos ou indiretos. O **terceiro** aspecto chave da presente definição é uma unidade social, ou seja, cada componente da rede é um elemento distinto do grupo.

A noção de rede social e métodos para análise de redes sociais têm atraído consideravelmente a comunidade das ciências sociais e comportamentais a fim de analisar os relacionamentos entre entidades sociais e os padrões e implicações desses relacionamentos (Wasserman, 1994). Para estudar como os comportamentos ou as opiniões dos indivíduos dependem das estruturas nas quais eles se inserem, as unidades de análise deixam de ser os atributos individuais (classe, sexo, idade), passando a ser o conjunto de relações que os indivíduos estabelecem através de suas interações uns com os outros através da rede (Marteletto, 1994).

Em contrapartida temos o conceito de rede social *online* que é geralmente utilizado para descrever um grupo de pessoas que interagem através de qualquer mídia de comunicação. Consequentemente, baseado nessa definição, podemos afirmar que redes sociais *online* existem desde a criação da Internet, muito embora, como já citamos anteriormente, o conceito de redes sociais exista muito antes do surgimento deste meio (Stanley 1994, Marteletto 2001). Ainda, por tratar-se de redes de interação entre pessoas, muitos conceitos já consolidados por outras disciplinas como Antropologia, Sociologia e Psicologia podem ser usados para analisar o comportamento de cada indivíduo e da rede de relacionamento como um todo, mesmo sendo redes virtuais.

No presente trabalho utilizaremos uma definição um pouco mais restrita, semelhante a definição adotada no trabalho de Boyd e Ellison (2007): Rede social *online* é um serviço Web que permite a um indivíduo **(1)** construir perfis públicos dentro de um sistema, **(2)** articular uma lista de outros usuários com os quais ele compartilha conexões e **(3)** visualizar e percorrer suas listas de conexões assim como outras listas criadas por outros usuários do sistema.

Após a devida conceituação e explanação de diferenças e, por motivos de simplificação, a partir deste momento citaremos o termo redes sociais *online* como simplesmente “redes sociais”. Com base nessa definição, existem várias redes sociais disponíveis na *web*, que variam de acordo com seus objetivos primários. A **Tabela 11** apresenta uma lista com várias redes sociais populares atualmente, juntamente com seus principais propósitos.

TABELA 11 - Principais Redes Sociais.
Fonte: Adaptado de Benevenuto *et al* (2011).

Nome	Propósito	Endereço
Orkut	Amizades	http://www.orkut.com
Facebook	Amizades	http://www.facebook.com
MySpace	Amizades	http://www.myspace.com
LinkedIn	Profissionais	http://www.linkedin.com
YouTube	Compartilhamento de vídeos	http://www.youtube.com
Flickr	Compartilhamento de fotos	http://www.flickr.com
LiveJournal	Blogs e diários	http://www.livejournal.com
LastFM	Rádio/músicas	http://www.last.fm
Twitter	Troca de mensagens curtas	http://twitter.com

3.3.2 Teoria dos grafos

Diversas situações do mundo real podem ser convenientemente modeladas através de uma rede que são representadas através dos **grafos**. Esta estrutura é composta por um conjunto de pontos, juntamente com linhas que unem alguns pares de pontos. No caso específico das redes sociais temos que os pontos podem representar pessoas e as linhas podem representar uma ligação de amizade, ou ainda os pontos podem ser centros de comunicação e as linhas podem representam o meio de comunicação entre eles (Mika, 2007).

A teoria dos grafos surge com Leonhard Euler no ano de 1736, quando ele demonstrou ser impossível atravessar as sete pontes de Königsberg. As sete pontes de Königsberg era um problema clássico que consistia em ter que atravessar as sete pontes, que ligavam duas ilhas (**Figura 9**), uma única vez durante um passeio pela cidade (Bondy and Murty, 1982). Euler modelou o problema em um grafo, onde as ilhas foram transformadas em arestas e as ligações em vértices, o que se acredita ter sido o primeiro grafo da história, e através desse grafo provou que não era possível realizar tal feito.

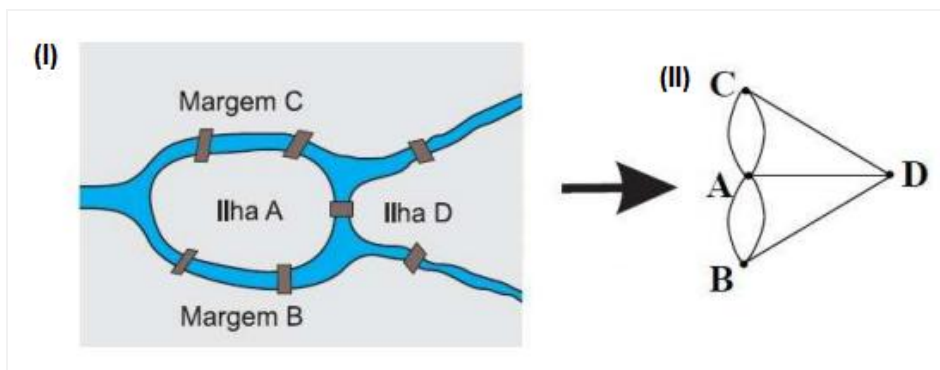


FIGURA 9 - (I) Problema da ponte de Königsber (II) Representação do Problema através de grafo.

Fonte: Adaptado de Bondy e Murty (1982).

Podemos definir um grafo como sendo uma representação gráfica que descreve os elementos e suas conexões. Formalmente, um grafo G é uma tripla ordenada de $(V(G), E(G), G)$, consistindo de um conjunto não vazio $V(G)$ de vértices que representam os elementos de um sistema, podendo este ser dos mais variados tipos e domínios, um conjunto $E(G)$ (disjunto de $V(G)$) de arestas que mostram os relacionamentos existentes, onde cada aresta consiste em fazer a conexão de um par de vértices e uma função de incidência G que associa a cada aresta um par não ordenado (e não necessariamente distinto) de vértices de G (Bondy and Murty, 1982).

Em algumas situações é necessário saber em qual direção a conexão está sendo feita, assim as arestas apresentam uma ou duas “flechas” apontadas para uma determinada direção indicando que as arestas partem do primeiro vértice para o segundo vértice, sendo esse grafo chamado de grafo dirigido ou simplesmente dígrafo (**Figura 10**). Podem-se exemplificar os dígrafos através de chamadas telefônicas, ou mensagens de e-mails, pois tanto as chamadas telefônicas como as mensagens de e-mails são direcionadas, no caso, seguem de uma pessoa para outra (Newman, 2003).

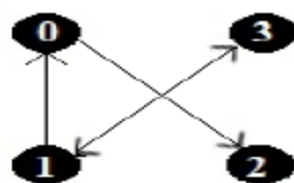


FIGURA 10 - Dígrafo Simples.

Fonte: Adaptado de Bondy e Murty (1982).

Um grafo simples, com n elementos (nós), pode ter no máximo $(n * (n-1)) / 2$ arestas conectadas, o que indica um grafo completamente conectado chamado de grafo completo (Caldarelli e Vespignani, 2007), e um grafo que não possui vértices é denominado de grafo vazio (Bondy and Murty, 1982). A **Figura 11** ilustra dois tipos de grafos.

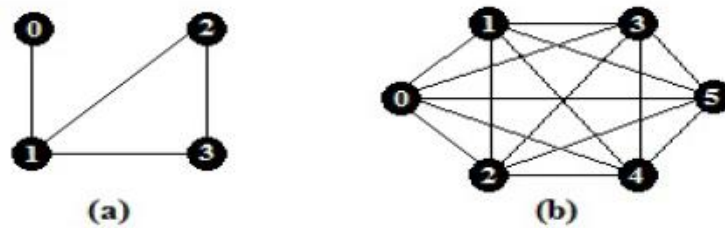


FIGURA 11 - Dois grafos. (a) Grafo simples. (b) Grafo completo.
Fonte: Adaptado de Bondy e Murty (1982).

Grafos podem ser representados de várias formas, e aqui são destacadas duas formas diferentes. A primeira possibilidade é chamada de lista de adjacências que consiste em um arranjo de V listas, uma para cada vértice em V . Cada lista de adjacência contém ponteiros para os nós que estão conectados com o vértice referido pela lista (Cormen *et al*, 2002). A segunda forma que é frequentemente utilizada para simbolizar os grafos é por meio de uma matriz de adjacência, A_{ij} , onde cada linha e coluna demonstra um vértice do grafo. A_{ij} é 1 quando houver relacionamento entre o vértice i e o vértice j , caso contrário terá o valor 0.

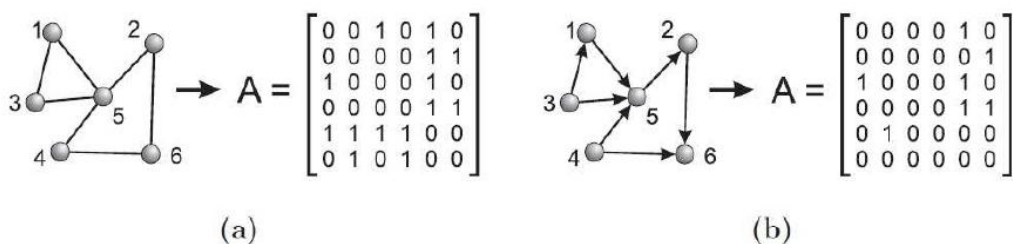


FIGURA 12 - As redes representadas por matrizes de adjacência.
Fonte: Adaptado de Balancieri (2010).

Vejamos os detalhes da **Figura 12** acima. Em (a) é apresentada uma rede não dirigida e em (b) uma rede dirigida. No caso (a), os elementos a_{ij} da matriz são iguais a 1 se há uma ligação entre os vértices i e j e iguais à zero, caso contrário. Já no caso (b), os elementos da matriz a_{ij} são iguais a 1 se existe uma conexão dirigida do vértice i para o vértice j .

3.3.3 Modelos de Redes

De forma geral representa-se uma rede complexa através de um grafo. É importante lembrar que nem todo grafo é considerado uma rede complexa, pois essa classificação só é possível se o grafo apresentar algumas propriedades que não estão presentes em grafos simples tais como coeficiente de aglomeração, distribuição de graus e resistência (Metz *et al*, 2007).

Diversos modelos de redes complexas já foram propostos, e alguns são frutos de grande interesse para o estudo e identificação de características e propriedades das topologias das redes complexas existentes. Neste capítulo destacamos os modelos de redes aleatórias, redes *small-world* (pequeno mundo) e redes *scale-free* (livres de escala).

3.3.3.1 Redes aleatórias

O modelo de redes aleatórias foi proposto por Erdos e Reny (1960). Esse é o modelo mais simples que uma rede complexa pode assumir. Nesse modelo, arestas não direcionadas são adicionadas aleatoriamente entre um número fixo de N vértices (Costa *et al*, 2007). Cada aresta é independentemente representada com base em alguma probabilidade p . O número de arestas que conectam cada vértice na rede (grau do vértice) segue a distribuição de Poisson com um limite máximo N . O grau esperado de um vértice qualquer é definido pela **Equação 1**, onde p é a probabilidade de um vértice se conectar a outro vértice qualquer, N representa o número de vértices da rede e k é o total de arestas que incidem em um vértice.

$$\langle k \rangle = p(N - 1)$$

EQUAÇÃO 1 - Equação do grau de um vértice.

Esse modelo gera grafos aleatórios com N vértices e k arestas, denominados grafo aleatório ER. Inicialmente com N vértices desconectados, o modelo ER é obtido conectando-se os vértices selecionados aleatoriamente até o número de arestas do grafo ser igual a k . Um exemplar desta modalidade de rede complexa pode ser visualizado na **Figura 13**.

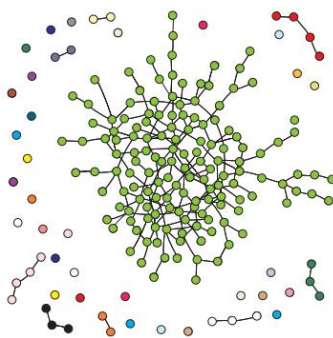


FIGURA 13 - Rede complexa aleatória.

Fonte: Strogatz, 2001.

Acredita-se que o processo de construção da rede seja aleatório no sentido de que vértices se agregam aleatoriamente. Com base nessa premissa, Erdős e Rény concluíram que todos os vértices de uma determinada rede têm aproximadamente a mesma quantidade de conexões e as mesmas chances de receberem novas ligações (Barabasi and Albert, 1999). Segundo os autores, quanto mais complexa for a rede, maiores serão as chances dela ser aleatória.

3.3.3.2 Redes pequeno mundo

Segundo Watts e Strogatz (1998), muitas redes apresentam padrões altamente conectados, tendendo a formar pequenas quantidades de conexões em cada vértice. Assim, eles propuseram um modelo semelhante ao de Erdős e Rény, no qual grande parte das conexões é estabelecida entre vértices mais próximos, apresentando-se como um mundo pequeno. Nesse modelo, a distância média entre quaisquer dois vértices de uma rede muito grande não ultrapassa um número pequeno de vértices. Para isso, basta que algumas conexões aleatórias entre grupos sejam estabelecidas (Buchanan, 2002).

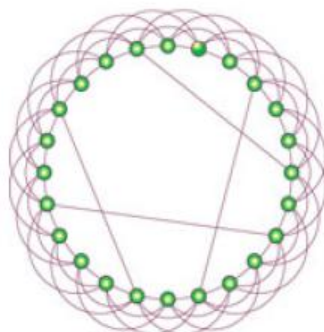


FIGURA 14 - Rede complexa de pequeno mundo.

Fonte: Strogatz, 2001.

O efeito pequeno-mundo (**Figura 14**) é observado nas redes em que a maioria dos vértices se conecta a outros através de um caminho mínimo. O caminho mínimo, também chamado de caminho geodésico ou distância geodésica, é aquele formado pelo menor número de arestas que conectam um vértice origem e um vértice destino. Para melhor ilustrar esse efeito, considere os indivíduos de uma sociedade qualquer. De acordo com o experimento conduzido por Stanley Milgram em 1960, se uma carta fosse entregue a um indivíduo, que não fosse o destinatário, e ele a repassasse a outro e, assim, por diante, em aproximadamente seis passagens ela chegaria ao destinatário. Esse resultado é uma demonstração direta do efeito pequeno-mundo, em que o caminho percorrido pela carta, partindo de um indivíduo qualquer até o destinatário, é mínimo. O comprimento do caminho mínimo médio entre pares de vértices em um grafo não direcionado é dada pela **Equação 2**, onde d_{ij} é a distância geodésica do vértice i até o vértice j .

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$$

EQUAÇÃO 2 - Equação do comprimento do caminho mínimo médio.

O efeito pequeno-mundo tem implicações óbvias na dinâmica de processos em redes. Por exemplo, um boato pode se espalhar muito mais rápido se, ao invés de mil passos, levarem apenas seis para chegar de um indivíduo qualquer a outro.

3.3.3.3 Redes livres de escala

Barabasi and Albert (1999) demonstraram que algumas redes apresentam uma ordem na dinâmica de estruturação, com características bem específicas. Uma das principais características, denominada conexão preferencial, é a tendência de um novo vértice se conectar a um vértice da rede que tem um grau elevado de conexões. Essa característica implica em redes com poucos vértices altamente conectados, denominados hubs, e muitos vértices com poucas conexões.

As redes com essas características são denominadas livres de escala devido à representação matemática da rede. Ela segue uma função $f(x)$ que permanece inalterada com um fator multiplicativo sob um reescalonamento da variável independente x . Em outras palavras, isso significa que as redes livres de escalas são aquelas em que a distribuição de graus segue a Lei de Potência, desde que exista uma solução somente para $f(ax) = bf(x)$. Conforme apresentado em (Newman, 2003), essas redes têm sido observadas em vários sistemas, por exemplo, na internet, em redes de metabolismos e em redes criadas a partir de citações de artigos científicos. Na **Figura 15** é apresentado um exemplo de rede livre de escala.

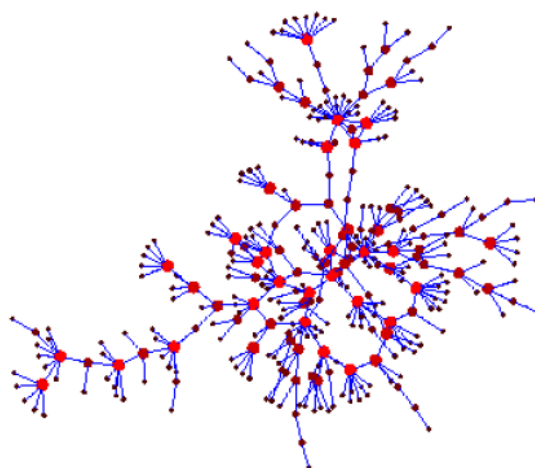


FIGURA 15 - Rede complexa livre de escala.
Fonte: Strogatz, 2001.

3.3.4 Métricas para o Estudo de Redes Complexas

Redes complexas podem apresentar as mais diferentes topologias. Para quantificar a estrutura das ligações, várias medidas foram (e ainda estão sendo) desenvolvidas. Com estas métricas as redes podem ser analisadas, classificadas e modeladas com os mais diversos propósitos (Costa *et al*, 2007).

Nas seções seguintes serão descritas algumas métricas para estudo de redes complexas que são de interesse do presente trabalho. Além das métricas supracitadas, outras foram propostas e podem ser aplicadas em face às características da rede complexa a ser analisada. Uma lista mais extensa de tais métricas podem ser encontradas nos trabalhos de Barbieri (2010), Metz *et al* (2007) e Newman (2003).

3.3.4.1 Modularidade (*Modularity*)

A maioria das redes complexas possui estrutura modular, isto é, as conexões são densamente distribuídas entre vértices que pertençam a um mesmo grupo e espalhados em várias direções entre os vértices de grupos distintos. Temos ainda o conceito de comunidades (Danon *et al*, 2005), que são definidas por vértices altamente conectados entre si e pouco conectados com o restante da rede.

A identificação das comunidades tem ainda várias aplicações práticas, como no caso da Teia Mundial, onde páginas relacionadas ao mesmo assunto são altamente conectadas por links. Neste caso, a identificação das comunidades pode ajudar na busca por informações, já que os algoritmos de busca podem se basear nesta divisão.

O processo de identificação das comunidades constitui um problema NP-completo de difícil solução, pois, geralmente não se tem ideia de quantas comunidades formam a rede. Além disso, em outros casos, comunidades podem ser definidas hierarquicamente, quando se tem comunidades dentro de outras comunidades (Ravasz e Barabási, 2003). Apesar dessas dificuldades, muitos métodos têm sido propostos para identificar tais estruturas, mas o seu uso depende dos resultados desejados e das limitações existentes quanto ao poder computacional disponível, já que o tempo de execução dos algoritmos atualmente existentes varia de $O(N^2 \log(N))$ a $O(N^4)$ (Danon *et al*, 2005), sendo N proporcional ao tamanho da rede.

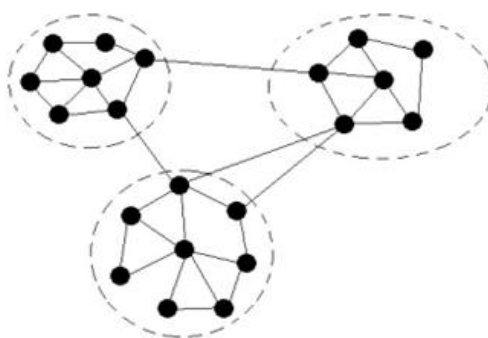


FIGURA 16 - Exemplo de uma rede com estrutura modular.

Fonte: Danon *et al*, 2005.

Na **Figura 16** podemos identificar as comunidades através das linhas tracejadas. Para se determinar a qualidade de uma divisão particular, é

utilizada a medida de modularidade, Q , que foi proposta por Newman (Newman, 2004a). Para uma dada divisão de uma rede em c comunidades, constrói-se uma matriz E , $c \times c$, cujos elementos ao longo da diagonal principal, e_{ii} fornecem a fração das conexões entre os vértices na mesma comunidade e os elementos e_{ij} $j \neq i$ representam a fração de conexões entre as comunidades i e j (Newman, 2006). A modularidade Q é calculada da seguinte forma:

$$Q = \sum_i [e_{ii} - (\sum_j e_{ij})^2] = Tr E - ||E^2||.$$

EQUAÇÃO 3 - Equação para o cálculo da modularidade.

$Q = 1$ identifica uma rede formada por módulos desconectados. Valores altos de Q determinam redes com estrutura modular definida. Uma metodologia bastante popular usada na identificação de comunidades é baseada no grau de intermediação (*betweenness centrality*) das conexões da rede (Girvan e Newman, 2001). Neste caso, são removidos os vértices de maior grau de intermediação, que representam os gargalos na rede, ou seja, aqueles que estão entre as comunidades. Existe ainda o método aglomerativo proposto inicialmente por Newman (2004b) e depois otimizado por Clauset *et al* (2004). Neste caso, as comunidades são aglomeradas de forma a obter o maior valor possível da modularidade.

As duas abordagens citadas constituem os principais métodos para o cálculo da métrica de modularidade resultando em algoritmos distintos principalmente no que tange a complexidade de tempo. A escolha do método a ser utilizado, depende de dois fatores básicos: precisão e tempo de processamento (Brandes, 2001).

3.3.4.2 Pagerank

O Pagerank é um algoritmo iterativo que assinala um peso numérico para cada nodo com o propósito de estimar sua importância relativa no grafo. O algoritmo foi inicialmente proposto por Brin e Page (1998) para ordenar resultados de busca do protótipo de máquina de busca da Google. A intuição por trás do Pagerank é que uma página Web é importante se existem muitas

páginas apontando para ela ou se existem páginas importantes apontando para ela. A equação que calcula o Pagerank (PR) de um nodo i , $PR(i)$, é definida da seguinte forma:

$$PR(i) = (1 - d) + d \sum_{v \in S(i)} \frac{PR(v)}{N_v}$$

EQUAÇÃO 4 - Equação para o cálculo do Pagerank.

Na equação acima definimos o $S(i)$ como o conjunto de páginas que apontam para i , N_v denomina o número de arestas que saem do nodo v , e o parâmetro d é um fator que pode ter valor entre 0 e 1. Embora o algoritmo do Pagerank tenha sido idealizado para utilização nas páginas web, tem sido aplicado em outros contextos, como por exemplo, para encontrar usuários influentes no Twitter. No trabalho de Weng *et al* (2010) foi criado o *Twitterrank* (uma extensão do algoritmo do Pagerank) que é utilizado para mensurar a influência dos usuários desta rede social.

3.3.4.3 Centralidade (Centrality)

De acordo com Marteleto (2001) calcular a centralidade de um ator significa identificar a posição em que ele se encontra em relação às trocas e à comunicação na rede. Embora não se trate de uma posição fixa, hierarquicamente determinada, a centralidade trás consigo a ideia de poder. Quanto mais central é um ator, mais bem posicionado ele está em relação às trocas e a comunicação, o que aumenta a sua influência na rede.

A respeito desta métrica Marteleto (2001) adverte para o fato de que muito embora alguns atores possuam mais contatos diretos em uma rede, estes podem não ser necessariamente aqueles que ocupam as posições mais centrais e esta ocorrência pode ser explicada por meio do conceito de abertura estrutural. Um ator com poucas relações diretas pode estar muito bem posicionado em uma rede por meio da utilização estratégica de suas aberturas estruturais, o que interfere diretamente na centralidade de um ator.

3.3.4.3.1 Centralidade de grau (*Centrality Degree*)

A centralidade de grau conta o número de arestas, ou grau $d(n_i)$ (a letra d vem do inglês *degree*), que cada ator i tem, ou seja, $CD(n_i) = d(n_i)$. Podemos ainda tratar diferentemente o grau de entrada e o grau de saída no caso de redes dirigidas (Wasserman e Faust, 1994).

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

EQUAÇÃO 5 - Equação para o grau de centralidade.

Na **Equação 5**, 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 Donniger (1986), a **Equação 6**, baseada nos modelos gráficos probabilísticos de Erdos e Reyni (1960), poderá ser utilizada para determinar a Centralidade de um nó. Onde $d(n_i)$ é o grau do nó a ser calculado e g é o total de nós do grafo.

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

EQUAÇÃO 6 - Equação normalizada para o grau de centralidade.

No caso de um grafo direcionado, considera-se o grau de entrada (*In-Degree*), $dl(n_i)$, e o grau de saída (*Out-degree*), $do(n_i)$, de acordo com a direção dos arcos que chegam ou partem do nó. De modo análogo a **Equação 7**, podemos obter distintamente o graus de saída e o grau de entrada conforme as equações a seguir:

$$\text{(I)} \quad C'_{D_O}(n_i) = \frac{d_O(n_i)}{g-1} \quad \text{(II)} \quad C'_{D_I}(n_i) = \frac{d_I(n_i)}{g-1}$$

EQUAÇÃO 7 - Cálculo do grau de saída (I) e grau de entrada (II).

Cabe a ressalva de que os cálculos para a métrica centralidade de grau descritos nas **Equações 6 e 7** encontram-se devidamente normalizados. A Normalização serve para que sejam utilizados valores em uma determinada faixa (geralmente 0 e 1) visando uma maior padronização dos valores obtidos.

