

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**

O presente trabalho investiga a viabilidade da utilização de fontes de dados abertos para o mapeamento de redes sociais para estudo da comunicação e mediação da informação entre as organizações públicas brasileiras. Estando a eficácia das decisões das organizações públicas vinculada às suas publicações, o fluxo de informações estabelecido entre os órgãos da administração pública visando coordenação e cooperação se manifesta nessas publicações e assim, a rede social que sustenta essa comunicação pode ser mapeada por meio da adequada extração e modelagem desses dados. Apresenta-se uma proposta de metodologia para coleta de dados usando como fonte principal as publicações veiculadas no Diário Oficial da União e discute-se, à luz das teorias da Ciência da Informação, 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.

Sumário

[1. Introdução 5](#_Toc365923992)

[1.1. Justificativa 6](#_Toc365923993)

[1.2. Pergunta da pesquisa 8](#_Toc365923994)

[1.3. Objetivo geral 9](#_Toc365923995)

[1.3.1. Objetivos específicos 9](#_Toc365923996)

[1.4. Hipótese 9](#_Toc365923997)

[2. Comunicação da informação 11](#_Toc365923998)

[2.1. Comunicação e mediação da informação 12](#_Toc365923999)

[2.2. A Comunicação da informação para negócios 14](#_Toc365924000)

[2.3. Os processos da comunicação de Le Coadic 14](#_Toc365924001)

[2.4. Modelo de comunicação helicoidal de Dance 17](#_Toc365924002)

[2.5. Princípio da Publicidade, Imprensa oficial e manifestações da comunicação entre organizações públicas 17](#_Toc365924003)

[2.6. Modelo de comunicação entre organizações públicas 17](#_Toc365924004)

[3. Extração de redes sociais a partir de dados abertos 18](#_Toc365924005)

[3.1. Mineração de dados abertos para construção de redes sociais 18](#_Toc365924006)

[3.2. Análise Co-Word 19](#_Toc365924007)

[3.3. Diversidade lexical e carga informacional das portarias 19](#_Toc365924008)

[4. Base Empírica, Ferramentas e Metodologia 19](#_Toc365924009)

[4.1. O Diário Oficial da União 19](#_Toc365924010)

[4.2. Outras fontes de dados abertos sobre organizações públicas brasileiras 21](#_Toc365924011)

[4.3. Extração das redes 23](#_Toc365924012)

[4.4. Análise dos dados 28](#_Toc365924013)

[5. Análise de Redes sociais 30](#_Toc365924014)

[5.1. Métricas de coesão 31](#_Toc365924015)

[5.2. Métricas de mediação 32](#_Toc365924016)

[6. Resultados e análises preliminares 33](#_Toc365924017)

[6.1. Visão Geral 33](#_Toc365924018)

[6.2. Redes por Ministério 35](#_Toc365924019)

[6.3. Redes temáticas 38](#_Toc365924020)

[6.4. Difusão 39](#_Toc365924021)

[2. Riscos e restrições 42](#_Toc365924022)

[3. Cronograma proposto 44](#_Toc365924023)

[Referencias 45](#_Toc365924024)

## Introdução

Por força do princípio da publicidade, consagrado na Constituição Federal de 1988 (C.F.88), em seu artigo 37, as atividades desempenhadas pelos organismos públicos e seus agentes devem ser publicadas. Essa publicação “abrange toda a atuação estatal, não só sob o aspecto da divulgação oficial de seus atos, como também de propiciação de conhecimento da conduta interna de seus agentes” (MEIRELLES 2013).

Como ensina MORAES (2002), "a publicidade faz-se pela inserção do ato no Diário Oficial (...) para conhecimento do público em geral e início da produção de seus efeitos”, e, portanto, podemos assumir que a atuação dos órgãos (ou, pelo menos, os resultados dessa atuação) é projetada em publicações no Diário Oficial da União, caso contrário não produziria efeitos.

Assim, todas as decisões da administração publica brasileira, desde a promulgação de leis até a aquisição de insumos militares, passando por contratação de pessoal e destinação de recursos, e os atos necessários para coordenar os diversos órgãos visando cumprir decisões, implementar projetos e executar estratégias, estão disponíveis para acesso livre e irrestrito não somente por cidadãos brasileiros, como assegura a C.F.88, mas por qualquer individuo ou entidade no planeta que esteja conectado à internet, já que o D.O.U. é publicado eletronicamente além da edição impressa vendida em bancas de jornal.

Toda essa massa de dados abertos e de acesso facilitado pode conter indicações sobre o funcionamento interno da administração publica brasileira, especialmente no que diz respeito à comunicação e mediação da informação entre as entidades que a compõem.

O trabalho investiga comunicação e mediação da informação em organizações públicas brasileiras por meio da análise do conteúdo das publicações do Diário Oficial da União (D.O.U.) visando identificar indicadores dos padrões de fluxo de informação entre as organizações por meio de redes sociais mapeadas com base nas portarias veiculadas nesse meio de comunicação oficial.

Uma vez que se extraia de fontes de dados abertos as redes sociais formadas pelos órgãos públicos brasileiros, pode ser possível visualizar estruturas de funcionamento que extrapolam a hierarquia formal, relações de cooperação, grupos coesos, entidades chave no processo de mediação da informação, padrões de difusão da informação entre outras características importantes da administração pública, úteis para a implementação de políticas e para a identificação de vulnerabilidades.

### Justificativa

A questão do principio da publicidade, discutida anteriormente, por si só já seria o suficiente para entendermos que a publicidade dos atos administrativos é uma realidade irretratável. A isso se soma o fato de que se desenvolve nos últimos anos um forte movimento de abertura dos dados governamentais em todo o mundo. A organização *Open Government Initiative Group (OGIG)* (CIVIC COMMONS, 2013) organiza uma lista extensa de países cujos governos oferecem fontes de dados sobre suas atividades. Importantes economias como os Estados Unidos, o Japão e Inglaterra, além de países com histórico político/ideológico particulares como a Rússia e a China já disponibilizam fontes de dados que contam, inclusive, com facilidades de integração de sistemas, para maior facilidade no processamento das informações. O Brasil também faz parte desse movimento e, além do Diário Oficial da União, Mantem um portal de dados abertos hospedado em *dados.gov.br*.

A descoberta de informações por meio da mineração de dados abertos pode trazer benefícios para a sociedade e, inclusive, é encorajada pelos portais governamentais que disponibilizam esses dados. Em quase todos os portais divulgados pela OGIG há pontos de acesso para a integração de softwares clientes para tratamento desses dados. Muitos disponibilizam um catálogo com os aplicativos desenvolvidos pelos cidadãos com suas abordagens e recortes dos dados, que vão desde o perfil de votação de parlamentares até destinação de recursos orçamentários.

Contudo, a quantidade de dados exposta é grande o suficiente para que deles possam ser inferidas informações que exponham vulnerabilidades do país que as disponibiliza. A recentemente promulgada Lei de acesso à informação determina a “observância da publicidade como preceito geral e do sigilo como exceção”, o que, na prática, torna pública toda informação não classificada como reservada, secreta ou ultrassecreta.

No caso de informações classificadas, a Lei de acesso à informação determina, em seu artigo 23, inciso V, que serão classificadas “informações cuja divulgação ou acesso irrestrito possam: prejudicar ou causar risco a planos ou operações estratégicos das Forças Armadas”. Entretanto, as compras do exercito são publicadas no D.O.U. e disponibilizadas no portal de dados nacional de forma aberta e irrestrita. Uma análise dos insumos e valores associados às transações das forças armadas poderiam dar fortes indícios do tipo de operação em curso e, portanto, comprometê-las. Eis um exemplo de como a análise de dados abertos pode revelar mais do que inicialmente se pretendia comunicar.

Fernandes (2012) discute a diferença entre segurança e defesa. O autor argumenta que a segurança é composta de controles arranjados de forma hierárquica que tentam prover a garantia de funcionamento de um sistema. Defesa, por sua vez, é um conceito mais abrangente, “uma ação necessária e essencialmente visceral (de dentro para fora), que deve ser desenvolvida por todo indivíduo ou ser” (FERNANDES 2012). Sob esse aspecto, podemos assumir que, embora as classificações previstas pela Lei de Acesso à Informação possam ser suficientes para garantir a segurança das informações, um mecanismo de defesa adequado deverá necessariamente ser implementado com uma postura mais holística e abrangente, que considere aspectos menos explícitos do sistema em questão.

Nesse sentido, o estudo desses dados abertos por meio da análise de redes sociais pode contribuir para a identificação de padrões de comunicação e operação diferentes do previsto e intuitivamente esperado por aquele que analisa apenas as relações hierárquicas da administração pública. Cross e Parker (2004) apresentam em seu trabalho diversos casos em que as redes sociais identificadas são bastante distintas das estruturas formais em organizações.

“Uma visão calibrada da real rede social contribui para a tomada de decisão gerencial e fomenta os esforços em promover a colaboração efetiva. Ao invés de negligenciar o funcionamento interno da rede, executivos podem potencializar os insights da análise de redes sociais para identificar desconexões críticas ou rigidez na rede e aprimorar a capacidade de resposta de forma profunda.” Cross e Parker (2004, p7, tradução do autor)

As publicações de ações conjuntas entre os órgãos são indícios de que as organizações públicas se coordenam para atingir os objetivos propostos, e essa coordenação, e consequente fluxo de informação, resulta em um arranjo em uma rede de comunicação (informal e posteriormente formal, como exposto acima) que reflete mais as necessidades de cooperação do que relações hierárquicas.

“Muitas redes se iniciam a partir de situações de mobilização mais amplas. Criam-se, nas redes, formas institucionais próprias associadas aos direitos, responsabilidades e tomadas de decisão.” (MARTELETO, 2004)

Ainda nesse sentido, Sugahara e Vergueiro (2010) consideram que o estudo do fluxo da comunicação numa rede pode trazer novos elementos para a análise do funcionamento do sistema estudado.

“O estudo sobre o “caminho” percorrido pela informação na rede social e o modo como ocorrem os fluxos de informação pode resultar em novas reflexões conceituais em relação à estrutura deste fluxo e aos vínculos relacionais que se configuram e reconfiguram na rede. (Sugahara e Vergueiro, 2010)

Por fim, observa-se que já existem experiências bem sucedidas na aplicação dos métodos da Análise de Redes sociais no estudo do funcionamento das organizações públicas com resultados que apontam para o descobrimento de características internas do Estado que se mostram úteis, identificando oportunidades e ameaças:

A análise de redes sociais é um campo de estudo amplo e recente, embora seja pouco desenvolvido no Brasil até o momento. Os raros estudos existentes, entretanto, têm permitido explorar de forma analiticamente inovadora certos fenômenos, em especial as dinâmicas internas ao Estado e as relações entre esse último e a sociedade mais ampla, no interior de comunidades de política pública específicas. Os resultados encontrados sugerem que os padrões de relação estruturados pelas redes influenciam os resultados da dinâmica política, assim como das políticas públicas. (...) Essa estratégia analítica parte da ideia de que **é importante conhecer detalhadamente os padrões de organização no interior do Estado** e a sua inserção no ambiente político e econômico mais amplo em que está inserido, pois esses elementos explicam uma parte importante da coesão do Estado, **assim como da sua fragilidade.** (Marques 2004)

As conclusões de Marques (2004) confirmam a tese de que as organizações publicas formam uma rede cuja configuração é distinta da estrutura formal e que o estudo dessa estrutura de comunicação entre os elementos da rede é importante e pode contribuir não somente barra o bom funcionamento da administração mas para a defesa do estado.

### Pergunta da pesquisa

Como analisar a comunicação da informação entre organizações públicas por meio da análise das publicações oficiais?

### Objetivo geral

Demonstrar a viabilidade de se analisar a rede de comunicação da informação entre as organizações publica brasileiras por meio da extração de informações disponíveis em fontes de dados abertas.

