UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Caracterização de redes complexas de transações comerciais brasileiras construídas mediante notas fiscais eletrônicas

Bruno Angeli Calza

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Bruno Angeli Calza

Caracterização de redes complexas de transações comerciais brasileiras construídas mediante notas fiscais eletrônicas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
EXEMPLAR DE DEFESA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho

Coorientador: Prof. Dr. Luiz Gustavo de Andrade Alves

USP – São Carlos
Novembro de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

A582c Angeli Calza, Bruno
 Caracterização de redes complexas de transações
 comerciais brasileiras construídas mediante notas
 fiscais eletrônicas / Bruno Angeli Calza; orientador
 André Carlos Ponce de Leon Ferreira de Carvalho;
 coorientador Luiz Gustavo de Andrade Alves. -- São
 Carlos, 2019.
 90 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
 em Mestrado Profissional em Matemática, Estatística
 e Computação Aplicadas à Indústria) -- Instituto de
 Ciências Matemáticas e de Computação, Universidade
 de São Paulo, 2019.

 1. Nota fiscal eletrônica. 2. Redes complexas. 3.
 Redes comerciais. I. Carlos Ponce de Leon Ferreira
 de Carvalho, André, orient. II. Gustavo de Andrade
 Alves, Luiz, coorient. III. Título.

Bruno Angeli Calza

Characterization of complex networks of Brazilian business transactions built through electronic invoices

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho

Co-advisor: Prof. Dr. Luiz Gustavo de Andrade Alves

USP – São Carlos
November 2019

AGRADECIMENTOS

Gostaria de agradecer ao professor André pela confiança e a me motivar a seguir em frente; ao Luiz Gustavo pelos seus direcionamentos e colocações pertinentes; a Arquivêi, em especial o Bruno Oliveira, pela oportunidade; aos times de Engenharia de Dados e Ciência de Dados da Arquivêi, base para o desenvolvimento deste trabalho; a minha família por terem me proporcionado condições para meus estudos e desenvolvimento; e a minha namorada Maria Laura pelo carinho, atenção e todo apoio. Sem vocês nada disso teria sido possível.

RESUMO

CALZA, B. A. **Caracterização de redes complexas de transações comerciais brasileiras construídas mediante notas fiscais eletrônicas**. 2019. 90 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Através de um banco de dados de oitenta e quatro milhões de notas fiscais eletrônicas coletadas através de um sistema de auxílio ao *compliance* tributário digital, são modeladas os seguintes tipos de redes comerciais brasileiras: redes de transações entre empresas, redes de transações entre grupos empresariais, redes de transações entre ramos de atividade econômica e redes de transações entre cidades. Através de técnicas de análise de rede como menor caminho médio, distribuição de graus, coeficiente de agrupamento, medidas de centralidade e detecção de comunidades, procura-se entender as propriedades estruturais que rege a economia brasileira. Este trabalho serve como alicerce para o desenvolvimento de novos estudos relacionados ao entendimento da economia brasileira por meio de redes complexas, abrindo portas, por exemplo, para o estudo de fenômenos como instabilidade de mercado e surgimento de monopólios.

Palavras-chave: NF-e, Nota fiscal eletrônica, Redes complexas, Redes comerciais, Redes econômicas.

ABSTRACT

CALZA, B. A. **Characterization of complex networks of Brazilian business transactions built through electronic invoices.** 2019. 90 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

From a database of eighty four million electronic invoices collected through a software system that helps with digital tax compliance, the following types of Brazilian commercial networks are modeled: inter-company transaction networks, inter-group transaction networks business networks, cross-industry transaction networks and inter-city transaction networks. Through network analysis techniques such as shortest average path, degree distribution, clustering coefficient, centrality measures and community detection, we seek to understand the structural properties that govern the Brazilian economy. This work serves as a foundation for the development of new studies related to the understanding of the Brazilian economy through complex networks, opening doors, for example, to the study of phenomena such as market instability and the emergence of monopolies.

Keywords: e-invoice, Electronic invoice, Complex networks, Business networks, Corporate networks.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de desdobramento do CNAE em subclasse do CNAE	30
Figura 2 – Composição do NCM	31
Figura 3 – Como o Arquivel consulta as notas	31
Figura 4 – A principal tela para gestão de NF-e do sistema Arquivel	32
Figura 5 – Exemplos de grafos não direcionado, direcionado e ponderado	34
Figura 6 – Distribuição de grau de uma rede aleatória com 10 000 vértices com probabilidade $p = 0.1$	37
Figura 7 – Evolução do grafo regular até o grafo de pequeno mundo	39
Figura 8 – Distribuição de grau de uma rede <i>Watts-Strogatz</i> com 10 000 vértices com probabilidade $p = 0.5$, que se inicia com grau $k = 100$	40
Figura 9 – Distribuição de grau de uma rede <i>Barabási-Albert</i> com 10 000 vértices e 100 arestas conectadas a cada passo.	42
Figura 10 – O maior componente da rede de colaboração do <i>Santa Fe Institute</i>	47
Figura 11 – Exemplo do cálculo da modularidade para a rede considerando duas comunidades C_1 e C_2 em que $C_1 = \{A, B, C, D\}$ e $C_2 = \{E, F, G, H, I\}$. O grafo possui 12 arestas, dos quais 5 pertencem a C_1 e 6 a C_2 ; e 11 meia arestas incidem em nós de C_1 e 13 meia arestas incidem em nós de C_2	50
Figura 12 – As 4 modelagens propostas para o estudo de redes comerciais brasileiras	57
Figura 13 – Análise de cobertura territorial	60
Figura 14 – Gráfico de barra de CNAEs e divisão de CNAEs mais frequentes, tanto para fornecedores quanto para compradores; e a porcentagem acumulada.	62
Figura 15 – Distribuição de graus das redes 3 (linha 1), 6 (linha 2), e 7 (linha 3), para redes direcionadas e não direcionadas. A primeira coluna traz a distribuição da rede não direcionada. A segunda coluna traz a distribuição dos graus de saída. A terceira coluna traz a distribuição dos graus de entrada. As últimas três colunas trazem as funções de distribuição de probabilidade e distribuição acumulada complementar resultantes da utilização da biblioteca (ALSTOTT; BULLMORE; PLENZ, 2014). Os eixos estão na escala <i>log-log</i>	65
Figura 16 – Cálculo da centralidade de grau de autovetores para as redes 6 e 7. Expõe as 10 cidades e os 10 CNAEs mais centrais para ambos aspectos.	66
Figura 17 – Análise dos tamanhos das comunidades, obtidos pela aplicação do algoritmo de detecção de comunidades para a rede comercial empresarial (1º trimestre 2019)	67

LISTA DE QUADROS

Quadro 1 – Exemplos de diferentes tipos de redes econômicas estudadas na literatura .	54
Quadro 2 – As cinco maiores comunidades identificadas e suas características. Contém o identificador da comunidades, a quantidade de membros, o valor total (em R\$) comercializado dentro da comunidade (e a porcentagem do total comercializada internamente, excluindo as transações entre comunidades), os NCMs e capítulos de NCMs mais comercializados (e suas porcentagens acumuladas) e os CNAEs mais frequentes.	69

LISTA DE TABELAS

Tabela 1 – Amostra real que exemplifica cinco registros selecionados para o estudo. Cada registro implica uma transação comercial de um produto entre duas empresas: a empresa emissora da nota que é quem vendeu o produto; e o destinatário da nota que é quem comprou o produto.	56
Tabela 2 – Estatísticas simples da base de dados sob estudo	59
Tabela 3 – Os 10 capítulos de NCMs mais comercializados	61
Tabela 4 – Características das redes comerciais brasileiras, rede de coautoria de matemáticos e rede de atores de filmes	64
Tabela 5 – Lista de interesse de CFOPs separada por um especialista fiscal com objetivo de restringir as operações que qualificassem transação comercial envolvendo recebíveis. Em outras palavras, exclui-se todas as operações referentes a manipulação fiscal sem rendimentos diretos (e.g. realocação de estoque, bonificação, doação, e etc.), restando apenas operações com natureza rentável (compras, vendas, industrialização, e afins).	81

LISTA DE ABREVIATURAS E SIGLAS

CFOP	Código Fiscal de Operações e Prestações
CNAE	Classificação Nacional de Atividades Econômicas
CONFAZ	Conselho Nacional de Política Fazendária
DANFE	Documento Auxiliar de Nota Fiscal Eletrônica
e-gov	do inglês <i>eletronic government</i>
EFD	Escrituração Fiscal Digital
ENAT	Encontro Nacional de Administradores Tributários
GTIN	<i>Global Trade Item Number</i>
NCM	Nomenclatura Comum do Mercosul
NF-e	Nota Fiscal Eletrônica
ONU	Organização das Nações Unidas
SaaS	Software as a Service
SPED	Sistema Público de Escrituração Digital
TIC	Tecnologias de Informação e Comunicação
UF	Unidade Federativa
XML	eXtensible Markup Language

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Objetivos	22
1.2	Estrutura da dissertação	22
2	COMPLIANCE TRIBUTÁRIO DIGITAL	23
2.1	Considerações iniciais	23
2.2	Motivação do governo	24
2.3	Obrigações fiscais	26
2.4	Detalhamento da Nota Fiscal Eletrônica	28
2.5	A Arquivei	30
2.6	Considerações finais	32
3	REVISÃO SOBRE REDES COMPLEXAS	33
3.1	Conceitos básicos	33
3.2	Modelos de redes complexas	34
3.2.1	<i>Modelo Erdős-Rényi</i>	35
3.2.2	<i>Modelo Watts-Strogatz</i>	38
3.2.3	<i>Modelo Barabási-Albert</i>	40
3.3	Medidas e métricas	42
3.3.1	<i>Centralidade de grau</i>	42
3.3.2	<i>Centralidade de proximidade (closeness)</i>	43
3.3.3	<i>Centralidade de intermediação (betweenness)</i>	43
3.3.4	<i>Centralidade do autovetor</i>	43
3.3.5	<i>PageRank</i>	44
3.3.6	<i>Excentricidade</i>	46
3.3.7	<i>Coeficiente de clusterização</i>	46
3.3.8	<i>Assortatividade</i>	46
3.4	Comunidades	46
3.4.1	<i>Deteccção de comunidades</i>	47
3.4.2	<i>Métodos para detecção de comunidades</i>	48
3.4.2.1	<i>Algoritmos divisivos</i>	48
3.4.2.2	<i>Algoritmos espectrais</i>	48
3.4.2.3	<i>Algoritmos dinâmicos</i>	49

3.4.2.4	<i>Métodos baseados em inferência estatística</i>	49
3.4.2.5	<i>Métodos baseados em otimização</i>	49
3.5	Considerações finais	50
4	CARACTERIZAÇÃO DE REDES COMERCIAIS	53
4.1	Considerações iniciais	53
4.2	Metodologia	54
4.3	Análise descritiva	59
4.4	Resultados da aplicação de técnicas de redes complexas	60
4.4.1	<i>As métricas $\langle k \rangle$, L e C</i>	61
4.4.2	<i>Distribuição de graus e centralidades</i>	63
4.4.3	<i>Comunidades</i>	67
4.5	Considerações finais	68
5	CONCLUSÃO	71
5.1	Trabalhos futuros	72
REFERÊNCIAS		73
APÊNDICE A	LISTA DE INTERESSE DE CFOPS	81

INTRODUÇÃO

É difícil não se desconcertar com o crescente aumento de complexidade das tarefas realizadas em nosso cotidiano. Mesmo que não se tenha definido o que é complexidade, pode-se dizer que sistemas como interações atômicas, relações pessoais e transações financeiras são complexos. O que torna estes sistemas similares é o fato de que cada um é uma coleção de partes que interagem entre si. E o nome que se dá a disciplina que estuda os fenômenos que emergem destas coleções é sistemas complexos ou ciência da complexidade (JOHNSON, 2007).

Espera-se, devido sua natureza abrangente, que o estudo de sistemas complexos envolva diversas disciplinas da ciência, e até mesmo, de áreas tão distintas como engenharia, medicina e administração. Existe, porém, uma disciplina cujo foco principal é o estudo de sistemas complexos modelados como redes (grafos). A tais sistemas é dado o nome de redes complexas. Esta disciplina que estuda redes complexas, chamada de ciência das redes, foi resultado de esforços realizados em diversas áreas, como teoria dos grafos, ciências sociais (WELLMAN, 2008) e física (ALBERT; BARABÁSI, 2002). Em 2005, o *National Research Council* definiu o termo ciência das redes como “o estudo da representação de fenômenos físicos, biológicos e sociais como redes que levam a modelos preditivos desses fenômenos”¹ (COUNCIL, 2005, tradução nossa).

Dentro do contexto econômico existem diversos estudos realizados recentemente que envolvem a aplicação de sistemas complexos, o que indica um grande interesse de pesquisadores pelo tema. Em (GARAS *et al.*, 2010) há uma discussão de como se dá a propagação de crises econômicas mundiais; em (BENEDICTIS; TAJOLI, 2011) é feita uma comparação ao longo do tempo da estrutura da rede mundial de comércio, onde se detecta e interpreta os padrões de laços comerciais entre os países; e em (BHATTACHARYA *et al.*, 2008) são feitas análises de robustez, do crescimento da força comercial e do coeficiente *rich-club*. Tais estudos se referem a

¹ Network science consists of the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena

rede internacional de comércio.

Um ótimo exemplo de sistema complexo, dentro do contexto da ciência econômica e redes comerciais, é o conjunto de transações realizadas entre empresas. No Brasil, tais transações devem ser registradas através da Nota Fiscal Eletrônica (NF-e). A NF-e é um documento fiscal de emissão obrigatória pela maior parte das empresas brasileiras, cujo propósito é documentar uma operação de circulação de mercadorias ou uma prestação de serviços. Este documento é um arquivo XML semi-estruturado que contém diversas informações que vão desde informações básicas das partes envolvidas na transação comercial, até detalhes como o valor de frete, informações sobre combustível, data de vencimento de medicamentos, impostos e detalhamento dos itens comprados.

1.1 Objetivos

Este trabalho se propõe, através da NF-e e da modelagem desta como um grafo, se fazer uso de análises de redes complexas para caracterizar redes comerciais da economia brasileira. Devido ao grande volume de documentos disponíveis para uso, muitos algoritmos se tornam inviáveis de serem aplicados. Um segundo objetivo é o estudo e aplicação de algoritmos que consigam escalar a grande quantidade de nós e arestas da rede. E por fim, se têm, também, como objetivo a aproximação da academia com o mercado de trabalho, que é um dos propósitos do programa de mestrado profissional.

1.2 Estrutura da dissertação

Este trabalho está estruturado da seguinte maneira: o [Capítulo 1](#) oferece uma contextualização de redes complexas e apresenta os objetivos deste trabalho; o [Capítulo 2](#) discorre sobre o universo tributário, as dores das empresas no que diz respeito a *compliance tributário*, a NF-e e o sistema Arquivo; o [Capítulo 3](#) faz uma revisão dos principais conceitos, técnicas e medidas utilizados para caracterizar redes complexas; [Capítulo 4](#) apresenta os resultados da aplicação das técnicas em redes comerciais; e, por fim, o [Capítulo 5](#) oferece conclusões.

COMPLIANCE TRIBUTÁRIO DIGITAL

2.1 Considerações iniciais

O termo *compliance* é derivado do verbo da língua inglesa *to comply*, cujo significado é obedecer a uma ordem, regra ou pedido (COMPLY, 2019). Um esclarecimento maior é dado por (MANZI, 2008), que se refere a *compliance* como o dever de cumprir, de estar em conformidade e fazer cumprir regulamentos internos e externos que foram impostos às atividades de uma determinada empresa. As práticas de *compliance* realizadas pelas empresas acabam por, não só mitigar os riscos relacionados aos descumprimentos de regulações e leis, mas também, reduzir os riscos relacionados a integridade, ética e imagem da empresa frente a sociedade. Existem diversas frentes regulatórias em que o *compliance* se faz necessário, como por exemplo: ambiental, trabalhista, segurança, e outras; este trabalho foca no *compliance* tributário, que é o conjunto de práticas aplicadas pelas empresas para estar em conformidade com relação a tributação.