3.3.4.3.2 Centralidade de intermediação (*Betweenness Centrality*)

Para medir o tráfego que passa em um dado vértice (ou aresta), é usada a medida chamada centralidade de intermediação (*betweenness centrality*) (Freeman, 1977; Anthonisse, 1971), que mede o quanto um vértice ou aresta está no caminho entre outros vértices. Seja g_{jk} o número de caminhos geodésicos (caminhos mais curtos e de mesmo tamanho) que ligam os vértices j e k , e $g_{jk}(n_i)$ o número de tais caminhos, no total de g_{jk} , que passa pelo vértice n_i . O índice de centralidade de intermediação é:

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

EQUAÇÃO 8 - Equação para centralidade de intermediação.

A equação mede, para um vértice n_i , a soma de probabilidades de o mesmo estar no caminho geodésico entre todos os demais vértices da rede. Para normalizar tal índice, deve-se dividi-lo pelo seu máximo possível, que é o número de pares de vértices na rede que não incluem n_i , ou seja, $(g - 1)(g - 2) / 2$, chegando-se ao seguinte índice de centralidade de intermediação normalizado por ator:

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

EQUAÇÃO 9 - Equação normalizada para centralidade de intermediação.

Exemplificando, na **Figura 17** os vértices A, B, C e D são os que possuem maior grau de centralidade de intermediação.

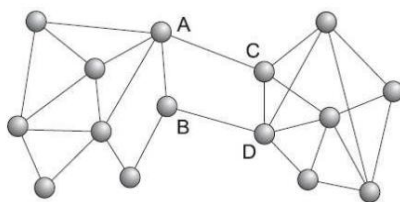


FIGURA 17 - Nós com maior grau de intermediação.

Portanto, a centralidade de intermediação é o potencial daqueles atores que servem de intermediários. Representa o quanto um ator atua como “ponte”, facilitando o fluxo de informação em uma determinada rede (Marteleteo, 2001).

3.3.4.3.3 Centralidade de proximidade (*Closeness Centrality*)

A centralidade de proximidade de um ator mede o quanto o vértice, que representa o ator, está próximo de todos os demais da rede (Freeman, 1979; Sabidussi, 1966). Para calcular a centralidade de proximidade é somada a distância geodésica do vértice em relação a todos os demais vértices da rede e depois inverte, uma vez que quanto maior for a distância, menor a proximidade, chegando-se à seguinte fórmula:

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

EQUAÇÃO 10 - Equação para centralidade de proximidade.

A normalização do índice, para que o mesmo fique na faixa de valores entre 0 e 1, pode ser feito multiplicando-se $C_C(n_i)$ por $g-1$, obtendo-se:

$$C'_C(n_i) = \frac{(g-1)}{[\sum_{j=1}^g d(n_i, n_j)]} = (g-1)C_C(n_i)$$

EQUAÇÃO 11 - Equação normalizada para centralidade de proximidade.

Portanto, a centralidade de proximidade mede o quanto um ator é tão mais central quanto menor o caminho que ele precisa percorrer para alcançar os outros atores da rede, isto é, a sua independência em relação aos outros atores da rede (Marteleteo, 2001).

3.3.4.4 Autoridade (*Authority*)

Hyperlink-Induced Topic Search (HITS) é um algoritmo que calcula os valores das métricas *hubs* e *Authority* (autoridade) utilizadas para análise de ligação. Este algoritmo foi precursor do Pagerank e criado por Kleinberg (1999) com o objetivo inicial de classificar as páginas da Web.

A ideia por trás Hubs e Autoridades decorre de uma visão especial para a criação de páginas da web quando a internet estava se formando originalmente, ou seja, certas páginas da web, conhecidos como hubs, serviam como grandes diretórios que não eram detinham a informação original, mas eram usados como compilações de um amplo catálogo de informações que levavam os usuários diretamente para outras páginas autorizadas (Kleinberg, 1999). Em outras palavras, um hub representava uma página que apontava para outras páginas, e uma autoridade representava uma página que estava ligada por diferentes hubs. Veja o detalhe dessas estruturas na Figura abaixo:

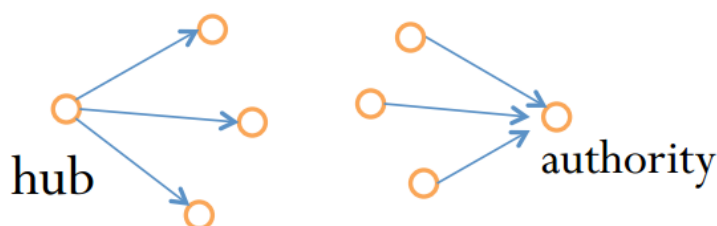


FIGURA 18 - Hub e Autoridade.

Assim, o esquema atribui duas pontuações para cada página: a sua autoridade, que estima o valor do conteúdo da página, e seu valor de hub, que estima o valor de suas ligações para outras páginas. Os Valores de autoridade e hub são definidos em termos de um ao outro em uma recursão mútua. O algoritmo executa uma série de iterações, cada um consistindo de dois passos básicos:

- **Atualização da Autoridade:** A atualização da pontuação de autoridade de cada nó deve ser igual à soma dos índices de *Hub* de cada nó que aponta para ele. Ou seja, a um nó é dada uma pontuação de alta autoridade por estar ligado por páginas que são reconhecidos como *hubs* de informação.

- **Atualização do Hub:** A atualização da pontuação de *Hub* de cada nó deve ser igual à soma dos índices de *Autoridade* de cada nó que ele aponta. Ou seja, a um nó é dada uma pontuação elevada hub por estar ligado a nós que são considerados autoridades no assunto.

Assim como no caso do Pagerank, este algoritmo também pode ser utilizado em outros contextos. No trabalho de Silva (2012a) verificamos a utilização destas métricas no contexto das redes sociais no intuito de identificar as “Autoridades” que propagam a informação por disseminação ou formação de opinião e os também conhecidos por “hubs” que são os atores que interligam várias autoridades.

3.3.4.5 Subgrupos (*Clique*)

Emirbayer e Goodwin (1994) denominam cliques como um grupo de atores no qual cada um está direta e fortemente conectado a todos os outros. A definição de clique é um ponto de partida útil para especificar a propriedade coesiva de subgrupos. Segundo essa definição, deve haver no mínimo três vértices para compor um clique (Freitas e Pereira, 2005).

A **Figura 19** mostra um exemplo de uma rede com três subgrupos (cliques). Os cliques podem representar uma instituição, um subgrupo específico e mesmo identificar a movimentação em torno de um determinado problema (Marteleto, 2001). É nos cliques que existe uma densidade maior de comunicação, ou seja, é mais eficiente compartilhar informações dentro de um grupo (Freitas e Pereira, 2005).

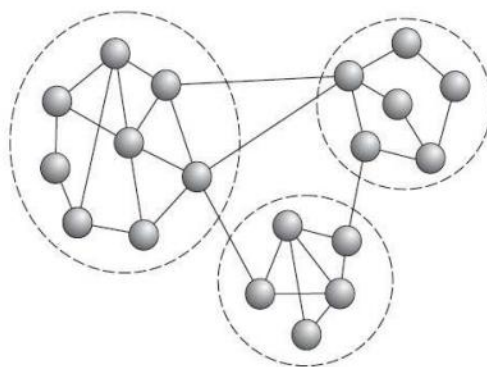


FIGURA 19 - Rede com três subgrupos (cliques).

3.3.5 Ferramentas para Análise de Redes Sociais

Um software ou API para análise de redes sociais pode ser utilizado para identificar, representar, analisar, visualizar ou simular entidades sociais (vértices) e seus respectivos relacionamentos (arestas). Essas ferramentas trabalham com vários formatos de arquivos de entrada e saída de dados. Além disso, elas permitem aos seus usuários investigarem representações de redes independente do seu tamanho.

As representações visuais das redes sociais são importantes para entender os dados da rede e transmitir o resultado da análise. Neste sentido iremos analisar separadamente os programas e API's para ARS que foram utilizados neste trabalho.

No trabalho de Huisman e Van Duijn (2004) podemos encontrar uma lista extensa de softwares e ferramentas utilizadas prioritariamente para análise de redes sociais. Algumas destas ferramentas encontram-se descritas em maiores detalhes no [Anexo F](#) e [Anexo G](#) do presente trabalho.

3.3.5.1 Programas para ARS

3.3.5.1.1 Gephi

O Gephi¹ é uma plataforma interativa que tem como objetivo 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: grau de intermediação, grau de proximidade, coeficiente de agrupamento, PageRank entre outras.

Por intermédio desta ferramenta o usuário pode interagir com as estruturas de dados através dos diversos *layouts* existentes. Desta maneira, através de diversos algoritmos os usuários poderão alterar as configurações de visualização do grafo durante a execução do mesmo, aumentando desta forma a capacidade de entendimento das informações, fazendo com que as propriedades mais ocultas passem a ser visualizadas e compreendidas.

¹ <https://gephi.org/>

Um dos objetivos do Gephi é auxiliar os analistas de dados de modo que a partir da criação de hipóteses, sejam descobertos intuitivamente padrões, singularidades e isolamentos estruturais das informações representadas no grafo. Ainda, este software pode 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.

3.3.5.1.2 Pajek

O software *Pajek*¹ tem como funcionalidades a visualização e análise de redes. Devido à facilidade de instalação e uso, esta ferramenta vem sendo amplamente utilizada por docentes em aulas relacionadas a grafos. A geração de uma rede 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.

Embora esta ferramenta permita uma ótima visualização dos dados através de um grafo e, possua algumas métricas de ARS implementadas, a mesma tornou-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.

3.3.5.2 API's para ARS

3.3.5.2.1 Prefuse

O *Prefuse*² 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.

¹ <http://pajek.imfm.si/doku.php?id=download>

² <http://prefuse.org/>

Destaca-se ainda neste framework, a base arquitetural com seus componentes distribuídos no 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.

3.3.5.2.2 JUNG

O *JUNG*¹ (*Java Universal Network / Graph Framework*) é 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.

Cabe ainda o comentário de que o mesmo fornece um quadro de visualização que facilita a construção de ferramentas para exploração interativa de dados da rede. Deste modo, os usuários poderão algum dos diversos algoritmos de *layout* definidos na biblioteca ou usar suas funcionalidades para desenvolver seus próprios *layouts* de acordo com a necessidade.

Como uma biblioteca de código aberto, *JUNG* estabelece um quadro comum para geração de grafos, visualização e análise de rede. Neste aspecto, ele torna-se um mecanismo facilitador uma vez que podemos criar as estruturas através de classes ou objetos ou através de arquivos (.) GraphML.

¹ <http://jung.sourceforge.net/>

3.3.5 Considerações

Uma rede é um conjunto de elementos, que chamamos de vértices ou nodos, com conexões entre eles, chamadas de arestas. A estrutura topológica de uma rede pode ser então modelada por um grafo. No caso específico do nosso estudo em que apresentamos como foco as redes sociais, podemos convenientemente representar as pessoas através dos nós e seus possíveis relacionamentos/interesses comuns através das arestas, caracterizando cada uma destas subestruturas com atributos de acordo com nossa conveniência.

Para analisarmos relações existentes dentro de um conjunto de informações, podemos fazer uso da teoria dos grafos. A teoria dos grafos se preocupa com o estudo dos relacionamentos entre objetos de um determinado sistema, podendo ser utilizada em qualquer área que necessite de organização de dados, como: sociologia, topologia, química, programação linear. Mesmo entendendo a grande importância da teoria dos grafos e de sua ampla utilização, não aprofundamos nesse assunto por motivos de simplicidade. Apenas introduzimos alguns conceitos básicos de grafos onde o restante que não foi devidamente abordado ao longo desta parte do trabalho poderá ser encontrado no trabalho de Bondy e Murty (1982).

De acordo com o que foi exposto nesta seção sabemos que as redes sociais são inerentemente redes complexas. Consequentemente, vários estudos recentes analisaram as características de diferentes redes sociais utilizando como base teorias existentes da área de redes complexas (Benevenuto *et al*, 2011). De fato, o estudo de redes complexas cobre um grande número de áreas e sua teoria tem sido utilizada como ferramenta para entender vários fenômenos, incluindo o espalhamento de epidemias, propagação de informação, busca na Web, consequências de ataques a redes de computadores entre outros trabalhos descritos na [seção 2.2](#).

A Análise de Redes Sociais (ARS) é baseada na descrição formal de redes através das diversas medidas utilizadas para caracterizar as redes complexas. Ela provê vocabulários e modelos para descrever estruturas sociais, assim como métodos para analisá-las. Sua formalidade permite precisar conceitos como "papel social" ou "grupo social", por exemplo, e, por

consequente, possibilita uma padronização que torna os estudos nela baseados mais coerentes e facilmente comparáveis (Chakrabarti, 2003). Ainda, de acordo com Simões (2011) através da ARS é possível compreender e acompanhar de forma mais eficaz a disseminação de informações e a interação entre as pessoas que compõem uma determinada rede.

Com relação aos programas para ARS elencados cumpre dizer que se optou pela utilização do Gephi devido a sua vasta gama de métricas definidas em seu laboratório de dados e aos diversos tipos de algoritmos de distribuição disponibilizados para modificar as formas de visualização da rede. Ainda, temos o fato de que esta ferramenta utiliza prioritariamente arquivos de extensão (.) graphML¹ em detrimento ao formato (.) net utilizado no Pajek.

No caso das bibliotecas de programação, mantemos a utilização do JUNG e do Prefuse. Ambas as bibliotecas fornecem meios para construção da rede através de objetos ou através do “carregamento” de arquivos (.) graphML (com isto podemos “abstrair” as formas de representação computacional dos grafos descritas na [seção 3.3.2](#) e nos preocupamos apenas com a criação deste tipo de arquivo). Contudo o prefuse não traz consigo embutido as métricas necessárias para análise de redes sociais, em contrapartida, fornece uma boa variedade de layouts de visualização das redes. Para suprir a deficiência do prefuse, o JUNG traz uma boa gama de métricas prontas para uso através das simples chamadas de métodos.

¹ <http://graphml.graphdrawing.org/>

3.4 Rede Social Twitter

Nesta seção apresentaremos as principais funcionalidades da rede social Twitter que se aplicam ao nosso estudo. Assim, a presente seção encontra-se dividida em quatro subseções: Em Conceitos Básicos ([Seção 3.4.1](#)) iremos discorrer sobre os principais aspectos da rede social em questão; Em Twitter no debate político ([Seção 3.4.2](#)) ampliaremos a discussão sobre os modos de utilização do *microblog* na disseminação de informações de cunho político; Em Acesso aos dados do Twitter ([Seção 3.4.3](#)) abordaremos a forma mais convencional de realizar busca e coleta de dados; Por fim em considerações ([Seção 3.4.4](#)) iremos avançar o debate em alguns tópicos do twitter relacionado com eleições.

3.4.1 Conceitos Básicos

O Twitter é um *microblog* que permite atualizações rápidas e curtas a partir de diferentes modalidades. Existe possibilidade de atualizar o Twitter pela *Web* através de *instant messaging* (IM) ou mesmo pelo celular através de mensagens SMS (*short message service*). Existem outras ferramentas de *microblog* na *Web*, como o Tumblr (comprado pela Yahoo em maio de 2013) e o Jaiku (comprado pelo Google e encerrado em janeiro de 2012).

A ferramenta surgiu em 2006 com o simples objetivo de que as pessoas respondessem à pergunta: “o que você está fazendo agora?”. Diante de uma pergunta tão simples, as respostas foram limitadas a 140 caracteres (O’reilly, 2009). Esta imposição no tamanho das mensagens parecia ser uma limitação no início do serviço, mas hoje representa um dos seus pontos fortes.

Desenvolvido pelo americano Jack Dorsey, o Twitter tem sua plataforma aberta, o que significa que qualquer um pode utilizar o código do sistema para buscar usos criativos. Há uma infinidade de *sites* e programas derivados do Twitter. Em entrevista ao jornal *El País*¹, Jack Dorsey afirmou que o Twitter não é apenas uma “rede social”, mas sim uma “ferramenta de comunicação”. O referido autor salientou ainda que:

¹ Jornal *El País*. Disponível em: <<http://migre.me/f9XY4>>. Acesso em junho de 2013.

“O Twitter é uma forma de comunicar com as pessoas muito portátil e acessível. Qualquer um com um telefone Nokia pode entrar na minha conversa. O Twitter permite controlar o que tem e aquilo que recebe e as pessoas podem optar por ler ou não. Aquilo que o Twitter faz é simplificar a comunicação entre as pessoas, expressar as tendências daquilo que se está a passar em cada cidade e no mundo, para além de forçar a interação entre os indivíduos que anteriormente se relacionavam cara a cara”.

Mesmo assim, a rede pode ser considerada uma “rede social” por conectar pessoas. É certo que estas pessoas não têm a “obrigação” de adicionar-se mutuamente, o que torna o ato de seguir alguém no Twitter uma ação, por vezes, unilateral. O Twitter além de ser uma rede social também é um lugar de auto-publicação, onde pessoas colocam *links*, falam de seus projetos, divulgam notícias. “Não há regra. Uns falam sobre a vida ou postam links; outros “*tittam*” das férias. É um serviço que está se definindo. A tecnologia é nova. Todos os dias as pessoas acham novas utilidades”, complementou Dorsey.

O Twitter se tornou um *hype* na Internet brasileira no ano de 2009, oportunidade em que registrou um volume exponencial de criação de perfis no sistema. Mesmo sendo um sucesso, a baixa compreensão de como “funciona” o sistema ou, mesmo, a questão que norteia o serviço (“o que você está fazendo agora?”) faz com que muitos usuários se cadastrem, mas não utilizem o serviço com frequência ou, ainda, nem voltem a utilizá-lo (Cardoso, 2009).

Em matéria concedida à *Folha On-line* no ano de 2009, o diretor de *insights* da Edelman Digital, Steve Rubel¹, afirmou que: “o Twitter está atingindo um ponto alto porque atraiu a maioria das pessoas que quer viver em público. O Facebook, por sua vez, permite mais intimidade”. Rubel revelou acreditar que o Twitter deve crescer bruscamente por seis a nove meses e, depois disso, ficar em nível estável. Contudo, nos dias atuais o fato que atrai mais usuários é a possibilidade de postar a partir de múltiplas plataformas denominadas de dispositivos móveis.

¹ *Folha On-Line*. Disponível em: <<http://migre.me/f9XQX>>. Acesso em junho de 2013.

3.4.2 Twitter no debate político

O Twitter pode ser considerado uma das ferramentas recentes que melhor sistematiza o processo de interação e participação política nas novas mídias, sendo que nas eleições de 2010 e, particularmente nas eleições de 2012, ele foi alvo de atenção de diversos candidatos como forma de alcançar uma parte do seu eleitorado, bem como dos próprios eleitores no sentido de expor a sua opinião política a respeito dos candidatos.

Mesmo que a rede ainda tenha acesso restrito, o debate produzido no *microblog* não se limita a este espaço, pois os usuários que mencionam algum dos candidatos atuam também como mediadores, repassando as informações aos demais usuários (eleitores) ou interagindo com fontes tradicionais de informação. Dessa forma, aqui se pretende discutir como o twitter atua em processos eleitorais, assim como o seu desenvolvimento nos últimos anos como meio para deliberação política.

O *microblog* pode ser utilizado para as mais diversas finalidades, desde postagens de questões banais até informações de interesse público. Para os que não conhecem o mecanismo, no twitter a relação entre os usuários se dá baseada na lógica do “siga-me” (*follow*). O usuário x pode escolher seguir o membro y, sendo que receberá as mensagens postadas por este último. O mesmo vai ocorrer caso y deseje seguir o indivíduo x. E quanto mais pessoas seguir (*follow*), mais mensagens receberá em sua página. Os candidatos que possuem página no twitter geralmente têm um número grande de seguidores (*followers*), o que permite que suas postagens cheguem às páginas de uma quantidade relevante de eleitores.

Segundo Lemos (2009), esse processo permite uma conversação, por meio dos mecanismos de resposta, possibilitando maior interação e uma alternativa aos meios massivos. Além da conversação entre os usuários, o twitter permite uma instantaneidade que o difere de outras mídias. De acordo com Orihuela (2007), os perfis das postagens são: **a)** trivialidades cotidianas; **b)** conversas em pequenas comunidades; **c)** compartilhar informação e endereços de sites; e **d)** difundir notícias e opiniões.

Sobre os usos da ferramenta, Recuero (2009) aponta para uma grande tendência dentre os usuários em favor da publicação de informações, embora ela também seja usada para a conversação. O objetivo é buscar informações qualificadas e também espalhá-las em suas redes sociais.

Tratando-se especificamente do uso do twitter pela elite política, essa ferramenta é a versão *online* da relação corpo a corpo que os candidatos possuem na disputa política (Spyer, 2009). “O atual presidente dos Estados Unidos, Barack Obama, usou com habilidade essa ferramenta durante sua campanha em 2008, aproveitando o canal direto com os cidadãos para avisar sua agenda e compartilhar perspectivas sobre assuntos que são debatidos durante a eleição” (Spyer, 2009, p. 65).

Em relação ao crescimento do uso da ferramenta, Spyer (2009) mostra que uma pesquisa feita em junho de 2009 pela *ComScore* revelou que o *Twitter* cresceu junho de 2008 e 19% em relação ao mês anterior, atingindo 44,5 milhões usuários únicos. No entanto destaca-se que desses usuários 45% estão nos EUA. Isso explica a relevância que a ferramenta obteve na campanha de Obama, pois grande parte da população possui o serviço e pode manter contato quase que permanente com o candidato.

Já no caso brasileiro, a *Semiocast*¹ aponta o país como segundo maior usuário em número de contas ativas com 33,3 milhões de contas (quase 10% do total), superando os 29,9 milhões do Japão. Constatou ainda que 27% dos usuários do Twitter em todo o mundo são considerados ativos (aumento de 10% em relação a 2009). Esse índice no Brasil foi de 25% de usuários ativos e, no Japão, 30%. Embora os japoneses sejam mais atuantes, em 2011, com o lançamento da versão em português, a participação brasileira deu contribuição importante para o crescimento.

Quanto ao perfil das pessoas que passaram a usá-la no Brasil, Spyer (2009) descreve como sendo predominantemente usuários avançados da Web, jovens, estudantes e com renda mensal alta, Essas características também representam o público atingido diretamente pela propaganda eleitoral no twitter.

¹ Pesquisa *Semiocast*. Disponível em: <<http://migre.me/fa1Kd>>. Acesso em maio de 2013.

3.4.3 Acesso aos dados do Twitter

Em um passado recente, redes sociais eram um domínio de sociólogos e antropólogos, que utilizavam pesquisas e entrevistas com pequenos grupos de usuários como ferramentas de coleta de dados. Com o surgimento das redes sociais *online*, a obtenção de dados reais em larga escala se tornou possível, e pesquisadores de diversas áreas da computação começaram a realizar coletas de dados.

No trabalho de Benevenuto (2011) podemos verificar diversas formas para acesso/coleta de dados das redes sociais em geral. Na presente pesquisa optamos pelo acesso através de uma interface de programação fornecida pelo próprio Twitter que será discutido na seção seguinte.

3.4.3.1 Twitter API

Considerado um mecanismo de informação e comunicação de rede, o *microblog* produz mais de 200 milhões de mensagens (*tweets*) por dia (Cardoso, 2009). O Twitter oferece acesso aos seus dados, através das suas APIs. Cada API representa uma faceta que permite aos desenvolvedores construir suas aplicações de maneira simplificada. A partir disso, encontramos no site oficial do twitter¹ uma subdivisão da API 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** - Permite consultar o conteúdo do Twitter, isso inclui encontrar um conjunto de *tweets* com palavras-chave específicas, que referencia a algum usuário ou ainda que 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 grandes quantidades de palavras-chave para ser especificadas e rastreadas.

¹ <https://dev.twitter.com/start>

- **Rest API** - Possui um conjunto de ferramentas REST¹ para acesso a informações dos usuários do twitter, como seguidores, seguidos, quantidade de amigos, timeline, entre outros.

Utilizaremos a Twitter API com a finalidade de extrair os dados dos usuários e seus relacionamentos junto ao twitter bem como para realizar a busca de *tweets* através de palavras-chave. Para utilização de algumas facetas da API precisamos configurar uma conta de autenticação OAUTH² para realizar as requisições dos serviços das ferramentas descritas. Cabe ressaltar de que nossa aplicação além de tratar diretamente com os meios necessários para autenticação, também tem que prover os meios necessários de manipular as informações vindas da API do Twitter (em sua maioria no formato JSON).

Diante do anteriormente exposto, e com a restrição de que a ferramenta para busca e coleta de dados foi desenvolvido na linguagem Java (maiores detalhes na [seção 4.3](#)), buscamos uma API que definisse uma forma de tratar os meios de autenticação e de manipular o resultados das requisições de forma mais amigável. Resumidamente, esta API teria como premissa fornecer um meio de comunicação direta com a API do Twitter.

3.4.3.2 Twitter4J

Como solução do problema anteriormente citado surge a Twitter4J³. Esta ferramenta de software é uma biblioteca Java não oficial para a API do Twitter. A partir do seu uso, pode-se facilmente integrar uma aplicação Java com os diversos serviços do Twitter. Com isso o desenvolvedor pode usufruir do uso de uma linguagem de alto nível em conjunto com o acesso “praticamente ilimitado” de informações do twitter.

Temos uma infinidade de facilidades com a Twitter4J onde basicamente escrevemos código nativo da linguagem Java, ou seja, permanecemos no campo dos objetos com seus atributos e métodos em detrimento da

¹ **RE**presentantion **S**tate **T**ransfer (REST) - Atualmente pode ser entendido como um conjunto de regras usadas para descrever uma interface web que utiliza HTTP e XML (JSON,ATOM).