### Objetivos específicos

Demonstrar, por meio de teorias da ciência da informação, a correspondência entre as publicações do Diário Oficial da União e a comunicação e mediação da informação entre organizações públicas brasileiras.

Demonstrar a viabilidade de se extrair informações de relações entre organizações brasileiras disponíveis em fontes de dados oficiais abertos de forma automatizada.

Identificar estratégias para a identificação de relacionamentos entre organizações públicas projetadas em publicações no Diário Oficial da União.

Demonstrar que redes extraídas do DOU podem ajudar a identificar os papeis desempenhados por cada organização e as bases do funcionamento da rede de comunicação da informação usando como base as teorias de comunicação da informação e, principalmente, as ferramentas e conceitos disponíveis na Análise de redes Sociais, especialmente os conceitos de ‘grupos coesos’, intermediação e difusão da informação.

### Hipótese

As interações entre organizações públicas, o fluxo de informação e recursos entre elas, ocorre de maneira informal no dia a dia desses órgãos no desempenho de suas missões institucionais. Essas interações subsidiam e influenciam os processo de tomada de decisão nessas organizações.

Para que surtam efeitos, entretanto, é preciso que essas decisões sejam publicadas, caso contrario não teriam nenhum significado prático. Assim, podemos esperar que todas as decisões publicadas tenham sido precedidas de trocas de informação (conforme exposto nas seções anteriores) e que, no mesmo sentido, trocas de informação que não tenham gerado decisões, não geraram publicações e, portanto, não são relevantes.

Assim, tem-se nas publicações evidencias de fluxos de informação entre os órgãos públicos, e esses fluxos de informação podem ser estudados por meio das ferramentas de Análise de Redes Sociais.

Entretanto, conforme observam Karthika e Bose (2011), “Uma limitação da Análise de redes sociais é que ela não pode ser considerada como uma técnica apropriada de data mining, pois ela é capaz de descobrir padrões a partir de estruturas conhecidas, e não de estruturas desconhecidas”.

Isso posto, os resultados vislumbrados por essa pesquisa podem contribuir no sentido de verificar a viabilidade de se conhecer as estruturas da comunicação e mediação da informação na administração pública e daí extrair redes sociais baseadas em publicações oficiais.

Se as publicações oficiais indicam implicitamente as relações, hierárquicas ou não, entre as organizações publicas brasileiras, então é possível extrair dessas relações uma rede a qual pode ser estudada utilizando-se as ferramentas da Análise de Redes Sociais.

### Organização do trabalho

Este trabalho esta organizado em seções da seguinte maneira:

Seção 1 – Introdução – Apresenta o problema e seu contexto, justificativa para o estudo, hipóteses, objetivos e a organização do trabalho.

Seção 2 – Comunicação da Informação – Discute os principais conceitos e teorias relacionados à comunicação da Informação usados no trabalho. Apresenta conceitos básicos de comunicação e mediação, destacadamente os que tratam da comunicação formal e informal nas organizações, tais como a comunicação para negócios apresentada por Costa (2006) e os processos de comunicação introduzidos por Le Coadic (2004). Apresenta ainda o modelo de comunicação helicoidal de Dance (1967) e discute a publicidade dos atos da administração pública e sua relação com a comunicação entre organizações. Apoiado nesse arcabouço teórico esboça um modelo de comunicação entre organizações públicas brasileiras.

Seção 3 – Identificação de redes sociais a partir de fontes de dados abertas – Expõe alguns trabalhos da autores que mostraram a viabilidade de se extrair redes sociais da mineração de dados abertos, apresenta a técnica da análise co-word e o conceito de diversidade lexical, técnicas que, combinadas com a mineração dos dados do Diário Oficial da União, permitiram a construção das redes sociais das organizações públicas brasileiras.

Seção 4 – Base Empírica, Ferramentas e Metodologia - Detalha as características de algumas fontes de dados abertas disponibilizadas pela administração pública brasileira e justifica a escolha do Diário Oficial da União para o desenvolvimento deste trabalho. Apresenta as soluções de software desenvolvidas para a coleta e processamento dos dados e as ferramentas usadas para manipulação e análise das redes extraídas. Expõe detalhes da utilização da análise co-word para definição de relacionamentos entre atores, da diversidade lexical para atribuição de peso às relações, dos macrodesafios do PPA 2012 para definição de palavras-chave para a filtragem das relações na construção das redes temáticas e outras definições acerca do processo de mapeamento da rede.

Seção 5 – Análise de Rede Sociais – Apresenta os principais conceitos e métricas oferecidas pela Análise de redes Sociais que foram usados na análise da comunicação entre as organizações públicas brasileiras neste trabalho.

Seção 6 – Análise dos dados e discussão dos resultados – Discute os resultados das aplicações das técnicas da Análise de Redes Sociais diversas redes extraídas.

Seção 7 – Apresenta conclusões do estudo frente a seus objetivos, as contribuições da pesquisa e sugestões para trabalhos futuros.

## Comunicação da informação

As organizações apresentam uma diversidade de constituição, finalidade e modo de atuação, mas, ao mesmo tempo são submetidas a forças de nivelamento e homogeneização do comportamento. Daí surge a necessidade de comunicação da informação.

Segundo Lasswell, o processo de comunicação cumpre, entre outras, a função de estabelecer relações entre os componentes da sociedade para produzir uma resposta ao meio. (Lasswell, 1948). Essa ideia pode ser extrapolada para as comunicações entre organizações públicas que precisam produzir resultados para a sociedade brasileira. Assim, é de se esperar que sejam identificados os fluxos de informação entre os órgãos, caso contrario, segundo Lasswell, seriam evidenciadas disfunções no sistema e, finalmente, crises.

A teoria difusionista da comunicação ajuda a explicar fenômenos como a difusão de informação na rede, fenômeno que supostamente ocorre entre organizações publicas no Brasil. Lazarsfeld, em 1944, estabelece que o fluxo de comunicação acontece em duas etapas, na qual o papel do “líder de opinião” se revela. Na teoria do two-step-flow, as pessoas bem informadas e diretamente expostas à fonte de informação, absorvem o conteúdo e, num segundo momento, transferem para os demais. (Lazarsfeld, 1944)

No estudo da comunicação da informação entre organizações publicas, entretanto, é preciso considerar a opinião contestatória de Beltran, que ressaltam a importância de se considerar as hierarquias rígidas e as relações de força no interior das sociedades (Beltran 1976 apud MATTELART 2011). Embora a informação flua dentro da rede, a relação de hierarquia estabelecida entre os órgãos públicos certamente desempenha papel importante na difusão de ideias, práticas e informação.

A cibernética fornece conceitos importantes e úteis ao presente estudo. Dentre eles, o da “ecologia da comunicação” de Moles (1975 apud MATTELART 2011), inspirado tanto na teoria matemática da comunicação (Shannon 1945) quanto nas ideias de Wiener, principal referencia em cibernética. “A comunicação é definida como a ação de fazer participar um sistema situado em dado ponto R das experiências e estímulos do meio de um outro sistema situado em outro local e em outro tempo, utilizando os elementos de conhecimento que possuem em comum. A ecologia da comunicação é a ciência da interação entre espécies diferentes no interior de um dado campo” (Moles 1975 apud MATTELART 2011). Wiener também chama a atenção para o conceito de entropia, tendência da natureza de destruir o ordenado, e postula que a informação e seu fluxo em redes são ferramentas que ajudam a retarda-la. “A soma de informação em um sistema é a medida de seu grau de organização. A entropia é a medida de seu grau de desorganização.” (Wiener 1948).

A característica da comunicação que se pretende estudar nesse projeto é de uma comunicação fortemente influenciada por regras e procedimentos cuja intenção primária pode não ser especificamente a comunicação entre os órgãos envolvidos. Entretanto, tais evidencias não podem, apenas por isso, serem desconsideradas. Watzalawick argumenta que nem toda comunicação é consciente ou voluntária e sugere que a comunicação pode acontecer também sem que se tenha esse objetivo. (Watzalawick 1977).

O trabalho de Niklas Luhmann nos permite considerar a rede de organizações publicas brasileira como um “sistema autopoiético”, que em suas palavras é: “Um sistema operacionalmente fechado e estruturalmente acoplado ao meio”. “Tais sistemas regulam as relações sociais como variações e circulação de sentido. Existem perturbações da comunicação que se devem à rigidez no funcionamento dos meios de comunicação e à resistência à mudança. (Luhmann 1971 apud MATTELART 2011). Parece razoável esperar que se encontre esse tipo de rigidez na comunicação entre organizações públicas, uma vez que suas atividades são regidas por um corpo denso de leis e regulamentos.

### Comunicação e mediação da informação

Borko (1967), em uma das primeiras tentativas de se definir a Ciência da Informação expôs que essa ciência é:

“(...) a disciplina que investiga as propriedades e o comportamento da informação, as forças que governam o seu fluxo, e os meios para processamento visando otimização da acessibilidade e usabilidade.”. (BORKO 1967)

O presente trabalho apresenta forte sintonia com esse conceito, ao discutir a relação entre publicações e a troca de informação entre organizações (comportamento), a configuração em rede das trocas de informação entre as organizações (forças e fluxos) e a viabilidade de identificação de redes a partir de dados abertos (meios para processamento).

Posteriormente, Saracevic (1996) discorre sobre a comunicação como um dentre quatro campos que mantêm relações interdisciplinares e ajudam a moldar a Ciência da Informação. Esse autor defende que os estudos da comunicação em CI devem compreender a informação como um fenômeno e a comunicação como um processo e, dessa forma, estuda-los em conjunto. Esse mesmo autor sugere a definição de uma ecologia informacional, que teria origem na evolução da comunicação do conhecimento.

“Ecologia informacional envolve: os produtores de conhecimento (autores, inventores, pesquisadores, coletores...) e as instituições onde trabalham ou residem; os financiadores dessas instituições e trabalhos; os editores (em qualquer media), incluindo seus próprios mecanismos de seleção, editoração, julgamento, avaliação, publicação...; os canais de divulgação; os reelaboradores (por exemplo, produtores de bases de dados, em qualquer media também) incluindo seus mecanismos próprios de seleção, tratamento, disseminação...; as bibliotecas e serviços de informação, também com seus mecanismos; os usuários e suas instituições fechando a cadeia ecológica. (SARACEVIC 1996)

No caso da presente pesquisa a ecologia informacional é identificada e se manifesta na base empírica utilizada na figura nas pessoas identificadas nas portarias, instituições públicas a elas associadas, o Diário oficial da União como canal de publicação, a disseminação da informação através da análise da difusão promovida pela Análise de redes sociais, e os próprios órgãos públicos e ministérios agindo como produtores e consumidores de informação dentro da “cadeia ecológica”, ou rede social.

Ampliando a noção de que a ciência da informação tem uma abrangência consideravelmente grande, Zins (2007) propõe o mapa do conhecimento da Ciência da Informação. Em seu trabalho o autor identifica dez categorias básicas de elementos que compõem a Ciência da informação. Ao abordar a mediação da informação, Zins(2007) define os elementos que atuam entre as fontes de informação e os usuários, destinatários da informação: Trabalhadores do Conhecimento, Aplicações, Operações e Processos, Tecnologias, Ambientes, Organizações. Esses elementos são a base da comunicação da informação e estão presentes na comunicação entre organizações públicas na figura dos servidores e agentes executando suas atividades diárias sendo regidos pelas leis e regulamentos, tomando decisões e publicando-as.

Autores fazem distinção entre mediação e intermediação. Mediação diz respeito a um processo que conta com a intervenção humana para aproximar um sujeito de uma informação. Na mediação o sujeito mediador acaba por interferir no processo de acesso do usuário à informação. O intermediário, por sua vez, é um ator que tem um papel mais simples, fazendo as vezes de um conector entre a informação e o usuário, desempenhando uma tarefa que poderia ocorrer sem a intervenção humana. Ambos são exemplos de interações. (SOUTO, 2008). Em alguns casos, como o de cooperação técnica, temos uma relação de mediação entre os órgãos. Interações como transferência de recursos, parecem estar melhor enquadradas como intermediação.