O problema da conformidade tributária, do ponto de vista dos governos ao redor do mundo, caracterizando e explicando os padrões observados de descumprimento de impostos e, posteriormente, sendo capaz de encontrar maneiras de reduzir esse descumprimento, é de importância vital. Dentro deste contexto de conformidade tributária, é que se estão inseridos documentos fiscais como a NF-e, principal objeto de estudo deste trabalho.

Este capítulo trata da motivação do governo brasileiro por trás da implementação do sistema de NF-e, as principais obrigações que giram em torno deste sistema, um detalhamento dos principais dados contido na NF-e, e a solução proposta pela Arquivei para o auxílio ao *compliance* tributário. Esse contexto é essencial para que se entenda o universo NF-e, e consequentemente facilitar a exploração da NF-e como um dado no qual é desejado encontrar aplicações práticas através de análises de redes complexas.

2.2 Motivação do governo

É bem evidente a participação da tecnologia e o tamanho de sua influência no mundo moderno. Essa influência não se deu apenas na esfera privada, mas também na esfera pública. Há pelo menos quase duas décadas já vem sendo discutido o impacto da modernização de setores públicos (HAMMOND, 2001). Dessa modernização, ou mais especificamente do uso de Tecnologias de Informação e Comunicação (TIC), na administração pública e processo de tomada de decisão política, cunhou-se o termo governo eletrônico, do inglês *eletronic government* (e-gov). De acordo com (NATIONS, 2014), o governo eletrônico se refere ao uso e aplicação de tecnologias de informação na administração pública visando otimizar e integrar fluxos de trabalho e processos. Isso torna a gestão de dados e informações mais eficientes, melhorando a prestação de serviços públicos, bem como a expansão dos canais de comunicação para engajamento e empoderamento das pessoas. Vários governos mundiais possuem estratégias sólidas de como ampliar o uso de tecnologia em seus serviços. O governo brasileiro não é exceção, e pode-se afirmar que sua revolução digital iniciou-se em 2000, com o Decreto Presidencial de 3 de abril de 2000 (BRASIL, 2000), no qual foi criado um Grupo de Trabalho Interministerial com a finalidade de examinar e propor políticas, diretrizes e normas relacionadas com as novas formas eletrônicas de interação. De acordo com o relatório da Organização das Nações Unidas (ONU) de 2018 (NATIONS, 2018), o Brasil ocupa a 44ª posição no índice geral de desenvolvimento de governo eletrônico. Uma descrição das estratégias de governo eletrônico implementadas ao longo dos anos pelo governo brasileiro podem ser encontradas na página Governo Digital do Ministério da Economia¹.

Dentro deste contexto de modernização, em 2003 a Emenda Constitucional nº 42 introduziu o Inciso XXII ao artigo 37 da Constituição Federal, visando uma atuação de forma integrada entre as administrações tributárias da União, Estados, Distrito Federal e Municípios (BRASIL, 2003):

XXII - as administrações tributárias da União, dos Estados, do Distrito Federal e dos Municípios, atividades essenciais ao funcionamento do Estado, exercidas por servidores de carreiras específicas, terão recursos prioritários para a realização de suas atividades e atuarão de forma integrada, inclusive com o compartilhamento de cadastros e de informações fiscais, na forma da lei ou convênio.

Tal emenda é considerada um marco do cenário de integração e modernização da Administração Tributária. A fim de atender este inciso, criou-se o Encontro Nacional de Administradores Tributários (ENAT), cujo objetivo está descrito em sua própria página *web* (ENAT,):

O objetivo principal do fórum é o de buscar soluções conjuntas das três esferas de governo que promovam uma maior integração administrativa,

¹ Disponível em: <<https://www.governodigital.gov.br/EGD/historico-1/historico>>. Acessado em 31 de Agosto de 2019.

a padronização e a melhoria na qualidade das informações; a racionalização de custos e da carga de trabalho operacional no atendimento; uma maior eficácia da fiscalização; a possibilidade de realização de ações fiscais coordenadas e integradas; melhoria no intercâmbio de informações fiscais entre as diversas esferas governamentais; o cruzamento de dados padronizados em larga escala e a uniformização de procedimentos.

Nos dias 25, 26 e 27 de agosto de 2005, aconteceu em São Paulo o II Encontro Nacional de Administradores Tributários (ENAT), onde se destacam as assinaturas dos seguintes protocolos de intenções: Sistema Público de Escrituração Digital (SPED) e NF-e. As motivações de tais protocolos são (ENAT, 2005):

Nota Fiscal Eletrônica

O primeiro protocolo visa facilitar o cumprimento das obrigações acessórias pelos contribuintes. Propiciar aumento da competitividade das empresas brasileiras pela racionalização das obrigações acessórias (redução do “custo Brasil”), em especial a dispensa da emissão e guarda de documentos em papel. Às administrações tributárias, a Nota Fiscal Eletrônica permite padronização e melhoria na qualidade das informações, racionalização de custos e maior eficácia da fiscalização.

Sistema Público de Escrituração Digital

Outro protocolo prevê o estabelecimento do Sistema Público de Escrituração Digital, que propicia aos contribuintes simplificação e racionalização de obrigações acessórias, agilização dos procedimentos sujeitos a controle das administrações tributárias e redução de custos com armazenamento de documentos em papel. Os fiscos ganham maior integração administrativa, padronização e melhor qualidade das informações, racionalização de custos e maior eficácia da fiscalização.

Após a assinatura desses protocolos de intenções, o Conselho Nacional de Política Fazendária (CONFAZ), órgão do Ministério da Fazenda, cuja finalidade é promover ações necessárias à elaboração de políticas e harmonização de procedimentos e normas inerentes ao exercício da competência tributária dos Estados e do Distrito Federal (CONFAZ, 2015), aprovou o Ajuste SINIEF² 07, de 5 de outubro de 2005, que institui a legislação base da NF-e, delineando as principais regras do documento (CONFAZ, 2005):

Cláusula primeira Fica instituída a Nota Fiscal Eletrônica - NF-e, modelo 55, que poderá ser utilizada pelos contribuintes do Imposto sobre Produtos Industrializados - IPI ou Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre a Prestação de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação - ICMS em substituição:

I - à Nota Fiscal, modelo 1 ou 1-A;

² SINIEF é Sistema Nacional Integrado de Informações Econômico-Fiscais gerido pelo CONFAZ

II - à Nota Fiscal de Produtor, modelo 4.

§ 1º Considera-se Nota Fiscal Eletrônica - NF-e o documento emitido e armazenado eletronicamente, de existência apenas digital, com o intuito de documentar operações e prestações, cuja validade jurídica é garantida pela assinatura digital do emitente e autorização de uso pela administração tributária da unidade federada do contribuinte, antes da ocorrência do fato gerador.

E dando sequência a esse processo, em 22 de janeiro de 2007, instituiu-se o Sistema Público de Escrituração Digital (SPED) através do Decreto Nº 6.022 ([BRASIL, 2007](#)):

Art. 1º Fica instituído o Sistema Público de Escrituração Digital - Sped.

Art. 2º O Sped é instrumento que unifica as atividades de recepção, validação, armazenamento e autenticação de livros e documentos que integram a escrituração contábil e fiscal dos empresários e das pessoas jurídicas, inclusive imunes ou isentas, mediante fluxo único, computadorizado, de informações.

Dado este contexto histórico de como foi o processo de implementação da NF-e e do SPED, fica mais que evidente as motivações do governo brasileiro para implementação de tais sistemas: desburocratização, integração entre as administrações tributárias, melhoria na qualidade da informação, eficácia na fiscalização, cruzamento de dados padronizados, uniformização de procedimentos, redução do “custo Brasil” e redução de custos de armazenamento.

2.3 Obrigações fiscais

São muitos os compromissos de uma empresa brasileira perante ao Governo. Tais compromissos são obrigações previstas em lei perante a órgãos como Fisco Federal, Ministério do Trabalho e Previdência Social. As inúmeras declarações que são feitas mensalmente, trimestralmente e anualmente informam ao Governo questões relacionadas a apuração de impostos, receitas, despesas, distribuição societária, quantidade de funcionários e outras. São muitas as regras aplicadas a esse conjunto de obrigações, e essa regras variam de empresa a empresa dependendo de fatores como regime tributário e ramo de atividade. Entender os detalhes e nuances deste universo e se manter atualizado com as alterações da legislação é uma tarefa árdua que compete mais a um profissional de contabilidade ou direito. De uma forma simplificada este trabalho apresenta algumas das obrigações referentes ao universo fiscal.

A obrigatoriedade mais evidente é a da própria emissão da NF-e, aprovada pelo CONFAZ através do Protocolo ICMS 10 de 18 de abril de 2007. Neste protocolo estabeleceu a obrigatoriedade de utilização da NF-e prevista no Ajuste SINIEF 07/05 em substituição a Nota Fiscal, modelo 1 ou 1-A, para os estabelecimentos de atividades econômicas listadas no protocolo, que em sua maioria incluíam estabelecimentos de fabricação de cigarros e distribuição de

combustíveis líquidos. Aos poucos mais ramos de atividades foram sendo adicionados a essa lista. Em 3 de julho de 2009, através do Protocolo ICMS 42, a obrigatoriedade deixou de ser do estabelecimento e passou a ser da empresa como um todo, englobando todos os estabelecimento de forma único. Para tal processo, passou a ser utilizado a Classificação Nacional de Atividades Econômicas (CNAE) e as operações com os destinatários como critério da obrigatoriedade. O protocolo apresenta uma relação de códigos CNAE e os respectivos inícios da obrigatoriedade. A partir de 2011 já se podia considerar que a grande maioria das empresas brasileiras, salvo algumas exceções, já eram obrigadas a realizar a emissão da NF-e. A emissão do documento se dá através de uma integração com *Web service* oferecido pela Secretaria da Fazenda. Para facilitar a adesão, o Governo disponibilizou um *software* de emissão gratuito.

Após a emissão do documento, as empresas são obrigadas a armazená-los por um período (geralmente 5 anos) como aponta a décima cláusula do próprio Ajuste SINIEF 07/2005 que instituiu a NF-e:

Cláusula décima O emitente deverá manter a NF-e em arquivo digital, sob sua guarda e responsabilidade, pelo prazo estabelecido na legislação tributária, mesmo que fora da empresa, devendo ser disponibilizado para a Administração Tributária quando solicitado.

§ 1º O destinatário deverá verificar a validade e autenticidade da NF-e e a existência de Autorização de Uso da NF-e.

§ 2º O destinatário da NF-e também deverá cumprir o disposto no caput desta cláusula e, caso não seja contribuinte credenciado para a emissão de NF-e, poderá, alternativamente, manter em arquivo o DANFE relativo à NF-e da operação, o qual deverá ser apresentado à Administração Tributária, quando solicitado.

§ 3º O emitente de NF-e deverá guardar pelo prazo estabelecido na legislação tributária o DANFE que acompanhou o retorno de mercadoria não entregue ao destinatário e que contenha o motivo do fato em seu verso.

Fica evidente, também, pela cláusula, que a obrigatoriedade não se refere apenas ao emissor, mas também ao destinatário. Então, é preciso que de alguma forma o destinatário tenha acesso ao documento. E também, explicita que em algumas circunstância pode-se utilizar do Documento Auxiliar de Nota Fiscal Eletrônica (DANFE) que é um papel que acompanha o transporte do produto, representando a NF-e. O *blog* da empresa Oobj, uma das pioneiras a desenvolver soluções em torno de documentos fiscais eletrônicos, explica o motivo do armazenamento ([OOBJ, 2007](#)):

Porque é durante esse período que o fisco pode solicitar os documentos para conferência. Caso a empresa não possua as notas solicitadas, ela pode receber multas. Além disso, é dentro desse prazo (5 anos) que o Governo pode cobrar as dívidas atrasadas.

Outra obrigação importante que fica a cargo do destinatário é informar às unidades federadas o que ocorreu com a operação. Isso fica bem descrito no Ajuste SINIEF 5 ([CONFAZ, 2012](#)), de 30 de março de 2012, que descreve sobre a obrigatoriedade, por parte do destinatário, de registrar o acontecimento de pelo menos um dos seguintes eventos:

- I. confirmação do recebimento da mercadoria;
- II. declaração do não recebimento da mercadoria;
- III. desconhecimento da operação.

A princípio tais registros não são obrigatórios em todas as unidades federativas, e só rege estabelecimentos distribuidores de combustíveis, postos de combustíveis e transportadores revendedores retalhistas ([CONFAZ, 2013](#)). Tais registros devem ser feitos via *Web service*.

Por fim, é importante destacar a obrigatoriedade do processo de escrituração digital chamado Escrituração Fiscal Digital (EFD), instituído pelo CONFAZ através do Ajuste SINIEF 2, de 3 de abril de 2009 ([CONFAZ, 2009](#)). Pela EFD as empresas informam ao Fisco as mercadorias que entraram e saíram, cadastro de produtos, cadastro de fornecedores, dados de produção, estoque e a apuração dos impostos. Como aponta a própria lei, o arquivo digital da EFD deverá ser enviado até o quinto dia do mês subsequente ao encerramento do mês da apuração.

Vale ressaltar que são muitas as obrigações acessórias de uma empresa. Aqui foram destacadas apenas as mais relevantes para o contexto deste trabalho.

2.4 Detalhamento da Nota Fiscal Eletrônica

A Nota Fiscal Eletrônica é um arquivo digital eXtensible Markup Language (XML) semi-estruturado que contém diversas informações que vão desde informações básicas das partes envolvidas na transação comercial, como detalhes de valor de frete, informações sobre combustível, data de vencimento de medicamentos, impostos, detalhamento dos itens comprados, entre muitas outras. Todas informações que uma NF-e pode conter e a estrutura dos campos podem ser encontradas no Manual de Orientação do Contribuinte (MOC) ([ENCAT, 2015](#)), que define a NF-e da seguinte maneira:

A Nota Fiscal Eletrônica (NF-e) é um documento de existência exclusivamente digital, emitido e armazenado eletronicamente, com o intuito de documentar uma operação de circulação de mercadorias ou prestação de serviços, no campo de incidência do ICMS, cuja validade jurídica é garantida por duas condições necessárias: a assinatura digital do emissor e a Autorização de Uso fornecida pela administração tributária do domicílio do contribuinte.

Uma NF-e pode conter centenas de campos, porém sua maior parte é facultativa e seu preenchimento depende do contexto da natureza daquela operação. Nesta seção apenas será abordado as informações mais relevantes para a finalidade deste trabalho. A NF-e é composta por:

- **chave de acesso:** toda NF-e possui um campo que a identifica unicamente que é a chave de acesso. A chave de acesso é uma cadeia de caracteres de 44 dígitos formada pelas seguintes informações:
 - código da Unidade Federativa (UF) do emitente do documento fiscal;
 - ano e mês de emissão da NF-e;
 - CNPJ do emitente;
 - modelo do documento fiscal;
 - série do documento Fiscal;
 - número do documento Fiscal;
 - forma de emissão da NF-e;
 - código numérico que compõe a chave de acesso;
 - dígito verificador da chave de acesso.
- **data de emissão:** a data e hora em que o documento fiscal foi emitido.
- **tipo de operação:** se é uma operação de entrada ou saída de mercadoria.
- **valor:** valor total da NF-e.
- **dados do emitente:** dados cadastrais de quem emitiu o documento fiscal, como CNPJ, razão social, inscrição estadual, regime tributário, endereço e telefone.
- **dados do transportador:** dados cadastrais de quem é o transportador da mercadoria caso o mesmo já tenha sido definido no momento da emissão do documento.
- **dados do destinatário:** dados cadastrais de quem é o destinatário da mercadoria, como CNPJ, razão social, inscrição estadual, regime tributário, endereço e telefone.
- **CNAE:** A Classificação Nacional de Atividades Econômicas é um campo referente ao emissor e destinatário que requer uma certa atenção. A CNAE é uma forma de padronizar, em todo o território nacional, as atividades econômicas e facilitar a administração tributária do país (CONTA AZUL, 2018). A CNAE é um código de 7 dígitos, estruturado de forma hierárquica em partes como indica a figura 1, que delimita quais operações uma determinada empresa pode realizar. É uma informação extremamente relevante, pois permite analisar as relações entre os ramos de atividades. Essa informação é opcional na NF-e mas é possível obtê-la de fontes externas ao documento fiscal.