² <http://oauth.net/>

³ <http://twitter4j.org>

manipulação direta de URL's e arquivos estruturados (JSON e XML). A Twitter4J fornece todas as funcionalidades através de classes, assim, quando da sua utilização em projetos já existentes, teoricamente temos um impacto mínimo na manutenção do código, além de facilitar a criação de código por parte de especialistas na linguagem Java.

3.4.4 Considerações

No início do nosso trabalho, especificamente na [seção 2.4](#) mostramos que uma das principais falhas nas pesquisas que utilizavam o Twitter com fins a “prever” algum acontecimento estava no fato de não considerar que nem todo segmento social encontra-se representado neste tipo de mídia. Este é um fator que difere de nação para nação por fatores que vão desde aspectos culturais até aspectos socioeconômicos. Discutimos brevemente esse assunto na [seção 3.2.1](#) trazendo como contexto os aspectos relacionados ao Brasil.

Contudo, à medida que o tempo passa, encontramos evidências de que uma maior representatividade dos segmentos sociais (independentemente do país em questão) tem sido verificada pelo exponencial crescimento dos usuários das redes sociais.

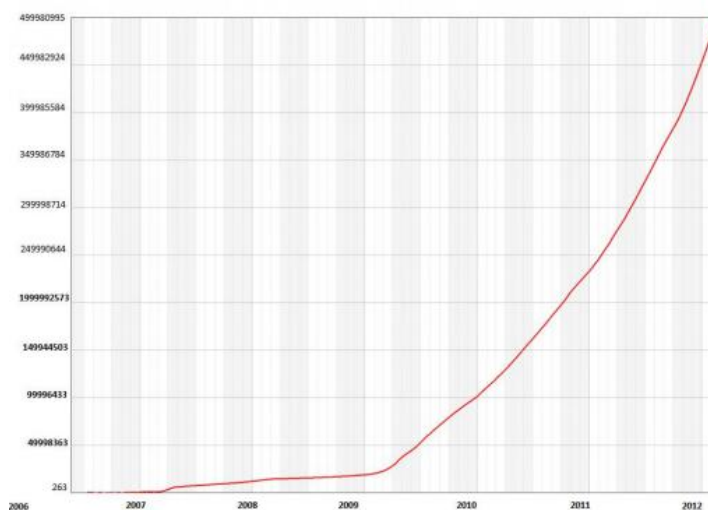


FIGURA 20 - Crescimento dos usuários do Twitter (em milhões).
Fonte: Site Wearesocial¹

¹ Site Wearesocial. Disponível em: <<http://migre.me/fhEB9>>. Acesso em maio de 2013.

No gráfico apresentado na **Figura 20**, a curva ascendente de 2009 a 2012 indica um crescimento quase exponencial – de 2009 a 2010 cresce cinco vezes; de 2010 a 2011, mais de duas vezes; de 2011 a 2012, mais de duas vezes. Embora em relação aos dados oficiais estejam destoantes, a tendência acima apresentada fortalece a ideia da popularização das redes sociais no meio digital.

Com isso, ainda que não seja possível afirmar que se trate de uma tendência constante ou irreversível, é inegável que significa uma mudança expressiva nas relações entre as pessoas. O processo em curso de digitalização da comunicação não opera à margem das mudanças nas instituições formais e consolidadas, mas as instrumentaliza. Ainda que aquele seja mais veloz que estas. Isso significa que famílias, igrejas, partidos políticos (entre outros) têm alterado seu padrão de relacionamento comunicacional em função das novas possibilidades de interação e troca de informações.

Conforme vimos, a ferramenta twitter é utilizada por um público diferenciado das demais redes sociais (veja detalhes na [seção 3.2.4](#)) que utiliza prioritariamente para disseminar suas opiniões. O twitter pode ser considerado uma das ferramentas recentes que melhor sistematiza o processo de interação e participação política nas novas mídias onde o usuário simplesmente faz uso dos *tweets* para expor suas preferências políticas ou mesmo para realizar críticas as demais linhas políticas. Desta forma ele pode influenciar diretamente seus seguidores ou mesmo ser influenciados através dos *tweets* publicados pelos seus seguidos. Esta afirmação encontra sua lógica nos conceitos elencados na [seção 3.4.5](#) onde temos que a maioria dos eleitores reúne uma gama de informações (o Twitter pode ser considerado como uma das fontes) que o auxiliam para o processo decisório de qual candidato votar.

Utilizamos a API do Twitter devido à facilidade de uso e a extensa documentação encontrada em seu site oficial. Utilizamos a Twitter4J para fornecer um meio para que a ferramenta de busca e coleta de dados pudesse comunicar diretamente com a API do Twitter. Maiores detalhes sobre a utilização de ambas APÍ's será descrito na [seção 4.5](#).

4. Metodologia

O objetivo deste capítulo é esclarecer os caminhos percorridos para atingir os objetivos propostos para esta pesquisa. Desta forma o presente capítulo encontra-se dividido em cinco seções: Em metodologia de análise ([Seção 4.1](#)) encontram-se dispostos de modo genérico as principais etapas necessárias para realização da pesquisa; Em arquitetura da *AppSNAElection* ([Seção 4.2](#)) estão descritas as tecnologias utilizadas para a construção da ferramenta de coleta de dados junto ao Twitter; Em Funcionamento da *AppSNAElection* ([Seção 4.3](#)) é demonstrado o fluxo de dados da ferramenta de busca e coleta de dados; Em modelagem da rede ([Seção 4.4](#)) expomos a heurística envolvida para definição dos nós e seus relacionamentos visando a construção da rede que será objeto de estudo; Finalizando iremos expor as principais considerações sobre o referida metodologia ([Seção 4.5](#)).

4.1 Metodologia de Análise

Conforme descrito na [seção 3.1.4](#) nossa pesquisa teve como cenário principal as eleições municipais da cidade de João Pessoa, onde, a partir de algumas regras, usuários do *microblog* Twitter serão identificados e relacionados a algum (ou alguns) candidato(s) a partir do conteúdo publicado ao longo do período compreendido entre 09/09/2012 e 30/10/2012. Cabe a ressalvada de que o período supracitado engloba as atividades eleitorais referentes ao 1º e 2º turno das eleições.

Para cumprir esta atividade foi proposta uma metodologia para análise de redes sociais, em especial no *microblog Twitter*, baseada num conjunto de atividades que tem como objetivo identificar e mapear a relação entre os diversos usuários (participantes) de uma rede. Deste modo, nossa metodologia terá como fundamento principal o estudo e aplicação sistematizada das seguintes atividades ilustradas na **Figura 21**: planejamento, monitoramento, modelagem da rede, métricas e análises.

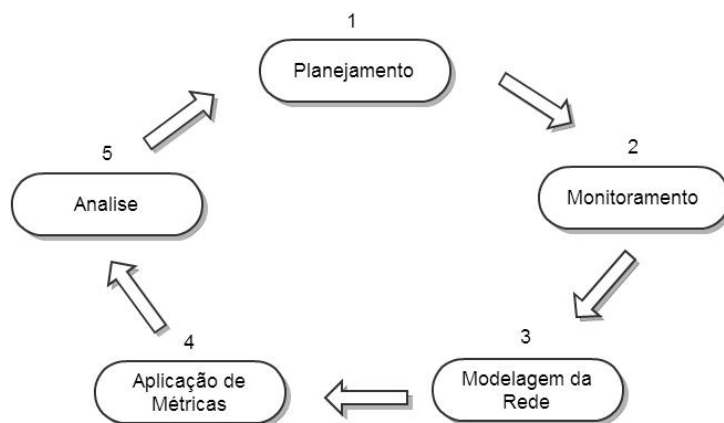


FIGURA 21 - Metodologia de ARS.

Os cinco estágios da “metodologia de ARS”, demonstram as atividades fundamentais para obtenção do objeto de estudo da nossa pesquisa. A seguir estas atividades são devidamente detalhadas.

- **Planejamento** - Neste estágio são traçados os objetivos, ou seja, as indagações referentes a um determinado ponto de vista, que buscam respostas, podendo estas ser positivas ou negativas. Nesta etapa são definidos os critérios utilizados no monitoramento tais como os grupos a serem pesquisadas, as palavras-chave que serão utilizadas para distinguirem as mensagens relevantes dentre outros.

- **Monitoramento** - Fase destinada à coleta de dados que serão a base de realização do experimento. Sendo assim, os atributos definidos como parâmetros no planejamento são inseridos na ferramenta de coleta de dados, de maneira que a mesma retorne os dados em arquivo XML ou CSV ou ainda inserindo estes dados diretamente em um banco de dados.

- **Modelagem da rede** - A partir dos dados obtidos 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, atributos em seus nós e arestas.

- **Métricas de ARS** - Esta fase objetiva-se a aplicação de métricas de análise de redes sociais, visando reconhecer padrões, 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.

Percebemos que a realização das atividades presentes na metodologia segue conforme um ciclo de maneira que, caso ocorra uma falha/insucesso nos resultados do 1º ciclo, inicia-se um novo a partir da atividade de planejamento e nesta fase serão ratificados/retificados os passos aplicados no ciclo anterior.

Para elaboração da presente metodologia foi utilizada a noção de ciclo introduzida através da atividade de PDCA (Moen e Norman, 2011) que é um método iterativo de gestão em quatro passos, utilizado para o controle e melhoria contínua de processos. O acrônimo é baseado nas palavras *Plan*, *Do*, *Check* e *Act* que significam (em tradução livre) Planejamento, Execução, Verificação e Ajuste. O conceito de PDCA por sua vez baseia-se no método científico, desenvolvido a partir do trabalho de Francis Bacon entitulado *Novum Organum*¹ no ano de 1620. O método científico pode ser escrito como "hipótese" - "experimento" - "avaliação" ou o Planejar, Executar e Verificar se comparado com as atividades do ciclo em questão.

4.2 Arquitetura da *AppSNAElection*

De acordo com as etapas da metodologia ARS proposta no item anterior, três elementos principais fazem parte são resultados de cada interação no ciclo: Dados, Informação e Conhecimento.

Os **dados** são os elementos conhecidos e determinados e no nosso contexto são representados basicamente pela interação dos usuários nos ambientes das redes sociais; A **informação** é obtida através do processamento e interpretação do item anterior e no nosso contexto pode ser, por exemplo, a informação do candidato mais citado; O **Conhecimento** é um conceito que transcende o item anterior, pois ele além de ter um significado tem uma aplicação em um determinado no nosso contexto pode ser, por exemplo, as implicações em um candidato receber muitas menções.

¹ Uma versão traduzida deste trabalho pode ser obtida em: <<http://migre.me/fe3kj>>

Trazendo os conceitos de dado, informação e conhecimento com vistas a materializar uma proposta para construção do software de busca e coleta de dados, a **Figura 22** abaixo descreve esta ideia onde, a partir da coleta de dados de determinado grupo de usuários, sendo estes obtidos através da interação destes com redes sociais *online*, podemos obter algumas informações importantes de cunho qualitativo ou mesmo quantitativo e, a partir desta última, gerarmos conhecimento que pode ser aplicado no contexto das eleições municipais da cidade de João Pessoa.

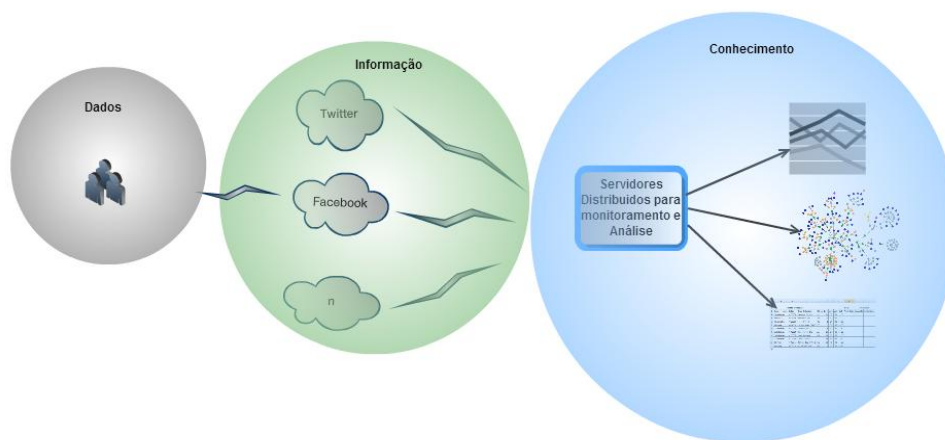


FIGURA 22- Proposta inicial da arquitetura.

Embora a figura acima nos apresente de forma concreta como a aplicação foi construída, ele tem grande utilidade no sentido de ilustrar o verdadeiro fluxo de informação existente durante seu funcionamento. Para iniciarmos a descrever, embora que de modo sucinto, a metodologia envolvida na construção da ferramenta de busca e coleta dos dados necessários denominada *AppSNAElection*, precisamos compreender um pouco mais a respeito das tecnologias que foram empregadas.

Na construção deste sistema foi necessário um “*Tool Chain*” ou “Cadeia de Ferramentas” (em tradução livre). A partir deste conceito as ferramentas podem ser utilizadas em cadeia, de modo que a saída de cada uma das ferramentas torna-se a entrada para a outra, embora o termo seja utilizado amplamente para se referir a qualquer conjunto de ferramentas interligadas visando o desenvolvimento de uma ferramenta de propósito único.

Desta maneira, para a construção da ferramenta de busca e coleta de dados do presente trabalho foram utilizadas as seguintes tecnologias listadas na **Tabela 12** abaixo:

TABELA 12 - Tecnologias utilizadas na construção da arquitetura.

Descrição	Tecnologia	Ferramenta
Linguagem de Programação	Java	Jdk 1.6.0_22
Camada de Persistência	Banco de Dados	MySQL 5.1
	Connector	MySQL Connector Java 5.1.6
	Framework	JDBC 1.4
Camada de Integração	API Twitter	Twitter 1.0
	API Java Twitter	Twitter4J 3.0.0
	API Geocode	Geocoding v3
	Métricas em Grafos	JUNG 2.0.1
Camada de Visualização	Visualização de Grafos	Prefuse beta e JUNG 2.0.1

Percebemos que, de acordo com a divisão proposta na Tabela acima, o *software* construído encontra-se separado em três camadas (*layers*): A **camada de persistência** ou de acesso aos dados é a parte da aplicação responsável por comunicar-se com o banco de dados (ou com o framework de persistência). Para sua construção, utilizamos basicamente um banco de dados relacional junto com seu *connector* para a linguagem Java. Ainda, como framework para persistência foi utilizado o JDBC (*Java Database Connectivity*) visando um maior desempenho.

Na **camada de integração** definimos os casos de uso e a lógica de negócios utilizando os recursos da linguagem Java. Basicamente foram utilizados os serviços da API do twitter auxiliados pela Twitter4J no sentido de captar os dados no microblog e a ferramentas JUNG que, além de serem usadas na camada de visualização, forneciam meios de programação para realizar os cálculos de métricas SNA.

Na **camada de visualização** figuram os arquivos de extensão graphml. Na camada anterior eram realizadas as heurísticas necessárias para geração deste um arquivo que era consumido nesta camada para uso nas APIs (descritas na [seção 3.3.5.2](#)) para fornecer meios de visualização dos grafos de acordo com o layout mais apropriado para visualizar determinada estrutura.

Esta solução tornou nossa aplicação independente de ferramentas de visualização, pois, de posse destes arquivos, poderemos utilizar outras APIs ou mesmo ferramentas de autoria (tais como as descritas na [seção 3.3.5.1](#)) uma vez que este formato estruturado de dados é extremamente consagrado perante as ferramentas de visualização e análise de redes sociais.

4.3 Funcionamento da *AppSNAElection*

Basicamente podemos descrever o funcionamento da aplicação com base no fluxo de dados descrito na **Figura 23**. Inicialmente os dados são produzidos pelos usuários da rede **(1)** e disponibilizados em suas respectivas *timelines* **(2)**, Com base em filtros (caso estivessemos usando a Streaming API, por exemplo, poderíamos atribuir o filtro “Brasil Inglaterra” para capturar textos como “Brasil venceu a Inglaterra por 3X0” ou “Jogo entra Brasil e Inglaterra foi cancelado pela Justiça”) ou mesmo parâmetros **(3)** (caso estivessemos usando a REST API, por exemplo, poderíamos entrar com um grupo de usuários no intuito de conhecer seus seguidos/seguidores, biografia, localização entre outros) que são repassados para Twitter API no sentido de buscar esses dados **(4)**.

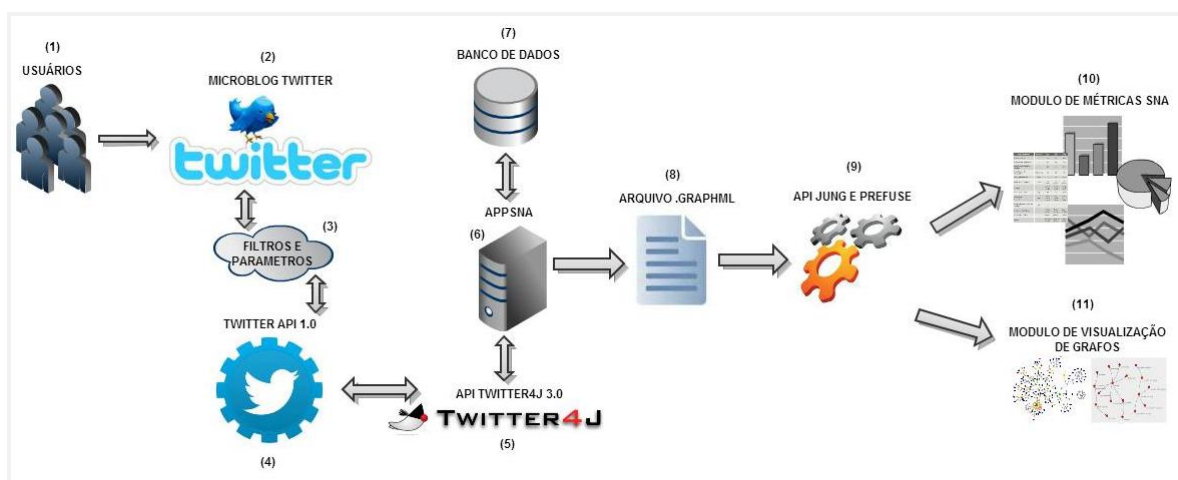


FIGURA 23 - Fluxo de dados da *AppSNAElection*.

Continuando os aspectos do funcionamento detalhados na **Figura 23** acima, temos que a Twitter4J realiza a comunicação direta com os dados provenientes da Twitter API **(5)** de modo que esta realiza os devidos tratamentos de modo que estes sejam disponibilizados para aplicação **(6)** em forma de objetos, sendo estes dados persistidos na nossa base de dados de

acordo com a conveniência **(7)**. A partir de heurísticas determinadas em nossa aplicação são gerados os arquivos (.) graphML **(8)**. Estes são manipulados diretamente pelas APIs do JUNG e Prefuse **(9)** de modo que ambas podem gerar dados estatísticos relativos as métricas de ARS, cabendo a ressalva de que a API Prefuse não contém este módulo disponibilizado de forma nativa **(10)** ou ainda gerar meios de visualização dos grafos a partir dos dados componentes do arquivo **(11)**.

4.4 Modelagem da Rede

Para obtenção dos dados necessários a realização da presente pesquisa foram monitorados prioritariamente os 04 (quatro) principais candidatos à prefeitura do município de João Pessoa de acordo com as informações obtidas através de pesquisas de intenção de voto registradas junto ao Tribunal Regional Eleitoral (TRE-PB, 2012) e os seus prováveis eleitores (representados neste contexto pelos usuários do Twitter) que os mencionaram durante o período das eleições de 2012.

Baseado no funcionamento da faceta do *Streaming* API, temos que no momento que algum usuário do Twitter realizava menção a algum dos termos associado ao candidato, a *AppSNAElection* armazena no banco de dados este resultado. Outros dados também são de interesse para o nosso de estudo de maneira que a *AppSNAElection* utilizando a faceta da REST API também coleta e armazena em banco os dados: Id (identificador) do usuário, *Screen Name* (Nome atribuído a conta, por exemplo, @fulano), *status* (representa os 140 caracteres do *Tweet*), data/hora da ocorrência, geolocalização (latitude e longitude) e o termo que referenciou algum dos candidatos.

A partir disto, a *AppSNAElection* realiza a modelagem da rede com base nos dados previamente armazenados. Este *software* permaneceu em operação durante o período de 09/09/2012 a 30/10/2012 (período do 1º e 2º turnos), conseguindo coletar aproximadamente de 90.000 (noventa mil) mensagens (*tweets*) contendo algum tipo de referência a algum candidato político realizado por aproximadamente 8.500 (oito mil e quinhentas) pessoas distintas.

A heurística envolvida para tal atividade é a seguinte: Realizamos a leitura do banco de dados objetivando criar nosso arquivo (.) graphML, a cada usuário identificado, criamos um nó e associamos ao candidato o qual este usuário faz menção. Caso o usuário realize menção a dois candidatos, será criada apenas uma nova relação para este nó. Em cada relacionamento da rede (aresta) temos o atributo peso representando a quantidade de menções realizadas pelo usuário (nó) a algum candidato (outro nó) A **Figura 24** traduz de maneira gráfica a ideia exposta anteriormente.

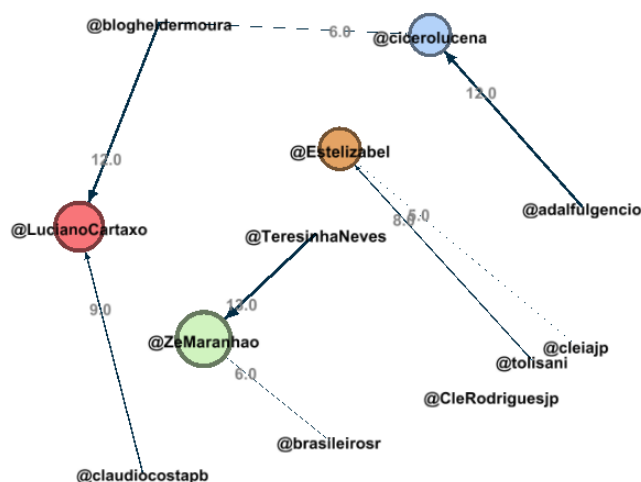


FIGURA 24 - Metodologia para a construção da rede.

Em relação ao estudo do 1º turno da eleição (apresentado em detalhes na [seção 5.1](#)), foram utilizados apenas os dados do período de 09/09/2012 até 07/10/2012. Com isso temos mais de 45.000 (quarenta e cinco mil) menções a candidatos, gerando um grafo com 4.257 (quatro mil duzentos e cinquenta e sete) nós e 5.851 (cinco mil oitocentos e cinquenta e um) arestas representando as ligações entre estes nós tendo como atributo o total de menções de cada usuário.

Em relação ao estudo do 2º turno da eleição (apresentado em detalhes na [seção 5.2](#)), foram utilizados apenas os dados do período de 08/10/2012 até 30/10/2012. Com isso temos mais 42.000 (quarenta e duas mil) menções a candidatos, gerando um grafo com 6.208 (seis mil duzentos e oito) nós e 7.318 (sete mil trezentos e dezoito) arestas representando as ligações entre estes nós tendo como atributo o total de menções de cada usuário.

4.5 Considerações

A metodologia de análise apresentada na [seção 4.1](#) se mostrou eficiente para a nossa pesquisa, além de servir de base para a construção da ferramenta de busca e coleta de dados denominada *AppSNAElection*. Para montagem da referida ferramenta optou-se por bibliotecas de softwares com “código-aberto” (veja a **Tabela 12**) de modo que pudessem ser modificadas de acordo com os requisitos inicialmente propostos.

Ainda sobre a ferramenta, vimos que ela utilizou as facetas do Streaming e REST API. Ambas necessitam de autenticação para uso, no entanto a REST API impõe um limite de requisições (350 requisições por hora) na versão utilizada pela ferramenta. Em virtude disto foi proposta uma solução denominada *Carousel* objetivando elevar a quantidade de requisições, viabilizando deste modo à coleta de dados. Basicamente foram criadas dezenas de contas autenticadas e as colocamos em duas listas (espera e pronto) de modo que, findada as requisições de uma conta, dinamicamente o contexto era passado para outra de modo que não houvesse solução de continuidade na busca dos dados.

Em relação aos aspectos relacionados à geolocalização descritos na [seção 5.1.4](#) e [seção 5.2.4](#), foram utilizados os dados da longitude e latitude dos *Tweets* para verificação dos bairros de origem de tais mensagens. Para Tal a *AppSNAElection* utilizou o serviço do *Google Geocoding API (v3)*¹ que retornava os dados de localização (logradouro, bairro, cidade, estado) que em seguida eram armazenados no banco de dados para posterior análise.

¹ Geocoding API. Disponível em: <<http://migre.me/fixAx>>. Acesso em maio de 2013.

5. Análise dos resultados

Neste capítulo iremos expor os resultados obtidos em nossa pesquisa de acordo com os estudos de caso realizados ao longo do período eleitoral em questão. Desta maneira o presente capítulo encontra-se dividido em três seções: Em Resultados do 1º turno ([Seção 5.1](#)) e Resultados do 2º turno ([Seção 5.2](#)) serão expostos os dados referente aos supracitados períodos comparando-os com os dados apresentados pelas nossas hipóteses; Finalmente em Publicações ([Seção 5.3](#)) iremos enumerar os principais artefatos científicos gerados a partir da presente pesquisa.