### A Comunicação da informação para negócios

### Os processos da comunicação de Le Coadic

Ao discutir a comunicação da informação LeCoadic (2004 p. 35) argumenta que o processo de comunicação pode ser desmembrado em dois outros: um processo escrito, formal, e um processo oral, informal. O autor ensina que o processo formal produz informações estáveis enquanto o processo informal entrega informações mais sujeitas a modificações e repetições. LeCoadic apresenta uma comparação entre os elementos formais e informais, estabelecendo, entre outros, que os formais apresentam: caráter predominantemente público, armazenamento permanente, relativa obsolescência e ausência de interação direta. Os elementos informais, por sua vez, se apresentam do lado oposto, com caráter predominantemente não público, efemeridade, predominantemente atual e interação direta. Outro aspecto importante a se considerar é a cronologia dos processos de comunicação. O autor defende que os processos informais antecedem os processos formais.

Analogamente à teoria apresentada por LeCoadic, podemos entender o processo de comunicação da informação que acontece nas organizações públicas brasileiras como um processo que também apresenta um componente formal e um componente informal. As decisões são tomadas internamente após análises conjunturais, discussões, estudos, e diversos outros processos suplementares que, dependendo do assunto a ser tratado, envolvem outros órgãos que compõem a administração. Algumas decisões de colaboração e de criação de grupos de trabalho ou implementação de projetos são tomadas após meses ou anos de negociações e intercâmbios informais de informação entre as organizações envolvidas.

Esse pode ser caracterizado como o processo informal da comunicação da informação entre órgãos públicos. Conforme exposto, para produzir efeitos, os atos e decisões da administração pública precisam, necessariamente, ser publicados no Diário Oficial da União. Tem-se, assim, o processo formal da comunicação, o processo complementar e cronologicamente posterior ao informal, como sugerido por LeCoadic.



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

A Tabela 1 mostra uma adaptação do quadro apresentado por LeCoadic (2004, p37) onde o autor apresenta e confronta os elementos formais e informais da comunicação. Na Tabela, inseriu-se duas colunas para oferecer exemplos de cada elemento proposto por LeCoadic no estudo da comunicação entre organizações públicas manifestado no Diário Oficial da Uniao.

A analogia feita entre os processos formal e informal da comunicação de LeCoadic e as negociações e posterior publicação no D.O.U. parece adequada sob os aspectos exemplificados acima, quais sejam, audiência, armazenamento, atualidade e interatividade.

Podemos perceber que uma publicação no Diário Oficial da União, processo de comunicação formal, representa a manifestação de um fenômeno de comunicação informal anterior que a provocou, e sem o qual não existiria.

Tavares (2011) expõe quatro categorias de classificação da informação de acordo com o objetivo, e, dentre elas, destaca-se a informação para negócios. Essa categoria, segundo a autora, “contempla tanto empresas privadas e indústrias como também **organizações públicas**, associações e o terceiro setor”. Na mesma direção da teoria de LeCoadic (2004), Tavares (2011) divide a informação para negócios em dois grupos: formais e informais.

“A primeira, em meio impresso, é constituída de correspondências oficiais e regulamentos editados, entre outros. A segunda compreende a comunicação verbal, que se realiza, por exemplo, por telefone e em conversas e reuniões. Com o advento da internet, a comunicação eletrônica encontra-se entre a comunicação formal e informal, pendendo para um lado ou outro de acordo com o objetivo da mensagem e o público ao qual se destina.” (TAVARES, 2011).



Figura 1 – Representação da comunicação formal e informal entre organizações públicas

Assim, os argumentos de Tavares (2011) vão ao encontro da hipótese desse trabalho, qual seja, as publicações no Diário Oficial da União são o componente formal de uma comunicação que acontece de maneira informal e prévia entre as organizações públicas. Essa comunicação informal que precede a formal acontece visando a cooperação entre os órgãos com a finalidade do atingimento de metas e o cumprimento de suas missões institucionais. O seguinte trecho do trabalho da autora confirma: “(...) a informação de negócios subsidia o processo decisório e denota a comunicação interna e externa das organizações relacionada aos seus objetivos.” (TAVARES, 2011).

A Figura 1 mostra como a comunicação entre organizações públicas pode ser modelada à luz da teoria proposta por LeCoadic (2004) e dos conceitos apresentados por Tavares (2011). Nela representa-se de forma didática e simplificada a comunicação interna e informal entre as organizações que produzem uma ou mais publicações no diário Oficial da União.

A partir dessa perspectiva, é possível estudar a comunicação e mediação da informação entre organizações públicas por meio da análise das publicações no Diário Oficial da União, já que as publicações são uma projeção da comunicação prévia ocorrida entre os órgãos envolvidos.

**Comunicação e mediação da informação entre organizações**

Costa e Pinheiro (2007), ao estudar o compartilhamento de informação e conhecimento em aglomerações produtivas argumenta que “A necessidade de cooperação tem deslocado a análise da empresa individual para as redes de relacionamentos que a envolve.”. Nesse trabalho os autores identificaram, usando ARS, alguns elementos na rede que funcionavam como peças fundamentais na mediação entre as empresas locais estudadas e o ambiente competitivo.

### Modelo de comunicação helicoidal de Dance

### Princípio da Publicidade, Imprensa oficial e manifestações da comunicação entre organizações públicas

### Modelo de comunicação entre organizações públicas

## Análise de Redes sociais

As relações sociais podem ser consideradas canais de transporte apara informação, serviço e bens entre pessoas e entre organizações. A análise de redes sociais permite que se identifique as características dessas relações que facilitem ou dificultem esse transporte e como esse fenômeno ocorre em uma determinada rede. Dentre as características de uma rede relevantes para essa análise, pode-se identificar dois grandes grupos: Métricas de coesão e métricas de mediação.

### Análise Exploratória dos dados

Embora não se trate de uma pesquisa exploratória clássica, esse trabalho utiliza a análise exploratória de redes sociais da forma definida por Nooy, Mrvar e Batagelj (2005, p 5).

### Pesquisa descritiva e análise exploratória

Não se pode considerar esse trabalho uma pesquisa exploratória clássica pois essas visam estudar um fenômeno sobre o qual há pouca informação, é normalmente aplicada a campos ainda pouco explorados pela ciência e é fortemente baseada em estudos de caso e entrevistas com o objetivo de entender melhor o fenômeno de interesse.

Sobre as pesquisas exploratórias, Gil (2008) entende que:

“Pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato. Este tipo de pesquisa é realizado especialmente quando o tema escolhido é pouco explorado e torna-se difícil sobre ele formular hipóteses precisas e operacionalizáveis.” (Gil, 2008, p.27).

Na presente pesquisa estudamos o fenômeno da comunicação e mediação da informação entre organizações e agentes públicos. Além das referencias sobre o tema que sustentaram as discussões deste trabalho, pode-se facilmente constatar que o número de textos que exploram o tema é considerável. A pesquisa pelos termos “comunicação organizações publicas” no repositório institucional da Universidade de Brasília retornou 3436 ocorrências (em consulta submetida em 07/09/2013) o que já parece suficiente para impedir que escritos sobre o tema sejam considerados “pouco explorados” como Gil sugere que devam ser os temas abordados pelas pesquisas exploratórias.

Sob este ponto de vista, a classificação mais adequada para a presente pesquisa, segundo o modelo proposto por Gil, parece ser a de pesquisa Descritiva.

Esse modelo classifica as pesquisas sociais em Exploratórias, Descritivas e Explicativas. Enquanto as primeiras, como exposto, tratam de um tema pouco explorado pela ciência e as ultimas se prestem a identificar causas para determinados fenômenos, as Descritivas são pesquisas que tentam detalhar os fenômenos estudados, expondo suas características principais, as variáveis envolvidas e o relacionamento entre elas.

Nas palavras do autor, esclarecimentos sobre o tipo de pesquisa Descritiva:

“As pesquisas deste tipo têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis. São inúmeros os estudos que podem ser classificados sob este título e uma de suas características mais significativas está na utilização de técnicas padronizadas de coleta de dados.” (Gil, 2008, p.28).

O que se faz nesse trabalho (extrair dados de fontes abertas tais como o Diário Oficial da União para investigar a comunicação e mediação da informação entre organizações públicas) parece estar relacionado ao que Gil identificou como “técnicas padronizadas de coleta de dados”. Além disso, aproxima essa pesquisa da classificação Descritiva o fato de que se busca aqui descrever o fenômeno da comunicação entre organizações públicas tentando estabelecer relações entre esse fenômeno e o fenômeno das publicações de portarias em comunicações oficiais.

Por fim, alguns aspectos do trabalho desenvolvido retomam a noção de exploração, principalmente no que diz respeito à análise dos dados coletados. Embora a fundamentação desenvolvida visando demonstrar que as publicações em portarias do Diário Oficial da União podem ser analisadas como manifestações de uma comunicação prévia entre as entidades envolvidas na publicação tenha caráter fortemente descritivo, a utilização da Análise de Redes Sociais para a investigação dessa comunicação parece ter um caráter mais aderente ao modelo Exploratório de pesquisa.

“há pesquisas que, embora definidas como descritivas a partir de seus objetivos, acabam servindo mais para proporcionar uma nova visão do problema, o que as aproxima das pesquisas exploratórias.” (Gil 2008 p. 28)

Se na primeira etapa este trabalho descreveu o fenômeno das publicações e sua relação com a comunicação entre órgãos federais, na segunda ele explora essas comunicações em busca de padrões e “hipóteses mais precisas e operacionalizáveis” acerca desse fenômeno e, ao mesmo tempo, oferece “uma nova visão do problema”, ou seja, uma maneira até então pouco utilizada de estuda-lo (qual seja, redes sociais mapeadas por meio das informações presentes no Diário Oficial), como sugere Gil que deva ser uma pesquisa Exploratória.

Também se classifica como Exploratória a metodologia de análise das redes sociais mapeadas desenvolvida neste trabalho.

A análise exploratória de redes sociais é uma técnica que permite que se estude redes sociais visando a identificação e interpretação de padrões estruturais das relações sociais entre os indivíduos da rede sem uma rigidez metodológica significativa e sem hipóteses claras previamente estabelecidas.

A análise exploratória de redes sócias proposta por Nooy, Mrvar e Batagelj (2005 p. 5) pressupõe que “a estrutura ou padrão das ligações em uma rede social é significativa para os membros da rede e, portanto, para o pesquisador.”. Os autores argumentam que, ao conduzir esse tipo de análise exploratória, “As hipóteses devem estar fundamentadas em teorias sócias e experiências de pesquisa anteriores”, uma vez que não há envolvimento de técnicas como os testes de hipótese fornecidos pela Estatística.

Assim, antes de se aplicar as técnicas e métricas da Análise de Redes Sociais essa pesquisa se dispôs a discutir os fundamentos teóricos que sustentam a utilização do Diário Oficial da União como fonte de dados para o mapeamento das redes sociais estabelecidas entre as organizações e agentes públicos federais. Essa discussão é importante para, seguindo a orientação dos autores, legitimar eventuais conclusões que surjam a partir da análise exploratória das redes.

São quarto as etapas da Análise exploratória de redes sociais proposta por Nooy, Mrvar e Batagelj: Definição da Rede, Manipulação da Rede, Identificação de elementos estruturais e Inspeção visual.

A primeira etapa da análise exploratória de redes sociais é a definição da rede. Nessa etapa toma-se decisões referentes ao objeto de estudo da análise, ou seja, a própria rede a ser mapeada. Deve-se eleger critérios para a inclusão ou exclusão de atores na rede bem como os fatores que determinam o mapeamento de uma conexão entre um par de atores.