Figura 1 – Exemplo de desdobramento do CNAE em subclasse do CNAE

- **dados sobre os produtos:** uma NF-e pode conter diversos produtos que estão sendo comercializados. Cada um deles é descrito por uma série de campos em que os mais importantes são:
 - **Global Trade Item Number (GTIN):** é um padrão definido pela GS1³ que identifica unicamente o produto. É o antigo código de barras. Sua forma mais comum é de 13 dígitos, mas também é encontrado códigos de 8, 12 ou 14 dígitos.
 - **descrição do produto:** um texto que descreve o produto.
 - **unidade comercial:** a unidade do produto que está sendo comercializado. Esta informação é definida pelo critério do emissor e contém unidade como: litro, peça, caixa, entre outras.
 - **quantidade comercial:** a quantidade que está sendo comercializada.
 - **valor do produto:** o valor do produto. Contém o valor unitário e o valor total multiplicado pelas quantidades.
 - **Nomenclatura Comum do Mercosul (NCM):** é um código de 8 dígitos convencionado pelos países do Mercosul para reconhecer os bens, serviços e fatores produtivos negociados entre si (CONTABILIZEI, 2018). O NCM é uma obrigatoriedade do governo no momento da emissão dos documentos fiscais e é composto pela estrutura hierárquica explicada na figura 2.
 - **Código Fiscal de Operações e Prestações (CFOP):** é um código que identifica a natureza de circulação do produto, e é por este meio que é definido se a operação fiscal terá recolhimento de impostos.

2.5 A Arquivei

A Arquivei é uma empresa que oferece como principal produto uma solução Software as a Service (SaaS) de gestão de documentos fiscais eletrônicos. A NFe é o documento mais

³ Sobre a GS1 Brasil: <<https://www.gs1br.org/sobre-a-gs1>>. Acessado em: 2019-09-14

Figura 2 – Composição do NCM

Figura 3 – Como o Arquivoi consulta as notas

Fonte: Elaborada pelo autor.

abundante no sistema e sua principal forma de entrada é através da integração com *Web service* oferecido pela Secretaria da Fazenda [Figura 3](#). A empresa cadastrada na Arquivoi, através de seu certificado digital, consegue consultar as notas emitidas contra seu CNPJ. A Secretaria da Fazenda não fornece *Web service* para consulta de notas emitidas. A forma de entrada das notas emitidas pela empresa na Arquivoi se dá por meio do *upload* dos documentos no sistema, pela instalação de um sistema no computador da empresa que faz o monitoramento de uma pasta que contém os documentos e envia automaticamente os documentos para Arquivoi ou pelo envio por *email*.

Sem um sistema centralizado como o Arquivoi, o custo para as empresas para gestão de todos seus documentos fiscais é muito alto, pois depende muito da colaboração dos fornecedores. Além disso, para empresas que possuem alto volume de documentos, ter o controle de todas suas notas é trabalhoso.

Depois que todos os documentos da empresa se encontram na plataforma, se torna muito mais simples e ágil para as empresas aplicarem suas práticas de *compliance*. A ferramenta possibilita auditorias, relatórios, filtros e diversas outras ações que facilitam o trabalho de profissionais de contabilidade que trabalham nas empresas ou escritórios contábeis que prestam serviços para as empresas. A [Figura 4](#) ilustra a principal tela do sistema de apoio ao *compliance*

Figura 4 – A principal tela para gestão de NF-e do sistema Arquivoi

Fonte: Elaborada pelo autor.

tributário digital.

2.6 Considerações finais

A cobrança de impostos por parte de um estado é tão antiga quanto o próprio estado em si, e existem evidências de sua ocorrência desde o ano 2500 a.C. na antiga Mesopotâmia (WHITE, 2002). Apesar de aparentemente não ser um problema tão recente, no Brasil, a questão da conformidade tributária ainda é um desafio tanto para os contribuintes, quanto para o Governo. Para as empresas o desafio é: manter-se atualizada perante as leis, ter todos os dados necessários acessíveis e prestar conta dentro de prazos; para o Governo o desafio é: fiscalizar, garantir apuração correta de impostos e autuar quando obrigações não são cumpridas. Fica evidente a necessidade, por ambas as partes, de modernização em seus processos. Alguns números apresentados em (ALMEIDA, 2018), apontam que o nível de assertividade em processos de fiscalização e autos de infração aumentou muito após a implementação do SPED. Em 2010, 19,06% dos processos de fiscalização e autos de infração foram julgados improcedentes, enquanto que em 2016, apenas 0,10%.

Este trabalho é feito em conjunto a Arquivoi que está providenciando os dados de forma segura e anonimizada, e assim junto com a universidade, avanços tecnológicos possam se desenvolver.

REVISÃO SOBRE REDES COMPLEXAS

3.1 Conceitos básicos

Nesta seção são revisadas algumas definições e conceitos básicos porém essenciais para o decorrer do capítulo. Tais conceitos abrangem teoria de grafos, redes complexas e algumas medidas simples. Uma observação a ser feita é a de que os termos grafo e rede são usados de mesmo modo neste texto.

Um **grafo** (rede complexa) é uma estrutura formada por um conjunto de vértices (ou nós) e um conjunto de arestas (ou *links*). As arestas são formadas por pares de vértices, o que indica estes vértices estão conectados. A [Figura 5a](#) representa um grafo de 5 vértices (v_1, v_2, v_3, v_4 e v_5) e 6 arestas. Grafos podem possuir nós isolados como o nó v_5 , arestas múltiplas como as arestas de (v_2, v_3) e laços como a aresta (v_4, v_4) . Um **grafo direcionado** é um grafo onde as arestas possuem direção, como representa a [Figura 5b](#). Além da direção, pode-se atribuir a uma aresta um valor chamado de peso. Tal grafo é denominado **grafo ponderado**. A [Figura 5](#) representa um grafo não direcionado ponderado.

O **grau** k_i de um vértice i , é o número de arestas conectadas ao vértice i . Como em grafos direcionados, as arestas possuem orientação, existem os graus de saída e de entrada. O **grau de entrada** k_i^{in} de um vértice i é o número de arestas incidentes a i . O **grau de saída** k_i^{out} de um vértice i é o número de arestas que saem de i . E o grau de um nó em um grafo direcionado é dado por $k_i = k_i^{in} + k_i^{out}$.

O **grau médio** $\langle k \rangle$ de uma rede é a média de k_i de todos os vértices,

$$\langle k \rangle = \frac{1}{n} \sum_i k_i. \quad (3.1)$$

O **caminho** entre os vértices i e j é a sequência alternada de vértices e arestas começando pelo vértice i e terminando pelo vértice j , em que cada aresta incide nos vértices que a precede e

Figura 5 – Exemplos de grafos não direcionado, direcionado e ponderado

Fonte: Elaborada pelo autor.

que a sucede na sequência. O **comprimento** de um caminho é o número de arestas da sequência. O **caminho mínimo** entre dois vértices é o caminho de menor comprimento dentre os caminhos possíveis. Uma medida importante em redes complexas, capaz de dizer o número de passos que leva para sair de um vértice e chegar a outro, é o **comprimento médio do caminho**. Suponha que d_{ij} é o comprimento do caminho mínimo entre os vértices i e j ($d_{ij} = 0$ caso não exista caminho entre i e j). O comprimento médio do caminho L é dado por

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \quad (3.2)$$

Em grafos não direcionados, um **componente conectado** é um subgrafo on qualquer dois vértices desse subgrafo estão conectados por um caminho, e não estão conectados a nenhum outro vértice do grafo.

3.2 Modelos de redes complexas

No mundo real as redes complexas emergem de diversas formas e possuem diferentes topologias. Para facilitar o estudo das propriedades dessas redes, foram propostos alguns modelos. Nas seções [Subseção 3.2.1](#), [Subseção 3.2.2](#) e [Subseção 3.2.3](#) são apresentados os três modelos de maior interesse e suas características.

3.2.1 Modelo Erdős-Rényi

Este modelo busca representar o mundo real através de grafos aleatórios. Pode-se argumentar que a justificativa de tal representação é o fato de que as relações entre as partes são tão complexas na realidade, que é bem plausível que as propriedades de tais relações se assemelhem as propriedades de um grafo aleatório.

O modelo de grafo aleatório, proposto por Paul Erdős e Alfred Rényi, considera quantidade de n vértices e m arestas de um grafo fixa. Um grafo escolhido de modo aleatório dentre os $\binom{n}{m}$ grafos de n vértices e m arestas possíveis de se formarem é considerado um grafo aleatório, ou seja, um grafo $G(n, m)$ é dito aleatório quando formado selecionando m arestas das $\binom{n}{2}$ aresta possíveis. Esta é a proposta original feita em (ERDŐS; RÉNYI, 1959), no qual se buscou responder as seguintes perguntas:

1. Qual é a probabilidade de $G(n, m)$ estar completamente conectado?
2. Qual é a probabilidade de que o maior componente conectado de $G(n, m)$ possuir efetivamente $n - k$ vértices? ($k = 0, 1, \dots$)
3. Qual é a probabilidade de $G(n, m)$ possuir exatamente $k + 1$ componentes conectados? ($k = 0, 1, \dots$)
4. Supondo que as arestas de um grafo são escolhidas sucessivamente de tal forma que a cada passo toda aresta que ainda não foi escolhida possui a mesma probabilidade de ser escolhida. Se seguirmos esse processo até que o grafo esteja completamente conectado, qual é a probabilidade de o número de passos v ser igual a um número l ?

Paul Erdős e Alfred Rényi publicaram muitos artigos a respeito de grafos aleatórios e obtiveram muito resultados com base no modelo $G(n, m)$, acabando-se por associar o modelo de grafo aleatório aos nomes Erdős e Rényi. Entretanto, pelo menos outros dois modelos de grafo aleatório foram propostos: os modelos de (AUSTIN *et al.*, 1959) e (GILBERT, 1959). O modelo de (AUSTIN *et al.*, 1959), que será denotado $G'(n, m)$, é muito similar ao de Erdős e Rényi, com a pequena distinção de que arestas paralelas podem se formar. No modelo proposto por (GILBERT, 1959), também chamado de modelo binomial e denotado como $G(n, p)$, o número de vértices se mantém n , porém as arestas deixam de possuir um número fixo e são arranjadas de forma independente com probabilidade p .

Os três modelos de grafos são bem similares entre si. Erdős e Rényi descrevem a relação do modelo $G(n, m)$ com os outros dois das seguintes formas: “A diferença entre as propriedades de $G(n, m)$ e $G'(n, m)$ são na maioria dos casos negligenciável. As probabilidades correspondentes são, no geral, assintoticamente iguais” (ERDŐS; RÉNYI, 1960, tradução nossa)

¹ e “Em muitos problemas tratado neste artigo, não causa nenhuma diferença se considerarmos ao invés de $G(n, m)$, o grafo aleatório $G(n, p)$ ” (ERDŐS; RÉNYI, 1960, tradução nossa) ². Apesar desses comentários, (KARONSKI; RUCINSKI, 2013) aponta que Erdős e Rényi não formularam nenhum teorema de equivalência e que ocasionalmente declaravam as partes equivalentes do modelo binomial de seus teoremas sem provas. Aparentemente o modelo $G(n, p)$ é o mais estudado por matemáticos e considerado o mais fácil de lidar, e muitos autores quando se referem a grafo aleatório sem citar o modelo estão implicitamente se referindo ao modelo binomial (NEWMAN, 2010).

Dependendo do modelo escolhido algumas propriedades são mais simples de serem derivadas. Tomemos por exemplo, o grau médio $\langle k \rangle$ de um grafo aleatório. No modelo $G(n, m)$, como o número de arestas e vértices são fixos o grau médio é dado pela soma total de graus $2m$ (cada aresta adiciona 2 graus no grafo) do grafo dividida pelo número de vértices n , ou seja,

$$\langle k \rangle = 2m/n.$$

Em contrapartida, o modelo $G(n, p)$, pelo fato do número de arestas ser variável, requer o uso da probabilidade para o cálculo de $\langle k \rangle$. A probabilidade de um grafo G de m arestas ser gerado, pelo modelo $G(n, p)$, é dada por

$$P(G) = p^m (1 - p)^{\binom{n}{2} - m};$$

como existem $\binom{\binom{n}{2}}{m}$ grafos de m arestas possíveis de serem gerados, a probabilidade de aparecer qualquer grafo com m arestas é de

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1 - p)^{\binom{n}{2} - m};$$

como essa é a expressão da distribuição binomial ³, o número médio de arestas pode ser calculado por

$$\langle m \rangle = \binom{n}{2} p.$$

A partir de $\langle m \rangle$ obtemos o grau médio

$$\langle k \rangle = 2 \frac{\langle m \rangle}{n} = \frac{2}{n} \binom{n}{2} p = (n - 1)p. \quad (3.3)$$

¹ The difference between the probable properties of $\Gamma_{n,N}$ ³ resp. $\Gamma_{n,N}^*$ ³ are in most (but not in all) cases negligible. The corresponding probabilities are in general (if the number N of edges is not too large) asymptotically equal

² In many (though not all) of the problems treated in the present paper it does not cause any essential difference if we consider instead of $\Gamma_{n,N}$ ³ the random graph $\Gamma_{n,N}^{**}$ ³

² $\Gamma_{n,N}$, $\Gamma_{n,N}^*$ e $\Gamma_{n,N}^{**}$ são as notações de Erdős e Rényi para os modelos $G(n, m)$, $G'(n, m)$ e $G(n, p)$

³ Esse é o motivo pelo qual é dado o nome modelo binomial para o modelo $G(n, p)$

A derivação do grau médio é feita apenas para exemplificar como cada modelo pode diferir ao se tentar encontrar as características de um grafo aleatório. Entretanto, a derivação das propriedades de um grafo aleatório para cada modelo não está dentro do escopo deste trabalho. De forma agnóstica ao modelo, aqui apenas se apresenta e se discute as características mais relevantes de um grafo aleatório dentro do contexto deste trabalho.

A primeira propriedade a se destacar é a distribuição de graus. Segundo (NEWMAN, 2010), em um grafo aleatório com probabilidade de conexão p , o grau k de um nó segue uma distribuição binomial

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Como, geralmente, se está interessado em estudar as características de grandes redes, quando $n \rightarrow \infty$, p_k se assemelha a distribuição de *Poisson*

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (3.4)$$

em que $\langle k \rangle$ é o grau médio definido na Equação 3.3. A Figura 6 exemplifica tal distribuição. A distribuição *Poisson* indica que a maioria dos vértices tem a mesma (ou um número próximo) quantidade de conexões do que a média dos vértices. Tal característica aplicada a redes do mundo real implicaria, por exemplo, nos seguintes cenários: a quantidade de amigos em uma rede social é aproximadamente a mesma para a maioria das pessoas; a quantidade de citações em publicações científicas são similares para a maior parte das publicações; e a quantidade de empresas que uma empresa negocia é próxima do mesmo número das outras empresas.

Figura 6 – Distribuição de grau de uma rede aleatória com 10 000 vértices com probabilidade $p = 0.1$.

Fonte: Elaborada pelo autor.

Uma segunda propriedade a se citar é o coeficiente de agrupamento. Como a probabilidade, em um grafo aleatório, de um nó ser vizinho de um outro nó é a mesma para todos os nós,

o coeficiente de agrupamento é dado por

$$C = \frac{\langle k \rangle}{n}. \quad (3.5)$$

Este resultado implica que conforme a rede cresce ($n \rightarrow \infty$) a chance da rede conter agrupamentos diminui ($C \rightarrow 0$). O que torna o modelo bem diferente de redes reais, que tipicamente costumam apresentar agrupamentos.

A propriedade mais importante e possivelmente o resultado de maior impacto de Erdős e Rényi, se refere as condições necessárias para um grafo aleatório conter um **componente gigante**⁴. De acordo com (NEWMAN, 2002) o modelo de grafo aleatório preve que existe uma frase de transição conforme $\langle k \rangle$ aumenta até que um componente gigante se forme. Quando $\langle k \rangle$ é baixo, existem muitos componentes desconectados de tamanhos pequenos. Contudo, existe um valor crítico de $\langle k \rangle$ em que o tamanho de um dos maiores componentes do grafo passa a crescer de maneira linear com o tamanho de todo o grafo. O resultado, em que essa fase de transição ocorre, encontrado é exatamente em $\langle k \rangle = 1$. Isso implica que é necessário apenas que cada vértice de um grafo aleatório possua em média um único vizinho, para que exista um caminho conectando a maior parte dos vértices.