5.1 Resultados do 1º Turno

Nesta seção será apresentado nosso primeiro estudo de caso. Para isso, a presente seção encontra-se dividida em cinco subseções: Em estatísticas do twitter e Pesquisas eleitorais ([Seção 5.1.1](#)) apresentaremos os números obtidos através da nossa ferramenta bem como através das pesquisas de intenção de voto; Em Resultado final da eleição ([Seção 5.1.2](#)) apontaremos os números finais do pleito; Em aplicação de métricas ARS ([Seção 5.1.3](#)) iremos expor os números obtidos através da ARS e compara-los com os números finais das eleições; Em aspectos sobre Geolocalização ([Seção 5.1.4](#)) e em Utilização do intervalo de Confiança ([Seção 5.1.5](#)) serão apresentados os detalhes da utilização destas técnicas; Finalizando iremos expor as principais considerações sobre o referido estudo de caso ([Seção 5.1.6](#)).

5.1.1 Estatísticas do Twitter e pesquisas eleitorais

Iniciamos esta parte do trabalho apresentando alguns dados relativos ao quantitativo final de *tweets* e pessoas distintas. Os dados foram obtidos através da *AppSNAElection* e encontram-se descritos na **Tabela 13** abaixo:

TABELA 13 - Estatísticas de Tweets e pessoas (1º Turno).

Candidato	Total Tweets		Nr Pessoas		Tweets/dia		Média Pessoa
	Nr	%	Nr	%	Nr	%	
Luciano Cartaxo	18890	42,06	2680	45,80	633	42,23	7,05
Cícero Lucena	4148	9,23	939	16,05	138	9,20	4,42
Estela Izabel	19732	43,95	1754	29,98	657	43,83	11,25
Zé Maranhão	2135	4,75	478	8,17	71	4,74	4,47

Os dados apresentados foram baseados no período de tempo compreendido entre o dia 09/09/2012 e o dia do pleito do 1º turno (07/10/2012). Neste período temos cerca de quatro semanas entre os meses de setembro e outubro de 2012: a primeira entre os dias 08 e 16, a segunda entre os dias 17 e 23, a terceira semana entre os dias 24 e 30 (as três primeiras semanas no mês de setembro) e a última entre os dias 1 e 7 de outubro.

A **Figura 25** nos traz a ideia do total de menções que cada candidato recebeu por semana. Em (a) temos o total de menções por semana de modo distinto. Em (b) temos o total de menções acumuladas por período.

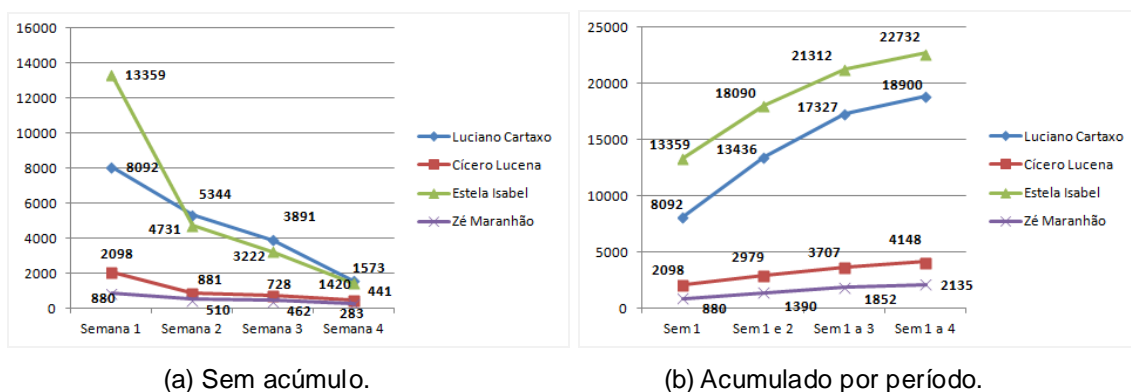


FIGURA 25 - Número de *tweets* por semana.

Em seguida temos os índices da quantidade de pessoas distintas que citam um dos candidatos. Na **Figura 26** temos esses dados acumulados por período (a) e, em (b) temos os dados divididos em períodos sem o acúmulo do número de pessoas.

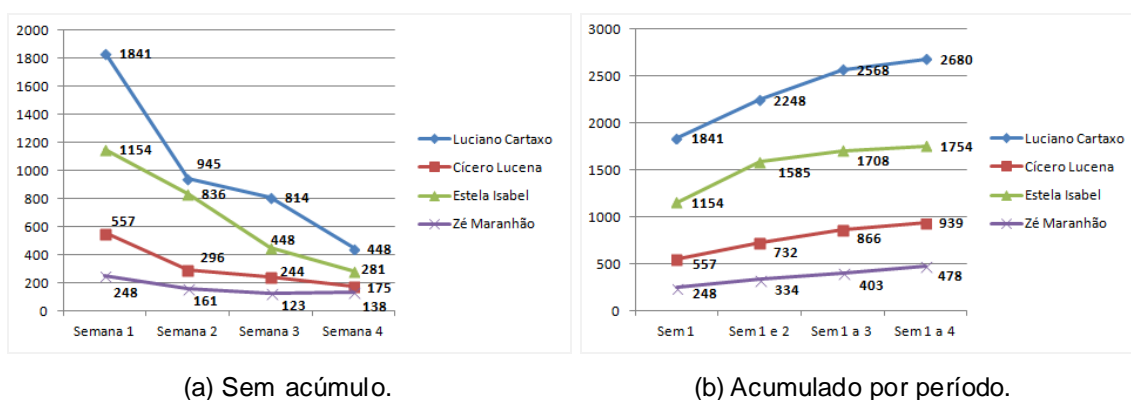


FIGURA 26 - Número de Pessoas distintas por semana.

A cada semana foi computada uma média simples da quantidade diária de *tweets* de cada candidato. Por fim temos os dados descritos na **Figura 27** abaixo que apresenta a evolução dos números de cada candidato.

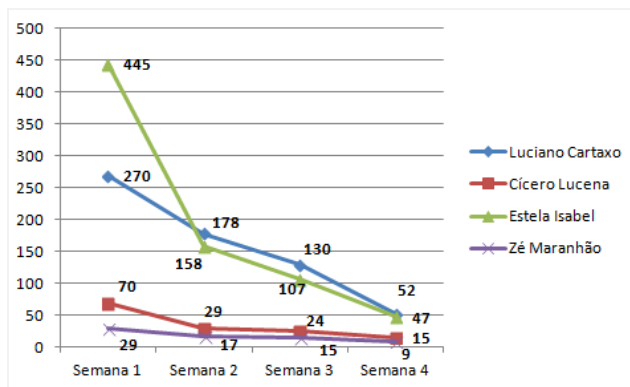


FIGURA 27 - Média de *tweets* diários por semanas.

O próximo índice verificado diz respeito aos dados das previsões apresentadas pelas pesquisas de intenção de voto registradas junto ao TRE-PB (2012). Os dados apresentados na **Figura 28** Erro! Fonte de referência não encontrada. foram obtidos através de pesquisas dos institutos IBOPE e Consult registradas durante o 1º turno.

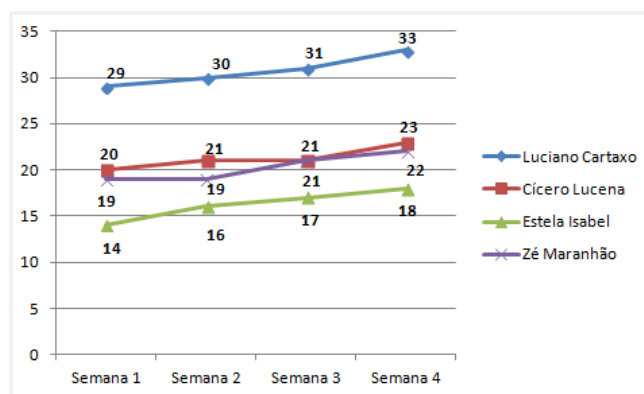


FIGURA 28 - Evolução das pesquisas eleitorais.

Por fim, a **Figura 29** apresenta o número de *tweets* recebidos diariamente. O número de partida do gráfico foi obtido através de uma média aritmética simples da quantidade de menções que os candidatos receberam entre os dias 08/09/2012 e 06/10/2012, exceto o dia do debate. A segunda informação diz respeito à quantidade de menções de cada candidato no dia 04/10/2012 (debate veiculado pela TV). Por fim, tem-se a quantidade que cada candidato obteve no dia 07/10/2012 (pleito eleitoral do 1º turno).

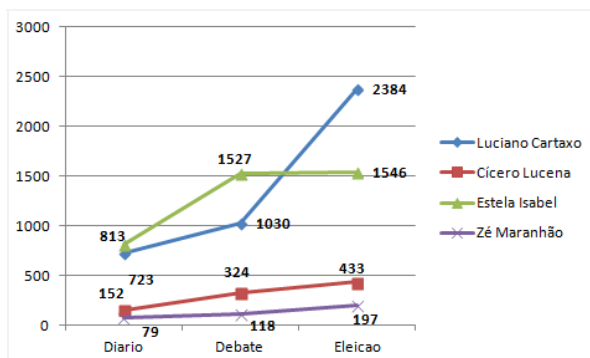


FIGURA 29 - Número de *tweets* diários.

5.1.2 Resultado final da eleição

Nesta parte do trabalho será apresentado o resultado oficial das eleições do município de João Pessoa de acordo com os dados do TRE-PB (2012). A **Tabela 14** expõe o resultado do 1º turno e o compara com os dados da pesquisa realizada pelo IBOPE registrada no site do TRE-PB em 06/10/2012.

TABELA 14 - Resultado do pleito eleitoral do 1º turno.

Candidato	Resultado do pleito		Resultado das Pesquisas	
	Votos	Percentual	Percentual	Margem
Luciano Cartaxo	142.158	38,32	33	31 - 35
Cícero Lucena	75.170	20,27	22	20 - 24
Estela Isabel	74.498	20,08	18	16 - 20
Zé Maranhão	69.978	18,87	23	21 - 25

Ainda, de acordo com a hipótese levantada sobre a relação entre usuários do Twitter e moradores dos bairros de maior poder aquisitivo, descrevemos na **Tabela 15** o resultado das eleições nos bairros de selecionados na [seção 3.1.4](#) deste trabalho.

TABELA 15 - Total de votos nos bairros de maior poder aquisitivo.

Bairro	Luciano Cartaxo	Estela Isabel	Zé Maranhão	Cícero Lucena
Cabo branco	462	288	225	210
Tambaú	4.431	2.910	2.267	2.081
Manaíra	3.090	1.798	1.733	1.340
Miramar	523	293	323	244
Tambauzinho	1.805	1.034	927	809
Bessa	5.664	3.653	2.420	2.457
Altiplano	1.826	1.110	1.169	914
Total	17.801	11.086	9.064	8.055
	38,7%	24,1%	19,7%	17,5%

5.1.3 Aplicação de métricas de ARS

Após a obtenção dos grafo através da metodologia exposta na [seção 4.4](#) foi utilizado o Gephi para visualização e aplicação de métricas de ARS. O presente estudo foi dividido em três fases: (1) Análise de acordo com a modularidade; (2) Análise de acordo com as métricas grau e grau ponderado; (3) Análise das métricas *PageRank*, *Betweenness Centrality*, *HITS* e *Closeness Centrality*.

Quando analisada de acordo com a métrica Modularidade, o Gephi foi capaz de encontrar quatro comunidades ou sub-redes (Blondel, 2008) na rede (Veja em maiores detalhes no [Anexo H](#) e [Anexo I](#)), indicando cada um dos principais candidatos na disputa eleitoral do 1º turno (**Figura 30**).

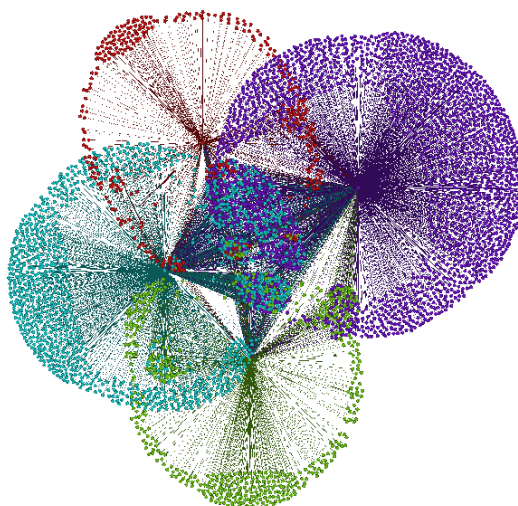


FIGURA 30 - Comunidades encontradas na rede.

Cada cor tem uma correspondência em relação aos quatro candidatos: **Roxo** - Usuários que realizaram menção ao candidato Luciano Cartaxo; **Azul** - Usuários que realizaram menção a candidata Estela Izabel; **Verde** - Usuários que realizaram menção ao o candidato Cícero Lucena; **Vermelho** - Usuários que realizaram menção ao candidato Zé Maranhão.

Percebe-se ainda que no centro do grafo existe uma junção entre os nós nas diversas cores. Nesta porção encontram-se os nós (usuários do Twitter) que citaram mais de um candidato de modo que a sua coloração é obtida através do atributo peso (número de citações), ou seja, caso ele menciona mais o candidato Luciano Cartaxo, recebe a cor roxa e assim sucessivamente.

Na **Tabela 16** podemos comparar os dados obtidos a partir da aplicação desta métrica com os resultados finais das eleições e das pesquisas de intenção de voto do supracitado período eleitoral.

TABELA 16 - Comparação dos resultados com a métrica Modularidade.

Candidato	Resultado do pleito		Resultado Bairros*		Modularidade	
	Votos	%	Votos	%	Nós	%
Luciano Cartaxo	142.158	38,32	17.801	38,7	2479	42,38
Cícero Lucena	75.170	20,27	8.055	17,5	1014	17,33
Estela Izabel	74.498	20,08	11.086	24,1	1549	26,48
Zé Maranhão	69.978	18,87	9.064	19,7	807	13,81

* Bairros selecionados de acordo com o poder aquisitivo

Quando analisada de acordo com a métrica Grau Ponderado (*Weighted Degree*), observa-se que 77,3% dos nós realizam entre 1 e 5 menções (nós com tons de verde) e que 14,7% realizam entre 6 e 10 menções (nós em tons de vermelho). Os demais 8% realizam mais de 10 menções (nós em tons de preto) a aos vários candidatos em questão.

Em face disto podemos analisar em cada uma das quatro comunidades o comportamento dos nós em relação a quantidade de menções. De acordo com a **Figura 31** abaixo, os usuários que citam o candidato Luciano Cartaxo (comunidade com nó centra **azul**), em sua maioria são em tons de verde, contudo verificamos uma maior presença de nós em tons de vermelho e em tons de preto do que as outras comunidades.

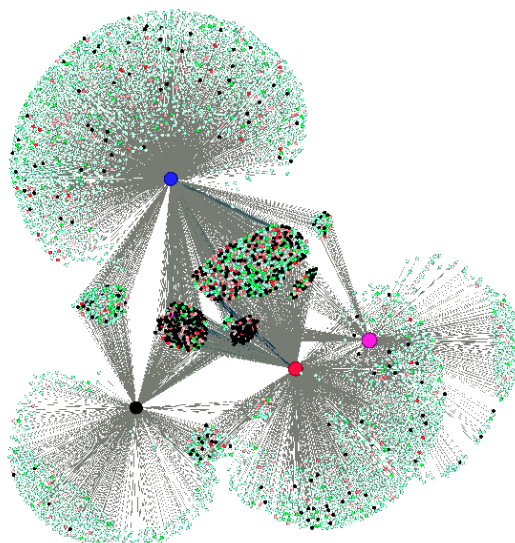


FIGURA 31 - Partição do grafo de acordo com o grau ponderado.

O grafo da figura acima pode ser visto em maiores detalhes no [Anexo J](#) deste trabalho. Para efeito de referência temos que a comunidade cujo nó central é **preto** corresponde ao candidato Cícero Lucena. A comunidade cujo nó central é **vermelho** corresponde a candidata Estela Izabel. Por fim, a comunidade cujo nó central é **rosa** corresponde ao candidato Zé Maranhão.

Finalizamos a segunda etapa analisando a rede de acordo com a métrica **Grau** (*Degree*). Esta métrica nos traz informações importantes a respeito da quantidade de ligações que cada nó possui, sem levar em conta a quantidade de menções feita por cada nó, ou seja, no caso anterior (grau ponderado) cada nó contribuía para a métrica com valor igual ao total de menções realizadas e neste caso, cada nó contribui apenas com uma unidade por ligação. No nosso contexto isto significa se um usuário cita/ou é citado por uma ou várias pessoas ao mesmo tempo.

Na **Figura 32** verifica-se que 74% dos nós do grafo (cor **azul**) possuem peso 1, ou seja, citaram apenas um dos candidatos. Ainda, os demais 17,5% dos nós (cor **verde**) possuem peso 2, ou seja, realizaram menção a mais de um candidato.

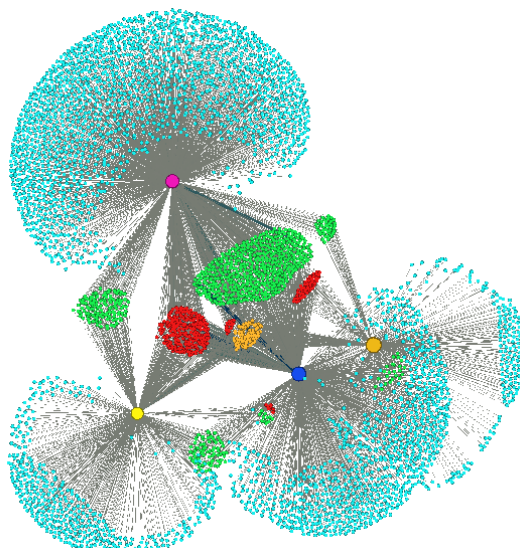


FIGURA 32 - Partição do grafo de acordo com o grau.

Na região central da figura podemos notar a maior frequência de nós **vermelhos** (6,6%) e **laranjas** (2,3%) que citaram respectivamente 3 ou 4 candidatos de modo concomitante. Esta figura encontra-se no [Anexo K](#) deste trabalho, podendo ser visualizada maior riqueza de detalhes.

Na **Tabela 17** a seguir estão expostos os resultados da métrica grau e grau ponderado em comparação com o resultado oficial do pleito bem como com o resultado de acordo com os bairros de maior poder aquisitivo*.

TABELA 17 - Comparação dos resultados com as métricas de grau.

Candidato	Resultado do pleito		Resultado Bairros*		Grau		Grau ponderado	
	Votos	%	Votos	%	Nós	%	Nós	%
Luciano Cartaxo	142.158	38,32	17.801	38,7	2680	45,80	18890	42,06
Cícero Lucena	75.170	20,27	8.055	17,5	939	16,05	4148	9,23
Estela Izabel	74.498	20,08	11.086	24,1	1754	29,98	19732	43,95
Zé Maranhão	69.978	18,87	9.064	19,7	478	8,17	2135	4,75

Na terceira etapa da análise foram analisados os resultados conforme as demais métricas. Como pode ser observado na Erro! Fonte de referência não encontrada.17, Luciano Cartaxo apresenta o maior índice em “*Betweenness*” que indica o quão frequentemente o nó que o representa aparece no caminho mais curto entre demais nós da rede. Ainda, mostrou-se mais próximo dos demais nós da rede por apresentar o menor “*Closeness*” e obteve os valores de “*PageRank*” e “*HITS*” elevados, indicando certo grau de autoridade ao nó que o representa uma vez que o valor destas métricas é baseada nos valores de outros nós que compõem a rede.

TABELA 18 - Comparação entre as demais métricas ARS.

Candidato	<i>Closeness</i>	<i>Betweenness</i>	<i>HITS</i>	<i>PageRank</i>
Luciano Cartaxo	1,740	6.341.336	0,168	0,219
Estela Isabel	2,175	3.707.639	0,112	0,135
Cícero Lucena	2,558	1.817.719	0,059	0,078
Zé Maranhão	2,775	1.044.550	0,031	0,036

5.1.4 Aspectos sobre Geolocalização

A *AppSNAElection* conseguiu verificar a existência de dados de geolocalização em aproximadamente 700 (setecentos) *Tweets*. Este número pode ser considerado extremamente pequeno (cerca de 1%) em relação ao montante obtido durante o período de operação do software (correspondente ao 1º turno das eleições). Na **Tabela 19** a seguir verifica-se a distribuição dos mesmos de acordo com o candidato o qual a mensagem (*Tweet*) realizava algum tipo de menção.

TABELA 19 - Menções (Tweets) com dados de geolocalização.

Candidato	Menções	%
Luciano Cartaxo	290	41,61
Cícero Lucena	82	11,76
Estela Isabel	259	37,16
Zé Maranhão	66	9,47
Total	697	100,00

Em seguida procurou-se realizar uma correspondência entre latitude e longitude (dados componentes dos *Tweets*) e os bairros de maior poder aquisitivo elencados na [seção 3.1.4](#) deste trabalho.

TABELA 20 - Distribuição das Menções (Tweets) por bairro (1).

Bairro	Luciano Cartaxo	Cícero Lucena	Estela Isabel	Zé Maranhão
Cabo branco	19	3	23	3
Tambaú	27	4	9	1
Manaíra	14	4	11	1
Miramar	43	3	20	4
Tambauzinho	9	1	3	3
Bessa	29	13	24	5
Altiplano	2	2	16	2
Total	143	30	106	19
	47,99	10,07	35,57	6,38

Finalmente foi realizada a mesma correspondência acima para outros bairros relevantes para o nosso estudo (veja seção [seção 3.1.4](#)).

TABELA 21 - Distribuição das Menções (Tweets) por bairro (2).

Bairro	Luciano Cartaxo	Cícero Lucena	Estela Isabel	Zé Maranhão
Mangabeira	34	2	4	10
Cristo	9	35	60	13
Castelo Branco	13	1	7	17
Bairro dos Ipês	6	2	3	3
Total	62	40	74	43
	28,31	18,26	33,79	19,63

Se for realizada uma comparação entre os percentuais obtidos em ambas as tabelas acima e os resultados oficiais expostos na [seção 5.1.2](#), percebemos que os resultados elencados extrapolam uma margem de erro superior a 10 pontos percentuais.

5.1.5 Utilização do Intervalo de Confiança

Uma técnica interessante que poderia ser utilizada para nosso estudo de caso seria a obtenção de intervalos de confiança (IC). Este conceito foi apresentado brevemente na [seção 3.2.3](#) necessita de alguns dados para obtenção desta medida. Na **Tabela 22** a seguir encontram-se dispostos os valores necessários para o cálculo da medida.

TABELA 22 - Dados para o cálculo do Intervalo de confiança.

Dado	Luciano Cartaxo	Cícero Lucena	Estela Isabel	Zé Maranhão
Tamanho do Arranjo	2680	939	1754	478
Significância (Alpha)	0,05	0,05	0,05	0,05
Desvio Padrão	30,41	16,75	54,03	13,50
Média tweet/pessoa	7,10	4,41	11,24	4,50
Margem	1,15	0,52	2,25	1,20
Intervalo arredondado	6 - 8	4 - 5	9 - 13	3 - 5

Cabe comentário que a significância é uma medida obtida de acordo com o nível de confiança esperado. Como os valores mais usados para confiança são: 0,90 (90%), 0,95 (95%) e 0,99 (99%), conseqüentemente para significância são: 0,10 (10%), 0,05 (5%) e 0,01 (1%) respectivamente. Desta forma estimamos uma confiança de 95% logo temos uma significância de 0,05.

Ainda analisando os dados da **Tabela 22**, temos que após a realização do cálculo do IC obtivemos o valor da margem que deverá ser utilizado com a média para obtenção dos respectivos intervalos. Exemplificando, temos que com 95% de certeza, cada pessoa que menciona o candidato Zé Maranhão realiza entre 3 e 5 menções. A partir do IC podemos obter e analisamos outros valores que se encontram descritos na **Tabela 23**:

TABELA 23 - Análise dos dados do Intervalo de confiança.

Candidato	Tweets		Pessoas		Média Tweet/Pessoa
	Nr	%	Nr	%	
Luciano Cartaxo	1622	56,54	230	55,96	7,10
Cícero Lucena	278	9,71	63	15,33	4,41
Estellzabel	731	25,50	65	15,82	11,24
Zé Maranhão	237	8,26	53	12,90	4,50

A coluna pessoas é obtida a partir da contagem de todos os usuários que estão contidos no IC proposto. A coluna Tweets é obtida através da multiplicação simples dos valores da coluna Nr/Pessoas pela coluna Média Tweet/Pessoa. Percebemos que com o uso do IC não conseguimos diminuir a margem de erro obtida em métodos descritos anteriormente.

Contudo, se realizarmos uma análise mais aprofundada nas contas de usuários que realizam menção aos candidatos de maneira que fosse realizado um filtro, eliminando as contas utilizadas para campanha eleitoral e contas “robôs”, podemos chegar a novos dados para o cálculo do IC.

TABELA 24 - Novos dados para o cálculo do Intervalo de confiança.

Dado	Luciano Cartaxo	Cícero Lucena	Estela Isabel	Zé Maranhão
Tamanho do Arranjo	2672	933	1742	472
Contas eliminadas	8	6	12	6
Significância (Alpha)	0,05	0,05	0,05	0,05
Desvio Padrão	19,78	12,84	44,26	13,34
Média tweet/pessoa	6,11	3,79	10,33	4,24
Margem	0,75	0,24	2,07	1,20
Intervalo arredondado	5 e 7	3 e 4	8 - 12	3 - 5

A partir da eliminação das contas (Veja a segunda linha da **Tabela 24**) podemos chegar a uma nova média de Tweet/pessoa (Veja na última coluna da **Tabela 25**). Perceba que estes valores são diferentes daqueles descritos na **Tabela 23**, isto deve basicamente pela redução da quantidade de pessoas que contém a quantidade de menções compatível com o intervalo de confiança) e a um novo número de pessoas dentro da margem do intervalo de confiança.