A etapa seguinte, manipulação da rede, é um recurso útil especialmente para lidar com redes de tamanho elevado, com muitos atores e conexões entre eles. Operações comuns nessa etapa incluem a remoção de elementos menos significativos, como linhas com valores baixos, atores com poucas conexões ou grupos de atores cuja característica especifica permita que sejam descartados sem prejuízo da análise que se deseja desenvolver. Também podem ser usados recursos como o agrupamento de um subconjunto de atores em um único vértice e a transformação de conexões direcionadas em não-direcionadas e vice versa.

Na etapa de Identificação de elementos estruturais é quando são utilizadas as métricas da Análise de Redes Sociais com maior intensidade. Nesse momento são feitos os cálculos quantitativos das características estruturais do fenômeno social de interesse. Essas métricas oferecem indicadores das mais diversas propriedades, relacionadas a características como a coesão entre grupos e subgrupos, o papel de cada ator na rede em relação a sua posição estrutural, o prestígio e o poder que cada ator pode exercer sobre os demais integrantes da rede entre outras medidas referentes não só aos elementos mas à rede como um todo. Nessa etapa já é possível extrair algumas conclusões acerca das hipóteses levantadas no estudo da rede.

A ultima etapa da análise exploratória de redes sociais é a inspeção visual. Nooy, Mrvar e Batagelj (2005 p. 14) argumentam que “o olho humano é treinado para reconhecer padrões. Assim a visualização da rede ajuda a rastrear e apresentar padrões das conexões”. A dificuldade que pode existir em analisar os valores numéricos extraídos na etapa anterior justificam a utilização da inspeção visual.

Além disso, argumentam os autores, os algoritmos de desenho automático de redes buscam aproximar atores com características semelhantes estruturalmente o que ajuda a identificar o que eventualmente não tenha sido explicitado na Identificação de elementos estruturais. Além disso, hipóteses e impressões surgidas nessa etapa podem orientar a condução de uma nova bateria de cálculos estruturais, refinando a análise e fazendo-a caminhar em direção a uma conclusão.



Figura X – Ilustra o processo de análise exploratória de redes sociais

Assim, podemos entender a análise exploratória de redes sociais como um processo circular, na qual as etapas são executadas sem rigidez metodológica e podem ser revisitadas, servindo de feedback umas às outras no processo de construção e verificação de hipóteses acerca da rede estudada. A Figura X ilustra esse processo.

//TODO: Esse texto não deve estar aqui.

A proposta dessa pesquisa é analisar os dados coletados na forma de redes longitudinais. As redes extraídas dos dados coletados possuem a informação temporal e serão analisadas usando o software Pajek. O pajek é um software que auxilia na análise de redes sociais e possui, entre outras, a funcionalidade de visualização e outras operações de redes longitudinais. Esse aspecto é fundamental para observar a dinâmica das relações entre as organizações públicas estudadas.

Esse software também será usado para a extração das métricas de coesão e de mediação que servirão para o estudo das redes. Essas métricas serão usadas para investigar a interação entre os órgãos e identificar os padrões de comunicação e mediação da informação nesse cenário.

Esses procedimentos podem ser realizados com a rede disposta de forma que cada vértice represente uma organização, mas também se aplicam à rede simplificada, na qual são agrupados os vértices em clusters ou sub-redes. Esse agrupamento pode simplificar o estudo e trazer novas revelações sobre a comunicação da informação nesse ambiente. Albagli e Maciel (2004, p. 11) advertem que “as interações não devem ser estudadas apenas entre os agentes de uma mesma rede, mas sobretudo entre redes”. Assim, por exemplo, o estudo da comunicação entre as sub-redes de universidades, sub-redes de embaixadas e de organizações militares, pode revelar informações que talvez não fossem detectadas se o estudo fosse concentrado nos vértices individualmente.

Outro recorte a ser feito nas redes consiste em construir subredes para cada ministério aos quais os órgãos identificados estão vinculados. Isso vai possibilitar uma visão da comunicação e mediação da informação internamente em cada órgão autônomo.

O estudo também será feito a partir de redes temáticas. Para extrair tais redes serão consideradas apenas as portarias que mencionem termos pré definidos. Essa metodologia de análise e processamento de texto para extração de informação é útil para diminuir o escopo e direcionar a investigação. Sobre ela, Schiessl (2007) considera:

“No momento em que se transforma o texto em termos individuais ou compostos observa-se que alguns aparecem muitas vezes, outros medianamente e outros raramente. A utilização de alguns termos em detrimento de outros é uma escolha feita pelos analistas que conduzem o processo de descoberta e, para tanto, são criados dicionários especializados e listas de termos que apoiam o trabalho de escolha dos termos que serão utilizados pelos algoritmos de mineração de texto.” (SCHIESSL 2007)

Uma vez que se supõe que as publicações do DOU refletem o resultado da comunicação entre os órgãos acerca das politicas públicas por eles conduzida, elegeu-se o Plano Plurianual (PPA 2012) como fonte de palavras chave para a mapeamento das redes temáticas. O PPA 2012 define “Onze Macrodesafios que orientarão as políticas públicas federais nos próximos 4 anos, consubstanciadas nos programas governamentais.” (MPOG 2012, p 77). Esses macrodesafios são detalhados no PPA 2012 e dessa descrição foram extraídas palavras chave que se supõe que ocorram em portarias do DOU relacionadas à implementação das politicas referentes à atuação visando o enfrentamento desses macrodesafios. É possível que a configuração da rede esteja relacionada ao tema e que dependendo do que esteja sendo tratado, os papeis na rede sejam desempenhados por agentes diferentes.

Assim, no que diz respeito às redes, a análise dos dados compreende:

Análise global (todas as ligações) e temática (somente ligações relacionadas a um tema) da rede de organizações públicas completa e reduzida por meio da eliminação de ligações pouco frequentes e vértices com pouca interação.

Análise das redes de cada ministério com os mesmos recortes mencionados no item anterior (global x temática, completa x reduzida).

Análise da rede encolhida (comando shrink do pajek) por ministérios visando estudar os padrões de comunicação entre os órgãos autônomos da administração publica brasileira.

Análise das redes formadas por pessoas, em confronto com os padrões encontrados na análise das redes dos ministérios.

Nessas análises serão avaliadas, no que couber, as seguintes métricas: Densidade, grau médio, centralização, centralidade dos indivíduos, ocorrência de cliques, cores. Além disso, identificar-se-á os papeis de mediação desempenhados pelos indivíduos na rede e investigar-se-á o fenômeno da difusão na rede.

### Métricas de coesão

As métricas de coesão são úteis para distinguir grupos de indivíduos (nesse contexto o termo ‘individuo’ remete às organizações publicas estudadas) que compartilham características e trabalham em um nível de cooperação mais intensa do que o restante da rede. WASSERMAN e FAUST (1994, p. 250) esclarecem que quanto mais imerso na rede mais o indivíduo será afetado pelos padrões de comportamento do grupo. Para os autores, a análise dos grupos coesos pode evidenciar o nível de isolamento (ou integração) dos membros desse grupo em relação à rede. Esse tipo de informação pode contribuir para a formulação de politicas publicas semelhantes para órgãos de um mesmo grupo e adaptadas para cada grupo identificado.

Dentre as métricas de coesão, destacam-se as medidas de densidade, componentes, cliques e cores.

A medida de densidade reflete o quão conectados estão os elementos de uma rede. Espera-se encontrar, em um grupo mais coeso, um maior número de conexões entre os indivíduos.

“Intuitivamente, coesão significa que a rede social possui muitas ligações. Mais ligações entre pessoas levam a uma estrutura mais apertada, que é, presumidamente, mais coesa. Em análise de redes sociais, a densidade de uma rede captura essa ideia.”. (NOOY, MRVAR E BATAGELJ, 2005, p. 62)

Componentes são subconjuntos de vértices conectados de tal forma que seja possível traçar um caminho entre quaisquer pares desses vértices. Pode-se esperar que os componentes formem ‘ilhas’ de comunicação independentes entre si, uma vez que não existe comunicação entre os elementos de diferentes componentes.

Cliques são grupos de vértices nos quais todos os elementos estão conectados a todos os demais integrantes do clique. Por definição, “um clique é uma subrede maximal completa contendo três ou mais vértices”. (NOOY, MRVAR E BATAGELJ, 2005).

Cores são subgrupos de vértices que compartilham o mesmo numero de conexões dentro de uma rede (grau). “Um k-core é uma subrede maximal na qual cada vértice tem, pelo menos, grau k nessa subrede”. (NOOY, MRVAR E BATAGELJ, 2005).

### Métricas de mediação

As métricas de mediação, ao contrario das métricas de coesão, visam destacar indivíduos dentro da rede para determinar os mais importantes sob o ponto de vista da intermediação. Para tanto, essas métricas oferecem maneiras de eleger quais os vértices mais relevantes levando-se em conta sua posição na rede e a quantidade de conexões que possui, bem como as conexões de seus vizinhos. Há várias métricas de mediação oferecidas pela ARS, e todas “tentam descrever e medir propriedades da localização do ator em uma rede social. Atores que são mais importantes ou mais proeminentes estão geralmente localizados em posições estratégicas dentro da rede.” (WASSERMAN e FAUST 1994, p. 169). Dentre as principais métricas de mediação pode-se citar centralidade, intermediação/pontes e difusão.

As medidas de centralidade identificam o centro e a periferia da rede. “A maioria das redes sócias contêm pessoas ou organizações centrais. Devido a sua posição eles têm mais acesso à informação e mais oportunidade de espalhar informação. [...] Em uma rede com alta centralização, há uma fronteira clara entre centro e periferia. Nesse tipo de rede a informação se espalha facilmente, mas o centro é indispensável para a comunicação.” (NOOY, MRVAR E BATAGELJ, 2005). A centralidade pode considerar o grau dos vértices, a distância entre os vértices ou os caminhos dos quais o vértice participa.

Medidas de intermediação comparam e medem o proveito que um indivíduo pode tirar de usa posição na rede. Dependendo das ligações que possui e do restante da rede, um indivíduo pode influenciar significativamente no fluxo de informação dentro da rede.

“Uma pessoa com muitos contatos tem mais chances de conseguir ajuda ou informação. Entretanto o tipo de conexão é relevante. O argumento geral é que ligações fortes (frequentes/intensas) com pessoas que são relacionadas entre si levam a informações menos úteis que ligações fracas com pessoas que não se relacionam. Ter muita ligação em um grupo expõe a pessoa à mesma informação reiteradamente, enquanto ligações para fora do grupo oferecem ao individuo informação diversificada que pode ser retida ou transmitida para gerar vantagem. (NOOY, MRVAR E BATAGELJ, 2005, p 138). (tradução do autor)

Finalmente, o conceito de difusão configura um processo social importante. Dentro de uma rede o processo de comunicação é determinado pelo arranjo das conexões entre as pessoas. “Difusão é um caso especial de mediação que considera a dimensão tempo” (NOOY, MRVAR E BATAGELJ, 2005, p 161). Assim, as métricas de difusão tratam da maneira com que a informação (ou qualquer outro elemento ‘transmissível’) é transmitida dentro da rede, de um indivíduo ao outro, durante um intervalo de tempo.

As métricas de mediação podem revelar aspectos importantes da comunicação e mediação da informação entre as organizações publicas pois permite identificar os atores mais relevantes, o papel de cada um deles no processo e apontar estratégias para a implantação de programas ou a comunicação de informações de forma mais eficiente.

## Identificação de redes sociais a partir de dados abertos

### Mineração de dados abertos para construção de redes sociais

Alguns autores têm usado dados públicos para a construção de redes sociais. A grande contribuição desse tipo de trabalho é a demonstração da viabilidade de coleta de dados para mapeamento de redes sociais sem a necessidade de procedimentos clássicos de coleta como entrevistas ou observação de campo.