Durante um bom tempo o modelo Erdős-Rényi foi a única opção disponível para a representação em rede da realidade. Porém existem diversas características em redes do mundo real das quais o modelo é incapaz de explicar, o que o torna inadequado. As principais críticas ao modelo, feitas por (NEWMAN, 2010) são: o modelo não possui agrupamentos; não existe nenhuma correlação entre os graus de vértices adjacentes; não existem comunidades; e a forma da distribuição de grau não nos ajuda a explicar interessantes fenômenos como resiliência, propagação epidêmica e percolação.

3.2.2 Modelo Watts-Strogatz

Em 1967, Stanley Milgram reportou em (TRAVERS; MILGRAM, 1967) os experimentos em que se buscou examinar o número médio de pessoas necessárias para conectar dois indivíduos selecionados arbitrariamente. Através do envio de 160 cartas a um conjunto de americanos selecionados aleatoriamente para participar do experimento e uma metodologia que permitisse o rastreo das pessoas pela qual a carta passou, Milgram encontrou a surpreendente média de 5.5. Tal resultado originou o popular termo “seis graus de separação”. Pelo que indica (BARABÁSI, 2003), John Guare foi o criador da expressão em 1991 em sua peça *Six Degrees of Separation*.

Além do fenômeno de pequeno mundo descrito por Milgram na rede de indivíduos (característica comum a redes aleatórias), muitas redes do mundo real apresentam, também, alto coeficiente de agrupamento (característica comum a redes reticuladas). Para explicar este

⁴ Componente conectado que contém a maior parte dos nós

Figura 7 – Evolução do grafo regular até o grafo de pequeno mundo

Fonte: Adaptada de [Watts e Strogatz \(1998\)](#).

fenômeno, ([WATTS; STROGATZ, 1998](#)) propuseram o modelo redes de mundo pequeno⁵ como alternativa intermediária entre os modelos altamente regulares ou altamente aleatórios, existentes até então.

A construção de um grafo de pequeno mundo se dá através de uma interpolação entre os grafos regulares e aleatórios. Inicia-se com um grafo regular (em forma de anel como mostra a [Figura 7a](#)) de n vértices e grau médio $\langle k \rangle = K$, onde cada nó está conectado com K vizinhos ($K/2$ de cada lado). Para cada aresta de cada nó n_i , religue-a com probabilidade p a qualquer outro nó escolhido aleatoriamente de forma uniforme. De acordo com ([WATTS; STROGATZ, 1998](#)), as redes de interesse devem possuir muitos vértices e conexões esparsas, mas não tão esparsas que tornam o grafo desconectado. Tais características são atingidas quando $n \gg K \gg \ln(n) \gg 1$. Tal algoritmo de geração introduz $p \frac{nK}{2}$ arestas não reticulares. Variando p é possível intercalar entre um grafo regular reticulado ($p = 0$) e um grafo aleatório $G(n, p^*)$ ($p = 1$), onde $p^* = \frac{K}{n-1}$.

Como um dos grandes propósitos de Watts e Strogatz era a criação de um modelo que possuisse alta tendência dos nós se agruparem, o alto coeficiente de agrupamento é a propriedade que mais se destaca. Para a situação em que $p = 0$, o coeficiente de agrupamento é dado por $C(0) = \frac{3(K-2)}{4(K-1)}$. Como estamos interessado em valores para o qual $p > 0$, temos que dois nós vizinhos de um nó i que estão conectados em $p = 0$ ainda são vizinhos de i e conectados por uma aresta com probabilidade $(1 - p)^3$. Logo,

$$C(p) \approx C(0)(1 - p)^3. \quad (3.6)$$

A principal observação a ser feita sobre a [Equação 3.6](#), é o fato do coeficiente de agrupamento não depender do número de nós na rede, diferente do modelo Erdős-Rényi em que o coeficiente de agrupamento é dado pela [Equação 3.5](#).

A outra propriedade de destaque é a propriedade de mundo pequeno, que matematicamente pode ser medida pelo comprimento médio do caminho denotado $L(p)$. Para situações extremas, temos $L(0) = n/2K \gg 1$ e $L(1) = \ln(n)/\ln(K) \ll 1$. Para $0 < p < 1$, ([NEWMAN, 2010](#)) aponta que a fórmula fechada é tão difícil de se calcular que ainda não foi encontrada.

⁵ Watts e Strogatz utilizaram o termo small-world networks

Apesar disso, (WATTS; STROGATZ, 1998) conseguiram mostrar, através de simulações para vários valores de p , que $L(p)$ é tão pequeno quanto $L(1)$ mantendo $C(p) \gg C(1)$, para um intervalo abrangente de p . Isso implica que existe um conjunto vasto de redes de pequeno mundo.

Por fim, a distribuição de grau de uma rede de mundo pequeno segue uma distribuição similar a de *Poisson*, com a seguinte variação (NEWMAN, 2010):

$$p_k = e^{-cp} \frac{(cp)^{k-c}}{(k-c)!}, \quad (3.7)$$

para $k \geq c$, em que c é o grau dos nós do grafo regular correspondente. A Figura 8 ilustra um exemplo da distribuição de grau de uma rede simulada.

Figura 8 – Distribuição de grau de uma rede *Watts-Strogatz* com 10 000 vértices com probabilidade $p = 0.5$, que se inicia com grau $k = 100$.

Fonte: Elaborada pelo autor.

A grande crítica a rede de mundo pequeno é que ela falha em tentar simular o comportamento de redes do mundo real no que diz respeito a conectividade dos nós, assim como redes aleatórias. Na seção seguinte é visto que as redes do mundo real costumam seguir um outro tipo de distribuição. Mas como é colocado por (NEWMAN, 2010), o objetivo nunca foi simular redes do mundo real nesse aspecto; e sim em relação ao coeficiente de clusterização e ao comprimento médio do caminho.

3.2.3 Modelo Barabási-Albert

Apesar dos modelos Erdős-Rényi e Watts-Strogatz se apresentarem como ótimas ferramentas no estudo e análise das características de redes, eles acabam por não oferecer uma explicação do porquê as redes contêm tais características. Tal natureza se deve ao fato de que as redes geradas por esses modelos são estáticas. Isso significa que parâmetros da rede, como número de nós, número de arestas ou grau médio, são definidos no início da modelagem e se

mantêm fixos ao longo do tempo. Com o propósito de explicar os mecanismos que levam as redes a possuir certas características, surgiu uma gama de modelos denominados de modelos generativos. O modelo generativo de maior destaque foi o proposto por Albert-László Barabási e Réka Albert, que explica como se dá o surgimento das distribuições *power-law*, característica muito comum em redes do mundo real.

É amplamente aceito, apesar de alguma controvérsia (BROIDO; CLAUSET, 2018), que muitas redes do mundo real seguem a distribuição *power-law*. Em (BARABÁSI; BONABEAU, 2003) são citados vários estudos apontando *World Wide Web*, redes sociais, redes de pessoas conectadas por email, redes de citações em artigos científicos, redes de atores de Hollywood e redes formadas pela interação protéica entre células como redes livres de escala.

A primeira característica apontada por (ALBERT; BARABÁSI, 2002) é a de que redes do mundo real são dinâmicas e crescem com a adição contínua de novos nós, diferente das modelagens vistas até então. E a segunda é a de que a probabilidade de um nó se conectar ao outro está relacionada com os graus dos nós. Tal comportamento também não é capturado pelos outros modelos, no qual novas conexões surgem aleatoriamente. Essas duas características, crescimento e conexão preferencial, formam a base do algoritmo do modelo *Barabási-Albert* (ALBERT; BARABÁSI, 2002):

1. **crescimento:** se inicia com um número pequeno de vértices n_0 , e a cada passo, se adiciona um novo vértice e m arestas ($m \leq n_0$) que conectam o novo vértice a m diferentes vértices que já constam no sistema.
2. **conexão preferencial:** existe uma tendência do novo vértice que será adicionado a rede a se conectar com vértices bem conectados; e isso é dado pela probabilidade P do novo vértice se conectar a um vértice existente i ,

$$P_i = \frac{k_i}{\sum_j k_j}.$$

Desse algoritmo decorre a principal propriedade do modelo: a distribuição de grau segue uma distribuição *power-law*,

$$p_k \sim k^{-\gamma}.$$

É apresentado em (ALBERT; BARABÁSI, 2002) que várias redes do mundo real apresentam essa distribuição, e é colocado o valor encontrado do parâmetro γ para cada uma das redes. Simulações indicam que para o modelo teórico o valor de γ é 3. A Figura 9 ilustra a forma da distribuição de grau de uma rede livre de escala.

Em uma rede *Barabási-Albert*, tanto o caminho médio mínimo quanto o coeficiente de clusterização evidenciam comportamentos distintos dos outros modelos. O caminho médio

Figura 9 – Distribuição de grau de uma rede *Barabási-Albert* com 10 000 vértices e 100 arestas conectadas a cada passo.

Fonte: Elaborada pelo autor.

mínimo se apresenta com valores menores em redes *Barabási-Albert*, quando comparado a redes aleatórias; e costuma aumentar logaritmicamente, conforme aumenta n (ALBERT; BARABÁSI, 2002). O coeficiente de clusterização, aponta (ALBERT; BARABÁSI, 2002), não possui uma fórmula fechada como nos modelos de *Watts-Strogatz* e *Erdős-Rényi*, porém existem evidências de que o coeficiente é cinco vezes maior do que em grafos aleatórios, e aumenta vagarosamente conforme o número de nós aumenta.

3.3 Medidas e métricas

Diferentes medidas e métricas podem ser calculadas para redes complexas. Tais quantidades são muito úteis quando se deseja estudar a estrutura de redes complexas. A seguir, as principais medidas e métricas comumente utilizadas são brevemente descritas.

3.3.1 Centralidade de grau

É a medida de centralidade mais elementar, sendo simplesmente o grau do vértice definido na seção Seção 3.1. A centralidade de grau é local e depende apenas dos nós vizinhos.

Apesar de ser uma medida simples, em determinados contextos pode ser útil. É de se esperar que nós com graus altos possuem maior influência sobre os outros nós, ou um grau de importância maior. Por exemplo, em uma rede de citações de artigos científicos pode-se assumir que o artigo mais referenciado (maior grau de entrada) é o que mais influenciou determinada área do conhecimento.

Quando se é necessário comparar a centralidade de graus em redes de diferentes tamanhos, costuma-se usar a centralidade relativa de grau, que é a centralidade de grau dividida pelo número

máximo de arestas incidentes. Isso nada mais é do que uma normalização do valor, que varia de 0 (o nó não tem vizinhos) a 1 (o nó é um ponto central).

3.3.2 Centralidade de proximidade (*closeness*)

Essa centralidade informa o quão perto um nó está dos outros nós, e é definida como o inverso da soma das distâncias de um vértice v_k aos outros vértices, ou seja,

$$C(v_k) = \frac{1}{\sum_{j=1}^n \text{dist}(v_j, v_k)}$$

.

A forma normalizada da centralidade de proximidade é dada por

$$C(v_k) = \frac{n-1}{\sum_{j=1}^n \text{dist}(v_j, v_k)}$$

.

Essa medida calcula valores baixos para nós mais centrais e valores altos para nós menos centrais. A maior crítica a essa medida, como aponta (NEWMAN, 2010), é o fato de que os valores de centralidade dos nós tendem a abranger um intervalo muito reduzido de valores do menor para o maior. Isso implica que é muito difícil distinguir entre nós centrais e nós menos centrais.

3.3.3 Centralidade de intermediação (*betweenness*)

Essa centralidade informa o grau em que um vértice se encontra no meio do caminho dos outros vértices. A centralidade de *betweenness* de um vértice v é definida por

$$C(v) = \sum_{s,t; s,t \neq v} \frac{\sigma_v(s,t)}{\sigma(s,t)},$$

em que $\sigma_v(s,t)$ é o número de caminhos mínimos entre s e t que passam por v e $\sigma(s,t)$ é o número de caminhos mínimos totais entre s e t .

3.3.4 Centralidade do autovetor

É uma extensão da centralidade de grau. A centralidade de grau leva em consideração que todos os nós vizinhos de um vértice contribuem de forma igualitária para o cálculo de centralidade. Entretanto, essa suposição nem sempre é válida, pois os nós vizinhos, geralmente, possuem graus de importância distintos. Assim, a centralidade do autovetor provê para cada vértice um cálculo de centralidade que é proporcional a soma das centralidades dos vizinhos.

Uma alta centralidade de autovetor implica que o nó está conectado a outros nó que também possuem alta centralidade.

A centralidade de cada nó é dada pelos componentes do autovetor da matrix de adjacência M em que é o vetor v que satisfaz a equação

$$Mv = \lambda v, \quad (3.8)$$

em que λ é uma constante.

3.3.5 PageRank

O *PageRank* é uma medida de centralidade proposta por Lawrence Page e Sergei Brin em (PAGE *et al.*, 1999), que se tornou a parte central do motor de busca do *Google*. Muito útil para determinar a melhor posição de uma página em um resultado de busca na *Web*, essa medida é uma variante da centralidade do autovetor.

Quando se analisa a centralidade do autovetor dentro do contexto da *Web*, uma rede de documentos interligados através de *hyperlinks*, nota-se algumas inconveniências. Lawrence Page e Sergei Brin precisavam endereçar essas inconveniências para que pudessem oferecer um buscador de melhor qualidade na época. A primeira delas é o fato do vértice de alta centralidade transferir automaticamente sua centralidade para todos os vértices que aponta. É de se esperar que uma página de notícias muito relevante que aponta para milhares de páginas possa redistribuir sua relevância entre as páginas e não transferir automaticamente sua relevância para todas elas. A segunda é o fato da *Web* ser uma rede modelada através de um grafo direcionado. Algumas páginas podem não possuir *links* o que atrapalha o cálculo da centralidade. E a terceira limitação é o fato da rede poder conter ciclos internos fazendo com que o cálculo da centralidade de alguns nós convirja para zero.

A primeira limitação é solucionada ao tornar a matriz M da Equação 3.8 em uma matriz estocástica, ou seja, todas as colunas tem soma igual a 1. Isso é feito dividindo cada termo m_{ij} de A por k_j^{out} , em que k_j^{out} é o número de conexões que saem de j . Ou seja,

$$m_{ij} = \begin{cases} \frac{1}{k_j^{out}} & \text{se tem uma conexão de } i \text{ para } j \\ 0 & \text{caso contrário} \end{cases} \quad (3.9)$$

Essa mudança implica que a importância de um nó será distribuída de forma proporcional ao número de conexões de saída que esse nó possui.

