

UNIVERSIDADE DE BRASÍLIA

Faculdade de Ciência da Informação

Programa de Pós-graduação em Ciência da Informação

Rafael Henrique Santos Soares

Modelo de análise da comunicação e mediação da informação em organizações públicas brasileiras através da Análise de Redes Sociais.

**Orientador:** Prof. Dr. Jorge Henrique Cabral Fernandes

Brasília – DF

2013

**Resumo**

Investiga a viabilidade da utilização de fontes de dados abertas para o mapeamento de redes sociais para estudo da comunicação e mediação da informação entre as organizações públicas brasileiras. A maioria as decisões das organizações públicas somente produzem efeito após sua publicação, em especial as veiculadas no Diário Oficial da União. O fluxo de informações estabelecido, portanto, entre os órgãos da administração pública e aquele veículo revela uma rede social que sustenta essa comunicação pode ser mapeada por meio da adequada extração e modelagem desses dados. A pesquisa apresenta uma proposta de metodologia para coleta de dados usando como fonte principal as publicações veiculadas no Diário Oficial da União e, à luz das teorias da Ciência da Informação, discute a modelagem da rede e as métricas mais adequadas para a condução de estudos sobre comunicação e mediação da informação usando essa base empírica.

**Palavras chave**: Análise de redes sociais. Mineração de dados abertos. Comunicação e mediação da informação.

Índice de Figuras

Figura 1 - Hierarquia de sistemas e subsistemas (FERNANDES 2012) 22

Figura 2- Classificação dos canais de comunicação para negócios (COSTA 2013) 28

Figura 3 – Representação da comunicação formal e informal entre organizações públicas 31

Figura 4 – Imagem oferecida por Dance para representação helicoidal da comunicação humana (DANCE 1967. p.374) 34

Figura 5 – Modelo de comunicação entre organizações públicas 42

Figura 6 – Representação gráfica e matricial de um grafo (FEOFILOFF; KOHAYAKAWA; WAKABAYASHI, 2004) 46

Figura 7 - Exemplos dos papeis de intermediação propostos por Nooy, Mvar e Batagelj (2005, p.151) 54

Figura 8 – modelo revisado do two-step-flow (Robinson 1976) 56

Figura 9 – Difusão acumulada a partir de um vértice central e um marginal (NOOY, MRVAR E BATAGELJ, 2005, p. 165) 58

Figura 10 – Categorização dos indivíduos com base no tempo de adoção da inovação. ROGERS (1995, p. 247) 60

Figura 11 - IDE - Um ambiente de desenvolvimento integrado - do GATE. 65

Figura 12 – Interface principal do Pajek 67

Figura 13 – Tela de visualização do sociograma no Pajek 69

Figura 14 – Ilustra o processo de análise exploratória de redes sociais 73

Figura 15 - Interação entre os módulos desenvolvidos 75

Figura 16 – Trecho da lista de identificação de órgãos do Gazetteer 77

Figura 17 – Trecho da lista de identificação de autoridades representantes de órgãos do Gazetteer 78

Figura 18 – Destaque da palavra-chave da natureza da portaria. 80

Figura 19 - Reprodução de trecho do DOU com identificação dos elementos da rede 83

Figura 20 – Diagrama de entidade e relacionamento. 86

Figura 21 – Visualização dos componentes fortes da rede manipulada. 92

Figura 22 – Ilustração da detecção dos cores da rede 94

Figura 23 - Destaque da quantidade de cliques que cada vértice participa 96

Figura 24 – Representação do prestigio de distancia na rede de ministérios 99

Figura 25 – Resultado da analise triádica do pajek 100

Figura 26 - Clusters identificados por meio do processo de Ranking 102

Figura 27 - Difusão acumulada da adesão ao SiSU na rede de instituições de ensino superior. 109

Figura 28 – Rede de Universidades com linhas com mais de 20 ocorrências 110

Figura 29 – Rede de ego da Presidente Dilma Rousseff 115

Figura 30 – Rede de ego com representação dos vetores de tempo no cargo e proximidade com a Presidente. 122

Lista de Tabelas

[Tabela 1 – Adaptação do quadro apresentado por LeCoadic (2004, p36). Interseções dos elementos da comunicação e o processo de publicação no DOU. 29](#_Toc378621486)

[Tabela 2 – Métricas dos ministérios e o orçamento 2014 105](#_Toc378621487)

[Tabela 3 – Informações básicas da rede inicial de universidades 107](#_Toc378621488)

[Tabela 4 – Distribuição das linhas na rede original. 112](#_Toc378621489)

[Tabela 5 – Métricas relacionadas a mediação na rede de ego 118](#_Toc378621490)

[Tabela 6 – Detalhes dos indivíduos da rede de Ego da Presidente Dilma Rousseff 120](#_Toc378621491)

[Tabela 7 – Correlação entre tempo de cargo e proximidade com a Presidente 124](#_Toc378621492)

Índice de Listagens

[Listagem 1 – Regra do Tokeniser para identificação de inicio de endereço de sitio da internet. 67](#_Toc378621511)

[Listagem 2 – Regra para identificação de nomes de pessoas 80](#_Toc378621512)

[Listagem 3 – Regras do Tokeniser para identificação de inicio de portaria 81](#_Toc378621513)

[Listagem 4 – Pseudo-codigo para identificação dos relacionamentos 85](#_Toc378621514)

[Listagem 5 – Palavras chave usadas para filtro das portarias no banco de dados 89](#_Toc378621515)

[Listagem 6 – comandos do Pajek para manipulação da rede 92](#_Toc378621516)

[Listagem 7 – comandos do Pajek para extração de métricas estruturais 93](#_Toc378621517)