Um exemplo de trabalho nesse sentido foi desenvolvido por KREBS (2002). Ao estudar a rede social dos sequestradores dos aviões do atentado de 11 de setembro nos Estados Unidos, o autor recorreu aos artigos publicados pela imprensa e a “procedimentos jurídicos públicos” para reproduzir a rede criminosa. Sobre a coleta de dados, o autor ressalta a importância de se verificar a confiabilidade da informação coletada e de se esperar que algumas ligações existentes na rede real não serão capturadas pela coleta e, portanto, a rede extraída pode não ser completa.

Aparentemente inspirado no trabalho de KREBS, o trabalho desenvolvido RODRIGUEZ(2004) estudou a rede social de terroristas responsável pelo atentado de 11 de Março na Espanha. A ponderação feita pelo autor no que diz respeito a utilização de fontes de dados públicas para a construção de redes tange o viés político das informações veiculadas. No contexto do atentado espanhol havia um componente eleitoral que, na visão do autor, pode ter influenciado na qualidade e quantidade de informações divulgadas acerca do evento estudado.

Como ultimo exemplo de estudo nesse formato, pode-se citar o trabalho de SAGEMAN(2007) que estudou, exclusivamente através de fontes de dados abertas, o movimento islâmico Salafi Jihad, usando redes sociais. Na seção onde aborda os problemas enfrentados na coleta dos dados, o autor expõe a impossibilidade da mínima validação dos modelos devido à lealdade dos eventuais entrevistados, sugere que a classificação da confiabilidade das fontes de informação sensibilizem o peso das ligações extraídas para a rede e a falta de simetria entre a publicidade de eventos “opostos”, que, na rede, deveriam ter a mesma relevância. (A captura de um suspeito recebe mais atenção do que a eventual constatação de sua inocência e posterior liberação). (SAGEMAN, 2007, Cap 3 p. 64 - 70)

Resumidamente o que se vê nas obras desses autores é a preocupação com a confiabilidade e completeza dos dados coletados. O primeiro aspecto atinge parcialmente a presente pesquisa pois, nesse caso, lida-se com publicações da imprensa oficial, e não da imprensa comum, o que garante a veracidade dos dados, mas a falta de meios de validação é observada. O segundo aspecto é relevante pois aborda a perda de informações no processamento ou simplesmente a não captura da informação durante a coleta dos dados, que, no caso desse trabalho, é mitigada pelo aprimoramento da extração por meio do emprego de esforços de engenharia textual.

### Análise Co-Word

### Diversidade lexical e carga informacional das portarias

## Base Empírica, Ferramentas e Metodologia

Neste trabalho sugere-se que a análise da rede de comunicação da informação entre as organizações publicas brasileiras seja feita por meio de métricas fornecidas pela disciplina Análise de Redes Sociais (ARS). Conforme discutido no item 2.2, o Diário Oficial da União apresenta manifestações das relações dos relacionamentos existentes entre as organizações públicas que podem ser extraídas e estruturadas de forma que seja possível construir modelos de redes sociais sobre as quais as métricas da ARS podem ser aplicadas.

Essa seção, portanto, discute (1) a escolha do Diário Oficial da União, (2) as técnicas propostas para o processamento dos textos do Diário Oficial da União, (3) a identificação das organizações e dos nomes de pessoas no texto bem como dos limites das publicações, (4) os critérios propostos para a identificação de relacionamentos, (5) as métricas da ARS mais adequadas para estudo da comunicação e mediação da informação em redes extraídas da forma proposta.

### O Diário Oficial da União

Esse trabalho investiga a o uso de fontes de dados abertas para o mapeamento de redes sociais visando o estudo da comunicação e mediação da informação entre organizações públicas brasileiras. Para tanto, sugere que seja usado o Diário Oficial da União (DOU).

Publicado diariamente pela Imprensa nacional, órgão vinculado à Casa Civil da Presidência da República, o Diário Oficial da União, que circula há mais de 200 anos, tem seu conteúdo regulado pelo decreto 4520 de 16 de dezembro de 2002.

Esse decreto, em seus artigos 2º a 5º, define que os seguintes atos oficiais serão obrigatoriamente publicados:

* Leis e demais atos resultantes do processo legislativo do Congresso Nacional;
* Tratados, as convenções e outros atos internacionais aprovados;
* Medidas provisórias, os decretos e outros atos normativos.
* Atos dos Ministros de Estado, baixados para a execução de normas;
* Pareceres do Advogado-Geral da União e respectivos despachos presidenciais;
* Dispositivos e ementas das ações direta de inconstitucionalidade, das ações declaratórias de constitucionalidade e das argüições de descumprimento de preceito fundamental decorrente da Constituição;
* Julgamentos do Tribunal de Contas da União;
* Atos de caráter normativo do Poder Judiciário.
* Os atos de caráter judicial do Poder Judiciário e dos órgãos auxiliares da Administração da Justiça.
* Os atos relativos ao pessoal civil e militar do Poder Executivo, de suas autarquias e das fundações públicas, bem assim dos servidores do Poder Legislativo e do Poder Judiciário, cuja publicação decorrer de disposição legal
* Atas e decisões de tribunais e de órgãos colegiados dos Poderes da União;
* Pautas, Editais, avisos e comunicados, Contratos, convênios, aditivos e distratos, Despachos de autoridades administrativas, relacionados a interesses individuais;
* Atos oficiais que autorizem, permitam ou concedam a execução de serviços por terceiros.

Da enumeração acima depreende-se que boa parte dos atos mais relevantes das organizações públicas brasileiras são registrados no DOU. Esse atos são a materialização a atuação essas organizações, que quando agem cooperativamente, comunicam-se e coordenam-se a fim de atingir um objetivo que supostamente culminará em uma publicação na qual os órgãos envolvidos são citados. São essas as publicações que podem ser consideradas ligações e que podem ajudar a modelar a rede social das organizações públicas brasileiras.

O portal da Imprensa Nacional (in.gov.br) esclarece que o Diário Oficial da União é publicado em três seções distintas:

* + - SEÇÃO 1 - Leis, decretos, resoluções, instruções normativas, portarias e outros atos normativos de interesse geral.
    - SEÇÃO 2 - Atos de interesse dos servidores da Administração Pública Federal.
    - SEÇÃO 3 - Contratos, editais, avisos e ineditoriais.

O conteúdo de cada seção é determinante para o tipo de relação manifestada em cada publicação. Assim, a metodologia proposta nas seções seguintes conserva essa informação como um atributo da associação entre as organizações, para que possa ser considerada nas análises.

### Outras fontes de dados abertos sobre organizações públicas brasileiras

Após a publicação da lei de acesso à informação, Lei no 12.527, de 18 de novembro de 2011, e sua regulamentação pelo decreto 7724 de 16 de maio de 2012, uma infinidade de dados sobre a atuação das organizações públicas brasileiras têm sido disponibilizadas à sociedade, destacadamente em meio eletrônico na internet.

Entretanto, várias outras fontes de informação dessa natureza já estavam disponíveis antes de maio de 2012. Essas fontes permitem que o cidadão acompanhe as atividades dos órgãos sob os mais diferentes aspectos.

Dentre as principais fontes de dados sobre as organizações publicas brasileiras disponíveis atualmente podemos destacar:

Portal da transparência ([www.portaltransparencia.gov.br](http://www.portaltransparencia.gov.br)): Trata-se de um sítio eletrônico mantido pela Controladoria Geral da União desde 2004 que disponibiliza dados relacionados aos recursos públicos sob responsabilidade do Governo federal. Segundo a seção ‘sobre’ do sítio, “O objetivo é aumentar a transparência da gestão pública, permitindo que o cidadão acompanhe como o dinheiro público está sendo utilizado e ajude a fiscalizar.”. Os dados estão disponíveis em formato estruturado, sendo a maioria em arquivo tipo csv (*comma separated values* – valores separados por virgula), e disponibilizam informações sobre transferências de recursos, gastos diretos, diárias, salários de servidores, receitas, contratos da administração pública entre outras relacionadas a recursos públicos.

IpeaData (http://www.ipeadata.gov.br/) : A Fundação Instituto de Pesquisa Econômica Aplicada (Ipea) é uma fundação pública federal vinculada à Presidência da República e mantém um portal onde disponibiliza dados, séries históricas e indicadores de dezenas de fontes, classificados em três grandes categorias: Macroeconômico - Dados econômicos e financeiros do Brasil em séries anuais, mensais e diárias na mesma unidade monetária. Regional - Dados econômicos, demográficos e geográficos para estados, municípios, regiões administrativas e bacias hidrográficas brasileiras. Social - Dados e indicadores sobre distribuição de renda, pobreza, educação, saúde, previdência social e segurança pública. Dados brutos e agregados podem ser acessados em formato html (necessita de algum esforço de extração antes de ser usado por programas de análise).

Dados abertos legislativo (<http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo>): “O serviço de Dados Abertos - Legislativo disponibiliza uma coleção de funcionalidades que permite o acesso direto aos dados legislativos produzidos na Câmara dos Deputados como dados sobre deputados, órgãos legislativos, proposições, sessões plenárias e reuniões de comissões.”. Os dados estão disponíveis em formato WebService, extremamente amigável para acesso por máquinas e de fácil integração por sistemas de análise automatizada.

Sitios oficiais de Autarquias e Ministérios: A maioria das autarquias e ministérios disponibiliza dados brutos e agregados sobre sua atuação. Esses dados normalmente estão disponíveis por meio de consultas viabilizadas por formulários que não foram projetadas para acesso automatizado, e os dados são apresentados em formato que requer processamento prévio antes de ser analisados automaticamente. Não há nenhuma padronização para a disponibilização desses dados e cada órgão o faz da maneira que melhor lhe parece, fazendo com que haja uma variedade grande de tipos de informação disponíveis e formas diferentes de consulta e apresentação.

Portal Brasileiro de Dados Abertos (<http://dados.gov.br/>): Trata-se de um sítio mantido pelo Ministério do Planejamento, Organização e Gestão, que, em atenção à Lei de acesso à informação, “... tem o objetivo de disponibilizar todo e qualquer tipo de dado.”. Encontram-se nele dados das mais diversas naturezas, desde informações sobre cadastro de eleitores a censo de educação passando por informações sobre segurança pública e gastos governamentais. Podem ser encontrados dados das esferas Federal, Estadual e Municipal, disponíveis em sua maioria em arquivos tipo csv.

A lista apresentada acima, longe de ser exaustiva, ilustra a grande quantidade e diversidade de fontes de dados acerca da atuação das organizações públicas brasileiras disponível na internet.

Trata-se de fontes de dados abertos comparáveis ao Diário Oficial da União, no sentido de que (1) trazem informações sobre a atuação das organizações públicas brasileiras, (2) Estão disponíveis de forma livre na internet e (3) apresentam-se em formato cujo processamento por computadores é viável. Isso posto, pode-se considerar que, em linhas gerais, o modelo apresentado neste trabalho pode ser aplicado, com os devidos ajustes, a quaisquer dessas fontes de dados, individualmente ou combinadas.

Entretanto, considerando a dinâmica da atuação das organizações públicas brasileiras que, conforme discutido anteriormente, têm a publicação dos seus atos no Diário Oficial da União como requisito para produção de efeitos, podemos considerar as demais fontes de dados como secundárias, colocando o DOU, portanto, como a fonte mais completa dentre as disponíveis, justificando sua escolha para o desenvolvimento deste trabalho.

### GATE – General Archteture for Text Engeneering

### PAJEK

### Mapeamento das redes

A identificação de redes sociais em fontes de dados abertas foi discutida por RESSLER (2006) em um trabalho no qual o autor classifica os processos de análise de redes sociais em dois grupos: (a) coleta de dados e (b) modelagem. As semelhanças entre os objetos de estudo do presente trabalho e do trabalho de RESSLER (quais sejam, redes sociais extraídas de fontes de dados abertos para estudo de mediação da informação e redes sociais extraídas de fontes de dados públicos para estudo do terrorismo), sinalizam a possibilidade de aplicação dessa classificação no presente trabalho.