TABELA 25 - Análise dos novos dados do Intervalo de confiança.

Candidato	Tweets		Pessoas		Média Tweet/Pessoa
	Nr	%	Nr	%	
Luciano Cartaxo	1024	39,57	180	39,39	5,69
Cícero Lucena	399	15,42	118	25,82	3,38
Estelzabel	968	37,40	105	22,98	9,21
Zé Maranhão	197	7,51	53	11,82	3,65

Percebemos claramente que os números apresentados na **Tabela 25** possuem um percentual diferente daqueles descritos na **Tabela 23**. A partir disto podemos afirmar que, com refinamentos sucessivos do universo amostral, podemos alcançar índices cada vez menores na margem de erro.

5.1.6 Considerações sobre o 1º Turno

Na **Tabela 26** a seguir apresentamos resumidamente todas as comparações realizadas ao longo da [seção 5.1](#).

TABELA 26 - Comparação geral dos resultados.

Método	1º Lugar		2º Lugar		3º Lugar		4º Lugar		Margem
	C	%	C	%	C	%	C	%	%
Resultado Oficial	LC	38,32	CL	20,27	EI	20,08	ZM	18,87	-
Pesquisas Eleitorais	LC	33	ZM	23	CL	22	EI	18	5
Bairros	LC	38,7	EI	24,1	ZM	19,7	CL	17,5	4
Modularidade	LC	42,38	EI	26,48	CL	17,33	ZM	13,81	6
Grau	LC	45,8	EI	29,98	CL	16,05	ZM	8,17	10
Grau Ponderado	LC	42,06	EI	43,95	CL	9,23	ZM	4,75	14
IC-Pessoas distintas	LC	39,38	CL	25,82	EI	22,97	ZM	11,81	7

* C – Candidato, LC – Luciano Cartaxo, CL – Cícero Lucena, ZM – Zé Maranhão, EI – Estela Isabel.

Um dado interessante obtido foi com relação a métrica modularidade. Se os dados obtidos a partir desta métrica forem comparados com os números finais das eleições temos uma margem de erro em torno de 6% (no pior caso). Contudo se realizarmos a mesma comparação dos valores desta métrica com os da votação nos bairros de maior poder aquisitivo podemos perceber que a margem de erro cai para menos de 4% (no pior caso).

Ainda, a métrica modularidade conseguiu apontar com certo grau de fidelidade os candidatos que ocuparam o 1º e o 4º lugares no pleito eleitoral do município de João Pessoa, contudo, não obteve o mesmo grau de fidelidade ao prever o posicionamento dos demais postulantes. Isso pode ter sido influenciado por vários fatores, uma vez que não fizemos a análise do conteúdo das mensagens postadas uma vez que muitas menções a um determinado candidato podem, na verdade, ser de cunho negativo e com isso aumentar o nível de rejeição dos eleitores contra ele.

Em relação aos candidatos que ocuparam a 2º e 3º posições temos alguns indícios que podem ser utilizados de modo preliminar no intuito de justificar o erro em nossas previsões. Com relação à candidata Estela Izabel (apontada de acordo com a modularidade como 2º colocada e obteve a 3º colocação no resultado oficial) analisamos que a referida candidata investiu maciçamente em propaganda política junto as mídias sociais. Esta atividade pode ser comprovada através do estudo do intervalo de confiança (IC). Esta candidata teve 12 contas de usuários eliminadas e após a análise desta métrica, percebemos que o candidato Cícero Lucena conseguiu ganhar a 2ª colocação.

Analisando o desempenho do candidato Cícero Lucena (apontada em nosso trabalho como 3º colocado e obteve a 2º colocação no resultado oficial) temos o fato de que as pesquisas de intenção de voto embora participem, contudo não determinam de maneira inequívoca a opinião pública em relação à escolha de candidatos nas eleições. Esta determinação tende a ser maior quanto menos qualitativo for o grau de informação do público, mas nunca é absoluta. Isto pode ser comprovado no desempenho do referido candidato que apesar de ser o 2º lugar nas pesquisas eleitorais e no resultado oficial das eleições, apenas foi 4º lugar na votação dos bairros de maior poder aquisitivo.

Com relação a modelagem da rede, apesar desta utilizar dados capturados a partir da rede social Twitter e possuir diversas variáveis que possam influenciar o resultado das análises iniciais, obtivemos uma boa aproximação também quando confrontamos as nossas previsões com o resultado da eleição nos bairros ditos de classe média/alta, uma vez que este é considerado o provável público utilizador desta rede social. É importante salientar que com isso pode-se excluir vários grupos de pessoas com acesso à Internet e de uso constante das Redes Sociais, como é o caso de estudantes que não fazem parte da camada mais elevada da sociedade. Também temos que levar em conta que uma parte das citações a um candidato não vem apenas de usuários “comuns”, mas sim de veículos oficiais de campanha, pessoas que trabalham diretamente com os candidatos, o que aumenta apenas a visibilidade do candidato, mas não indica necessariamente aumento de votos.

5.2 Resultados do 2º Turno

Nesta seção apresentaremos nosso segundo estudo de caso. Para isso, a presente seção encontra-se dividida em cinco subseções: Em estatísticas do twitter e Pesquisas eleitorais ([Seção 5.2.1](#)) apresentaremos os números obtidos através da nossa ferramenta bem como através das pesquisas de intenção de voto; Em Resultado final da eleição ([Seção 5.2.2](#)) apontaremos os números finais do pleito; Em aplicação de métricas ARS ([Seção 5.2.3](#)) iremos expor os números obtidos através da ARS e compara-los com os números finais das eleições; Em aspectos sobre Geolocalização ([Seção 5.2.4](#)) e em Utilização do Intervalo de Confiança ([Seção 5.2.5](#)) serão apresentados os detalhes da utilização destas técnicas; Finalizando na iremos expor as principais considerações sobre o referido estudo de caso ([Seção 5.2.6](#)).

5.2.1 Estatísticas do Twitter e pesquisas eleitorais

Iniciamos esta parte do trabalho apresentando alguns dados relativos ao quantitativo final de tweets e pessoas distintas que realizaram menção a algum candidato. Os dados foram obtidos através da *AppSNAElection* e encontram-se descritos na **Tabela 27** abaixo:

TABELA 27 - Estatísticas de Tweets e pessoas (2º Turno).

Candidato	Total Tweets		Nr Pessoas		Tweets/dia		Média Pessoa
	Nr	%	Nr	%	Nr	%	
Luciano Cartaxo	25.676	61,05	4.740	64,77	1.284	61,05	5,42
Cícero Lucena	16.385	38,95	2.578	35,23	819	38,95	7,16

Os dados apresentados foram baseados no período de tempo compreendido entre o dia 08/10/2012 e o dia do pleito do 2º turno (28/10/2012). Neste período temos três semanas no mês de outubro de 2012: a primeira entre os dias 08 a 14, a segunda entre os dias 15 a 21 e a última entre os dias 22 a 28.

A **Figura 33** apresenta os números relativos ao total de menções que cada candidato recebeu por semana (considerar os períodos descritos acima). Em (a) temos o total de menções por semana de modo distinto. Em (b) temos o total de menções acumuladas por período.

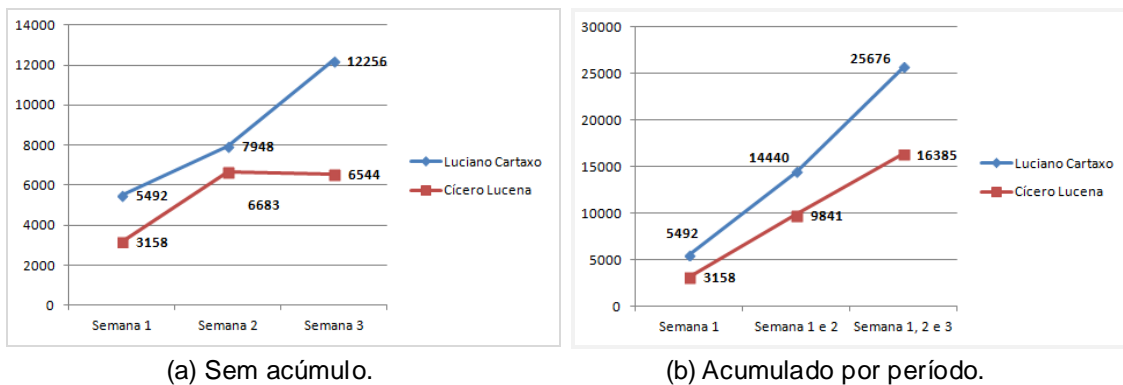


FIGURA 33 - Número de Tweets por semana.

Em seguida são apresentados os índices da quantidade de pessoas distintas que citam um dos candidatos. Na **Figura 34** temos os dados divididos em períodos sem o acúmulo do número de pessoas (a) e, em (b) temos esses dados acumulados por período.

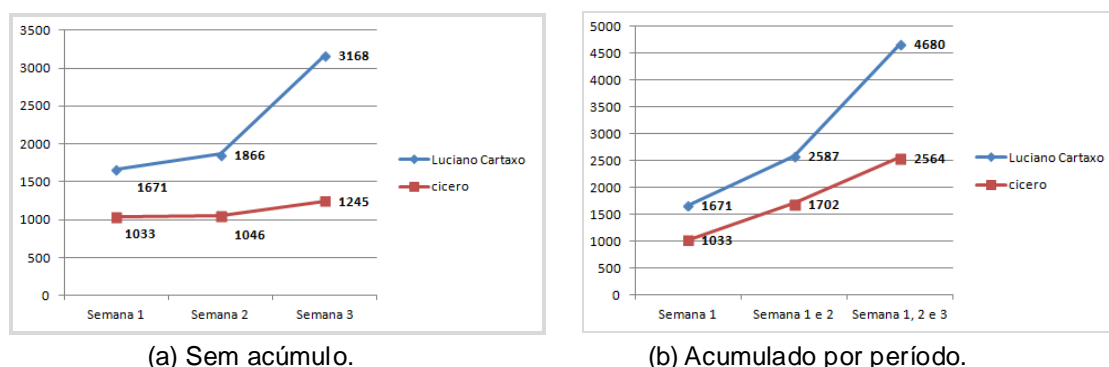


FIGURA 34 - Número de Pessoas distintas por semana.

A cada semana foi computada uma média simples da quantidade diária de *Tweets* de cada candidato. Por fim temos os dados descritos na Figura abaixo apresenta a evolução dos números de cada candidato.

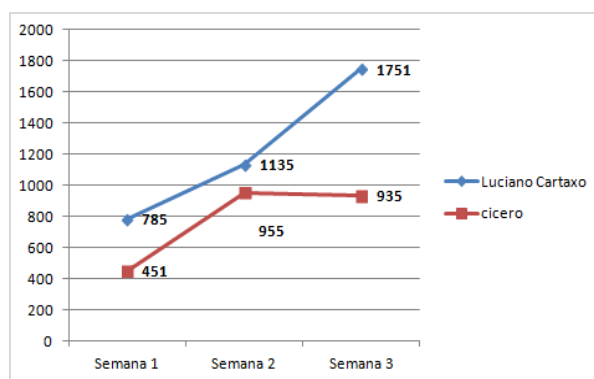


FIGURA 35 - Média de Tweets diários por semanas.

Outro índice verificado diz respeito aos dados das previsões apresentadas pelas pesquisas de intenção de voto registradas junto ao TRE-PB. Os dados apresentados na **Figura 36** foram obtidos através de pesquisas dos institutos Ibope e Consult devidamente registradas junto ao TRE-PB durante o período do 2º turno.

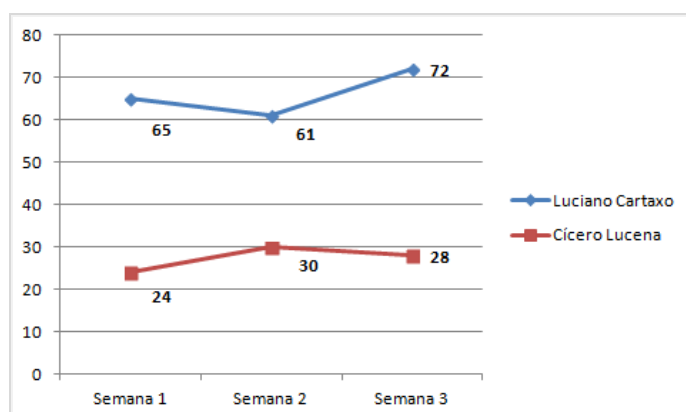


FIGURA 36 - Evolução das Pesquisas Eleitorais.

Por fim, a **Figura 37** apresenta o número de menções recebidas em momentos distintos durante a corrida eleitoral do 2º turno. O número de partida do gráfico foi obtido através de uma média aritmética simples da quantidade de menções que cada candidato recebeu no período compreendido entre 07/10/2012 e 25/10/2012. A segunda informação diz respeito à quantidade de menções que cada candidato recebeu durante o dia 26/10/2012 (debate veiculado pela TV Cabo Branco – Afiliada da emissora globo na Paraíba). Por fim, tem-se a quantidade que cada candidato obteve durante o dia 28/10/2012 (pleito eleitoral do 2º turno).

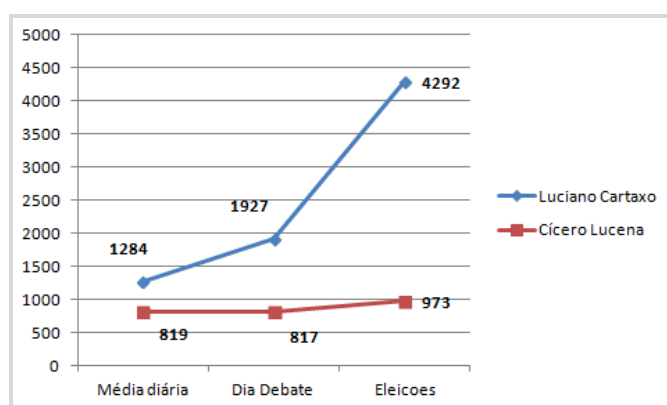


FIGURA 37 - Número de tweets diários.

5.2.2 Resultado final da eleição

Nesta parte do trabalho serão apresentados os resultados oficiais das eleições do município de João Pessoa de acordo com os dados obtidos junto ao TRE-PB. A **Tabela 28** apresenta o resultado do 2º turno e o compara com os dados da pesquisa do IBOPE registrada no TRE-PB em 27/10/2012.

TABELA 28 - Resultado do pleito eleitoral do 2º turno.

Candidato	Resultado do pleito		Resultado das Pesquisas	
	Votos	Percentual	Percentual	Margem
Luciano Cartaxo	246.581	68,13	72	70 - 74
Cícero Lucena	115.369	31,87	28	26 - 30

Ainda, de acordo com uma hipótese levantada sobre a relação entre usuários do Twitter e moradores dos bairros de maior poder aquisitivo, descrevemos na **Tabela 27** o resultado das eleições nos bairros de selecionados na [seção 3.1.4](#) deste trabalho.

TABELA 29 - Total de votos nos bairros de maior poder aquisitivo.

Bairro	Luciano Cartaxo	Cícero Lucena
Cabo branco	817	332
Tambaú	8.052	3.253
Manaíra	5.535	2.205
Miramar	925	448
Tambauzinho	3.128	1.357
Bessa	9.804	3.937
Altiplano	3.389	1.463
Total	31650	12995
	69,08 %	30,92 %

5.2.3 Aplicação de métricas de ARS

Após a obtenção dos grafo através da metodologia exposta na [seção 4.4](#) foi utilizado o Gephi para visualização e aplicação de métricas de ARS. De modo análogo ao que ocorreu na [seção 5.1.3](#), o presente estudo também foi dividido em três fases distintas destacadas a seguir: **(1)** Análise de acordo com a modularidade; **(2)** Análise de acordo com as métricas grau e grau ponderado; **(3)** Análise das métricas *PageRank*, *Betweenness Centrality*, *HITS* e *Closeness Centrality*.

Quando analisada de acordo com a métrica Modularidade o Gephi foi capaz de encontrar 02 (duas) comunidades na rede representadas na **Figura 38** (Veja em maiores detalhes no [Anexo L](#) e [Anexo M](#)), indicando cada um dos principais candidatos na disputa eleitoral do 2º turno.

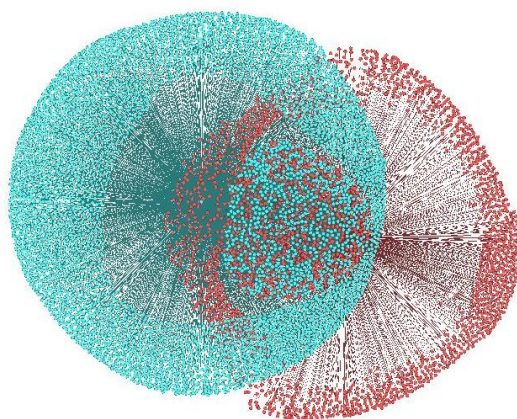


FIGURA 38 - Comunidades encontradas na rede.

Os nós **azuis** correspondem aos usuários que realizaram menção ao candidato Luciano Cartaxo. Os nós **vermelhos** correspondem aos usuários que realizaram menção ao candidato Cícero Lucena.

Percebe-se que no centro do grafo existe uma junção entre os nós **vermelhos** e **azuis**. Nesta porção estão os nós (pessoas) que citaram os dois candidatos de modo que a sua coloração é obtida através do maior número de citações, ou seja, caso ele menciona mais o candidato Cícero Lucena, recebe a cor vermelha e vice-versa. Na **Tabela 30** podemos comparar os dados obtidos com essa métrica com os resultados das eleições e pesquisas.

TABELA 30 - Comparação dos resultados com a métrica modularidade.

Candidato	Resultado do pleito		Resultado Bairros*		Modularidade	
	Votos	%	Votos	%	Nós	%
Luciano Cartaxo	246.581	68,13	31.650	69,08	4.182	67,28
Cícero Lucena	115.369	31,87	12.995	30,92	2.026	32,62

* Bairros selecionados de acordo com o poder aquisitivo

Quando analisada de acordo com a métrica **Grau Ponderado** (*Weighted Degree*), observa-se que 84,4% dos nós realizam entre 1 e 5 menções (nós com tons de **verde**) e que 14,7% realizam entre 6 e 10 menções (nós em tons de **vermelho**). Os demais 8% realizam mais de 10 menções (nós **pretos**).

Em face disto podemos analisar em cada comunidade o comportamento dos nós em relação a quantidade de menções. De acordo com a **Figura 39** abaixo ([Anexo N](#)), os usuários que citam o candidato Luciano Cartaxo (comunidade com maior número de nós), em sua maioria são em tons de **verde**, contudo verificamos uma maior presença de nós em tons de **vermelho** e em tons de **preto** do que na outra comunidade.

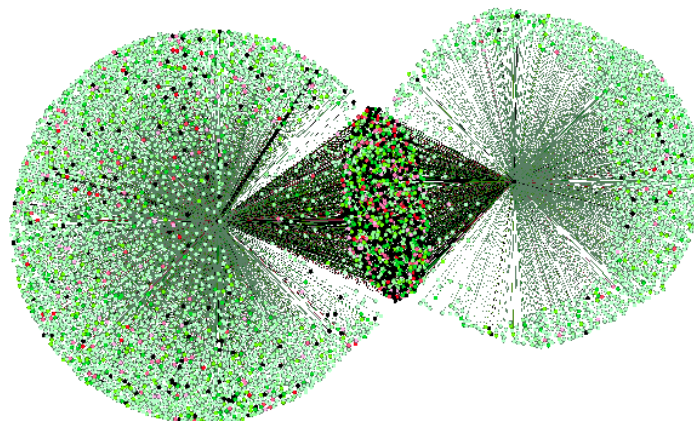


FIGURA 39 - Partição do grafo de acordo com o grau ponderado.

Na segunda etapa da análise verificaremos a rede de acordo com a métrica **Grau** (*Degree*). Esta métrica nos traz informações importantes a respeito da quantidade de ligações de cada nó. No nosso contexto isto significa se um usuário cita/ou é citado por uma ou várias pessoas. Na **Figura 40** ([Anexo O](#)) constatamos que 82% (nós **azuis**) dos nós possuem peso 1, ou seja, citaram apenas um dos candidatos. Ainda, os demais 18% (nós **verdes**) são nós com peso 2, ou seja, citando mais de um candidato.

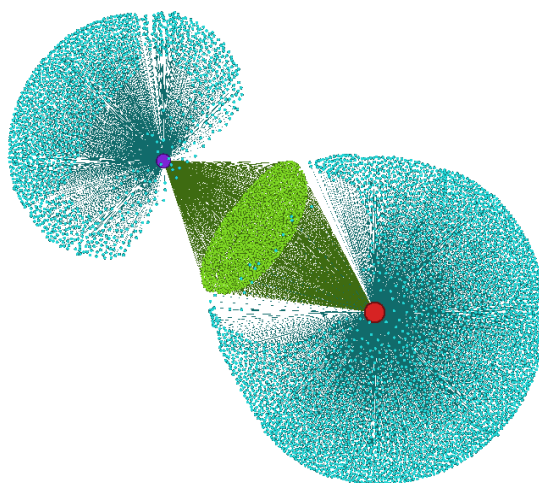


FIGURA 40 - Partição do grafo de acordo com o grau.

Diferentemente da modularidade onde um nó contribui apenas no valor final de um candidato, com esta métrica temos que cada nó pode contribuir no aumento do índice dos dois candidatos concomitantemente. Em virtude disto, se somarmos os valores desta métrica dos dois candidatos teremos um numero superior ao da quantidade de nós. Na **Tabela 31** podemos comparar os dados obtidos nesta métrica com os resultados das eleições e pesquisas.

TABELA 31 - Comparação dos resultados com as métricas de grau.

Candidato	Resultado do pleito		Resultado Bairros*		Degree		Weighted Degree	
	Votos	%	Votos	%	Nós	%	Peso	%
Luciano Cartaxo	246.581	68,13	31.650	69,08	4.740	64,77	25.678	61,05
Cícero Lucena	115.369	31,87	12.995	30,92	2.578	35,23	16.385	38,95

* Bairros selecionados de acordo com o poder aquisitivo

Por fim, foram analisados os resultados conforme as demais métricas. Como pode ser observado na Erro! Fonte de referência não encontrada.**32**, Luciano Cartaxo apresenta o maior índice em “*Betweenness*” que indica o quão frequentemente o nó que o representa aparece no caminho mais curto entre demais nós da rede. Ainda, mostrou-se mais próximo dos demais nós da rede por apresentar o menor “*Closeness*” e obteve os valores de “*PageRank*” e “*HITS*” elevados, indicando certo grau de autoridade ao nó que o representa na rede uma vez que o valor destas métricas é baseada nos valores de outros nós que compõem a rede.

TABELA 32 - Comparação entre as demais métricas ARS.

Candidato	<i>Closeness</i>	<i>Betweenness</i>	<i>HITS</i>	<i>PageRank</i>
Luciano Cartaxo	1,473	16.244.848	0,227	0,304
Cícero Lucena	2,169	8.333.099	0,124	0,155

5.2.4 Aspectos sobre Geolocalização

A *AppSNAElection* conseguiu verificar a existência de dados de geolocalização em aproximadamente 600 (seiscentos) *Tweets*. Este número pode ser considerado extremamente pequeno (cerca de 1%) em relação ao montante obtido durante o período de operação do software (correspondente ao 2º turno das eleições).

Na **Tabela 33** a seguir verifica-se a distribuição dos mesmos de acordo com o candidato o qual a mensagem (*Tweet*) realizava algum tipo de menção.

TABELA 33 - Menções (Tweets) com dados de geolocalização.

Candidato	Menções	%
Luciano Cartaxo	362	61,99
Cícero Lucena	222	38,01
Total	584	100,00

Em seguida procurou-se realizar uma correspondência entre latitude e longitude (dados componentes dos *Tweets*) e os bairros de maior poder aquisitivo elencados na [seção 3.1.4](#) deste trabalho.

TABELA 34 - Distribuição das Menções (Tweets) por bairro (1).

Bairro	Luciano Cartaxo	Cícero Lucena
Cabo branco	20	4
Tambaú	31	11
Manaira	53	14
Miramar	39	6
Tambauzinho	21	23
Bessa	34	23
Altiplano	2	1
Total	200	82
	70,92	29,08

Finalmente foi realizada a mesma correspondência acima para outros bairros relevantes para o nosso estudo (veja [seção 3.1.4](#)).

TABELA 35 - Distribuição das Menções (Tweets) por bairro (2).

Bairro	Luciano Cartaxo	Cícero Lucena
Mangabeira	40	13
Cristo	9	12
Castelo Branco	4	15
Bairro dos Ipês	9	32
Total	62	72
	46,27	53,73

Comparando os percentuais obtidos em ambas as tabelas acima e os resultados oficiais expostos na [seção 5.2.2](#), percebemos que os resultados elencados extrapolam uma margem de erro superior a 10 pontos percentuais. Resultados semelhantes foram apresentados na [seção 5.1.4](#).

5.2.5 Utilização do intervalo de Confiança

De igual forma ao que foi apresentado na [seção 5.1.5](#), buscamos realizar o cálculo do Intervalo de confiança a partir da “eliminação” das contas utilizadas pela campanha e contas de usuários robôs dos dados descritos na **Tabela 36** a seguir:

TABELA 36 - Dados para o cálculo do Intervalo de confiança.