A centralidade do autovetor de uma rede que possui um nó sem conexões de saída é zero. Tais cenários surgem principalmente em grafos direcionados, tornando a centralidade autovetor apenas útil para grafos não direcionados. Tais nós, por definição, possuem grau de saída k^{out} igual a zero e podem ser identificados pela matriz de adjacência como o nó correspondente a

coluna nula. Esses nós, não só são problemáticos para o cálculo da centralidade do autovetor como também para o cálculo da matriz de adjacência proposta na [Equação 3.10](#), pois k^{out} é zero. Para solucionar esse dilema, é proposto fazer um ajuste que substitua todas as colunas nulas da matriz M ajustada, pela coluna

$$\begin{bmatrix} 1/N \\ 1/N \\ \vdots \\ 1/N \end{bmatrix}.$$

O argumento para tal ajuste é a de que em uma navegação *web*, um internauta que se depara com uma página que não possua *links*, irá eventualmente acessar aleatoriamente qualquer uma das outras páginas. Para solucionar, então, as duas primeiras limitações deve-se considerar uma nova matriz S tal que,

$$s_{ij} = \begin{cases} \frac{1}{k_j^{out}} & \text{se tem uma conexão de } i \text{ para } j \\ \frac{1}{N} & \text{se } k_j^{out} = 0 \\ 0 & \end{cases} \quad (3.10)$$

Um grafo é irredutível se para qualquer par de nós é possível partir de um nó e chegar até o outro, através dos outros nós e arestas do grafo. Um grafo que não é irredutível é chamado de redutível. Um grafo redutível possui ciclos em seu interior. Um navegante do grafo que atinge esse ciclo nunca mais consegue sair. Para o cálculo do *PageRank* tais ciclos são nocivos pois não se consegue propagar a importância dos nós de dentro do ciclo para fora, fazendo com que nós fora do ciclo terminem com a centralidade zerada. É preciso, então, tornar a matriz S irredutível. Em outras palavras, precisamos tornar o grafo associado a S , um grafo fortemente conectado. Em ([GALLARDO, 2007](#)) é citado um estudo que apresenta que 90% das páginas *Web* sob estudo se encontravam dentro de um gigante componente fracamente conectado. Para solucionar essa questão, ([PAGE et al., 1999](#)) propuseram uma constante $d \mid 0 < d < 1$, chamada de fator de amortecimento, que pondera a contribuição do nó baseado na probabilidade d de um navegante do grafo continuar seguindo pelas ligações. Com esse parâmetro estamos dando a possibilidade de um navegante aleatório se entediar de seguir os *links* e poder pular para uma páginas aleatória. Para adicionar esse comportamento ao algoritmo é necessário a definição de uma nova matriz que denotaremos G , a matriz *Google*, tal que:

$$G = dS + \frac{1-d}{S} \cdot 1$$

Como aponta ([LANGVILLE; MEYER; FERNÁNDEZ, 2008](#)), a matriz *Google* G , é artificialmente criada através de modificações da matriz de *hyperlinks* M para adicionar as características de um navegador aleatório em uma rede de páginas *Web*. Curiosamente, essas características fazem da matriz G uma matriz especial, tal que através das garantias oferecidas

pelo teorema Perron-Frobenius, é possível calcular o *PageRank* através do método das potências, da seguinte forma: dado um vetor inicial $\mathbf{x}$, o cálculo do *PageRank* é obtido pela multiplicação de $\mathbf{x}$ com M diversas vezes até que $\mathbf{x}$ convirja.

3.3.6 Excentricidade

Uma medida de quão longe do centro um nó está. A excentricidade é a distância máxima de um nó para todos os outros nós da rede. Algumas propriedades são definidas a partir da excentricidade:

- **Diâmetro:** é a maior excentricidade da rede
- **Raio:** é a menor excentricidade da rede
- **Centro:** conjunto de nós em que a excentricidade é igual ao raio
- **Periférico:** conjunto de nós em que a excentricidade é igual ao diâmetro

3.3.7 Coeficiente de clusterização

É uma medida do grau em que os nós de um grafo tendem a se agrupar. Foi primeiramente apresentada por (WATTS; STROGATZ, 1998). Suponha que um vértice v tenha v_k vizinhos; então, no máximo, $k_v(k_v - 1)$ arestas podem se formar entre esses vizinhos. O coeficiente C_v do nó v é a fração dessas arestas que realmente existem. Enquanto que o coeficiente de agrupamento C da rede é a média dos coeficientes de agrupamento individuais de cada nó. Para exemplificar, (WATTS; STROGATZ, 1998) aponta que em uma rede de amigos, C_v indica a extensão em que os amigos de v também são amigos entre si.

3.3.8 Assortatividade

A assortatividade é uma métrica que avalia a correlação entre os valores de atributos de um nó através das arestas. Um rede com atributos de nós correlacionados positivamente é chamada de assortiva.

3.4 Comunidades

Uma propriedade também muito comum buscada em redes complexas é a presença de estruturas de comunidades. Comunidades em redes complexas são grupos de nós em que os nós dentro do grupo estão densamente conectados, enquanto que os nós entre os grupos estão esparsamente conectados.

3.4.1 Detecção de comunidades

Dada a relevância da presença de comunidades em redes complexas, um das principais análises nessas redes é a detecção de comunidades. Um exemplo de redes complexas cuja detecção de comunidades pode trazer informações relevantes é a rede de coautorias do *Santa Fe Institute* indicada na Figura 10, em que cada nó representa uma pessoa e cada aresta indica os colaboradores daquele nó. Tal figura exemplifica o resultado da detecção de comunidades pelo algoritmo aplicado por (GIRVAN; NEWMAN, 2002)

Figura 10 – O maior componente da rede de colaboração do *Santa Fe Institute*

Fonte: Girvan e Newman (2002).

A detecção de comunidades é apontada por (NEWMAN, 2010) como um problema complementar à partição de grafos, problema clássico muito conhecido na Ciência da Computação. Este problema consiste em dividir os vértices de uma rede em um número de grupos, que não se sobrepõem, em que os tamanhos são sabidos *a priori*, de tal forma que o número de arestas entre os grupos é minimizado. O que diferencia a detecção de comunidades é o fato de que se busca por grupos que ocorrem naturalmente em uma rede, independentemente do seu número ou tamanho. Apesar da partição de grafos ser um problema bem conhecido e estudado, a detecção de comunidades acabou por se tornar um problema não tão claro, como aponta (FORTUNATO; HRIC, 2016). Não existem protocolos universais sobre os ingredientes fundamentais, como a definição da própria comunidade, nem sobre outras questões cruciais, como a validação de algoritmos e a comparação de seus desempenhos. É por tal motivo que se originou uma série de algoritmos com diferentes estratégias para se identificar comunidades. Na Subseção 3.4.2 se discute os principais métodos.

3.4.2 Métodos para detecção de comunidades

A seguir serão brevemente descritos alguns dos principais métodos utilizados para detecção de comunidades na literatura da área de análise de redes complexas.

3.4.2.1 Algoritmos divisivos

São métodos que procuram identificar as arestas que conectam comunidades diferentes e eliminá-las. O algoritmo mais conhecido é o proposto por (GIRVAN; NEWMAN, 2002) onde se define conceito de *edge betweenness* para uma aresta, que é número de caminhos mínimos entres pares de nó dessas aresta, como critério de identificação de arestas que separam comunidades. O algoritmo consiste em:

1. Calcular a *edge betweenness* de cada aresta da rede
2. Remover a aresta de maior *betweenness*
3. Recalcular a *betweenness* das arestas afetadas pela remoção
4. Repetir passos 2 e 3 até que não existam mais arestas restantes

Os algoritmos divisivos variam de acordo com a definição de *edge betweenness* e a forma de calcular.

3.4.2.2 Algoritmos espectrais

Partição de grafo através de métodos espectrais tradicionais são embasados pela teoria espectral de grafos, disciplina em que se estuda propriedades dos grafos a partir de suas representações matriciais e respectivos espectros (conjunto de autovalores de uma matriz). O algoritmo mais conhecido se utiliza do conceito de conectividade algébrica, apresentado por (FIEDLER, 1973), para encontrar os agrupamentos de vértices. A representação matricial que permite o cálculo da conectividade algébrica de um grafo G com n vértices, se dá através a matriz Laplaciana L , definida por $L = D - A$, onde D é a matriz de graus e A a matriz de adjacência de G . Se o grafo é conectado, o espectro de L será $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, onde λ_2 é a conectividade algébrica de G . Quanto maior λ_2 , mais conectado é G ; por outro lado, quanto menor λ_2 , melhor a qualidade do particionamento. Para encontrar uma bipartição de G , basta atribuir uma partição aos vértices que corresponda aos sinais do autovetor correspondente a λ_2 . Esse método foi estendido por (NG; JORDAN; WEISS, 2002), através do uso da matriz Laplaciana normalizada $L_{norm} = D^{-1/2} L D^{-1/2}$ e da clusterização *k-means*, para o encontro de k grupos ao invés de 2. O mesmo conceito é utilizado em outras representações matriciais como a matriz de modularidade, em (NEWMAN, 2006), ou a matriz de tranferência, em (ERIKSEN *et al.*, 2003), para detecção de comunidades.

3.4.2.3 Algoritmos dinâmicos

Detecção de comunidade dinâmica é o processo de encontrar comunidades relevantes em redes que alteram ou evoluem ao longo do tempo. Redes sociais são bons exemplos de redes que possuem estrutura de comunidades dinâmicas. Ao longo do tempo pessoas entram e saem dessas comunidades, porém isso não impede a existência da comunidade. Por essa natureza, a comunidade dinâmica é definida por um par de conjuntos de nós e de períodos. Em (CAZABET; ROSSETTI; AMBLARD, 2017) é apresentado as três abordagens principais (*Instant Optimal*, *Temporal Trade-off* e *Cross-time*) e os algoritmos publicados até então.

3.4.2.4 Métodos baseados em inferência estatística

A detecção de comunidades pode ser abordada como um problema de inferência. Tratando o grafo como um conjunto de dados, tais métodos procuram encontrar o melhor ajuste de um modelo hipotético. A inferência Bayesiana é uma técnica conhecida em tal cenário, como aponta (FORTUNATO, 2010). A inferência Bayesiana se utiliza de observações sobre um sistema para estimar a probabilidade de uma hipótese ser verdadeira. Para o contexto de redes complexas, as observações são a própria estrutura do grafo, e assume-se que cada vértice possui uma classe, que é a comunidade a ser descoberta. A ideia de boa parte dos métodos de detecção de comunidade baseados em inferência é encontrar os parâmetros do modelo, supostamente responsável pela classificação, que maximizam a probabilidade desse modelo explicar a qual comunidade (ou classe) cada vértice pertence. Outras abordagens apontadas por (FORTUNATO, 2010) são: *Blockmodeling*, seleção de modelos e teoria da informação.

3.4.2.5 Métodos baseados em otimização

O objetivo principal das técnicas baseadas em otimização é encontrar o valor máximo para uma função de qualidade. Uma função de qualidade atribui um número a uma comunidade que quantifica a força daquela estrutura de comunidade. A função mais utilizada é a de modularidade definida em (NEWMAN; GIRVAN, 2004), como

$$Q = \sum_i (e_{ii} - a_i^2),$$

em que e_{ii} é a fração de arestas pertencentes a comunidade i e a_i^2 é o número esperado de arestas pertencentes a comunidade i para um grafo aleatório com a mesma distribuição de graus que a rede estudada. Sucintamente, a modularidade mede o quanto a rede original difere de um grafo aleatório com as mesmas propriedades. Portanto, para uma rede em que a fração de arestas dentro da comunidade não é maior do que o que seria esperado por acaso, Q será 0. Valores maiores que 0 indicam desvios da aleatoriedade. (NEWMAN, 2004b) aponta que $Q > 0.3$ indica estrutura de comunidade significativa. A Figura 11 exemplifica o cálculo da modularidade. Diferentes variações da modularidade para diferentes tipos de redes são listadas em (KHAN; NIAZI, 2017).

Figura 11 – Exemplo do cálculo da modularidade para a rede considerando duas comunidades C_1 e C_2 em que $C_1 = \{A, B, C, D\}$ e $C_2 = \{E, F, G, H, I\}$. O grafo possui 12 arestas, dos quais 5 pertencem a C_1 e 6 a C_2 ; e 11 meia arestas incidem em nós de C_1 e 13 meia arestas incidem em nós de C_2 .

$$Q = \sum_i (e_{ii} - a_i^2)$$

$$Q = \left(\frac{5}{12} - \left(\frac{11}{24} \right)^2 \right) + \left(\frac{6}{12} - \left(\frac{13}{24} \right)^2 \right)$$

$$Q = 0.4131944$$

Fonte: Elaborada pelo autor.

Maximização da modularidade é *NP-hard* (FORTUNATO; HRIC, 2016), portanto podemos realisticamente encontrar apenas boas aproximações. O primeiro algoritmo para otimização de modularidade foi a técnica de otimização proposta por (NEWMAN, 2004b). É um algoritmo guloso aglomerativo que se inicia com n comunidades (cada nó pertencente a sua própria comunidade), e repetidamente se une as comunidades em pares, escolhendo a cada etapa o par que resultou no maior ganho de modularidade. O algoritmo para quando não há mais ganho de modularidade.

Muitas outras abordagens de otimização foram propostas. Dentre as mais populares se destacam: o algoritmo *Louvain* proposto em (BLONDEL *et al.*, 2008), em que se procura encontrar comunidades menores otimizando a modularidade localmente e depois agrupando essas comunidades menores em nós; o uso de *Simulated Annealing* como descrito em (GUIMERA; AMARAL, 2005); a técnica de (DUCH; ARENAS, 2005) que se utiliza da otimização extrema projetada por (BOETTCHER; PERCUS, 2001); e a otimização espectral como explicado na Subsubseção 3.4.2.2.

3.5 Considerações finais

Este capítulo apresentou os principais conceitos de redes complexas necessários para a compreensão do projeto realizado. Após apresentar os conceitos básicos, a partir de teoria dos grafos, descreveu os principais modelos de redes complexas encontrados na literatura consultada. Em seguida, apresentou as principais medidas e métricas utilizadas para caracterizar redes complexas. Em seguida, explicou como comunidades podem ser encontradas nestas redes. O próximo capítulo faz uso da aplicação dessas métricas na tentativa de caracterizar redes

comerciais brasileiras.

CARACTERIZAÇÃO DE REDES COMERCIAIS

4.1 Considerações iniciais

Na ciência econômica, o estudo de redes complexas vem se tornando cada vez mais popular, inclusive o número de artigos científicos e conferências que se utilizam do termo *network economics* como rótulo para endereçar assuntos relacionados a fenômenos organizacionais e redes físicas, cresceu significativamente durante a década de 90 (ERBER; HAGEMANN, 2002). Trabalhos como: (GOYAL, 2012), um tratado auto contido sobre a teoria econômica de redes; (KRANTON; MINEHART, 2001), uma proposta de modelagem de trocas baseadas em redes, e não em mercados; e (SCHWEITZER *et al.*, 2009), em que é feito uma revisão sobre pesquisas de redes complexas em economia, evidenciam tal popularidade.

Dentro do contexto da ciência econômica, redes podem ser vistas sob duas óticas (UMBHAUER, 1998). A primeira é a de um conjunto de conexões que se constroem a partir da interação entre agentes que buscam obter benefícios através dessas interações. E a segunda é a de um conjunto de agentes que adotam um comportamento similar para diferentes propósitos econômicos. Diretores, empresas, produtos, grupos de negócios, bancos, investidores, comerciantes, cidades e países são tipos de nós comumente encontrados em redes econômicas (EMMERT-STREIB *et al.*, 2018) (ALVES *et al.*, 2019) (ALVES *et al.*, 2018). O Quadro 1 apresenta os tipos de redes econômicas frequentemente estudadas na literatura.

A NF-e é um documento com um conjunto de informações muito vasto, que possibilita a criação de diversos tipos de redes econômicas. Neste trabalho restringiremos o estudo para redes comerciais entre empresas. A importância deste estudo é bem colocada em (SOUMA; FUJIWARA; AOYAMA, 2003), que argumentam que apesar do estudo das interações entre agentes estar bem avançado, as características topológicas das redes formadas por esses relacionamentos não é um assunto muito conhecido. Dessa forma, estudos sobre redes econômicas possibilitam expor aspectos topológicos de redes comerciais. A partir de uma rede complexa criada a partir

Quadro 1 – Exemplos de diferentes tipos de redes econômicas estudadas na literatura

Tipo de rede econômica	Significado da rede
Rede interbancária	Responsabilidade entre instituições (bancos); interconexão de bancos
Rede de investimento	Capital de risco ou investimento em empresas e corporações
Rede de diretores	Membros comuns do conselho de instituições; estrutura corporativa entre instituições
Rede de quadro societário sobre empresas	Influência em decisões corporativas
Rede financeira	Dependência entre ações; interconexão entre ações e mercado
Rede de produtos	Indivíduos cocompradores de produtos ou países coprodutores de produtos
Rede de comércio	Troca de bens e serviços; comércio entre nações

Fonte: Adaptada de Emmert-Streib *et al.* (2018).

dos dados da NF-e, esta seção busca, através de técnicas de análise de rede tradicionais, descrever aspectos de uma rede de transações comerciais entre empresas brasileiras.