### Coleta de dados

Sacerdote (2013), ao discutir o processo de coleta de dados para modelagem de redes sociais em seu estudo sobre mediação da informação em ambientes de aprendizagem virtual, argumenta que o esse processo é frequentemente baseado em questionários, entrevistas, observações e experimentos cuja viabilidade é comprometida à medida que os grupos estudados aumentam. Entretanto, o aumento da utilização de tecnologia nas comunicações contribui para a consolidação de uma fonte de dados para a identificação de redes de forma automatizada e em uma grande quantidade de dados.

“Com o uso das tecnologias de forma crescente na atualidade, surgem novos cenários sociais na internet, tais como as mídias sociais e os AVAs. (...) arquivos gerados por meio das interações nesses contextos tecnológicos de educação, podem ser considerados como novas fontes de dados para análise das interações sociais e mediação.” (SACERDOTE, 2013).

A coleta de dados para o desenvolvimento do estudo proposto foi feita usando como fonte primária as edições do Diário Oficial da União (D.O.U.) disponíveis no portal da imprensa nacional (IN) ([www.in.gov.br](http://www.in.gov.br/)). No portal da IN pode-se ter acesso a todas as edições do D.O.U. Desde março de 1997, entretanto, para os fins dessa pesquisa, limitou-se a análise às publicações do período entre 01/01/2007 e 01/01/2013.

Foram desenvolvidos dois softwares usando a linguagem java para executar a coleta dos dados. De forma complementar esses softwares descarregam e tratam a informação a fim de gerar dados dos quais se pode extrair redes sociais. A figura 1 ilustra a interação entre os componentes da solução desenvolvida. O leitor interessado em detalhes da implementação, códigos e bibliotecas utilizadas encontrará mais informações no ANEXO I deste trabalho.

O módulo batizado de DouDownloader acessa cada página de cada edição do D.O.U. e busca os arquivos correspondentes em formato pdf para depois transformá-los em arquivos texto e os armazena localmente. Essa operação é necessária para a exploração do conteúdo dos arquivos, que são disponibilizados em formato PDF, de difícil manipulação.

Como destacam Singhal e Srivastava (2013) ao discutir estratégias para a mineração de dados em artigos científicos, “Trabalhar diretamente com arquivos pdf não é uma solução viável, pois é difícil extrair texto de pdfs. Então, trata-se de uma pratica comum a conversão de pdf para arquivo texto para qualquer processamento que precise ser feito” (SINGHAL e SRIVASTAVA, 2013 – tradução do autor). Essa conversão é feita usando estratégias que, pela complexidade do formato pdf, introduzem ruídos nos dados convertidos. Na seção 5 deste trabalho são discutidas as limitações dessa operação.



Figura 1 - Interação entre os modulos desenvolvidos

O modulo chamado AnalisadorDou acessa cada arquivo gravado pelo DouDownloader e faz o processamento descrito a seguir:

1 - Para cada arquivo disponível para processamento:

1.1 - Identifique todos os órgãos mencionados no texto.

1.2 - Identifique todas os nomes de pessoas mencionados no texto.

1.3 - Identifique todos os inicios de publicação presentes no texto.

1.4 - Identifique todos as assinaturas de publicação presentes no texto.

fim para

O passo 1.2 foi implementado usando o gazzeter do GATE e uma lista de nomes extraída da internet (<http://www.listadenomes.com.br/>). Embora não seja exaustiva, os resultados preliminares indicam que o material usado é suficiente. Outro recurso do GATE usado nessa etapa foi o Tokenizer. Esse recurso permite que sejam criadas regras para a identificação de elementos no texto. Assim, criou-se uma regra que determina que qualquer sequência de palavras com inicial ou todas letras maiúsculas dentre as quais pelo menos uma delas seja um nome da lista identificado pelo gazzeter, configura um nome de uma pessoa.

O passo 1.4 foi implementado definindo uma assinatura como a sequencia de um nome de pessoa seguido de o nome de um órgão na linha imediatamente seguinte. Os elementos identificados nos passos anteriores e o tokenizer foram usados nessa etapa.

Assim, a Figura 2 mostra a tela do GATE onde pode-se identificar todos os elementos detectados pela execução dos passos descritos acima.As etapas descritas no laço 1 foram implementadas por meio da utilização de uma biblioteca especializada chamada GATE. O GATE (General Archtechture for Text Engeneering) é o resultado de um projeto mantido por empresas privadas e universidades desde 1995, no qual se investiu mais de cinco milhões de euros, e que oferece um software de código aberto capaz de resolver quase todos os problemas e processamento de texto e linguagem humana. (CUNNINGHAM, 2011. p 5).

O passo 1.1 foi implementado usando a lista de órgãos públicos do SIORG. “O SIORG é a fonte oficial de informações sobre a estrutura organizacional dos órgãos do Poder Executivo, Administração direta, Autarquias e Fundações.” (SIORG 2012). A função Gazzeter do GATE trata de identificar no texto dado todas as ocorrências dos elementos de uma lista. O passo 1.3 também foi implementado com o gazzeter, a lista, entretanto, foi elaborada manualmente, e contém palavras chave de inicio de publicação (Ex: ATA, PORTARIA, EDITAL, entre outras.)

Figura 2 - Tela do GATE com os elementos identificados.

Detalhes das implementações java e dos componentes de software utilizados, bem como as regras implementadas para o GATE podem ser encontrados no ANEXO I.

### Modelagem

A modelagem, segundo grupo de atividades do processo de análise de redes sociais segundo RESSLER (2006), pode ser dividida em dois outros grupos de atividades, conforme sugerem HANNEMAN e RIDDLE (2011). Esses autores identificam como atividades fundamentais na modelagem da rede: (a) a definição de fronteiras da rede e (b) o estabelecimento de critérios para a detecção de relacionamentos.

Com relação à definição de limites, Hanneman e Riddle argumentam que a saída mais comum é adotar um limite estabelecido pelos próprios atores. “Todos os membros de uma sala de aula, organização, clube, vizinhança ou comunidade” por exemplo. Nesse caso, sabe-se da existência da rede social a priori. Outra alternativa é considerar uma abordagem mais “demográfica” ou “ecológica” na definição de fronteiras. “Pessoas encontradas em uma determinada área ou que atendem a um critério objetivo como nível de renda”. Nesse caso, há razoes para suspeitar da existência de uma rede social entre os elementos, mas o estudo parte de uma abstração imposta pelo pesquisador. (HANNEMAN e RIDDLE,2011)

O presente estudo define os limites da rede como organizações públicas da administração direta e indireta e pessoas cujo nome foi mencionado no Diário Oficial da União, e o autor investiga se realmente há uma rede social entre esses elementos que possa ser extraída das informações do DOU.

A definição de critérios para a identificação de relacionamentos entre os indivíduos é discutida por Hanneman e Riddle, que chamam a atenção para a decisão da quantidade de informações a ser considerada e o custo que isso pode acarretar. Além disso, eles indicam que podem existir vários tipos de ligação entre indivíduos, e cada um desses tipos leva a uma rede diferente, com características próprias. É preciso ter em mente o tipo de estudo que se deseja desenvolver e como a escolha do tipo de relacionamento influencia nesse trabalho.

Essa pesquisa propõe que a definição dos relacionamentos seja feita de forma análoga à proposta em trabalhos de infometria que utilizam a técnica chamada de ‘*co-word analisys*’.

Alguns estudos abordam o tema utilizando a ‘*co-word analisys’*. Courtial (1986 apud POLANCO 2005) propôs a apresentação de clusters co-word no formato de grafos. A técnica do co-word determina que se estabeleça uma borda entre dois termos (vértices) sempre que houver uma co-ocorrência deles no texto. Estando os dois termos no mesmo excerto textual, observar-se-á uma ligação entre eles na rede social correspondente.

Nas palavras de COURTIAL e LAW (1984), o método co-word consiste em:

“(desenvolver) um estudo detalhado das palavras chave associadas à literatura cientifica em um domínio da ciência. Daí, o primeiro passo em um estudo co-word é selecionar, de uma base bibliográfica como a CNRS/INIST em Paris, a população de artigos relevantes para uma área em particular. Os artigos selecionados são indexados, por experts, por entre cinco e quinze palavras chave cientificamente relevantes. O método, então, conta o numero de artigos os quais estão indexados por cada par de palavras chave. Assume-se que se muitos artigos estão indexados por um par de palavras chave, então os conceitos, problemas, ou ideias associados a essas palavras chave estão fortemente ligados e intensamente associados. O produto desse estágio da análise co-word é, então, uma grande rede a qual expõe ligações entre conceitos problemas e ideias. (COURTIAL e LAW, 1989) (tradução do autor)

A implementação da aproximação descrita acima, para o presente trabalho, foi desenvolvida no mesmo módulo java citado na seção anterior, qual seja, AnalisadorDou. A seguinte sequencia de comandos é executada logo após o laço 1 descrito acima:

2 - Para cada inicio de publicação identificado

2.1 - Identifique a próxima assinatura.

2.2 - Considere todo texto entre o inicio e a assinatura como conteúdo de uma publicação

2.3 - Para cada ocorrência de uma entidade (Órgão ou Pessoa) presente no texto da publicação, crie uma associação entre a entidade e a publicação.

fim para

Os passos descritos no laço 2 foram implementados usando java e são constituídos basicamente de instruções de código que percorrem as listas dos elementos identificados pelo GATE no laço 1 e, em seguida, fazem inclusões no banco de dados de relações entre entidades e publicações.

## Resultados e análises preliminares

Os resultados apresentados a seguir são fruto da análise exploratória preliminar realizada até o presente momento e não se pretende que sejam considerados finais, podendo sofrer alterações relevantes ao final da pesquisa, bem como apresentarem inconsistências.

As métricas mencionadas no item 3 foram extraídas de uma rede construída com base em dados relativos às publicações do período compreendido entre 01/01/2012 e 01/03/2012, incluindo-se os três jornais que compõem o Diário Oficial da União utilizando-se o software Pajek.

### Visão Geral

“A inspeção visual da rede social tem sido uma ferramenta importante para os pesquisadores desde o principio da análise de redes sociais.” (NOOY, MRVAR E BATAGELJ, 2005). Assim, ao se analisar a visão geral da rede exposta na Figura 4, na qual o autor agrupou manualmente as partições, pode-se perceber alguns aspectos interessantes, que podem revelar padrões da comunicação no ambiente estudado.

Destaca-se a coesão do grupo de organizações representado no canto direito inferior da imagem. Trata-se de todos os órgãos da rede que pertencem à partição que representa o Ministério do Desenvolvimento Agrário e as conexões entre eles revelam a existência de um clique. Nesse clique é de se esperar que seja intensa a comunicação entre os órgãos e haja pouca mediação já que todos têm livre acesso a qualquer outro membro do grupo. Esse mesmo grupo apresenta um vértice que desempenha o papel de gatekeeper ou representative para o grupo. Gatekeeper é o mediador dentro da rede que regula a entrada de informação para dentro do grupo. Representative é o elemento que regula a saída de informação para fora do grupo. (NOOY, MRVAR E BATAGELJ, 2005, p. 151). Nota-se que o vértice 134 (Instituto Nacional de Colonização e reforma Agraria) desempenha o papel de mediador entre os membros do Ministério do Desenvolvimento Agrário e os demais elementos da rede.

De forma semelhante se comporta o vértice 22 (Caixa Econômica Federal) em relação aos membros do Ministério do Esporte, representados no canto esquerdo superior da imagem. Nesse caso, entretanto, deve-se considerar o baixo peso das ligações na análise, antes de se concluir que a caixa Econômica represente um mediador de informação entre o Ministério do esporte e os demais integrantes da administração pública.