[Listagem 8 – Comandos para extração de métricas de prestigio. 98](#_Toc378621518)

[Listagem 9 – comandos do Pajek para manipulação da rede 109](#_Toc378621519)

Sumário

[1. Introdução 9](#_Toc378621531)

[1.1. Justificativa 10](#_Toc378621532)

[1.1.1. Implicações no campo da segurança da informação 14](#_Toc378621533)

[1.2. Pergunta da pesquisa 15](#_Toc378621534)

[1.3. Objetivo geral 16](#_Toc378621535)

[1.3.1. Objetivos específicos 16](#_Toc378621536)

[1.4. Hipótese 16](#_Toc378621537)

[1.5. Organização do trabalho 17](#_Toc378621538)

[2. Referencial teórico 18](#_Toc378621539)

[2.1. Comunicação da informação 18](#_Toc378621540)

[2.2. Mediação da informação 20](#_Toc378621541)

[2.3. A Comunicação da informação Organizacional 22](#_Toc378621542)

[2.3.1. Abordagem sistêmica no estudo das organizações 23](#_Toc378621543)

[2.3.2. Tipos de comunicação entre as organizações 25](#_Toc378621544)

[2.3.3. Fluxos e canais de comunicação organizacional 27](#_Toc378621545)

[2.4. Os processos da comunicação de Le Coadic 30](#_Toc378621546)

[2.5. Modelo de comunicação helicoidal de Dance 34](#_Toc378621547)

[2.6. Princípio da Publicidade e manifestações da intersetorialidade das organizações públicas 37](#_Toc378621548)

[2.6.1. Outras fontes de dados abertos sobre a atuação das organizações públicas brasileiras 40](#_Toc378621549)

[2.7. Modelo de comunicação entre organizações públicas 42](#_Toc378621550)

[2.8. Mineração de dados abertos para o mapeamento de redes sociais 45](#_Toc378621551)

[2.9. Análise de Redes sociais 46](#_Toc378621552)

[2.9.1. Redes de um modo e dois modos 49](#_Toc378621553)

[2.9.2. Métricas de coesão 50](#_Toc378621554)

[2.9.3. Redes de Ego e suporte social 52](#_Toc378621555)

[2.9.4. Métricas de mediação 53](#_Toc378621556)

[2.9.4.1. Métricas de centralidade 53](#_Toc378621557)

[2.9.4.2. Papeis de Intermediação e pontes 55](#_Toc378621558)

[2.9.5. Métricas para estudo da difusão em redes sociais 57](#_Toc378621559)

[2.10. Elementos de Tecnologia da Informação 62](#_Toc378621560)

[2.10.1. Sistema de gerenciamento de banco de dados 63](#_Toc378621561)

[2.10.2. Linguagens de programação 63](#_Toc378621562)

[2.10.2.1. Java 64](#_Toc378621563)

[2.10.2.2. SQL – Structured Query Language 65](#_Toc378621564)

[2.10.3. Interface de Programação de Aplicações 65](#_Toc378621565)

[2.10.4. GATE - General Architecture For Text Engeneering 66](#_Toc378621566)

[2.10.5. Pajek – Programa para Análise de Grandes redes 69](#_Toc378621567)

[3. Metodologia 71](#_Toc378621568)

[3.1.1. Pesquisa descritiva e análise exploratória 72](#_Toc378621569)

[3.1.2. Coleta de dados - Automação do processamento do DOU 76](#_Toc378621570)

[3.1.2.1. Recuperação e armazenamento dos arquivos originais 77](#_Toc378621571)

[3.1.2.2. Identificação de elementos chave no texto 78](#_Toc378621572)

[3.1.3. Modelagem das redes sociais 83](#_Toc378621573)

[3.1.4. Extração das redes sociais a partir de consultas ao banco de dados 87](#_Toc378621574)

[4. Métodos para Investigação em Redes sociais mapeadas a partir de informações de fontes de dados abertas 90](#_Toc378621575)

[4.1. Fenômenos coesivos e grupos de destaque 91](#_Toc378621576)

[4.1.1. Definição e Manipulação da rede 92](#_Toc378621577)

[4.1.2. Métricas estruturais e inspeção visual 93](#_Toc378621578)

[4.1.3. Interpretações e análises complementares 105](#_Toc378621579)

[4.2. Fenômenos de difusão 108](#_Toc378621580)

[4.2.1. Definição e Manipulação da rede 109](#_Toc378621581)

[4.2.2. Métricas estruturais e Inspeção visual 110](#_Toc378621582)

[4.3. Fenômenos de mediação 113](#_Toc378621583)

[4.3.1. Definição e manipulação da rede 113](#_Toc378621584)

[4.3.2. Métricas estruturais e inspeção visual 115](#_Toc378621585)

[4.3.3. Interpretações e análises complementares 118](#_Toc378621586)

[4.3.3.1. Fluxo da informação na vizinhança da Presidente 118](#_Toc378621587)

[4.3.3.2. Proximidade estrutural com a presidente e tempo de permanência em cargo comissionado 121](#_Toc378621588)

[5. Discussão dos resultados 126](#_Toc378621589)

[5.1. Mineração de texto para mapeamento de redes sociais 126](#_Toc378621590)

[5.2. Coesão em redes sociais mapeadas a partir do DOU 129](#_Toc378621591)

[5.3. Difusão em redes sociais mapeadas a partir do DOU 131](#_Toc378621592)

[5.4. Mediação em redes sociais mapeadas a partir do DOU 133](#_Toc378621593)

[6. Conclusões e comentários finais 136](#_Toc378621594)

[Referencias 137](#_Toc378621595)