4.2 Metodologia

A NF-e, principal fonte de dados deste trabalho, é um documento XML semi-estruturado, como visto na Seção 2.4. Através de uma série de estratégias, explicadas sucintamente na Seção 2.5, o sistema Arquivoi obtém legalmente as notas eletrônicas de validade jurídica para seus clientes e as armazena de forma segura em servidores na nuvem. Para possibilitar o desenvolvimento das funcionalidades que hoje o Arquivoi oferece, é necessária uma grande operação. Hoje, são mais de milhões de documentos capturados diariamente. É preciso de um conjunto robusto de sistemas tolerantes a falha capazes de armazenar esses documentos, de extrair as informações de mais de 500 campos do XML, estruturá-las e indexá-las em bancos de dados de forma a otimizar os padrões de acesso ao dado. Isso é uma tarefa árdua, que requer horas de dedicação exclusiva por parte de vários profissionais do time de Engenharia da Arquivoi. Além disso, também é necessário um trabalho extenso de Engenharia de Dados para armazenar uma réplica de todas essas informações em uma estrutura analítica conhecida por *Data Warehouse*, que possibilite com que cientistas de dados explorem toda essa informação a fim de encontrar oportunidades de melhoria no produto Arquivoi. O *Data Warehouse* de NF-e da Arquivoi foi construída em cima do produto *Google BigQuery*¹ da *Google Cloud Platform*, onde os detalhes de implementação dessa tecnologia são descritos em (MELNIK *et al.*, 2010).

Com todos os dados de todos os documentos fiscais que já passaram pelo Arquivoi estruturados em um *Data Warehouse*, é feita uma projeção inicial dos campos e documentos que são relevantes para este trabalho. Como se deseja estudar as redes comerciais econômicas

¹ Página principal do produto: <<https://cloud.google.com/bigquery/>>

entre empresas, são consideradas apenas as seguintes informações das notas fiscais eletrônicas: emissor, destinatário, data de emissão, CFOP, GTIN, NCM, valor unitário, unidade, quantidade comprada, descrição do produto, valor do produto, unidade do produto e valor total da nota. A [Tabela 1](#) exemplifica as informações selecionadas.

Como se deseja também efetuar análises de nível mais agregado é feito um enriquecimento através de dados cadastrais de pessoas jurídicas publicados pela Receita Federal². É adicionado aos dados da [Tabela 1](#), o ramo de atividade, que se dá através da CNAE, e a cidade; tanto do emissor quanto do destinatário da nota.

Além da projeção e enriquecimento explicados anteriormente, é utilizado vários critérios de seleção a fim de tornar a análise viável computacionalmente, representativa, confiável e relevante para o contexto de redes econômicas. Alguns dos critérios de seleção utilizados requer alto conhecimento de domínio, e boa parte desse conhecimento foi sendo adquirido pela empresa ao longo do tempo. Foram selecionados todas as transações que ocorreram entre o segundo semestre do ano passado e o primeiro semestre desse ano. Dados mais antigos que esse período sofrem de representatividade. Como a Arquei trabalha com documentos de diversas origens que não possuem necessariamente validade jurídica, foram consideradas apenas as transações em que se sabe com certeza a validade do documento que representa a transação. Como se está interessado em construir uma rede comercial, se faz necessário excluir algumas transações como por exemplo: realocação de estoque, bonificação e doação. A [Tabela 5](#) oferece a lista de CFOPs de interesse utilizados como critério de seleção. Vale ressaltar que são consideradas apenas as operações de saída de mercadoria do emissor para o destinatário e são excluídas todas as transações canceladas, que são aquelas em que o produto não chegou ao seu destino.

O objeto de estudo deste trabalho são as redes complexas de transações comerciais. São feitas 4 modelagens, como indica a [Figura 12](#), a partir do conjunto de dados descrito anteriormente. Na primeira abordagem ([Figura 12a](#)) as transações comerciais dos itens da NFe são modeladas em um grafo em que o vértice do grafo é uma empresa e a aresta representa o total em R\$ transacionado pelas duas empresas conectadas pela aresta. A direção da aresta indica a relação de compra e venda entre as empresas. Se a aresta sai do nó, significa que aquela empresa é o fornecedor (ou emissor da nota fiscal). Temos então um grafo direcionado ponderado. Para a segunda abordagem ([Figura 12b](#)), os nós passam a representar um grupo econômico empresarial. A relação matriz-filial é dada pelos 8 primeiros dígitos do CNPJ; nesta modelagem então, um nó representa esses primeiros 8 dígitos. Na terceira abordagem ([Figura 12c](#)), modela-se os nós como ramos de atividades econômicas dado pelos códigos CNAEs das empresas. E na última modelagem ([Figura 12d](#)), o nó é representado por uma cidade. Perceba que as modelagens são muito similares, o que muda de uma para outra é o nível de agregação do nó, que se inicia em seu nível mais granular, que é o CNPJ, passa para o grupo econômico, ramo de atividade e chega

² Esses dados podem ser acessados em: <http://receita.economia.gov.br/orientacao/tributaria/cadastros/cadastro-nacional-de-pessoas-juridicas-cnpj/dados-publicos-cnpj>

Tabela 1 – Amostra real que exemplifica cinco registros selecionados para o estudo. Cada registro implica uma transação comercial de um produto entre duas empresas: a empresa emissora da nota que é quem vendeu o produto; e o destinatário da nota que é quem comprou o produto.

Data de Emissão	CFOP	GTIN	NCM	Valor Unitário	Quantidade	Descrição	Valor	Unidade	Valor NF-e
2019-01-17	5403	7896006268666	30049059	5.745	2	FLUCISTEIN 20MG/ML XPE X100ML	11.49	UN	1273.34
2019-02-09	6409	7896094903036	30049049	138.89	1	BENEGRIP 500MG 25X6	138.89	CX	3998.40
2019-03-06	5409	7891252822068	84841000	29.02	2	82206-JUNTA DO CABECOTE	58.04	UN	1393.51
2019-02-20	6404	SEM GTIN	84219999	437.37	1	FILTRO AR MO- TOR	437.37	UN	2177.07
2019-03-26	5152	2439172000002	2439172000002	6.55	4.99	TOMATE KG	32.68	KG	298.66

Fonte: Dados da pesquisa.

Nota – A identidade das empresas foram preservadas

Figura 12 – As 4 modelagens propostas para o estudo de redes comerciais brasileiras

Fonte: Elaborada pelo autor.

ao nível de cidade. Dessa forma podemos estudar a topologia das redes complexas comerciais da atividade econômica brasileira em diversos níveis.

Para a aplicação das técnicas de redes complexas são construídas sete redes utilizando as quatro modelagens descritas. São construídas 4 redes comerciais empresarias (modelagem 1), em que cada rede representa quatro trimestres subsequentes, em que o primeiro período é o terceiro trimestre do ano 2018 e o último é o segundo trimestre do ano 2019. As outras três são construídas utilizando as outras 3 modelagens, e é utilizado os dados do primeiro trimestre de 2019 para tais redes. A [Tabela 4](#) esclarece esses pontos. Ela também traz uma coluna de identificação numérica (n.^o) para facilitar a explicação do texto.

De início são aplicadas técnicas de análise descritiva ao conjunto de dados para que se adquira conhecimento básico e familiaridade sobre os dados. É realizado: uma sumarização de variáveis básicas; análise de cobertura; análise de produtos mais comercializados; e análise de quais são os ramos de atividade econômica mais frequentes. São utilizadas, também, técnicas de visualização aplicadas através da ferramenta *Tableau*.

Antes da execução das técnicas de redes complexas é feita uma análise de componentes nas redes. Como muitas métricas e algoritmos trazem resultados não confiáveis para um grafo não conectado, todas as métricas e algoritmos são aplicados no componente principal da rede, encontrado a partir da análise de componentes.

Com objetivo de entender a topologia de redes complexas do comércio brasileiro, foca-se em três métricas robustas e bem utilizadas ([ALBERT; BARABÁSI, 2002](#)): menor caminho médio, coeficiente de clusterização e distribuição de grau. Para os grafos menores se faz uso extensivo da ferramenta *Jupyter Notebooks* ([KLUYVER et al., 2016](#)) como ambiente de programação para manipulação dos dados e uso das bibliotecas em *Python* para representação em grafos e análises de redes complexas como *iGraph* ([CSARDI; NEPUSZ, 2006](#)) e *networkx* ([HAGBERG; SWART; CHULT, 2008](#)). Para os grafos de larga escala se faz necessário outra abordagem. É sabido que alguns algoritmos possuem complexidade computacional muito alta, como por exemplo o cálculo

do caminho mínimo, que leva $O(m+n)$ para encontrar o menor caminho entre dois vértices dentro de $\frac{1}{2}n(n-1)$ pares de vértices, e a centralidade *betweenness*, de complexidade $O(n^3)$. Algoritmos que escalam proporcional a $O(n^3)$ não possuem valor para grandes redes (NEWMAN, 2010). Não só se faz necessário o uso de algoritmos mais eficientes, mas também o uso da computação distribuída. Tanto para o cálculo do coeficiente de clusterização e do menor caminho médio dos grafos 1, 2, 3, 4 e 5, é utilizado a ferramenta *Apache Flink*³ configurada dentro do ambiente da *Google Cloud Platform*. O *Apache Flink* é um *framework* e um sistema de computação distribuída que permite performar computações em grandes volumes de dados. Para o cálculo do coeficiente de clusterização, fez-se o uso, através de uma implementação na linguagem *Scala*, de um biblioteca⁴ do próprio *Apache Flink*. Como existe uma relação direta entre o cálculo da centralidade de proximidade (*closeness*) e cálculo do menor caminho médio, esta pode ser usada para tal. Porém, o algoritmo clássico para seu cálculo também é de complexidade $O(n^3)$. Em (KANG *et al.*, 2011) é proposto o algoritmo *Effective Closeness* que através de um cálculo aproximado, porém muito preciso, dos caminhos mínimos realiza o cálculo da centralidade de proximidade em complexidade $O(dm)$, onde d é o diâmetro do grafo. Uma implementação em *Java* para ser executada no *Apache Flink* é oferecida por (CHAKKARADHARI, 2014) e pode ser encontrada publicamente no repositório de códigos *GitHub*⁵. Tal algoritmo foi executado em um *cluster* de 10 máquinas de configuração com 16 CPUs e 60 GB RAM.

É muito importante que se faça uma comparação com outros tipos de redes, não só para que se entenda as semelhanças e diferenças com redes bem estudadas, como também para ser ter uma referência para validação dos resultados. São feitas comparações com redes aleatórias, rede de coautoria matemática (ALBERT; BARABÁSI, 2002), rede de atores de filmes (BARABÁSI *et al.*, 2002) e redes corporativas japonesas (SUGIYAMA *et al.*, 2005).

Além das três métricas, se faz um teste de ajuste a curva *power-law* através da biblioteca (ALSTOTT; BULLMORE; PLENZ, 2014), em busca de estimativas para o expoente γ (tanto para as redes direcionadas quanto para as não direcionadas) e alguns cálculos de medidas de centralidade.

E por fim, uma análise se as redes possuem comunidades em sua estrutura é feita. Existem inúmeros algoritmos de detecção de comunidade (KHAN; NIAZI, 2017), que se baseiam em diferentes estratégias e medidas de similaridade. Quando se trata de grafos grandes, como no caso deste trabalho, a preocupação passa a ser o tempo de execução dos algoritmos. O algoritmo escolhido é o descrito por (CLAUSET; NEWMAN; MOORE, 2004) e implementado por (CSARDI; NEPUSZ, 2006). É mostrado por (CLAUSET; NEWMAN; MOORE, 2004) que o algoritmo consegue extrair informações de redes com mais de 400 mil vértices e 2 milhões de arestas. O algoritmo se baseia no conceito de modularidade, que para grafos ponderados é

³ <<https://flink.apache.org/>>

⁴ <https://ci.apache.org/projects/flink/flink-docs-stable/dev/libs/gelly/library_methods.html>

⁵ Implementação do algoritmo *Effective Closeness* no *Apache Flink* se encontra em <<https://github.com/JananiC/NetworkCentralities>>. Acessado em 12 de Novembro de 2019.

definido em (NEWMAN, 2004a) como :

$$Q = \frac{1}{2W} \sum_{i,j} \left[A_{ij} - \frac{s_i s_j}{2W} \right] \delta(C_i, C_j),$$

em que W é o peso total de todas as conexões na rede, W_{ij} é o peso entre os nós i e j , s_i é a força do vértice i , s_j é a força do vértice j , e $\delta(C_i, C_j) = 1$ se i e j pertencem a mesma comunidade, senão é igual a 0.

4.3 Análise descritiva

Antes de modelar os dados em uma rede complexa comercial, é necessário que se faça uma análise descritiva para que se entenda alguns aspectos relevantes da base selecionada para o estudo. Como explicado na Seção 4.2, a base de dados contém dados de notas fiscais emitidas dentro do período de 1 de julho de 2018 até 30 de junho de 2019. A Tabela 2 traz alguns números sobre algumas dimensões do conjunto de dados.

Tabela 2 – Estatísticas simples da base de dados sob estudo

Variável	Valor aproximado
Quantidade de notas fiscais	84 milhões
Quantidade de itens comercializados	510 milhões
Quantidade de emissores	1.40 milhão
Quantidade de destinatários	619 mil
Quantidade de CNPJs (emissores ou destinatários)	1.66 milhão
Quantidade de códigos de barras distintos (GTIN)	7.66 milhões
Quantidade de NCMs distintos	15 mil
Média de documentos emitidos por mês	7 milhões
Média de documentos emitidos (2º sem. 2018)	6.6 milhões
Média de documentos emitidos por dia	330 mil

Fonte: Dados da pesquisa.

Como se deseja explorar a estrutura topológica da rede comercial brasileira, é de interesse que se faça uma análise representativa do conjunto de dados. De acordo com a empresa (EMPRESÔMETRO, 2019), há aproximadamente 20 milhões de empresas ativas no Brasil; portanto, a base representa transações comerciais de aproximadamente 8% das empresas brasileiras. Outra estatística interessante é a média mensal que foi de aproximadamente de 6.6 milhões no segundo semestre de 2018. Comparando com a média nacional que foi de aproximadamente 205 milhões de notas de saída (SECRETARIA DA FAZENDA, 2019), implica que a base representa 2.9% das notas emitidas de saída do país no segundo semestre de 2018.

Pode-se olhar para a representatividade, também, sob a ótica territorial: quantas cidades e estados estão sendo abrangidos pela base e como se dá a concentração de compradores e vendedores por região. A [Figura 13](#) traz essa avaliação. Dos 5 570 municípios brasileiros, a base contém informação de 4 867 municípios que realizaram alguma venda e 5 305 municípios que realizaram alguma compra, isso representa 87.3% e 95.2%, respectivamente.

Figura 13 – Análise de cobertura territorial

Fonte: Elaborada pelo autor.

É interessante também que se entenda quais são os tipos de produtos mais comercializados e em quais ramos de atividade econômicas a maior parte das empresas se enquadram. A [Tabela 3](#) lista os 10 capítulos de NCMS (de um total de 97) mais comercializados e a [Figura 14](#) os 10 CNAES e divisões de CNAES mais presentes na base dados.

4.4 Resultados da aplicação de técnicas de redes complexas

Um resumo das principais métricas analisadas e investigadas para todas as redes sob estudo é colocado na [Tabela 4](#). Também é colocado as métricas para os respectivos grafos aleatórios e outras redes famosas do mundo real, como explicado na [Seção 4.2](#).

Tabela 3 – Os 10 capítulos de NCMs mais comercializados

Capítulo	Descrição	Quantidade	Acúmulo (%)
30	Produtos farmacêuticos	77.140.008	15,15
33	Óleos essenciais e resinoides; produtos de perfumaria ou de toucador preparados e preparações cosméticas	39.890.258	22,99
22	Bebidas, líquidos alcoólicos e vinagres	36.176.054	30,10
19	Preparações à base de cereais, farinhas, amidos, féculas ou leite; produtos de pastelaria	25.774.031	35,16
39	Plástico e suas obras	21.955.391	39,47
21	Preparações alimentícias diversas	18.868.935	43,18
84	Reatores nucleares, caldeiras, máquinas, aparelhos e instrumentos mecânicos, e suas partes	18.190.350	46,75
85	Máquinas, aparelhos e materiais elétricos, e suas partes; aparelhos de gravação ou de reprodução de som, aparelhos de gravação ou de reprodução de imagens e de som em televisão, e suas partes e acessórios	17.404.556	50,17
61	Vestuário e seus acessórios, de malha	17.145.068	53,54
87	Veículos automóveis, tratores, ciclos e outros veículos terrestres, suas partes e acessórios	14.774.351	56,44

Fonte: Dados da pesquisa.