Figura 4 – Visão geral da rede de organizações publicas

Por fim, percebem-se as organizações vinculadas ao Ministério da Pesca representadas no canto superior direito da Imagem e notam-se que eles mantêm conexões com apenas dois outros grupos: Ministério da Previdência Social e Ministério da Justiça. Isso indica que o Ministério da Pesca pode ter uma atuação desconectada do restante da administração, tendo, para suas atividades, o apoio de apenas dois ministérios.

### Redes por Ministério

Da rede original discutida acima foram extraídas subredes para prosseguimento da análise exploratória. Extraiu-se uma rede para cada partição da rede, o que, na prática, significa examinar a rede de cada Ministério, entretanto, foram excluídas as partições com menos de dez elementos, para simplificação e melhor condução das análises. Assim, as redes estudadas nessa etapa estão listadas na Tabela 1.



Figura 5 – Rede do Ministério da educação destacando os componentes identificados

Todas as redes estudadas apresentaram-se como um único componente (considerando apenas componentes com mais de 2 vértices). O que pode indicar alta coesão interna nos ministérios. A exceção ficou por conta do Ministério da Educação que, talvez por contar com um numero muito maior de órgãos, apresentou um grande componente com aprox. 80% da rede e outros quatro componentes com os demais membros. Nos menores componentes predomina a ocorrência dos Institutos federais de Educação Ciência e tecnologia de cada estado, sendo que estados próximos formaram componentes. Esse é uma pista importante da forma como esses institutos cooperam entre si.

Ao analisar os valores de densidade exibidos na Tabela 1, percebe-se que os Ministérios da defesa e da justiça apresentam um nível de coesão muito acima dos demais. Esse resultado sugere que esses ministérios são mais coesos do que o Ministério do Trabalho e emprego, que possui quantidade parecida de vértices. As comparações de densidade devem ser feitas com o cuidado de se observar os tamanhos das redes envolvidas.

“A densidade como ferramenta de síntese em análises de redes pode ser falaciosa, especialmente se os valores forem pequenos. Esse resultado frequentemente se deve ao fato de que, enquanto o tamanho dos grupos cresce, a densidade da rede decresce se os graus dos atores permanecem constantes”. (WASSERMAN e FAUST, 1994)

De fato, observa-se que o grau médio da rede mais densa (Min. Da Defesa) e a menos densa (Min. Da Educação) é bastante parecido, mas sua sua densidade é bem diferente assim como a quantidade de vértices na rede, conforme preveem WASSERMAN e FAUST. Assim, comparações entre o Ministério da defesa e da educação levando em conta apenas a densidade das suas redes não devem ser consideradas boas fontes de informação.



Tabela 1 – Métricas da rede de cada ministério

A centralização da rede é um indicador da configuração da rede em termos da distancia entre elementos centrais e periféricos. Essa característica tem influencia direta no fluxo de informação na rede, sendo que quanto maior for esse indicador, mais facilmente a informação se espalha, sendo o centro, contudo, mais importante para o funcionamento da rede.



Figura 6 – Centralidade na rede do ministério da Fazenda



Figura 7 – Centralidade na rede do Ministério do Trabalho e Emprego

Dos dados apresentados vê-se as redes do Ministério da fazenda e do Ministério da Integração social com os maiores valores de centralização, o que sugere que os órgãos que compõem esses subconjuntos tenham uma comunicação mais facilitada em função da existência de órgãos proeminentes. Por sua vez, os dados indicam a possibilidade de não existir um órgão destacadamente central no Ministério do trabalho e Emprego, e, provavelmente, a disseminação de informações nessa rede seja menos eficiente.

As Figuras 6 e 7 mostram as redes na qual o tamanho dos vértices é proporcional a sua centralidade. Nota-se claramente a existência de um elemento central (Receita federal) na rede do Ministério da Fazenda (Figura 6) e a ausência de elementos de destaque na rede do Ministério do Trabalho e Emprego (Figura 7).

### Redes temáticas

Um dos estudos que se pode desenvolver usando as redes sociais é a visão por temas. Trata-se simplesmente da extração de redes sociais segmentada por determinadas palavras chave, que no caso deste trabalho, forem observadas dentro das portarias que promovem a ligação entre os vértices.

A Figura 8 mostra a rede de órgãos cujas ligações são estabelecidas com base na co-ocorrência em portarias nas quais são mencionadas as palavras chave relacionadas a ciência tecnologia e inovação, extraídas do PPA conforme metodologia exposta na seção três deste trabalho.



Figura 8 – rede temática ciência tecnologia e inovação

A inspeção visual da rede nos permite perceber uma clara predominância de vértices de cor azul, correspondentes aos órgãos vinculados ao Ministério da Educação. Percebe-se que o tema “Ciência, tecnologia e inovação” tem bastante afinidade com as atividades de pesquisa desenvolvidas pelas universidades e institutos sob o comando do Ministério da Educação, o que indica correspondência da modelagem da rede com a realidade.

O tamanho dos vértices na figura reflete sua centralidade. Assim, apesar os membros da partição correspondente ao Ministério da Educação apresentarem-se, no geral, maiores que os demais, há alguns indivíduos com centralidade alta em outras partições, notadamente o departamento de Polícia Federal e a Assessoria de Comunicação Social, o que sugere que esses órgãos tenham participação relevante na implementação dos projetos relacionados ao macrodesafio estudado.

### Difusão

O fenômeno da difusão em redes sociais também pode ser estudado nas redes de organizações públicas extraídas de fontes de dados abertos. De maneira preliminar e exploratória apresenta-se a seguir uma análise da difusão da adoção do Sistema de Seleção Unificada, o SISU, pelas universidades federais brasileiras.

A Ministério da Educação define o Sistema de Seleção Unificada – SiSU como “sistema informatizado gerenciado pelo Ministério da Educação - MEC, por meio do qual são selecionados candidatos a vagas em cursos de graduação disponibilizadas pelas instituições públicas de educação superior participantes.” (MEC 2010). A adesão ao SISU é feita pela Instituição de ensino interessada por meio da assinatura do Termo de Adesão e trata-se de uma ação voluntária, embora incentivada pelo governo federal.

Por meio de consulta ao Ministério da Educação baseada na lei de acesso informação foi possível ter acesso à tabela reproduzida no ANEXO II, que revela o semestre e ano que cada instituição de ensino passou a utilizar o SiSU. Confrontando-se essa informação com as redes extraídas de acordo com a metodologia proposta neste trabalho, é possível estudar a difusão em redes sociais e verificar se o fenômeno da adoção do SiSU como mecanismo de seleção pelas instituições de ensino é um comportamento cuja propagação é influenciada pelas relações de comunicação estabelecidas entre essas instituições.

O estudo de difusão em redes sociais pode tratar de adoção de inovações, transmissão de doenças, compra de produto, opinião, intenção de voto, atitude e qualquer outro fenômeno no qual as relações sociais possam influenciar ao longo do tempo.



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

A figura 9 ilustra a curva de difusão esperada para um fenômeno propagado através de relações estabelecidas em uma rede social. Apresentado por NOOY, MRVAR E BATAGELJ (2005, p. 165), como um padrão para o estudo da difusão, o gráfico exibe a porcentagem de indivíduos da rede ‘contaminados’ ao longo do tempo. Os autores argumentam que O formato em ‘S’ característico da curva é requisito para a identificação do fenômeno como um processo de difusão em rede:

“Se encontramos uma curva de difusão que não apresenta o formato típico em S, é pouco provável que as conexões da rede sejam importantes para o processo de difusão, e a difusão é propagada predominantemente por outras forças, como campanhas de mídia.” (NOOY, MRVAR E BATAGELJ, 2005, p. 165)

A Figura 10, por sua vez, apresenta a curva de difusão construída com base nas informações fornecidas pelo MEC no ANEXO II. Ela apresenta um gráfico com a quantidade de universidades integrantes do SiSU em cada semestre desde sua criação no inicio de 2010.

A curva de adesão ao SiSU apresenta um formato semelhante ao apresentado na figura 9, aproximando-se do formato em S característico da curva de difusão em redes sociais. Entretanto, o formato não é perfeitamente apresentado como um S, o que pode sugerir que outros fatores como pressões políticas tenham influencia na decisão da instituição de ensino, e, portanto, embora influenciado pela rede de comunicação, a adesão ao SiSU seja resultado de um conjunto mais complexo de influências.

Figura 10 - Difusão acumulada da adesão ao SiSU na rede de instituições de ensino superior.

A rede de universidades que aderiram ao SiSU até o segundo semestre de 2013 é apresentada na figura 11. Na imagem os vértices foram organizados em colunas de acordo com o semestre de adesão ao SiSU e foram excluídas os relacionamentos cujo número de ocorrências não superaram 2000. Isso permitiu a visualização da rede de forma mais clara e considerando-se apenas os relacionamentos mais frequentes e assim supostamente mais relevantes.



Figura 11 – Rede de Universidades com linhas com mais de 2000 ocorrências

Outra transformação à qual a rede foi submetida diz respeito à orientação das linhas. Originalmente não orientadas, as conexões foram consideradas orientadas de acordo com o tempo em que ocorreu a adesão. Como a hipótese é de que haja influencia na adesão ao SiSU entre as instituições e essa influencia é transferida por meio das relações sociais, a orientação das ligações da forma proposta traduz essa semântica.

A figura 11, assim, revela que, de forma geral, universidades cuja adesão ao SiSU se deu em períodos separados por mais de um ano não estão (fortemente) conectadas. Na rede apresentada não se identifica, por exemplo, nenhuma conexão sequer entre instituições cuja adesão se deu em 2010 e instituições que aderiram ao SiSU a partir do segundo semestre de 2011, o que pode indicar que o grupo das instituições que aderiram ao SiSU no primeiro semestre de 2011 o fizeram sob influencia das que assinaram o termo de adesão em 2010 e, em seguida, influenciaram as que o fizeram no final de 2011.

Mais uma vez reforça-se a necessidade de uma investigação mais profunda sobre o fenômeno da adesão ao SiSU por parte das instituições públicas de ensino, que certamente envolve muitos outros componentes não considerados na análise apresentada. Entretanto, os dados estudados sugerem que há algum nível de relação entre a rede de comunicação da informação estabelecida entre instituições públicas de ensino e a adoção de práticas como a adesão ao Sistema de Seleção Unificado.

## Riscos e restrições

Os riscos e restrições identificados que podem afetar a execução desse trabalho estão relacionados principalmente às simplificações adotadas no desenvolvimento do software de apoio para a coleta dos dados. Detalhes da implementação podem ser encontrados no ANEXO I deste trabalho.

A lista de nomes utilizada para a identificação das pessoas mencionadas nas portarias não é exaustiva. Corre-se o risco de existirem pessoas cujo nome seja composto de palavras as quais nenhuma esteja listada no arquivo utilizado. Caso isso ocorra, essa pessoa não será detectada pelo software e não será considerada na análise.

A diagramação do Diário Oficial da União (DOU) não é uniforme e isso pode gerar perda de informação. A disposição padrão das informações no DOU é feita em três colunas de texto. A titulo de ilustração, levantou-se a quantidade exata de páginas que não obedecem a esse padrão na edição do DOU de 28/05/2013. 79 de 440 páginas não seguem o padrão de três colunas. (17%). Entretanto, a existência de outros padrões em algumas páginas provoca inconsistências na extração das informações. O efeito prático dessa diferença de diagramação é que o componente de software, por não ter condições de tratar essa diferença, extrai os textos de forma desordenada, podendo causar ruídos na posterior criação de relacionamentos e identificação de redes sociais.