Dado	Luciano Cartaxo	Cícero Lucena
Tamanho da amostra	4731	2572
Contas Eliminadas	9	6
Significância (<i>Alpha</i>)	0,05	0,05
Desvio Padrão	14,72	31,28
Média	4,78	7,02
Margem	0,42	1,20
Intervalo	4 - 5	6 - 8

A partir de cálculos simples (veja novamente na [seção 5.1.5](#)) conseguimos obter novos valores para análise. Veja a **Tabela 37**:

TABELA 37 - Análise dos dados do Intervalo de confiança.

Candidato	Tweets		Pessoas		Média Tweet/Pessoa
	Nr	%	Nr	%	
Luciano Cartaxo	1219	63,25	225	69,44	4,42
Cícero Lucena	708	36,75	99	30,56	6,96

Diferentemente dos resultados exposto no estudo de caso anterior, a utilização de intervalo de confiança nesta seção conseguiu apresentar uma margem de erro de pouco mais de 5% se comparado ao número de pessoas distintas. Se comparado a montante de tweets, também verificamos uma margem de erro de aproximadamente 5%.

Uma possível explicação para isto reside no fato de que devido a diminuição de opções de voto (no primeiro turno eram 4 opções) conseguimos obter uma menor disparidade entre a votação nos bairros de maior poder aquisitivo e nos demais bairros da cidade de João Pessoa.

5.2.6 Considerações sobre o 2º Turno

Na **Tabela 36** a seguir expomos de modo resumido as comparações realizadas ao longo deste trabalho. Cumpre dizer que identificamos um padrão relevante nos resultados obtidos pela métrica modularidade, bairros selecionados e o resultado oficial das eleições.

TABELA 38 – Comparação geral dos resultados.

Método	1º Lugar			2º Lugar			Margem
	C*	Valor	%	C*	Valor	%	
Resultado Oficial	LC*	246.581	68,13	CL*	115.369	31,87	X
Pesquisas Eleitorais	LC*	-	72,0	CL*	-	28,0	4
Bairros Seleção	LC*	31.650	69,08	CL*	12.995	30,92	1
Métrica <i>Modularity</i>	LC*	4.182	67,38	CL*	2.026	32,62	2
Métrica <i>Degree</i>	LC*	4.740	64,77	CL*	2.578	35,23	4
Métrica <i>W. Degree</i>	LC*	25.676	61,05	CL*	16.385	38,95	7
IC/Pessoa distintas	LC*	273	73,38	CL*	99	26,61	5

* C – Candidato, LC – Luciano Cartaxo, CL – Cícero Lucena

Do mesmo modo que o estudo de caso da [seção 5.1](#), a análise da métrica modularidade mostrou-se interessante. Se os dados obtidos a partir desta métrica forem comparados com os números finais das eleições temos uma margem de erro inferior a 1%. Da mesma forma se realizarmos a mesma comparação dos valores desta métrica com os da votação nos bairros de maior poder aquisitivo, podemos perceber que a margem também permanece baixa, ficando inferior a 2%.

Diferentemente do estudo de caso anterior, todos os métodos apresentados na **Tabela 38** conseguiram apontar a ordem dos candidatos mais bem votados durante o pleito do 2º turno. Ainda, conforme citamos na seção anterior, o método de intervalo de confiança conseguiu uma boa aproximação.

A partir das informações obtidas com o nosso trabalho podemos afirmar que quantitativamente, o twitter contribui claramente para a ampliação do debate público sobre aspectos relacionados a política. O grande número de *tweets* (mais de 40.000) no dado período de tempo atestam a sua utilização neste sentido. Percebemos ainda que grande parte dos comentários vindos do twitter são provocados em sua maioria através de eventos cobertos pela mídia convencional (radio, TV e jornais) tais como comícios políticos, debates ou mesmo eventos relacionados ao dia do pleito.

Em segundo lugar percebemos que a maioria dos usuários que citaram os candidatos são perfis relacionados a pessoas comuns e que a infinita maioria das contas, cerca de 80 %, realizaram entre 1 e 3 menções no máximo. Contudo, também verificamos a presença de contas “Robôs” que faziam grandes quantidades de menções e contas “anônimas” onde o modo com os quais estes usuários interagem e se relacionam com demais usuários, e em que medida a sua comunicação fica em distribuição mais ampla é um tema claro para uma análise mais aprofundada.

Finalmente, concluímos que a utilização deste tipo de estudo pode fornecer indícios fortes a respeito dos possíveis vencedores de um pleito eleitoral. Pode-se afirmar que de acordo com alguns dados deste estudo conseguiu apontar, com maior grau de fidelidade do que as pesquisas de intenção de votos, o posicionamento dos candidatos no 2º turno, muito embora nossos resultados no 1º turno não foram tão expressivos. A partir disto podemos concluir que a diminuição da polaridade de uma disputa pode ser benéfica em algumas técnicas utilizadas para este fim como a métrica *modularity* e a métrica *degree*.

5.3 Publicações

Até o presente momento nosso trabalho resultou em duas publicações e algumas submissões. A primeira publicação apresentou genericamente a proposta inicial de trabalho, revisão bibliográfica sobre redes sociais, as metodologias existentes para análise e os resultados preliminares obtidos através dos dados do 1º turno das eleições Municipais. O artigo foi denominado “*Análise da Influência das Redes Sociais sobre o Resultado de uma Eleição*”, apresentado no mês de novembro de 2012 durante o Workshop de Iniciação Científica da II Escola Paraibana de Informática (II EPI) em João Pessoa - PB e publicado como resumo expandido nos Anais deste evento.

Na segunda publicação relacionada ao nosso projeto de pesquisa conseguimos aprofundar os resultados obtidos no trabalho publicado descrito anteriormente e ainda, conseguimos levantar novas hipóteses a cerca da problemática em questão. O artigo foi denominado “*Estudo do uso do Twitter como Ferramenta de Análise de Opinião durante as Eleições Municipais de*

João Pessoa” e foi aceito e apresentado durante as atividades do II *Brazilian Workshop on Social Network Analysis and Mining* (II BRASNAM) componente do Simpósio da Sociedade Brasileira de Computação (CSBC) realizado no mês de julho do ano de 2013 na cidade de Aracajú-SE.

Temos a submissão de outro artigo denominado “*SONIC - Social Network Information’s Chain - Proposal Methodology for Social Network Analysis*”. Este artigo relata a metodologia utilizada para realizar a coleta de dados na rede social Twitter bem como detalha o modo da construção e funcionamento do software utilizado para este fim. O referido artigo foi submetido ao III Social Computing and Its Applications (II SCA) a ser realizado no mês de outubro do ano de 2013 no Karlsruhe Institute Technology (KIT) na cidade de Karlsruhe, Alemanha.

Em curto prazo temos a intenção de submeter o artigo denominado “*Acompanhando a Discussão Política nas Eleições Municipais de João Pessoa através da interação de usuários no Twitter*” que apresenta as conclusões obtidas a partir dos dados coletados durante o 2º turno. Este trabalho será submetido ao XII Workshop on Tools and Applications, evento componente do Simpósio de Multimídia e Web (WebMedia 2013) a ser realizado no mês de outubro do ano de 2013 nas instalações da Universidade Federal da Bahia (UFBA) na cidade de Salvador – BA. Ainda, iremos submeter o presente trabalho de dissertação ao XIII Workshop on going Thesis and Dissertations, evento este também componente do WebMedia 2013.

Finalmente, encerrando o ciclo de artefatos gerados com este tema de estudo, pretendemos compilar os resultados obtidos durante o 1º e 2º turno, tal qual fizemos no presente trabalho, em um artigo completo que deverá ser submetido para alguma revista ligada a área de análise de redes sociais. Ainda ampliaremos o estudo realizado com aspectos de geolocalização em um novo artefato científico.

6. Considerações

Neste capítulo serão discutidas as principais conclusões obtidas a partir dos estudos de caso apresentados nas seções [5.1](#) e [5.2](#). Desta forma o presente capítulo encontra-se dividido em quatro seções: Em Limitações do Trabalho ([Seção 6.1](#)) serão apontados os principais fatores que dificultaram o desenvolvimento da presente pesquisa; Em Discussão ([Seção 6.2](#)) ampliaremos o debate acerca das conclusões obtidas e expostas nas seções [5.1.5](#) e [5.2.5](#); Em Trabalhos Futuros ([Seção 6.3](#)) estão descritos algumas sugestões de pesquisas; Finalmente em considerações finais ([Seção 6.4](#)) realizamos os últimos comentários sobre a pesquisa em questão.

6.1 Limitações do trabalho

Com relação ao software desenvolvido, foram apresentados alguns óbices nos mecanismos para visualização e análise de redes sociais. As bibliotecas eleitas (descritas na [seção 3.3.5.2](#)) não apresentavam de modo satisfatório as duas funcionalidades de modo concomitante. Tal dificuldade foi resolvida a partir da utilização do Gephi ([seção 3.3.5.1](#)) que apesar de ser uma ferramenta de autoria, utilizava normalmente os arquivos com extensão (.) graphml.

Outro impedimento do nosso software foi a depreciação da versão 1.0 da REST API, onde o Twitter deixou de fornecer suporte a esta versão. Com isto a solução denominada *Carousel* inicialmente proposta ([seção 4.2](#)) teve que sofrer inúmeras modificações de modo que permanecesse cumprindo com seu papel de incrementar o número de requisições ao serviço desta API.

As principais limitações para este estudo são de ordens socioculturais. Iniciamos esta discussão na [seção 2.4](#) onde foram apresentadas as principais falhas ocorridas em pesquisas recentes relacionadas a este aspecto. Falhas estas que em médio ou longo prazo não serão resolvidas uma vez que não dependem única e exclusivamente do avanço da tecnologia, mas, sobretudo, de um maior acesso a este tipo de mídia por parte dos mais variados segmentos sociais que compõem uma dada população. Claramente isto depende, entre outros fatores, de uma melhor distribuição de renda.

Percebemos que existe uma enorme dificuldade em se analisar a intenção de voto em determinados segmentos sociais, uma vez que os usuários das redes sociais estão contidos em um maior ou menor número destes segmentos dependendo das características de cada sociedade. Este tipo de pesquisa consegue obter números expressivos em sociedades que possuem além de poder aquisitivo e nível de instrução para uso do computador e seus recursos, possuem, sobretudo, a cultura de se interessar e discutir sobre aspectos políticos. Contudo, se realizarmos uma análise sobre estes aspectos na sociedade brasileira (dados apresentados na [seção 3.2.1](#)) constata-se que os números não são muito favoráveis.

6.2 Discussão

Com base nos dados apurados, pode-se afirmar que a partir da construção e análise de uma rede composta por indivíduos envolvidos diretamente através da discussão política no contexto de uma eleição municipal, existe a possibilidade de fazer uso de métricas de SNA para obter informações importantes a respeito da evolução dos números reais de uma eleição. Este estudo se mostrou satisfatório, uma vez que conseguiu uma aproximação razoável em comparação com os resultados finais, mesmo analisando apenas a troca de mensagens, sem considerar outras variáveis que podem ter influência sobre o resultado do processo.

A partir das informações coletadas pela *AppSNAElection* podemos afirmar que quantitativamente, o twitter contribui de modo claro para a ampliação do debate público sobre aspectos relacionados a política. O grande número de *tweets* (aproximadamente de 90.000) no dado período de tempo atestam a sua utilização neste sentido. Percebemos ainda que grande parte dos comentários vindos do twitter são provocados em sua maioria através de eventos cobertos pela mídia convencional (radio, TV e jornais) tais como comícios políticos, debates ou mesmo eventos relacionados ao dia do pleito.

Nos estudos de caso descritos nas seções [5.1](#) e [5.2](#) podemos perceber que o uso da métrica modularidade conseguiu boas aproximações se comparadas com o resultado da votação nos bairros de maior poder aquisitivo

e também na votação final dos candidatos. Para o seu cálculo não foi levado em consideração o atributo *weight* (representa o total de menções) no intuito de minimizar a influência de contas que realizam um número exorbitante de citações a um candidato. Claramente este tipo de situação influencia negativamente nos resultados. Isto pode ser comprovado em dois momentos do trabalho: **(1)** Se compararmos os resultados apresentados pelas métricas grau e grau ponderado em ambos os estudos, percebemos que a margem de erro apresentada pela métrica grau ponderado é muito superior a apresentada pela métrica grau. **(2)** Analisando os resultados do intervalo de confiança percebemos claramente que após a eliminação das contas de campanha e contas “robôs” (que contribuíam com um número excessivo de menções) a margem de erro caiu consideravelmente.

Outro fator interessante da pesquisa foi a possibilidade de realizarmos uma comparação da votação dos candidatos nos bairros de maior poder aquisitivo e os números obtidos a partir do nosso estudo. No caso particular das eleições de João Pessoa não foi percebido uma diferença muito grande entre a votação total dos candidatos e a votação registradas nesse conjunto de bairros. Contudo, percebemos que os resultados obtidos a partir da nossa ferramenta tiveram uma maior aproximação quando comparado com a votação nos bairros de maior poder aquisitivo.

Podemos explicar isto a partir da suposição de que estes bairros possuem a população que se enquadra no perfil dos usuários do Twitter, ou seja, a população de maior poder aquisitivo e de maior nível de instrução. Este tipo de demografia poderia ser comprovado se tivéssemos um melhor universo amostral de *tweets* com informações geográficas. Nas seções [5.1.4](#) e [5.2.4](#) percebemos que menos de 1% das mensagens continham informações de latitude e longitude.

Finalmente, concluímos que a utilização deste tipo de estudo pode fornecer indícios fortes a respeito dos possíveis vencedores de um pleito eleitoral, em particular, os resultados apresentados nos bairros de maior poder aquisitivo. Pode-se afirmar ainda que, de acordo com alguns dados, este estudo conseguiu apontar com maior grau de fidelidade (principalmente no 2º

turno da eleição) do que as pesquisas de intenção de votos, o posicionamento dos candidatos. A partir disto podemos concluir que a diminuição da polaridade de uma disputa pode ser benéfica em algumas técnicas utilizadas para este fim tais como as métricas modularidade e grau.

6.3 Trabalhos Futuros

Para trabalhos futuros pretendemos realizar uma série de atividades. Inicialmente pretendemos verificar o conteúdo das mensagens para obtenção dos outros pontos de vista sobre o processo. Para isto sugerimos a realização da análise de sentimento ou da mineração de opinião que são técnicas bem definidas e aceitas na área. Além dos aspectos mencionados, também será analisada a rede sobre outros pontos de vista estruturais através da aplicação de outras métricas de ARS com intuito de verificar alguns padrões semelhantes aos resultados oficiais.

Percebemos que a utilização da demografia pode ser interessante para análise dos resultados. Contudo a utilização da rede social Twitter não tem condições de fornecer informações etárias sobre seus usuários. Com isso sugerimos a repetição dos passos metodológicos da presente pesquisa, tendo como objeto de estudo outra rede social que apresente nos metadados dos seus usuários informações etárias para que seja possível a realização de uma classificação a partir deste critério.

6.4 Considerações Finais

A presente pesquisa buscou estudar e analisar conceitos referentes ao processo de aplicação da análise de redes sociais a partir da modelagem de uma rede de usuários da rede social Twitter que realizaram algum tipo de menção aos candidatos postulantes ao cargo de prefeito da capital paraibana. Neste contexto, notou-se a importância de se propor um modelo de processo tanto para modelagem, quanto para sua análise, de modo que pudemos obter informações semelhantes daquelas obtidas através de pesquisas de opinião eleitoral, com vistas a prever o resultado final do pleito

Este trabalho foi desenvolvido no Laboratório de Análise de Redes Súcias – *LabSNA* (Acrônimo do Inglês *Laboratory of Social Network Analysis*) - da Universidade Federal da Paraíba (UFPB). O código-fonte da *AppSNAElection* bem como os dados obtidos durante o presente estudo poderão ser disponibilizado gratuitamente, desde que o solicitante assine um termo de comprometimento explicitando os motivos para uso do material. Para tal, entre em contato com o autor deste projeto através do email danylo_wagner@hotmail.com.

Referências Bibliográficas

ABEP, Associação Brasileira de Empresas de Pesquisa (2010). Disponível em: <http://www.abep.org>. Acesso em agosto de 2011.

- Albuquerque, D.W. *et al.* (2013). Estudo do uso do Twitter como Ferramenta de Análise de Opinião durante as Eleições Municipais de João Pessoa. // *Brazilian Workshop on Social Network Analysis and Mining*. XXXIII Congresso da Sociedade Brasileira de Computação.
- Aldé, A. (2004). Jornalistas e internet: a rede como fonte de informação política. Trabalho apresentado ao NP 02 – Jornalismo, do IV Encontro dos Núcleos de Pesquisa da Intercom, 2004.
- Anstead, Nick. O'Loughlin, Ben. Ampofo, Lawrence. (2012). *Trust, Confidence, Credibility: Citizen Responses on Twitter to Opinion Polls During the 2010 UK General Election*. In Loader, B. & Mercea, D. (Eds.), *Social Media and Democracy: Innovations in Participatory Politics*.
- Anthonisse, J. M. (1971). The rush in a directed graph. Technical Report BN 9/71, Stichting Mathematisch Centrum, Amsterdam.
- Balancieri, R. (2010). Um método baseado em ontologias para explicitação de conhecimento derivado da análise de redes sociais de um domínio de aplicação. Tese de Doutorado. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade federal de Santa Catarina.
- Barabasi, A. L.; Albert, R. (1999). *Emergence of scaling in random networks*. *Science*, pages 286–509.
- Benevenuto, F.; Almeida, J. M.; Silva, A. S. (2011). Explorando Redes Sociais Online: Da Coleta e Análise de Grandes Bases de Dados às Aplicações. Mini-curso do XXIX Simposio Brasileiro de redes de Computadores (SBRC).
- Beraldo, A. F. (2002). Pesquisas Eleitorais. In: Como ler pesquisas eleitorais, São Paulo: SBPM.
- Berkowitz, S. D. (1982). *An Introduction to Structural Analysis: The Network Approach to Social Research*. Toronto: Butterworth.
- Blondel, V. D., Guillaume, J., Lambiotte, R., Lefebvre, E. (2008) “Fast unfolding of communities in large networks”. In: *J. Stat. Mech.*
- Boyd, D., Ellison, N. (2007). *Social network sites: Definition, history, and scholarship*. *Journal of Computer-Mediated Communication*, 13(1-2).
- Bondy, J. A.; Murty, U. S. R. (1982). *Graph Theory With Applications*. North-Holland, p edition. Disponível em: <<http://migre.me/f6HJS>>. Acesso em maio de 2013.
- Brandes, U. (2001). *A Faster Algorithm for Betweenness Centrality*, in *Journal of Mathematical Sociology* 25(2):163-177.
- Bruns, A., Liang, E. (2012). *Tools and methods for capturing Twitter data during natural disasters*. *First Monday*. Disponível em: <<http://migre.me/f3T3I>>. Acesso em maio de 2013.
- Brin, S., Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the seventh International Conference on the World Wide Web*.
- Buchanan, M. (2002). *Nexus - small world and the groundbreaking science of network*. W. W. Norton Company.

- Caldarelli, G; Vespignani, A. (2007). *Large Scale Structure and Dynamics of Complex Networks*. World Scientific.
- Cardozo, M. L. (2009). Twitter: microblog e rede social, caderno.com, pp. 24-38. Disponível em:<<http://migre.me/f6H0w> >. Acesso abril de 2013.
- Castells, M. (1999). *A sociedade em rede*. São Paulo: Paz e Terra.
- Castells, M.; Cardoso, G. (2005). *A sociedade em rede do conhecimento à ação política*. Imprensa Nacional.
- Cervi, Emerson Urizzi. Massuchin, Michele Goulart (2011). *O uso do twitter nas eleições de 2010: o microblog nas campanhas dos principais candidatos ao governo do Paraná*. Contemporânea, Vol. 9, No 2: Dossiê Wikileaks - Cibercultura e Política.
- Centro de estudos sobre as Tecnologias de Informação e Comunicação – CETIC (2012). *Estudo sobre o uso da internet*. Disponível em: <http://www.cetic.br/usuarios/tic/2012/>. Acesso em maio de 2013.
- Chakrabarti, S. (2003) *Social Network Analysis, Mining the Web*, Morgan Kaufmann, pp. 203-254.
- Cheong, F.; Cheong, C. (2011). *Social media data mining: A social network analysis of tweets during the 2010-2011 Australian floods*. Proceedings of the Pacific Asia Conference on Information Systems, Paper 46.
- Cheng, J., Sun, A., Hu, D., Zeng, D. (2011). *An information diffusion-based recommendation framework for micro blogging*. Journal of the Association of Information Systems 12(7), 463-486.
- Clauset, A., Newman, M. E. J., Moore, C. (2004). *Inding community structure in very large networks*. Physical Review.
- Coleman, S. (1999). *Can the New Media Invigorate Democracy?*. The Political Quartely Publishing.
- Cormen, T. H.; Leiserson, C. E.; Rivest, R. L; Stein, C. (2002). *Algoritmos Teoria e Prática*. Editora Campus, 2ª edição.
- Costa, L. F., Rodrigues, F. A., Boas, P. R. V. (2007). *Characterization of complex networks: A survey of measurements*. Advances in Physics.
- Dahlgren, P. (2000). *The Internet and the Democratization of Civic Culture*. Political Communication, 17:335–340.
- Danon, L., Duch, J., Diaz-Guilera, A., Arenas, A. (2005). *Comparing community structure identification*. Journal of Statistical Mechanics: Theory and Experiment.
- Diakopoulos, N. A., Shamma, D. A. (2010). *Characterizing debate performance via aggregated twitter sentiment*. In Proceedings of the 28th international conference on Human factors in computing systems (CHI), 1195–1198.
- Donninger, C. (1986). *The distribution of centrality in social networks*. Social Networks, 8, 191-203.
- Erdos, P., Renyi, A. (1960). *On the evolution of random graphs*. Publications of the mathematical Institute of the Hungarian Academy of Sciences. 5, 17-61.

- Emirbayer, M., Goodwin, J. (1994). *Network Analysis, Culture and the Problem of Agency*. *The American Journal of Sociology*, v.99, n.6, p.1411-1454.
- Fernandes, L. C.; Carneiro, S. M. P. (2013). Divulgação de pesquisas eleitorais e decisão de voto. Pontifícia Universidade Católica de Goiás. Disponível em: <http://migre.me/fdgEE>. Acesso em maio de 2013.
- Filho, Walter dos Santos. Pappa, Gisele L. Júnior, Wagner Meira. Guedes, Dorgival (2010). Observatório da web: Uma plataforma de monitoração, síntese e visualização de eventos massivos em tempo real. Anais do XXXVII Seminário Integrado de Hardware e Software, SEMISH'10, pages 110–120.
- Freeman, L. C. (1977). *A set of measures of centrality based on betweenness*. *Sociometry*, 40:35–41.
- Freeman, L. C. (1979). *Centrality in social networks: Conceptual clarification*. *Social Networks*, 1:215–239.
- Freitas, M. C.; Pereira, H. B. B. (2005). Contribuição da análise de redes sociais para o estudo sobre os fluxos de informações e conhecimento. 6º Encontro Nacional de Ciência da Informação (CINFORM). Salvador-Bahia
- Gayo-Avelo, Daniel (2011a). Don't turn social media into another 'literary digest' poll. *Commun. ACM*, 54(10). Disponível em: <http://migre.me/f5QBU>. Acesso em abril de 2013.
- Gayo-Avello, Daniel. (2011b). *Limits of Electoral Predictions using Twitter*. In International AAAI Conference on Weblogs and Social Media. Disponível em: <http://migre.me/f49Qj>. Acesso em abril de 2013.
- Gayo-Avello, Daniel. (2012). *"I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper" - A Balanced Survey on Election Prediction using Twitter Data*. Disponível em: <http://migre.me/f3TRx>. Acesso em fevereiro de 2013.
- Girvan, M., Newman, M. E. J. (2001). *Community structure in social and biological networks*. In *Proceedings of National Academy of Sciences*, number 99, pages 8271–8276, USA.
- Glossário Eleitoral Brasileiro (2013). Glossário de termos eleitorais versão online. Disponível em: <http://www.tse.jus.br/eleitor/glossario>. Acesso em maio de 2013.
- Gomes, W. (2005). Internet e participação política em Sociedades Democráticas. *Revista FAMECOS*, Nr 2.
- Guimarães, P.R.B. (2008). Métodos quantitativos estatísticos. 1ª Edição. Editora IESDE Brasil S/A.
- Huisman, M., Van Duijn, M. A. J. (2004) *Software for social network analysis*. In: *Models and methods in social network analysis*, P. J. Carrington, J. Scott, S. Wasserman (eds.). Cambridge: Cambridge University Press.
- Ifukor, Presley. (2010). *Elections or Selections? Blogging and Twittering the Nigerian 2007 General Elections*. *Bulletin of Science, Technology & Society*, 30 (6), 398-414.