4.4.1 As métricas $\langle k \rangle$, L e C

A primeira observação a ser feita é em relação aos grafos aleatórios. Quando se compara as redes comerciais com seus respectivos grafos aleatórios, fica evidente que as redes comerciais possuem características distintas. A tendência de agrupamento é mais que quatro, em rdem de magnitude, para as redes comerciais empresariais. E também, para as outras redes comerciais, o número se apresenta bem superior. Também, nota-se a tendência dos nós estarem mais próximos para as redes comerciais ($L < L^{ER}$).

É interessante que se faça uma comparação das métricas das redes comerciais brasileiras com as métricas das redes comerciais japonesas (SUGIYAMA *et al.*, 2005). Ao comparar as duas, nota-se que as redes brasileiras possuem mais vizinhos em média do que as empresas japonesas; e também, as empresas estão mais próximas. Entretanto, as empresas japonesas possuem um grau de agrupamento levemente maior.

Outra observação que se faz, é o fato das métricas $\langle k \rangle$, L e C aumentarem gradativamente da rede 1 a rede 4 (exceto para o coeficiente de clusterização dos grafos 2 e 3, que são muito próximos). Ou seja, conforme o tempo avança, o grau médio, o menor caminho médio e o coeficiente de clusterização da rede comercial empresarial aumentam. O que diferencia os grafos de 1 a 4 é o período. A tendência é que períodos mais recentes possuam mais nós e mais

Figura 14 – Gráfico de barra de CNAEs e divisão de CNAEs mais frequentes, tanto para fornecedores quanto para compradores; e a porcentagem acumulada.

arestas, pois a quantidade de notas fiscais aumenta ao longo do tempo (por conta do aumento do número de clientes). Porém, o grafo 2 exibe um maior número de nós e arestas, isso se deve ao fato, muito provavelmente, de que o 4º trimestre do ano é o de maior atividade econômica. Seria interessante analisar mais redes ao longo do tempo para entender se essa tendência de crescimento se mantém. É interessante de se observar que este fato não se repete para rede comercial japonesa (SUGIYAMA *et al.*, 2005). O padrão observado é justamente o oposto. Perceba também que o tamanho das redes japonesas se mantém praticamente constante.

Uma comparação também é feita em relação aos níveis de agregação. Nota-se que quanto maior o nível de agregação do nó, maior o coeficiente de agregação e menor o comprimento do menor caminho médio. As redes 6 e 7 possuem alto grau de agrupamento. Essas tendência não se repete para os respectivos grafos aleatórios.

Quando se compara as métricas das redes comerciais com redes conhecidas na literatura, outras características interessantes emergem. Em (BARABÁSI *et al.*, 2002), é feito um estudo em que se advoga que a rede de coautores de publicações matemáticas (rede 8) possuem propriedade livre de escala. Barabási, em (ALBERT; BARABÁSI, 2002), advoga o mesmo para a rede que conecta atores que atuaram juntos em filmes (rede 9); apesar de em (WATTS; STROGATZ, 1998) ter considerado que esta rede possui propriedades de uma rede de pequeno mundo. As

redes comerciais sob estudo possuem um comportamento um pouco distinto das redes 8 e 9. O comprimento do menor caminho médio é menor nas redes comerciais, mesmo considerando a diferença de tamanho. Para as redes 6 e 7 que possuem coeficiente de agrupamento similar as redes 8 e 9, o mesmo comportamento se mantém. Em (HUMPHRIES; GURNEY, 2008), é definido uma característica comum de redes de pequeno mundo: uma rede G é dita de pequeno mundo se $L_G \geq L^{ER}$ e $C_G \gg C^{ER}$. Essa definição, acaba classificando as redes 8 e 9 como redes de pequeno mundo, contrariando (BARABÁSI *et al.*, 2002) e (ALBERT; BARABÁSI, 2002); e excluindo a possibilidade das redes comerciais serem redes de pequeno mundo, já que $L < L^{ER}$ é menor para as redes comerciais. Ver Tabela 4.

4.4.2 Distribuição de graus e centralidades

A distribuição de graus é, também, uma característica muito importante de se analisar quando se estuda topologia de redes. A Figura 15 apresenta a distribuição de graus das redes 3, 6 e 7. É evidente que as redes comerciais não apresentam uma distribuição semelhante a redes aleatórias e a redes de pequeno mundo, que seguem distribuições similares a distribuição de *Poisson*, como visto no Capítulo 3.

A distribuição das redes comerciais se assemelham muito mais com uma distribuição *power-law*, comum em redes livres de escala. É notável, também, o aumento do espelhamento conforme aumenta o nível de agregação da rede: a rede empresarial (rede 3) possui uma distribuição mais densa, a rede de cidades (rede 6) um pouco menos densa e a rede de ramos de atividade (rede 7) possui uma distribuição mais espalhada. Apesar de visualmente a distribuição se assemelhar a uma *power-law*, se faz necessário o uso de técnicas estatísticas para se assegurar de tal semelhança. Em (CLAUSET; SHALIZI; NEWMAN, 2009), é proposta uma técnica para tal, apesar de suas colocações sobre o quão complicado é fazer tais detecções.

Os resultados da aplicação da técnica de (CLAUSET; SHALIZI; NEWMAN, 2009), através da biblioteca (ALSTOTT; BULLMORE; PLENZ, 2014) são encontrados na Tabela 4 e na Figura 15. A Tabela 4 indica as estimativas dos expoentes (γ , γ_{out} e γ_{in}) da *power-law* para todas as redes comerciais (tanto as direcionadas, quanto as não direcionadas). Os expoentes são bem similares entres as cinco primeiras redes e aumenta um pouco tanto para a rede 6, quanto para a rede 7.

Não existe um consenso na literatura, porém a posição mais forte para que uma rede seja considerada livre de escala é a de (BARABÁSI; ALBERT, 1999), que argumenta que redes livre de escala possuem $\gamma \in [2, 3]$. Porém, (BROIDO; CLAUSET, 2018) coloca que muitos autores consideram que basta $\gamma > 1$ para classificar a rede como livre de escala. Em (SEYED-ALLAEI; BIANCONI; MARSILI, 2006) se estuda propriedades de redes livres de escala com $\gamma < 2$, e discute que tais redes possuem características distintas das redes livre de escala $\gamma > 2$; o que pode explicar a divergência nas métricas encontrada na Subseção 4.4.1, em que as redes 8 e 9 apresentam algumas características de rede de pequeno mundo, enquanto que as redes comerciais

Tabela 4 – Características das redes comerciais brasileiras, rede de coautoria de matemáticos e rede de atores de filmes

Tipo da rede	n	m	$\langle k \rangle$	$\langle k \rangle^{\text{ER}}$	γ	γ_n	γ_{out}	L	L^{ER}	C	C^{ER}	n.º
rede comercial empresarial (3º trim. 2018)	1 050 577	3 263 406	6.22	6.21	1.43	1.33	1.48	3.67	7.56	0.0160	5.9×10^{-6}	1
rede comercial empresarial (4º trim. 2018)	1 128 816	3 595 044	6.37	6.37	1.43	1.33	1.49	3.91	7.52	0.0194	5.6×10^{-6}	2
rede comercial empresarial (1º trim. 2019)	1 016 712	3 336 653	6.56	6.56	1.43	1.33	1.50	4.12	7.35	0.0193	6.5×10^{-6}	3
rede comercial empresarial (2º trim. 2019)	1 104 323	4 389 156	7.95	7.95	1.44	1.32	1.50	4.81	6.71	0.0204	7.2×10^{-6}	4
rede comercial de grupos empresariais (1º trim. 2019)	916 592	2 896 061	6.32	6.32	1.44	1.36	1.50	2.91	7.44	0.0374	6.9×10^{-3}	5
rede comercial de cidades	5 551	390 745	118.96	140.78	1.59	1.65	1.50	2.20	10.93	0.6922	0.0214	6
rede comercial de ramos de atividade econômica	1 449	198 447	221.12	273.91	1.76	1.76	1.63	1.93	11.07	0.7149	0.1526	7
rede de coautoria de matemáticos (BARABÁSI <i>et al.</i> , 2002)	70 975	-	3.9	-	2.5	2.5	2.5	9.5	8.21	0.59	3.6×10^{-6}	8
rede de atores de filmes (ALBERT; BARABÁSI, 2002)	212 250	-	28.78	-	2.3	2.3	2.3	4.54	3.65	0.79	1.6×10^{-7}	9

Fonte: Dados da pesquisa.

Nota – A marcação ER é para indicar os respectivos grafos aleatórios de Erdős e Rényi. O hífen (–) indica que não foi possível encontrar os valores na literatura.

Figura 15 – Distribuição de graus das redes 3 (linha 1), 6 (linha 2), e 7 (linha 3), para redes direcionadas e não direcionadas. A primeira coluna traz a distribuição da rede não direcionada. A segunda coluna traz a distribuição dos graus de saída. A terceira coluna traz a distribuição dos graus de entrada. As últimas três colunas trazem as funções de distribuição de probabilidade e distribuição acumulada complementares da utilização da biblioteca (ALSTOTT; BULLMORE; PLENZ, 2014). Os eixos estão na escala *log-log*.

Fonte: Elaborada pelo autor.

não. Outra consideração importante a se fazer é a de (GERLACH; ALTMANN, 2019), que argumenta que existe uma grande confusão em boa parte dos testes estatísticos (inclusive o teste de ajuste a *power-law*) em sistemas complexos, pois não se leva em consideração, em muitos casos, o fato de que as observações do conjunto de dados possam não ser independentes. A rede coporativa japonesa de (SUGIYAMA *et al.*, 2005), apresenta $\gamma = 2.06$, e o autor sugere que tal rede possui estrutura livre de escala.

Figura 16 – Cálculo da centralidade de grau de autovetores para as redes 6 e 7. Expõe as 10 cidades e os 10 CNAEs mais centrais para ambos aspectos.

Algumas medidas de centralidade são computacionalmente custosas de se calcular. Este trabalho foca em duas medidas simples; apenas para redes menores, como as redes 6 e 7. Existem trabalhos, como (KANG *et al.*, 2011), que descrevem algoritmos eficientes para o cálculo de centralidades como a centralidade de proximidade e intermediação, porém, decidiu-se não focar nesses aspectos. Outro ponto que se considera em relação a ausência dos cálculos de

centralidade para redes mais granulares, é que tais cálculos correm o risco de expor a identidade e características das empresas; algo não desejado neste trabalho.

A [Figura 16](#) traz os cálculos da centralidade para as redes 6 e 7. É interessante notar que o cálculo de centralidade de grau apresenta resultado distinto do cálculo da centralidade de autovetores. Tal fato é esperado, já que a centralidade de grau não considera as centralidades dos nós vizinhos e nem o peso das arestas. Como esperado, a cidade de São Paulo se destaca como cidade mais "central", e junto dela as cidades ao seu redor, que formam a região metropolitana de São Paulo. Também, são consideradas centrais, algumas capitais como: Rio de Janeiro, Manaus e Curitiba. Em relação ao CNAE, o comércio por atacado (divisão 46 do CNAE), domina a centralidade, liderado pelo CNAE 46.44-3-01, de comércio atacadista de medicamentos e drogas de uso humano. Aparecem como CNAEs centrais, também, o CNAE 21.21-1-01 (Fabricação de medicamentos alopáticos para uso humano) e o CNAE 47.71-7-01 (Comércio varejista de produtos farmacêuticos, sem manipulação de fórmulas). É interessante notar que os três CNAEs mais centrais se referem a cadeia farmacêutica: que se inicia pela fabricação do medicamento, passa pelo atacado e chega até varejo.

4.4.3 Comunidades

Figura 17 – Análise dos tamanhos das comunidades, obtidos pela aplicação do algoritmo de detecção de comunidades para a rede comercial empresarial (1º trimestre 2019)

(a) As comunidades e seus respectivos tamanhos

(b) Distribuição dos tamanhos das comunidades

A [Figura 17](#) e o [Quadro 2](#) apresentam os resultados da aplicação da técnica de detecção de comunidades. O tempo de execução do algoritmo foi de 11 horas, se estabilizando na modularidade 0.82. Detectou-se um total de 1 290 comunidades representadas com seus respectivos tamanhos no grafo da [Figura 17a](#), em que a maior delas contém um total de 120 550 membros. Pela [Figura 17b](#), nota-se uma distribuição de cauda longa, o que implica que inúmeras comuni-

dades pequenas coexistem com poucas comunidades grandes. Essa característica, é reportada por vários estudos de detecção de comunidade, como aponta (BARABÁSI *et al.*, 2016). Também é alegado, que tal distribuição não é resultado de algum algoritmo em particular e sim de propriedades inerentes de algumas redes.

Pelo [Quadro 2](#) temos que 47% da rede está contida dentro de uma das cinco maiores comunidades, o que corresponde a 36% do valor comercializado interno as comunidades (existe também o valor comercializado entre comunidades). Deve-se notar que as comunidades com mais membros, não são necessariamente as que comercializam mais valor. É muito difícil de saber o mecanismo da rede que a faz se agrupar em comunidades e quais são as características dessas comunidades. Pelo [Quadro 2](#), percebe-se que o que é comercializado interno as comunidades difere de uma para outra, assim como o ramo de atividade mais frequente. Tome como exemplo a comunidade 2, em que 68.39% do que foi comercializado são produtos farmacêuticos e o CNAE 46.45-1-01 mais comum se refere ao comércio atacadista de medicamentos e drogas de uso humano. Tal observação sugere que a comunidade 2 é formada por uma rede de farmácias e/ou empresas do setor de saúde. Outro exemplo é a comunidade 6 em que 35.57% do que foi comercializado é vestuário e seus acessórios de malha e o CNAE mais frequente é o de comércio varejista de artigos do vestuário e acessórios, o que sugere que a comunidade é formada por uma rede de lojas de roupas. Tais análises sugerem que cada comunidade representa um nicho de mercado. Se for possível fazer tal afirmação, uma possível aplicação para a detecção de comunidades é encontrar o tamanho de mercado de um determinado conjunto de produtos ou setor. Seria interessante analisar outras variáveis como cidade, tamanho das empresas e valor dos produtos, como também uma análise do que está sendo comercializado entre comunidades.

4.5 Considerações finais

Esta seção condensa as técnicas aplicadas em redes comerciais construídas a partir de um banco de dados de oitenta e quatro milhões de documentos fiscais. Tais técnicas são aplicadas a sete redes construídas a partir de quatro tipos de modelagens. Com os resultados da aplicação das técnicas, há uma tentativa de se entender a estrutura topológica de tais redes. Pelo que se pode notar, as redes comerciais estudadas apresentam algumas características de redes livres de escala como tendência a agrupamento, distribuição de grau similar a *power-law*, menor caminho médio pequeno e a presença de comunidades. É desejável que mais técnicas sejam aplicadas e métricas coletadas para que se construa um conhecimento sólido em que novos trabalhos do ramo da economia e redes complexas possam se pautar.

Quadro 2 – As cinco maiores comunidades identificadas e suas características. Contém o identificador da comunidade, a quantidade de membros, o valor total (em R\$) comercializado dentro da comunidade (e a porcentagem do total comercializada internamente, excluindo as transações entre comunidades), os NCMs e capítulos de NCMs mais comercializados (e suas porcentagens acumuladas) e os CNAEs mais frequentes.