RESSLER (2006) discute as falhas da coleta de dados e sua influencia nas redes sociais extraídas desses dados:

“Um problema comum para modeladores é a questão dos dados. Qualquer trabalho acadêmico é tão bom quanto os dados utilizados, não importando os tipos de métodos avançados aplicados. Modeladores frequentemente não contam com os melhores dados. [...] Muitos modelos são criados sem dados ou com dados incompletos. A implicação disso é que os resultados podem ser potencialmente enganosos.” RESSLER (2006) (tradução do autor)

Por fim, um risco associado à execução do software em si está relacionado aos recursos computacionais exigidos. Nos testes preliminares, o software exigiu mais memória RAM do que o disponível na máquina utilizada, fazendo o processo ser interrompido antes do fim. Esse comportamento foi detectado em dias nos quais houve um numero acima da média de publicações no Diário Oficial. Uma possível solução para esse problema consiste em um pré-processamento desses arquivos, no qual seria feita um particionamento das informações em vários arquivos menores, passiveis de serem processados usando o hardware atual. Outra alternativa seria a utilização de hardware mais poderoso, já que o atual conta com apenas 2GB de memória disponível.

## Cronograma proposto

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov |
| Levantamento Bibliográfico | |  |  |  |  |  |  |  |  |  |  |  |
| Desenvolvimento do software de mineração | |  |  |  |  |  |  |  |  |  |  |  |
|  | Modulo de download do D.O.U. |  |  |  |  |  |  |  |  |  |  |  |
|  | Modulo de identificação de publicações |  |  |  |  |  |  |  |  |  |  |  |
|  | Modulo de identificação de órgãos |  |  |  |  |  |  |  |  |  |  |  |
|  | Modulo de identificação de pessoas |  |  |  |  |  |  |  |  |  |  |  |
|  | Modulo de persistência em banco de dados |  |  |  |  |  |  |  |  |  |  |  |
|  | Modulo de geração dos arquivos do pajek |  |  |  |  |  |  |  |  |  |  |  |
| Extração das redes | |  |  |  |  |  |  |  |  |  |  |  |
|  | Definição dos cortes e filtros a serem usados |  |  |  |  |  |  |  |  |  |  |  |
|  | Execução do software |  |  |  |  |  |  |  |  |  |  |  |
|  | Geração dos arquivos do pajek |  |  |  |  |  |  |  |  |  |  |  |
| Análise das redes extraídas | |  |  |  |  |  |  |  |  |  |  |  |

## Referencias

ALBAGLI, S.; MACIEL, M. L. Informação e conhecimento na inovação e no desenvolvimento local. Ciência da Informação, Brasília, v. 33, n.3, p. 9-16, set./dez. 2004.

BUFREM, Leilah Santiago ; ARBOIT, A. E. ; SORRIBAS, T. V. . Diálogo entre a teoria do Círculo de Bakhtin e a ciência da informação. Ciência da Informação (Online), v. 40, p. 145-159, 2011. Disponivel em <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1939/1404> Acesso em 30/05/2013

CIVIC COMMONS. Sharing technology for public good. Disponivel em < <http://wiki.civiccommons.org/Initiatives#Model_Open_Government_Initiatives>> Acesso em 10/05/2013

CONFERENCE REPORT ON H.R. 2658, DEPARTMENT OF DEFENSE APPROPRIATIONS ACT, 2004 Disponível em: <<http://www.gpo.gov/fdsys/pkg/CREC-2003-09-24/html/CREC-2003-09-24-pt1-PgH8500.htm>> Acessado em 21/05/2013

COSTA, W. J. V. da; PINHEIRO, M. M. K. Redes sociais e compartilhamento de informação e conhecimento em aglomerações produtivas. Inf. Inf., Londrina, v. 12, n. especial, p. 1-23, dez. 2007.

COURTIAL, Jean-Pierre e LAW, John. Author(s): A co-word study of artificial intelligence. Social Studies of Science,Vol. 19, No. 2 (May, 1989), pp. 301-311. Disponível em <http://www.jstor.org/stable/pdfplus/285145.pdf?acceptTC=true> Acessado em 18/05/2013

CROSS, Rob and PARKER, Andrew. The Hidden Power of Social Networks. Understanding How Work Really Gets Done in Organizations. Harvard Business School Press, 2004

CUNNINGHAM, et al. Text Processing with GATE (Version 6). University of Sheffield Department of Computer Science. 15 April 2011. ISBN 0956599311 Disponível em <http://gate.ac.uk/sale/tao/tao.pdf>

Fernandes, Jorge Henrique Cabral . Segurança e Defesa Cibernéticas para Reduzir Vulnerabilidades nas Infraestruturas Críticas Nacionais. 2012. Disponível em <http://www.eme.eb.mil.br/ceeex/public/arquivos/nep2012/NEP\_CEEEx\_Jorge\_Fernandes\_2012.pdf>

KARTHIKA S e BOSE S, A comparative study of social networking approaches in identifying the covert nodes. International Journal on Web Service Computing (IJWSC), Vol.2, No.3, September 2001. Disponível em <<http://airccse.org/journal/jwsc/papers/0911wsc06.pdf>> Acessado em 18/02/2013

KREBS, Valdis E. “Mapping Networks of Terrorist Cells,” Connections, Vol. 24, No. 3,pp.43-52, 2002. Disponível em <http://vlado.fmf.uni-lj.si/pub/networks/doc/Seminar/Krebs.pdf> Acessado em: 19/05/2013

LAZARSFELD, Paul F., BERNARD BERELSON, and HAZEL GAUDET. The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign. 1944. Disponível parcialmente em < http://books.google.com.br/books/about/The\_People\_s\_Choice.html?id=S-lnIFR02FIC&redir\_esc=y>

LASSWELL, Harold D, Propaganda Technique in the World War (New York: Alfred A. Knopf, 1927), 1 14-1 5; Merriam, New Aspects. Disponível parcialmente em < http://books.google.com.br/books/about/Propaganda\_Technique\_in\_the\_World\_War.html?id=e5PtAAAAMAAJ&redir\_esc=y>

LE COADIC, Yves-François. A ciência da informação. 2. ed. Brasilia: Brinquet de Lemos, 2004. 124 p. Disponível em <http://www.restaurabr.org/siterestaurabr/CICRAD2011/M1%20Aulas/M1A3%20Aula/20619171-le-coadic-francois-a-ciencia-da-informacao.pdf>

MARQUES, Eduardo Cesar. Redes sociais e poder no Estado brasileiro. Trabalho originalmente apresentado no XXVIII Encontro da Anpocs. 2004

MARTELETO, Regina Maria; SILVA, Antônio Braz de Oliveira e . Redes e capitais sociais: o enfoque da informação para o desenvolvimento local. CIÊNCIA DA INFORMAÇÃO. Brasíilia, v. 33, n. 3, p.41-49, set./dez. 2004.

MEC, Ministério da Educação. Portaria Normativa MEC No - 02, de 26 de janeiro de 2010.

MEIRELLES, Hely Lopes. Direito Administrativo Brasileiro. 2013 pág. 654

MORAES, Alexandre de. Direito Constitucional Administrativo, 2002, p. 104

POLANCO Xavier. (2005) "Co-Word Analysis Revisited: Modelling Co-Word Clusters in Terms of Graph Theory", Proceedings of the 10° International Conference on Scientometrics an d Informe-tries, July 24-28, Stockholm, Sweden.

RESSLER, Steve. “Social Network Analysis as an Approach to Combat Terrorism: Past,Present, and FutureResearch,”Homeland Security Affairs,Vol. 2, No. 2 (July, 2006), 1-10. Acesso em <18/05/2013> Disponivel em <http://www.hsaj.org/?fullarticle=2.2.8>

RODRIGUEZ, Jose. 2004. The March 11th terrorist network: In its weakness lies its strength. Department of Sociology Analysis of Organizations: University of Barcelona.Disponivel em <http://www.ub.edu/epp/wp/11m.PDF> Acessado em 16/05/2013

HANNEMAN, Robert A.; RIDDLE, Mark. Introduction to Social Network Methods. Riverside, CA: University of California, Riverside Disponivel em <http://faculty.ucr.edu/~hanneman/>: 2005. Acesso em: 10/05/2013.

MPOG, Ministério do Planejamento, Organização e gestão. Plano Plurianual 2012-2015, Mensagem Presidencial. Disponível em <http://www.planejamento.gov.br/secretarias/upload/Arquivos/spi/PPA/2012/mensagem\_presidencial\_ppa.pdf> Acesso em 29/05/2013

NOOY, Wouter; MRVAR, Andrej; BATAGELJ, Vladimir. Exploratory Network Analysis with Pajek. Cambridge University Press, Cambridge: 2005.

SACERDOTE, Helena Célia de Souza. Análise da mediação em educação online sob a ótica da Análise de Redes Sociais: o caso do curso de Especialização em Gestão da Segurança da Informação e Comunicações. 2013. 145 f., il. Dissertação (Mestrado em Ciência da Informação)—Universidade de Brasília, Brasília, 2013.

SAGEMAN, Marc. Understanding Terror Networks. UPCC book collections on Project MUSE. University of Pennsylvania Press, 2011 Disponível parcialmente em <<http://books.google.com.br/books?id=iCoYDUv63L8C&dq=>> Acessado em 15/05/2013

SHANNON, C. e WEAVER, W., The mathematical Theory of Communication, University of Illinois Press, 1949.

SCHIESSL, José Marcelo. Descoberta de conhecimento em texto aplicada a um sistema de atendimento ao consumidor. 2007. 106 f. Dissertação (Mestrado em Ciência da Informação)-Universidade de Brasília, Brasília, 2007.

SIORG. Site oficial. Acesso em 18/05/2013. Disponível em <<http://www.siorg.redegoverno.gov.br/>>

SINGHAL Ayush, SRIVASTAVA Jaideep, "Data Extract: Mining Context from the Web for Dataset Extraction," International Journal of Machine Learning and Computing vol. 3, no. 2, pp. 219-223, 2013. Disponível em <<http://www.ijmlc.org/papers/306-K0019.pdf>>

SOUTO, Leonardo Fernandes. Mediação em serviços de disseminação seletiva de informações no ambiente de bibliotecas digitais federadas. 2008. Tese (Doutorado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2008. Disponível em: <http://www.teses.usp.br/teses/disponiveis/27/27151/tde-13072009-184314/>. Acesso em: 2013-06-03.

TAVARES, Rosemeire Barbosa. O uso de técnicas de pesquisa participatória na comunicação da informação em comunidades, levando em conta letramento informacional e trabalho colaborativo para desenvolvimento de cidadania. 2011. 257 f., il. Tese (Doutorado em Ciência da Informação)— Universidade de Brasília, Brasília, 2011. Disponível em <http://repositorio.unb.br/bitstream/10482/10410/1/2011\_RosemeireBarbosaTavares.pdf>

VERGUEIRO, W.; SUGAHARA, C. R. Aspectos conceituais e metodológicos de redes sociais e sua influência no estudo de fluxos de informação. Revista Digital de Biblioteconomia & Ciência da Informação, Campinas, v. 7, n. 2, p. 102-117, jan./jun. 2010.

WATZLAWICK, P., Beavin, J., & Jackson, D. D. (1967). Pragmatics of Human Communication. A study of interactional patterns, pathologies, and paradoxes. New York: Norton. Disponível parcialmente em < http://books.google.com.br/books/about/Pragmatics\_of\_Human\_Communication.html?id=OqFmcgAACAAJ&redir\_esc=y>

WASSERMAN, Stanley; FAUST, Katherine. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.

WIENER, N. Cybernetics or Control and Communication in the Animal and The Machine. Hermann. 1948. Disponivel parcialmente em < http://books.google.com.br/books/about/Cybernetics.html?id=NnM-uISyywAC&redir\_esc=y>

ZINS, Chaim. Knowledge map of information science: Research Articles. Journal of the American Society for Information Science and Technology. 58(4):526–535, 2007. 10pp.

ANEXO I



ANEXO II