- Instituto Data Popular – IDP (2010). Estudo sobre o uso das Redes Sociais. Disponível em: <<http://www.datapopular.com.br/>>. Acesso em novembro de 2012.
- Instituto Brasileiro de Geografia e Estatística - IBGE (2000). Censo Demográfico. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em novembro de 2012.
- Instituto Brasileiro de Geografia e Estatística - IBGE (2010). Censo Demográfico. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em novembro de 2012.
- Instituto Brasileiro de Opinião Pesquisa e Estatística – IBOPE (2013). Pesquisas Eleitorais. Disponível em: <<http://migre.me/dfdKk>>. Acesso em maio de 2013
- Jungherr, A., Jürqens, P., Schoen, H. (2011). *Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasian, A., Sprenger, T. O., Sander, P. G., & Welpke, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment"*. Social Science Computer Review.
- Kleinberg, J. M. (1999), Authoritative Sources in a Hyperlinked Environment, in Journal of the ACM 46 (5): 604–632.
- Lamos, V., Cristianini, N. (2010). *Tracking the flu pandemic by monitoring the social web*. In Cognitive Information Processing: 2nd International Workshop.
- Larsson, A. O., Moe, H. (2011). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. New Media & Society.
- Lemos, A. (2009). Nova esfera Conversacional. Em: Dimas A. Künsch, D.A, da Silveira, S.A., et al. Esfera pública, redes e jornalismo. Ed. E-Papers.
- Levy, P. (1999). Cibercultura. Tradução de Carlos Irineu da Costa. Editora 34.
- Lotan, G., Ananny, M., Gaffney, D., Boyd, d. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. International Journal of Communication 5, 1375-1405.
- Marteletto, 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.
- Marteletto, R. M., (2001). Análise de Redes Sociais – Aplicações nos estudos de transferência da informação. *Ci. Inf.* Brasília, Brasil, PP 1-10.
- Metz, Jean *et al.* (2007) Redes complexas: Conceitos e aplicações. Relatório técnico do Instituto de Ciências Matemáticas e de Computação (ICMC). Universidade Federal de São Carlos (UFSCar). Nr 290.
- Mika, P. (2007). *Social Networks and the Semantic Web*. Springer, 2007.
- Moen, R.; Norman, Clifford. (2011). *Evolution of the PDCA Cycle*. Associates in Process Improvement – Détroit, United States.
- Newman, M. E. J. (2003). *The structure and function of complex networks*. Cornell University library. Disponível em: <<http://migre.me/f5eJL>>. Acesso em fevereiro de 2013.

- Newman, M. E. J. (2004a). *Analysis of weighted networks*. Physical Review.
- Newman, M. E. J. (2004b). *Fast algorithm for detecting community structure in networks*. Physical Review.
- Newman, M. E. J. (2006). *Modularity and community structure in networks*. Proceedings of the National Academy of Sciences of the United States of America.
- Oliveira, R. F. (2013). *Análise Quantitativa das Relações entre Usuários no Twitter*. Dissertação de mestrado. Programa de Pós-Graduação em Informática da Universidade Federal da Paraíba.
- O'reilly, T (2009). *The Twitter Book*. Canadá.
- Orihuela. (2007). *Twitter y el boom del microblogging*. Perspectivas del mundo de la comunicación. Facultad de Comunicación, Universidad de Navarra.
- Park, J., Cha, M., Kim, H., Jeong, J. (2011). Managing bad news in social media: A case study on Domino's Pizza crisis. *Proceedings of the ICWSM*.
- Pereira, J. B., Albuquerque, D. W., Mendonça, M. L., Brito, A. V. (2012) "Análise da Influência das Redes Sociais sobre o Resultado de uma Eleição". II Escola Paraibana de Informática, 2012.
- Ravasz, E., Barabási, A. L. (2003). *Hierarchical organization in complex networks*. Physical Review.
- Recuero, R. (2009). *Redes Sociais na Internet*. Cibercultura.
- Silva, S. (2006). Participação política e Internet: propondo uma análise teórico-metodológica a partir de quatro conglomerados de fatores. Congresso Anual da Associação Brasileira de Pesquisadores de Comunicação e Política, Universidade Federal da Bahia – Salvador-BA.
- Silva, Pedro A. Leme. (2012a) *Análise de Redes Sociais aplicada à Engenharia Social*. Faculdade de Tecnologia de Guaratinguetá, São Paulo. Disponível em: <<http://migre.me/f7JSD>>. Acesso em junho de 2013.
- Silva, Talita Victor. (2012b) *O uso do twitter pelos deputados federais brasileiros: estudo sobre atuação e tendências de comportamento*. Monografia – Curso de Especialização em Instituições e Processos Políticos do Legislativo. Disponível em: <<http://migre.me/fa8G7>>. Acesso em julho de 2013.
- Sabidussi, G. (1966). *The centrality index of a graph*. Psychometrika, 31:581–603.
- Santos, R. O.; Lopes, L. F. D. (2002). *Cálculo do tamanho da amostra para uma pesquisa eleitoral*. Universidade Federal de Santa Maria (UFSM). Disponível em: <<http://migre.me/fdfWP>>. Acesso em maio de 2013.
- Simoës, S. J., et al. (2011). *Exploring Influence and Interests among Users within Social Networks*. Social Networks: Computational Aspects and Mining, Computer and Communication Networks Series, Springer.
- Skoric, M., Poor, N., Achananuparp, P., Lim, E. P., Jiang, J. (2012). *Tweets and Votes: A Study of the 2011 Singapore General Election*. In Proceedings of the 45th Hawaii International Conference on System Sciences.

- Spyer, J. (2009). Tudo o que você precisa saber sobre o twitter. Disponível em: <<http://migre.me/fa3Wa>>. Acesso em maio de 2013.
- Stanley, W., Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Stanley, J. W., Weare, C. (2003). *The Effects of Internet Use on Political Participation: Evidence from an Agency Online Discussion Forum*. DG.O.
- Stieglitz, S., Krüger, N. (2011): Analysis of sentiments in corporate Twitter communication – A case study on an issue of Toyota. Proceedings of the 22nd Australasian Conference on Information Systems, Paper 103.
- Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410:268–276. Disponível em: <<http://migre.me/f7sle>>. Acesso em fevereiro de 2013.
- Topografia Social de João Pessoa - TSJP (2009). Cedest/IEE/PUCSP. Disponível em: <<http://migre.me/f48nt>>. Acesso em novembro de 2012.
- Tribunal Regional Eleitoral da Paraíba – TRE/PB (2012). Estatísticas do eleitorado. Disponível em: <<http://www.tre-pb.gov.br/eleitor/estatisticas-eleitorado.html>>. Acesso em novembro de 2012.
- Tribunal Superior Eleitoral – TSE (2012). Estatísticas das eleições. Disponível em: <<http://www.tse.jus.br/eleicoes/estatisticas/estatisticas-eleicoes-2012>>. Acesso em outubro de 2012.
- Tumasjan, Andranik., Sprenger, Timm O., Sandner, Philipp G., Welpe, Isabell M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.
- Tumasjan, Andranik., Sprenger, Timm O., Sandner, Philipp G., Welpe, Isabell M. (2011). *Election Forecasts With Twitter How 140 Characters Reflect the Political Landscape*. Social Science Computer Review.
- Twitter. (2012). “Sobre o Twitter”. Disponível em: <<https://twitter.com/about>> Acessado em Novembro de 2012.
- Venturi, Gustavo. (1995). Pesquisas Pré-Eleitorais: legitimidade, influência e contribuições à cidadania. *Revista Opinião Pública*, v. 3, nr 2, Cesop Unicamp, Campinas – São Paulo.
- Vis, F. (2012). *Reading the riots on Twitter: Who tweeted the riots? Researching Social Media*. Disponível em: <<http://migre.me/f3Tar>>. Acesso em maio de 2013.
- Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, (393):440–442.
- Wasserman, S., Faust, K., (1994). *Social Network Analysis: Methods and Applications*. 8. Ed. Cambridge University Press, New York, EUA.
- Weng, Jianshu. Lim, Ee-peng. Jiang, Jing. He, Qi. *Twitterrank: finding topic-sensitive influential twitterers*. In ACM international conference on Web search and data mining (WSDM). Disponível em: <<http://migre.me/f7Gmt>>. Acesso em junho de 2013.

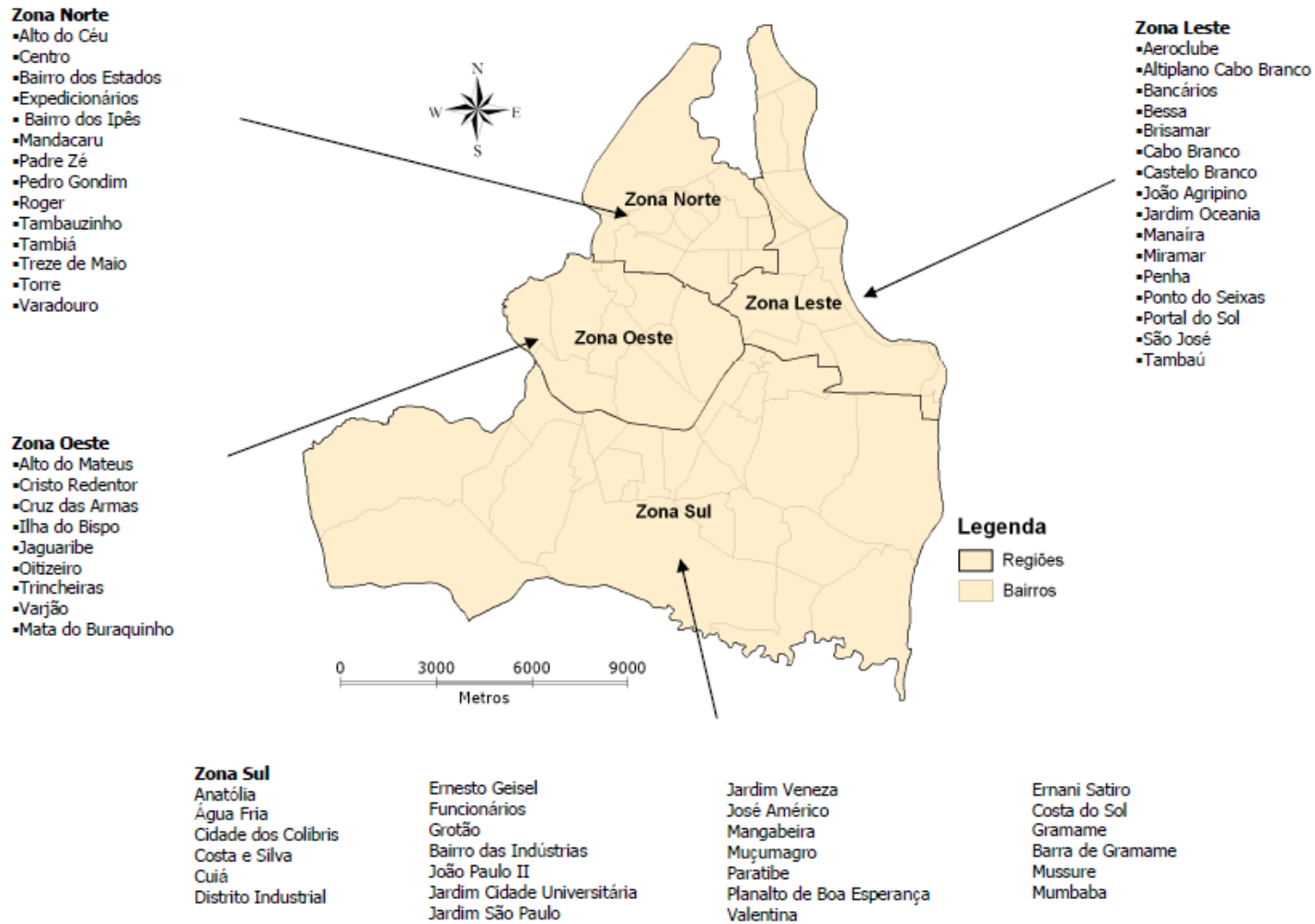
Anexos

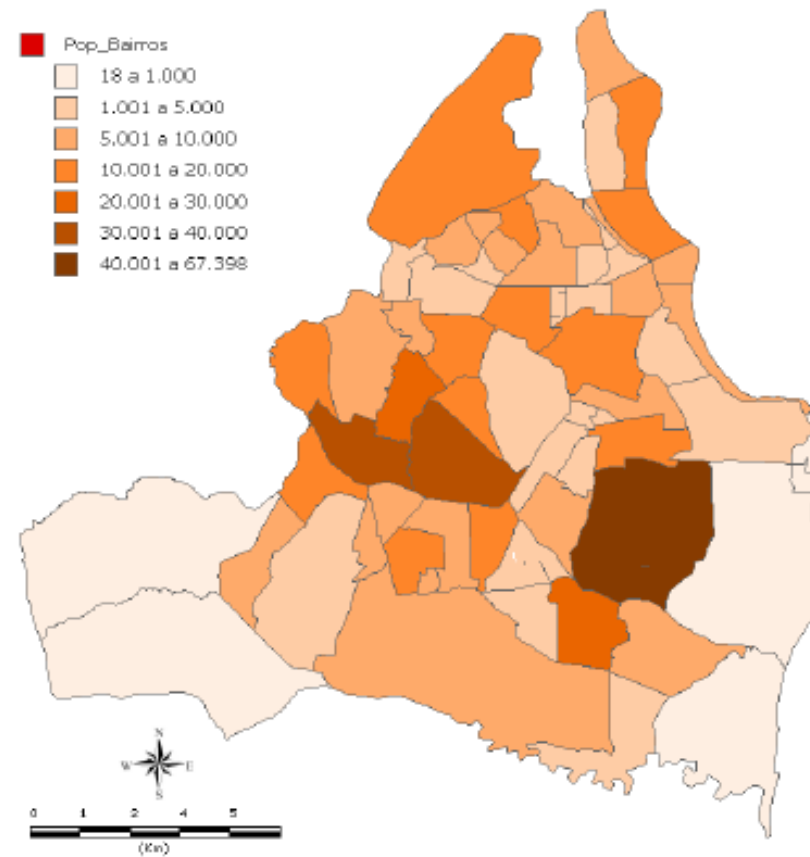
Anexo A – Seções eleitorais e eleitores por bairro (TRE-PB, 2012).

Bairro	Seções	Locais	Eleitorado	Bairro	Seções	Locais	Eleitorado
Altiplano Cabo Branco	17	3	6.730	Jaguaribe	48	3	14.033
Alto do Mateus	38	6	11.416	Jardim C. Universitária	9	2	4.189
Anatólia	12	2	5.558	Jardim Luna	3	1	1.148
Bairro das Indústrias	27	5	8.139	Jardim Planalto	12	1	3.746
Bancários	27	3	11.862	Jardim Veneza	13	2	4.387
Bessa	44	6	19.413	João Agripino	9	1	3.731
Cabo Branco	4	1	1.682	João Paulo II	2	1	634
Castelo Branco	20	4	8.007	José Américo	22	3	8.815
Castelo Branco I	4	1	1.753	Manaíra	27	5	10.949
Centro	64	12	18.912	Mandacarú	21	4	9.192
Conjunto José V. Diniz	8	1	2.639	Mangabeira I	65	7	25.788
Pedro Gondim	8	1	3.548	Mangabeira II	31	5	11.533
Costa e Silva	26	6	7.502	Mangabeira IV	5	1	1.841
Cristo Redentor	69	13	30.983	Mangabeira VI	6	1	2.351
Cruz das Armas	64	9	25.872	Mangabeira VII	11	2	3.873
Ernani Sátiro	19	3	5.706	Mangabeira VIII	12	3	4.742
Ernesto Geisel	43	6	16.304	Miramar	6	2	1.942
Esplanada I	8	1	2.414	Bairro dos Novais	39	5	12.287
Bairro dos Estados	20	4	8.389	Padre Zé	15	2	5.703
Expedicionários	14	3	6.033	Rangel	23	5	9.768
Funcionários I	26	4	7.780	Roger	14	3	5.706
Funcionários II	35	4	10.626	Tambaú	39	4	16.176
Gramame	6	1	2.340	Tambauzinho	17	3	6.190
Grotão	21	3	7.019	Tambiá	22	5	7.431
Ilha do Bispo	15	3	4.696	Torre	44	6	15.987
Bairro dos Ipês	35	5	14.224	Treze de Maio	17	4	6.929
Jaguaribe	48	3	14.033	Valentina	74	10	31.412

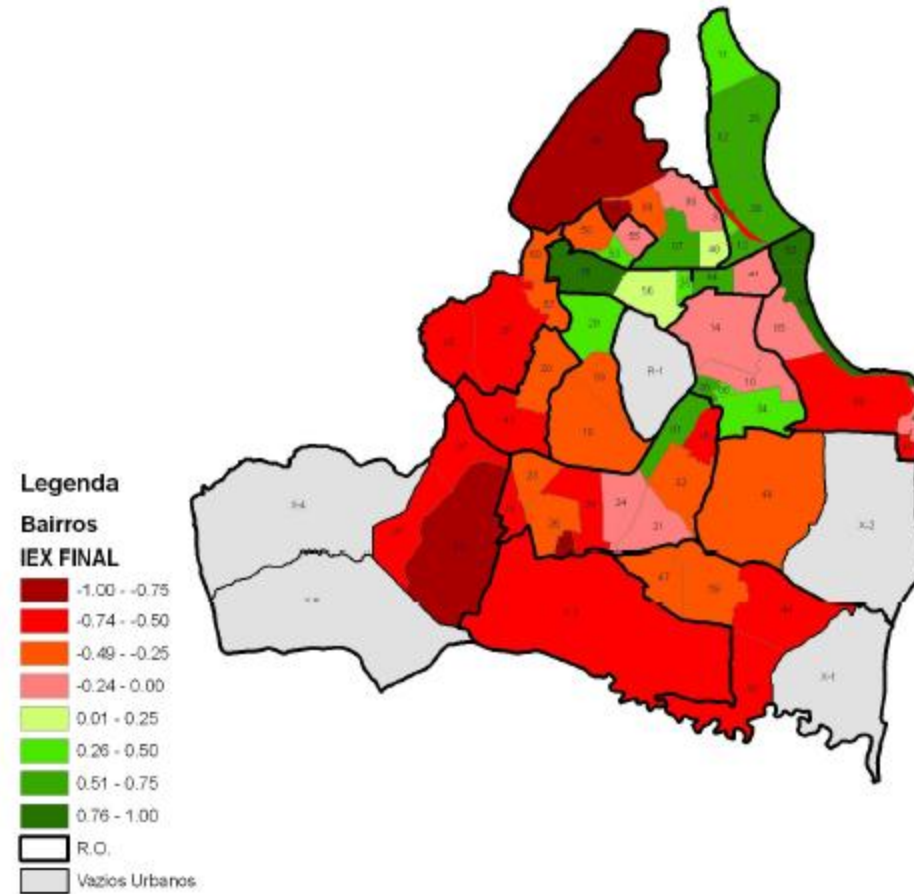
(*) Alguns bairros listados não são considerados oficialmente como tal, a não ser para o TSE. Também vale salientar que alguns bairros não possuem Seções Eleitorais específicas.

Anexo B – Divisão dos bairros de João Pessoa por região (TSJP, 2009).

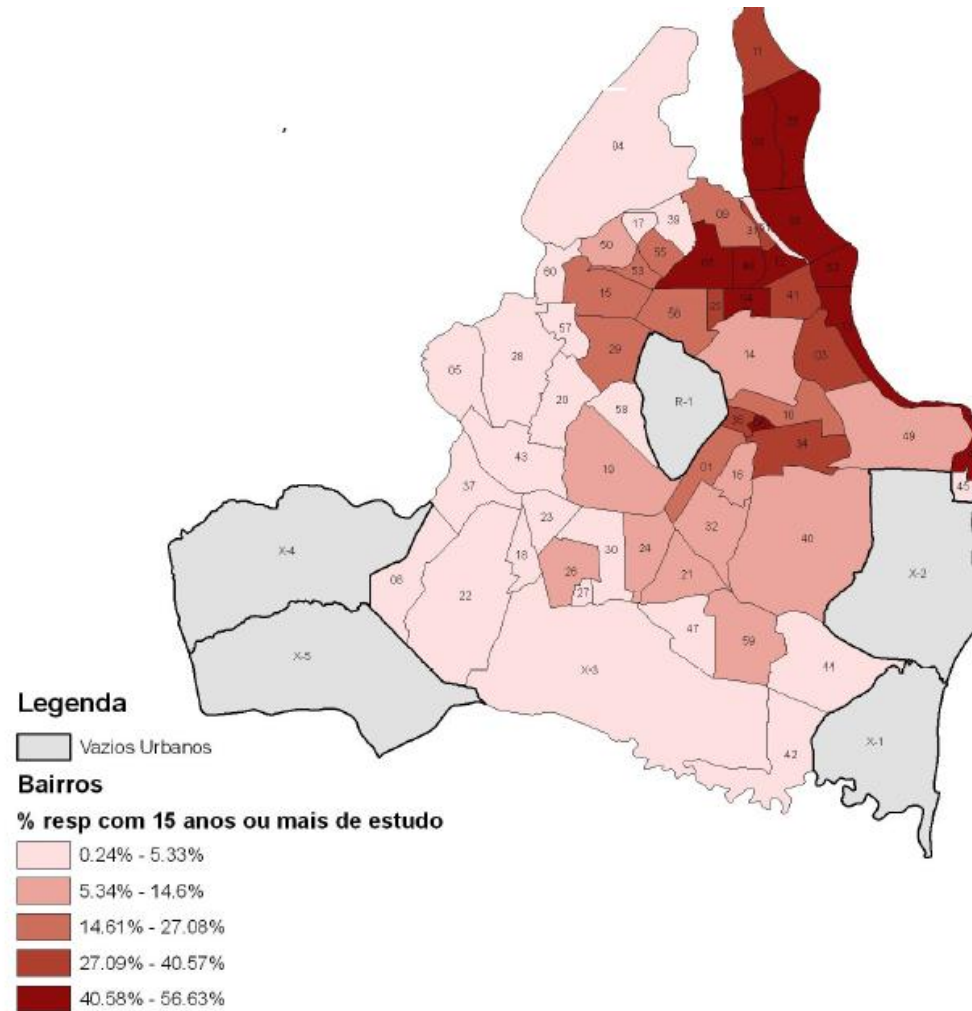


Anexo C – Intensidade de população dos bairros de João Pessoa (TSJP, 2009).

Anexo D - Mapa da exclusão/inclusão social de João Pessoa (TSJP, 2009).



Anexo E – Incidência % de responsáveis por domicílio com mais de 15 anos de estudo (TSJP,2009)



Anexo F - Programas para análise de redes sociais (Adaptada de Huisman e Van Duijn, 2004).

Programa	Versão	Objetivo	Dados		Funcionalidade		Suporte		
			Tipo ¹	Entrada ²	Visualização	Análises ³	Disponibilidade ⁴	Manual	Help
Agna	2.1.1	Geral	c	m	Sim	d, sl, sequencial	free	Sim	Sim
Blanche	4.8.1	Redes Dinâmicas	c	m	Sim	simulação	free	Sim	Sim
FATCAT	4.2 ⁵	Análise contextual	c	ln	Não	d, s	free ⁵	Não	Sim
GRADAP	2.0 ⁵	Análise gráfica	c	ln	Não	d, sl, dt	com ⁵	Sim	Não
Iknow	-	Redes de conhecimento	e	n	Sim	d, sl	free	Sim	Sim
InFlow	3.1	Mapeamento	c, e	ln	Sim	d, sl, rp	com	Sim	Sim
KliqueFinder	0.11	Subgrupos coesivos	c	m, ln	Sim	sl, s	-	Sim	Não
MultiNet	4.76	Análise contextual	c, l	ln	Sim	d, rp, s	free	Não	Sim
NEGOPY	4.30 ⁵	Subgrupos coesivos	c	ln	Sim	d, sl, rp	com ⁵	Sim	Sim
NetDraw	2.097	Visualização	c, e, a	m, ln	Sim	d, sl	free	Sim	Não
NetMiner III	3.4.0	Análise visual	c, e, a	m, ln	Sim	d, sl, rp, dt, s	com ⁷	Sim	Sim
NetVis	2.0	Exploração visual ⁶	c, e, a	m, ln	Sim	d, sl	free ⁶	Não	Sim
Pajek	2.00	Visualização dados	c, a, l	m, ln	Sim	d, sl, rp, dt	free	Não	Não
PermNet	0.94	Testes de permutação	c	m	Não	dt, s	free	Não	Sim
PGRAPH	2.7	Redes de parentesco	c	ln	Não	d, rp	free	Não	Sim
ReferralWeb	2.0	Cadeias de referências	e	ln	Sim	d	-	Sim	Sim

Continuação do Anexo F - Programas para análise de redes sociais (Adaptada de Huisman e Van Duijn, 2004).

Programa	Versão	Objetivo	Dados		Funcionalidade		Suporte		
			Tipo ¹	Entrada ²	Visualização	Análises ³	Disponibilidade ⁴	Manual	Help
SM LinkAlyzer	2.1	Populações escondidas	e	ln	Sim	d	com ⁷	Sim	Sim
SNAFU	2.0	Geral para MacOS ⁶	c	m, ln	Sim	d, sl	free ⁶	Não	Não
Snowball	- ⁵	Populações escondidas	e	ln	Não	s	free ⁵	Sim	Não
STOCNET	1.8	Análise estatística	c	m	Não	d, dt, s	free	Sim	Sim
STRUCTURE	4.2 ⁵	Análise estrutural	c, a	m	Não	sl, rp	free ⁵	Sim	Não
UCINET	6.289	Abrangente	c, e, a	m, ln	Sim	d, sl, rp, dt, s	com ⁷	Sim	Sim
Visone	2.62	Exploração Visual	c, e	m, ln	Sim	d, sl	free	Sim	Sim

1 c=completa, e=egocêntrica, a=afiliação, l = redes grandes.

2 m=matriz, ln=link/vértice, n=vértice.

3 d=descritiva, sl=estrutural e localização, rp=papéis e posições, dt=métodos para díades e tríades, s=estatística.

4 com=produto comercial, free=freeware/shareware.

5 Programa DOS que já não é atualizado.

6 Software Open Source.

7 Versão de demonstração/avaliação disponível.

Anexo G – API's para análise de redes sociais (Adaptada de Huisman e Van Duijn, 2004).

Programa	Versão	Objetivo	Dados		Funcionalidade		Suporte		
			Tipo ¹	Entrada ²	Ambiente	Análises ³	Disponibilidade ⁴	Manual	Help
JUNG	2.01	Modelagem gráfica	c	ln	Java	d, sl, vis	free	Sim	-
MatMan	1.0	Análise estrutural	c, a	m	Excel	d, sl	com	Sim	Sim
SNA	0.41	Geral	c	m	R/S	d, sl, rp, dt, s, vis	free	Sim	-
SNAP	2.5	Geral	c	m	Gauss	d, sl, rp, dt, s	com	Sim	-
yFiles	2.7.0.1	Exploração Visual	c	ln	Java	d, sl, vis	com	Sim	-

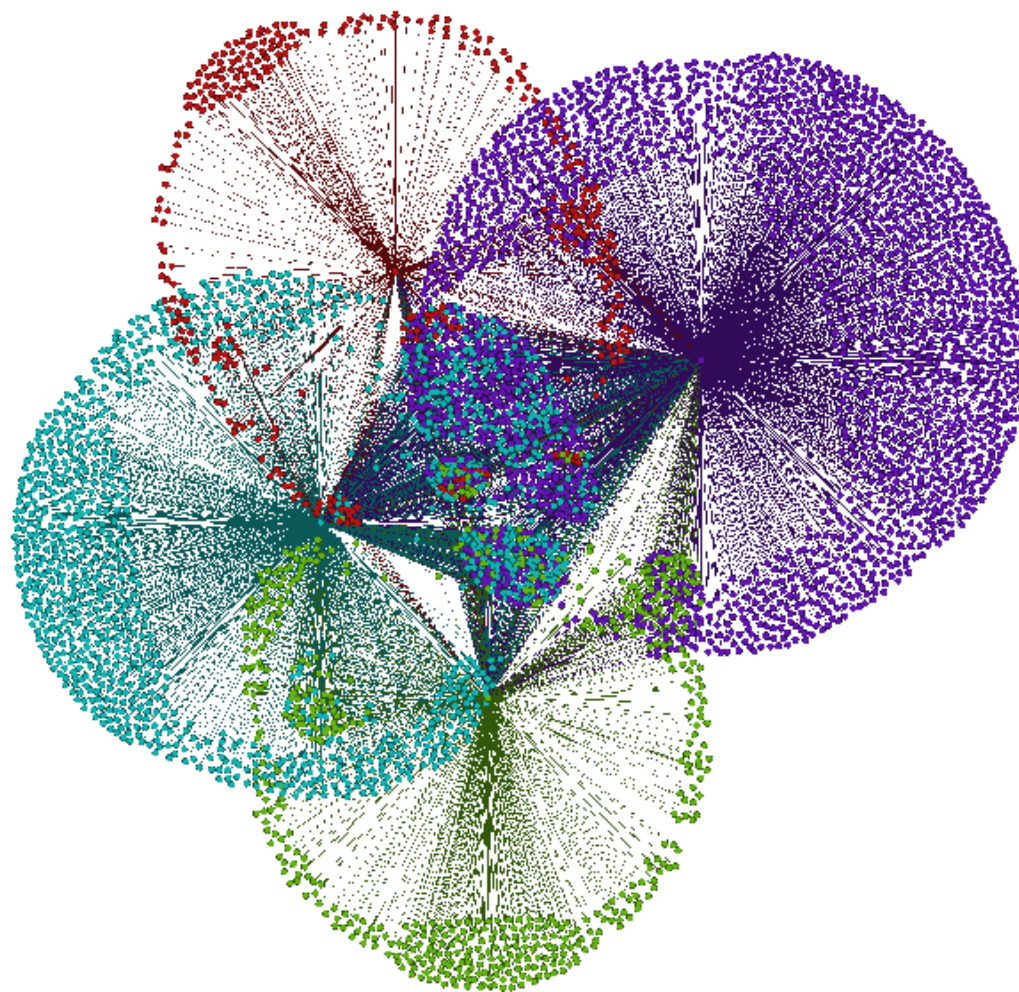
1 c=completa, e=egocêntrica, a=afiliação, l = redes grandes.

2 m=matriz, ln=link/vértice, n=vértice.

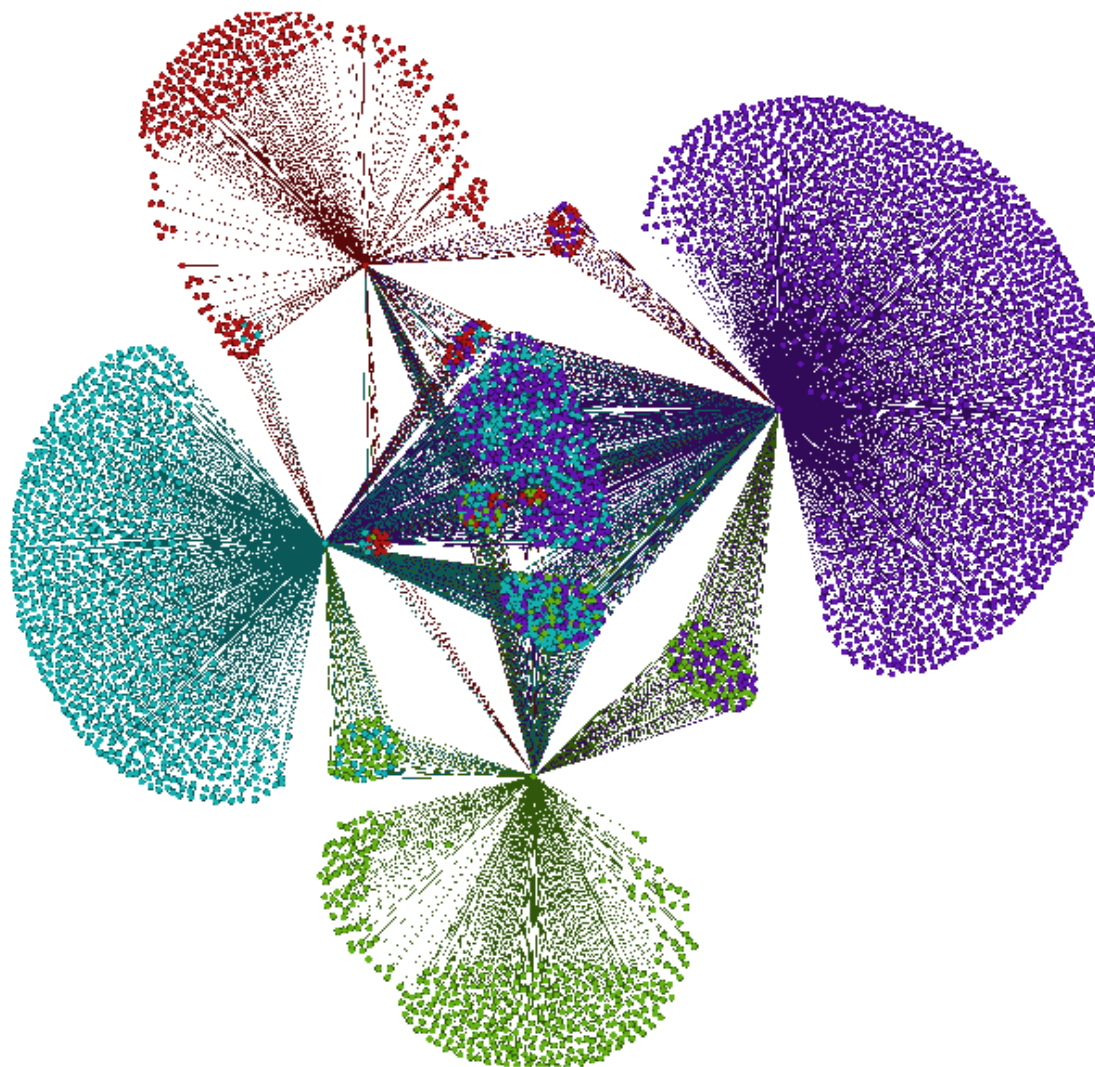
3 d=descritiva, sl=estrutural e localização, rp=papéis e posições, dt=métodos para díades e triades, s=estatística, vis=visualização.

4 com=produto comercial, free=freeware/shareware.

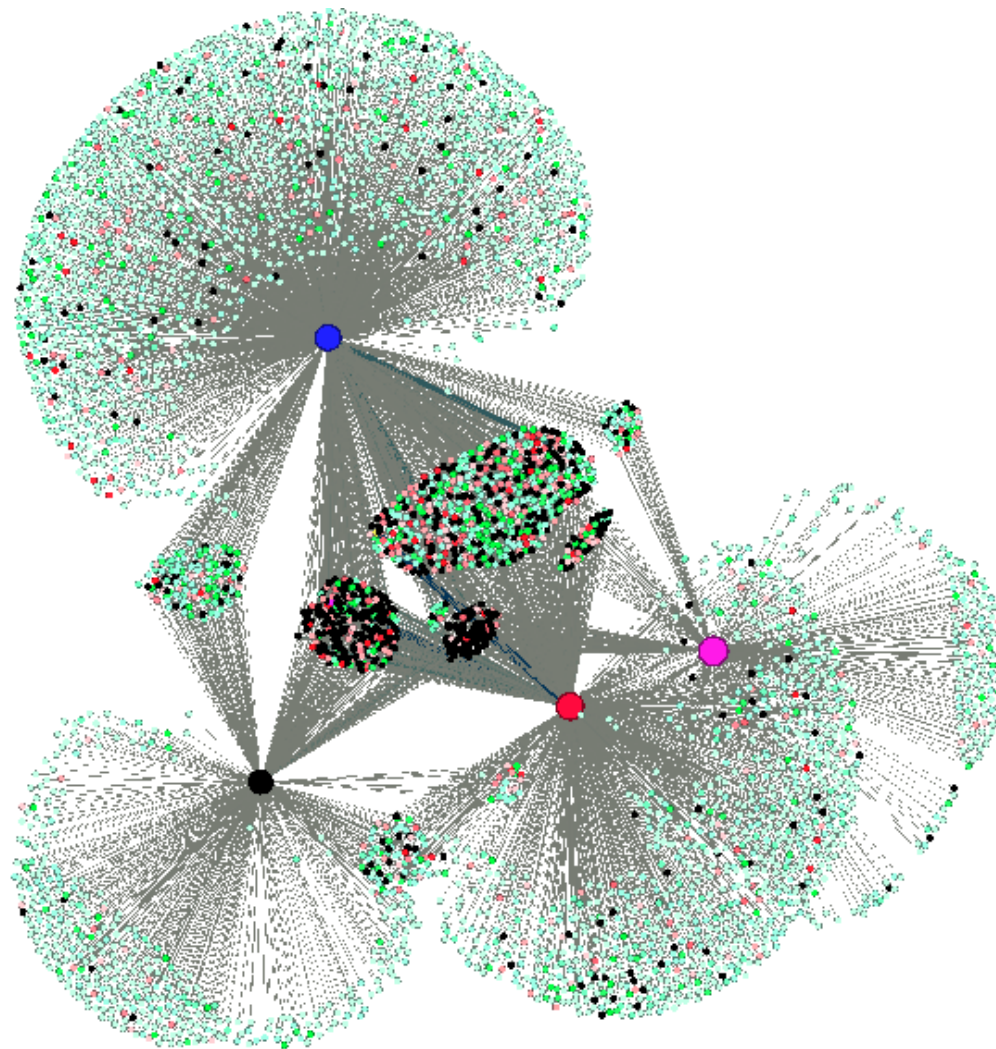
Anexo H – Rede com partição de acordo com Modularidade (1º Turno) – (1)



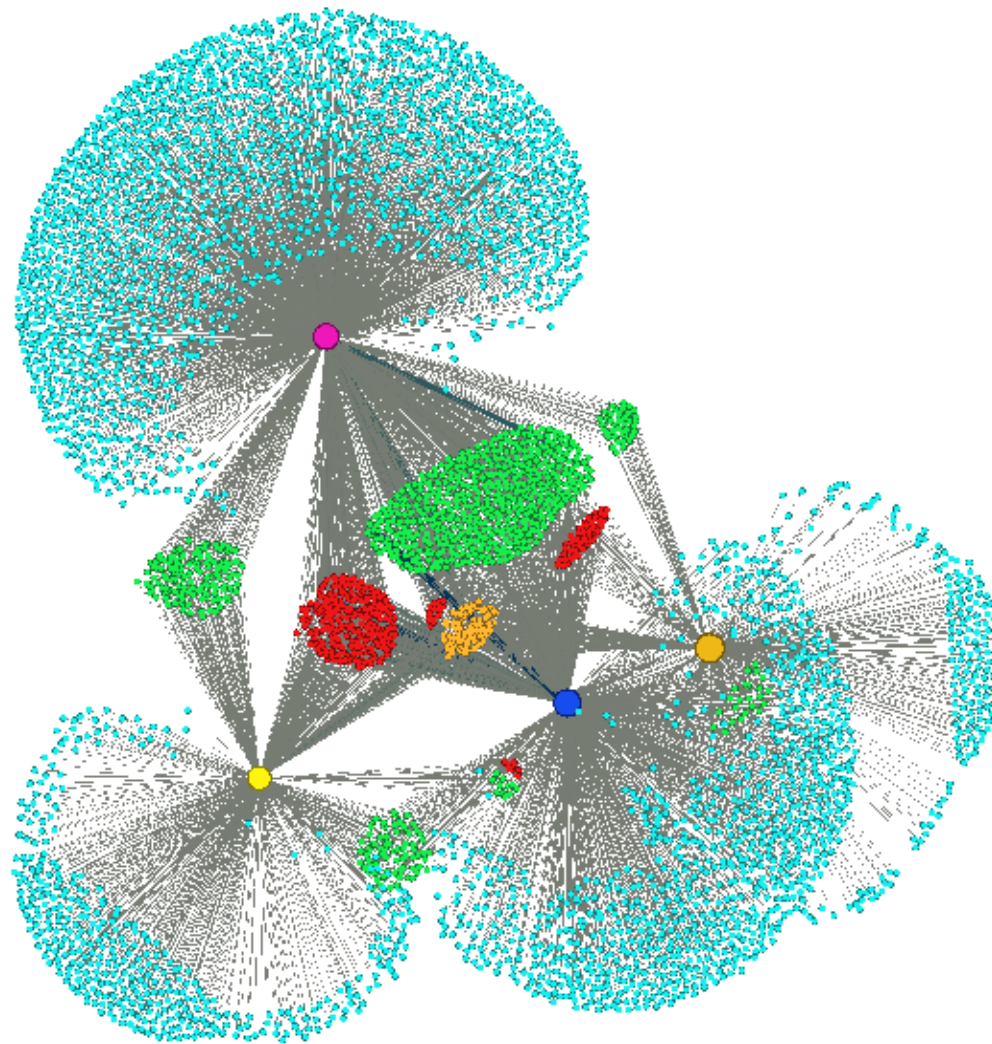
Anexo I – Rede com partição de acordo com Modularidade (1º Turno) – (2)



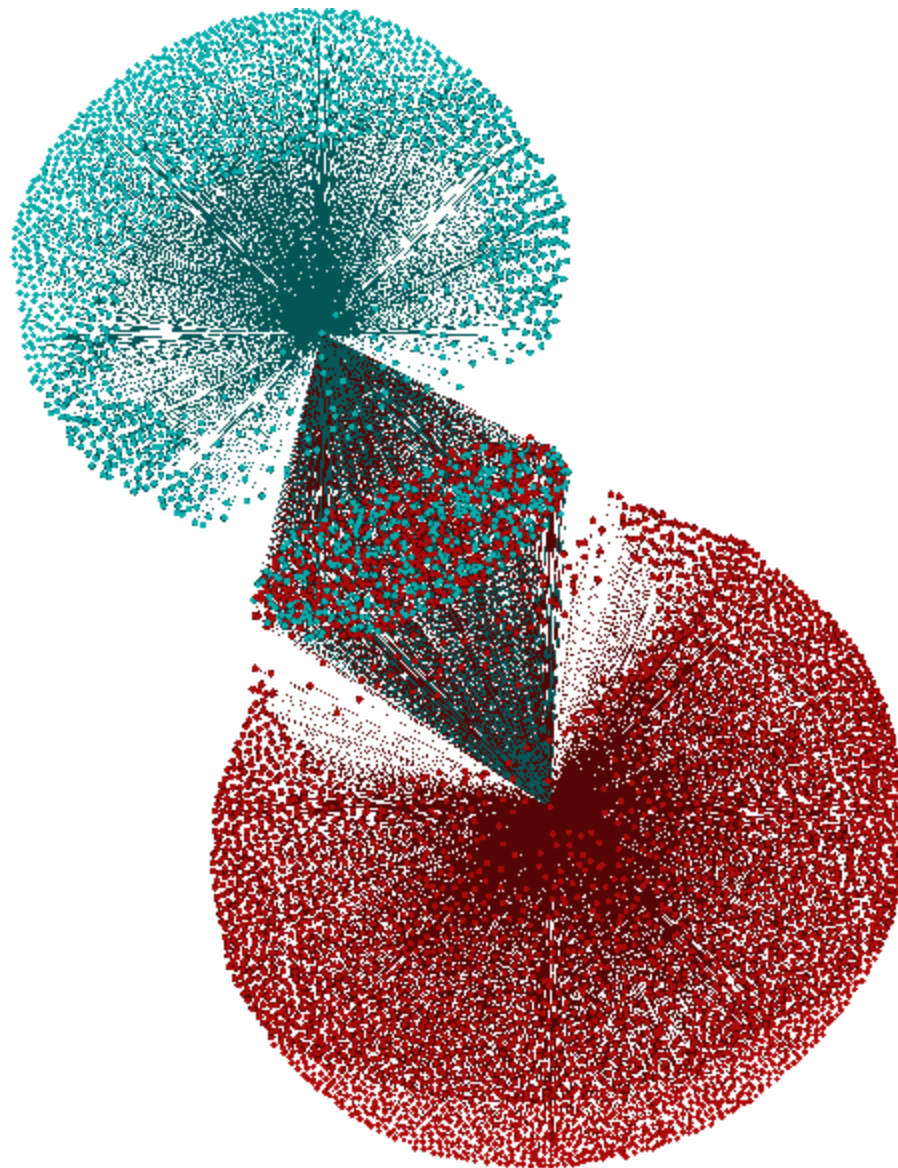
Anexo J – Rede com partição de acordo com Modularidade e Grau ponderado (1º Turno)



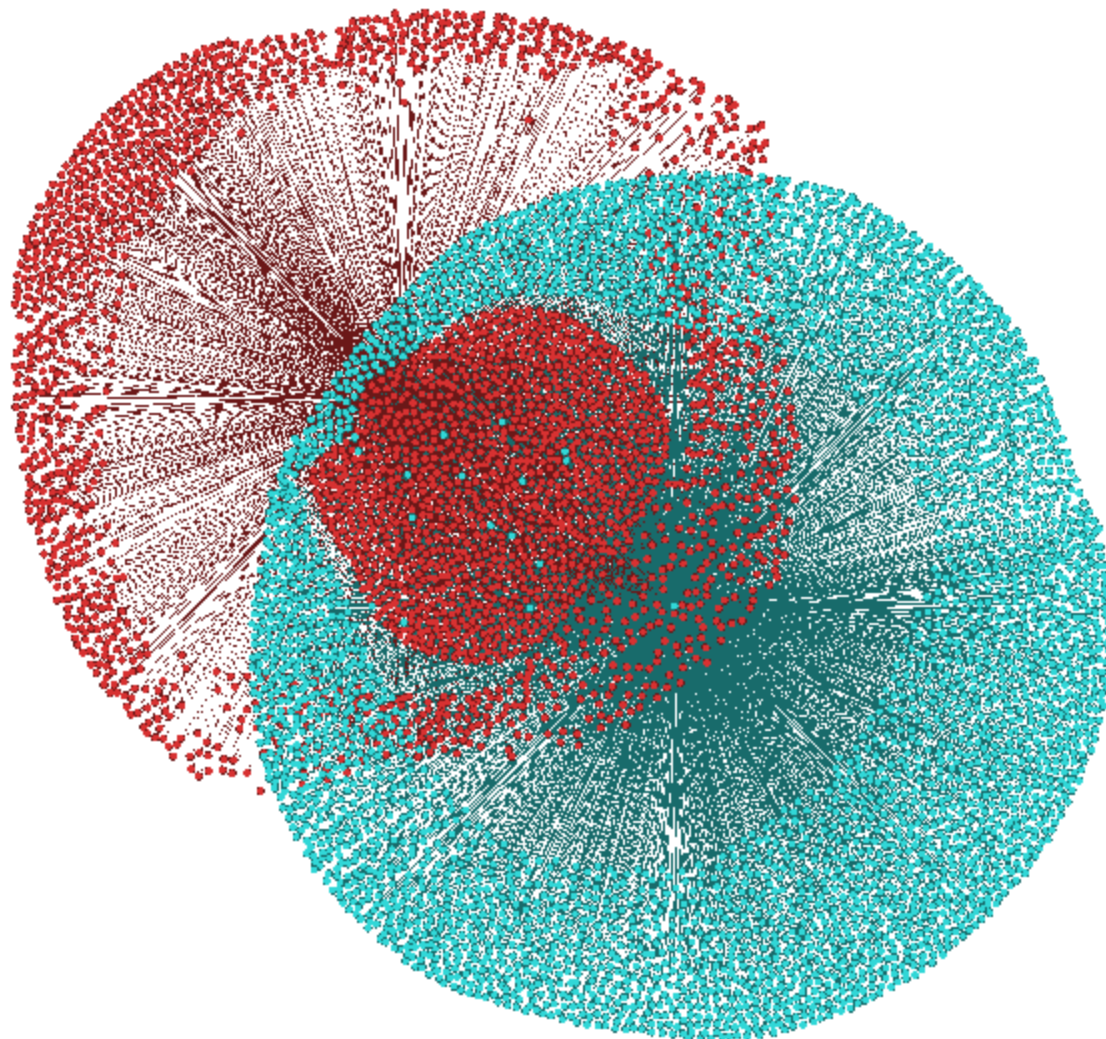
Anexo K – Rede com partição de acordo com Modularidade e Grau (1º Turno)



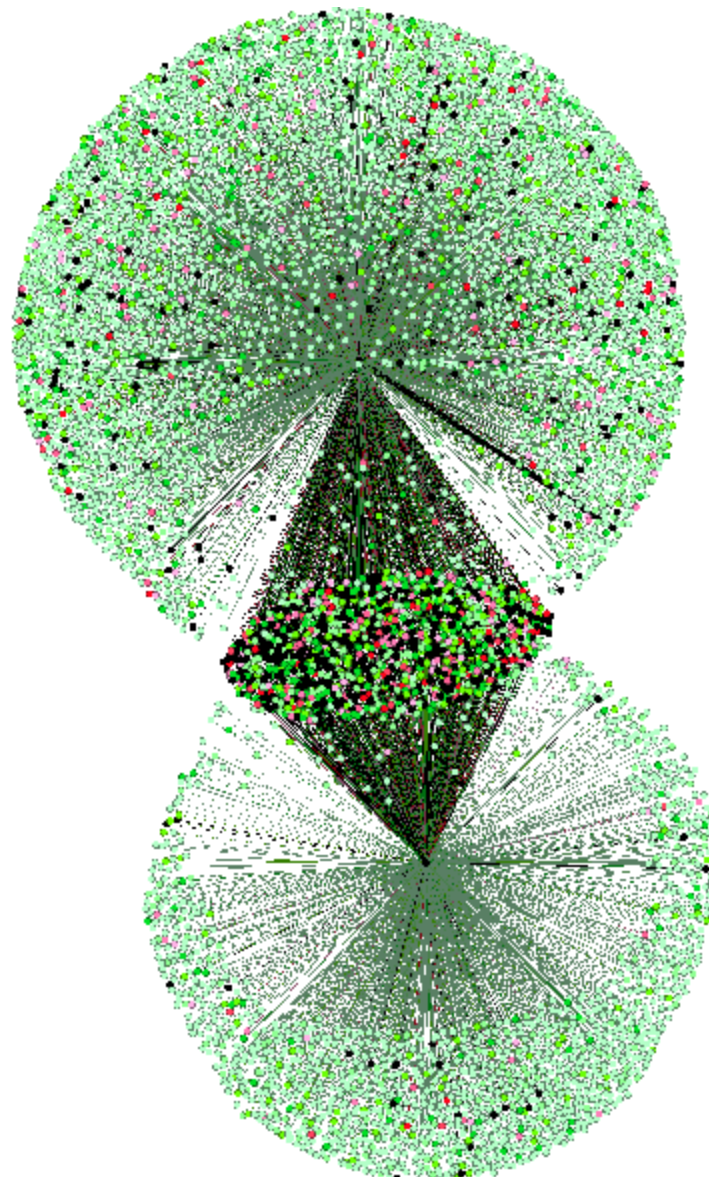
Anexo L – Rede com partição de acordo com Modularidade (2º Turno) – (1)



Anexo M – Rede com partição de acordo com Modularidade (2º Turno) – (2)



Anexo N – Rede com partição de acordo com Modularidade e Grau ponderado (2º Turno)



Anexo O – Rede com partição de acordo com Modularidade e Grau (2º Turno)