Comunidade	NCMs	Capítulos de NCMs	CNAEs
Identificador: 4 Membros: 120 550 (11.8%) Valores: 13.8 Bi (11.4%)	33059000 (6.81%)	33 (16.25%)	47.11-3-02
	64022000 (3.95%)	22 (9.44%)	47.12-1-00
	22021000 (3.40%)	19 (9.35%)	46.39-7-01
	22030000 (2.69%)	21 (7.77%)	46.91-5-00
	33051000 (2.68%)	34 (6.35%)	46.93-1-00
Identificador: 5 Membros: 120 510 (11.8%) Valores: 2.8 Bi (2.3%)	39174090 (7.55%)	39 (21.79%)	47.44-0-99
	84818019 (2.99%)	84 (11.38%)	47.44-0-05
	39222000 (2.58%)	85 (10.17%)	47.44-0-01
	69109000 (2.37%)	73 (6.66%)	46.79-6-99
	39259090 (2.15%)	82 (6.35%)	94.91-0-00
Identificador: 0 Membros: 114 301 (11.2%) Valores: 7.9 Bi (6.6%)	22021000 (6.75%)	22 (14.55%)	47.11-3-02
	04031000 (4.83%)	19 (13.16%)	56.11-2-01
	22030000 (4.02%)	07 (12.34%)	46.39-7-01
	19059090 (3.46%)	04 (11.74%)	47.12-1-00
	16010000 (2.81%)	02 (7.17%)	46.34-6-01
Identificador: 6 Membros: 72 348 (7.1%) Valores: 6.6 Bi (5.5%)	90015000 (12.57%)	61 (35.57%)	47.81-4-00
	64029990 (6.80%)	64 (19.21%)	14.12-6-01
	61091000 (4.91%)	90 (15.42%)	47.82-2-01
	64041900 (4.78%)	62 (12.73%)	47.74-1-00
	61062000 (2.60%)	63 (2.85%)	15.31-9-01
Identificador: 2 Membros: 56 404 (5.5%) Valores: 13.1 Bi (10.9%)	30049099 (16.88%)	30 (68.39%)	46.44-3-01
	30049069 (9.86%)	33 (11.82%)	47.71-7-01
	30049079 (4.03%)	21 (4.27%)	46.45-1-01
	30049039 (3.20%)	96 (4.27%)	86.10-1-01
	21069030 (3.01%)	90 (2.93%)	46.46-0-01

Fonte: Dados da pesquisa.

CONCLUSÃO

Neste trabalho, construiu-se, a partir de dados de notas fiscais eletrônicas, sete tipos de redes complexas que modelam o seguinte aspecto das transações econômicas realizadas na economia brasileira: comercialização de produtos entre empresas, entre grupos empresariais, entre cidades e entre ramos de atividade econômica. Tais redes são tratadas, aqui, como redes econômicas comerciais. Técnicas de análise de rede como menor caminho médio, coeficiente de clusterização, distribuição de graus, centralidade de grau, centralidade de autovetores e detecção de comunidades são aplicadas visando a caracterização destas redes comerciais.

Encontrou-se os seguintes fatos sobre as redes comerciais brasileiras: o menor caminho médio varia de 2.20 a 4.81; o coeficiente de clusterização para redes empresarias (1 a 5) variam de 0.0160 a 0.0374, e apesar de serem muito superiores as redes aleatórias, apresentam valores abaixo quando comparado a outras redes do mundo real; o coeficiente de clusterização para as redes 6 e 7 apresentam valores bem altos (0.6922 e 0.7149, respectivamente); a distribuição de graus das redes comerciais seguem a curva *power-law* com expoentes que variam de 1.43 a 1.76; as medidas de centralidade apontam a metrópole de São Paulo e as capitais Rio de Janeiro, Curitiba e Manaus como os grandes centros que movimentam a economia, e do ponto de vista do ramo de atividade econômica, o setor farmacêutico apresentou maior destaque; e, por fim, foram identificadas um total de 1290 comunidades ($Q = 0.82$) na estrutura das redes comerciais empresariais com características similares a estruturas encontradas em outras redes da literatura, com um leve indício de que o mecanismo de formação dessas comunidades se dá pelo tipo de produto sendo comercializado e a atividade econômica das empresas (por exemplo, a comunidade 2 contém várias empresas do ramo hospitalar e farmacêutico com comercialização de muitos remédios), fato que se melhor estudado, por ser usado como técnica para estudos de tamanho de mercado.

Apesar de ter sido feito um uso pequeno de técnicas de análise de rede dentre as inúmeras existentes, e das principais técnicas escolhidas (menor caminho médio, coeficiente

de clusterização e distribuição de graus) acabarem, por muitas vezes, produzindo análises incompletas (RODRIGUES; COSTA, 2007), acredita-se que o primeiro objetivo do trabalho foi atingido. Principalmente, dado que o estudo de redes complexas na economia é algo recente. E não foi encontrado por este autor, nenhum estudo de redes econômicas comerciais brasileiras.

Desde o início deste trabalho, sabia-se da dificuldade que seria em lidar com grafos de larga escala. Muitos algoritmos possuem complexidade computacional elevada inviabilizando seus usos. Contudo, tais desafios conseguiram ser contornados através do uso de técnicas otimizadas (como no caso do algoritmo de detecção de comunidades de (NEWMAN, 2004b)), implementações aproximadas (como no caso do algoritmo *Effective Closeness* de (KANG *et al.*, 2011)) e da utilização de um *cluster* de *Apache Flink*. Assim, julga-se que, devido o aprendizado destas técnicas e o uso de ferramentas para processamento de grafos em larga escala, o segundo objetivo foi alcançado com sucesso.

E por fim, destaca-se a realização do terceiro objetivo que é a aproximação do contexto profissional com o contexto acadêmico, na qual se deseja que a academia leve até o mercado de trabalho conhecimentos que visem inovação em processos, metodologias e técnicas, e o profissional do mercado leve a academia problemas relevantes e desafiadores. Não há dúvidas que este objetivo foi atingido, beneficiando ambos os lados.

5.1 Trabalhos futuros

Este trabalho foca em um conjunto reduzido de técnicas simples, porém essenciais, que formam um alicerce para trabalhos futuros. Existem diversas possibilidades a serem exploradas para que se dê continuidade a este trabalho. A princípio acredita-se que a direcionalidade e os pesos das arestas deveriam ser mais trabalhados. Foram aplicadas poucas técnicas em cima dessas duas características. Além disso, se faz necessário a utilização de mais métricas. Por exemplo, métricas como assortatividade e *rich-club* podem ser utilizadas para entender melhor as tendências de conectividade dentro da rede. Há possibilidade de se realizar análises de vulnerabilidade, robustez, aninhamento e *motifs*. Talvez seja interessante também, olhar para a rede do ponto de vista temporal, e entender seu dinamismo.

Ademais, todo este trabalho foi realizado em cima de um conjunto restrito de informações da NF-e. Não só existem outros campos dentro da nota fiscal eletrônica, como também, existem outros tipos de documentos fiscais relacionadas a nota fiscal eletrônica. E também, existem muitas outras modelagens de redes complexas possíveis dentro deste universo. São muitas as possibilidades, e este trabalho é só um começo de estudos que unem redes complexas e documentos fiscais.

REFERÊNCIAS

ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of modern physics**, APS, v. 74, n. 1, p. 47, 2002. Citado nas páginas [21](#), [41](#), [42](#), [57](#), [58](#), [62](#), [63](#) e [64](#).

ALMEIDA, F. d. S. e. Compliance tributário: Contexto, atualidades, riscos, e aplicação nas empresas. In: OLIVEIRA, F. R. de; ALMEIDA, F. da Silva e; MADRUGA, E. (Ed.). **Compliance tributário: práticas, riscos e atualidades**. [S.l.]: Realejo, 2018. cap. 1, p. 9. Citado na página [32](#).

ALSTOTT, J.; BULLMORE, E.; PLENZ, D. powerlaw: a python package for analysis of heavy-tailed distributions. **PloS one**, Public Library of Science, v. 9, n. 1, p. e85777, 2014. Citado nas páginas [11](#), [58](#), [63](#) e [65](#).

ALVES, L. A.; MANGIONI, G.; RODRIGUES, F.; PANZARASA, P.; MORENO, Y. Unfolding the complexity of the global value chain: Strength and entropy in the single-layer, multiplex, and multi-layer international trade networks. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 20, n. 12, p. 909, 2018. Citado na página [53](#).

ALVES, L. G.; MANGIONI, G.; CINGOLANI, I.; RODRIGUES, F. A.; PANZARASA, P.; MORENO, Y. The nested structural organization of the worldwide trade multi-layer network. **Scientific reports**, Nature Publishing Group, v. 9, n. 1, p. 2866, 2019. Citado na página [53](#).

AUSTIN, T.; FAGEN, R.; PENNEY, W.; RIORDAN, J. The number of components in random linear graphs. **The Annals of Mathematical Statistics**, JSTOR, v. 30, n. 3, p. 747–754, 1959. Citado na página [35](#).

BARABÁSI, A.-L. **Linked: The new science of networks**. [S.l.]: AAPT, 2003. Citado na página [38](#).

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999. Citado na página [63](#).

BARABÁSI, A.-L.; BONABEAU, E. Scale-free networks. **Scientific american**, JSTOR, v. 288, n. 5, p. 60–69, 2003. Citado na página [41](#).

BARABÁSI, A.-L.; JEONG, H.; NÉDA, Z.; RAVASZ, E.; SCHUBERT, A.; VICSEK, T. Evolution of the social network of scientific collaborations. **Physica A: Statistical mechanics and its applications**, Elsevier, v. 311, n. 3-4, p. 590–614, 2002. Citado nas páginas [58](#), [62](#), [63](#) e [64](#).

BARABÁSI, A.-L. *et al.* **Network science**. [S.l.]: Cambridge university press, 2016. Citado na página [68](#).

BENEDICTIS, L. D.; TAJOLI, L. The world trade network. **The World Economy**, Wiley Online Library, v. 34, n. 8, p. 1417–1454, 2011. Citado na página [21](#).

- BHATTACHARYA, K.; MUKHERJEE, G.; SARAMÄKI, J.; KASKI, K.; MANNA, S. S. The international trade network: weighted network analysis and modelling. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2008, n. 02, p. P02002, 2008. Citado na página 21.
- BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. **Journal of statistical mechanics: theory and experiment**, IOP Publishing, v. 2008, n. 10, p. P10008, 2008. Citado na página 50.
- BOETTCHER, S.; PERCUS, A. G. Extremal optimization for graph partitioning. **Physical Review E**, APS, v. 64, n. 2, p. 026114, 2001. Citado na página 50.
- BRASIL. Decreto de 3 de abril de 2000. institui grupo de trabalho interministerial para examinar e propor políticas, diretrizes e normas relacionadas com as novas formas eletrônicas de interação. 2000. Disponível em: <http://www.planalto.gov.br/ccivil_03/DNN/2000/Dnn8917.htm>. Acesso em: 31 ago. 2019. Citado na página 24.
- _____. Constituição (1988). emenda constitucional nº42, de 19 de dezembro de 2003. da nova redação ao art. 37 da constituição federal, introduzindo o inciso xxii. 2003. Disponível em: <http://www.planalto.gov.br/ccivil_03/Constituicao/Emendas/Emc/emc42.htm#art1>. Acesso em: 31 ago. 2019. Citado na página 24.
- _____. Decreto nº6.022, de 22 de janeiro de 2007 institui o sistema público de escrituração digital - sped. 2007. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6022.htm>. Acesso em: 1 set. 2019. Citado na página 26.
- BROIDO, A. D.; CLAUSET, A. Scale-free networks are rare. **arXiv preprint arXiv:1801.03400**, 2018. Citado nas páginas 41 e 63.
- CAZABET, R.; ROSSETTI, G.; AMBLARD, F. Dynamic community detection. In: **Encyclopedia of Social Network Analysis and Mining**. [S.l.: s.n.], 2017. p. 1–10. Citado na página 49.
- CHAKKARADHARI, J. **Large Scale Centrality Measures in Apache Flink and Apache Giraph**. Tese (Doutorado) — Technical University of Berlin, 2014. Citado na página 58.
- CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. **Physical review E**, APS, v. 70, n. 6, p. 066111, 2004. Citado na página 58.
- CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. Power-law distributions in empirical data. **SIAM review**, SIAM, v. 51, n. 4, p. 661–703, 2009. Citado na página 63.
- COMPLY. 2019. <<https://dictionary.cambridge.org/us/dictionary/english/comply>>. Acessada: 14 de setembro de 2019. Citado na página 23.
- CONSELHO NACIONAL DE POLÍTICA FAZENDÁRIA. Ajuste sinief 7, de 30 de setembro de 2005. institui a nota fiscal eletrônica e o documento auxiliar da nota fiscal eletrônica. 2005. Disponível em: <https://www.confaz.fazenda.gov.br/legislacao/ajustes/2005/AJ_007_05>. Acesso em: 1 set. 2019. Citado na página 25.
- _____. Ajuste sinief 2, de 3 de abril de 2009 . dispõe sobre a escrituração fiscal digital - efd. 2009. Disponível em: <https://www.confaz.fazenda.gov.br/legislacao/ajustes/2009/AJ_002_09>. Acesso em: 15 set. 2019. Citado na página 28.

_____. Ajuste sinief 5, de 30 de março de 2012. altera o ajuste sinief 07/05, que institui a nota fiscal eletrônica e o documento auxiliar da nota fiscal eletrônica. 2012. Disponível em: <https://www.confaz.fazenda.gov.br/legislacao/ajustes/2012/AJ_005_12>. Acesso em: 1 set. 2019. Citado na página 28.

_____. Ajuste sinief 11, de 26 de julho de 2013 . altera o ajuste sinief 07/05, que institui a nota fiscal eletrônica e o documento auxiliar da nota fiscal eletrônica. 2013. Disponível em: <https://www.confaz.fazenda.gov.br/legislacao/ajustes/2013/AJ_011_13>. Acesso em: 1 set. 2019. Citado na página 28.

_____. 2015. Descreve a finalidade do órgão. Disponível em: <<http://fazenda.gov.br/carta-de-servicos/lista-de-servicos/conselho-nacional-de-politica-fazendaria-confaz>>. Acesso em: 01 set. 2019. Citado na página 25.

CONTA AZUL. **CNAE: aprenda a classificar sua empresa**. 2018. Publicado por Carin Tom em 9 jul. 2018. Disponível em: <<https://blog.contaazul.com/cnae-o-que-e-confira-tabela/>>. Acesso em: 13 set. 2018. Citado na página 29.

CONTABILIZEI. **Tabela NCM Completa**. 2018. Disponível em: <<https://www.contabilizei.com.br/contabilidade-online/ncm/>>. Acesso em: 13 set. 2018. Citado na página 30.

COUNCIL, N. R. **Network Science**. Washington, DC: The National Academies Press, 2005. ISBN 978-0-309-10026-7. Disponível em: <<https://www.nap.edu/catalog/11516/network-science>>. Citado na página 21.

CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. **Inter-Journal**, Complex Systems, p. 1695, 2006. Disponível em: <<http://igraph.org>>. Citado nas páginas 57 e 58.

DUCH, J.; ARENAS, A. Community detection in complex networks using extremal optimization. **Physical review E**, APS, v. 72, n. 2, p. 027104, 2005. Citado na página 50.

EMMERT-STREIB, F.; TRIPATHI, S.; YLI-HARJA, O. P.; DEHMER, M. Understanding the world economy in terms of networks: A survey of data-based network science approaches on economic networks. **Frontiers in Applied Mathematics and Statistics**, Frontiers, v. 4, p. 37, 2018. Citado nas páginas 53 e 54.

EMPRESÔMETRO. 2019. Página Principal. O Empresômetro possui dados da 19 823 556 empresas ativas no Brasil. Disponível em: <<https://www.empresometro.com.br/>>. Acesso em: 13 novembro de 2019. Citado na página 59.

ENCONTRO NACIONAL DE ADMINISTRADORES TRIBUTÁRIOS. Disponível em: <http://www.enat.receita.fazenda.gov.br/pt-br/area_nacional/areas_interesse/foruns>. Acesso em: 31 ago. 2019. Citado na página 24.

_____. 2005. Encontro realizado dias 25, 26 e 27 de agosto de 2005, onde se protocolou o SPED e NF-e. Disponível em: <http://www.enat.receita.fazenda.gov.br/pt-br/area_nacional/areas_interesse/foruns/historico/ii-enat/ii-enat-sao-paulo-sp>. Acesso em: 31 ago. 2019. Citado na página 25.

ENCONTRO NACIONAL DE COORDENADORES E ADMINISTRADORES TRIBUTÁRIOS ESTADUAIS. **Manual de Orientação do Contribuinte**: Padrões técnicos de comunicação. 2015. Versão 6.0. Setembro 2015. Define as especificações e critérios técnicos necessários para a

integração entre os Portais das Secretarias de Fazendas dos Estados e os sistemas de informações das empresas emissoras de NF-e. Disponível em: <<http://www.nfe.fazenda.gov.br/>>. Acesso em: 13 set. 2019. Citado na página 28.

ERBER, G.; HAGEMANN, H. Economics of networks. In: ZIMMERMANN, K. F. (Ed.). **Frontiers in economics**. [S.l.]: Springer Science & Business Media, 2002. p. 235–268. Citado na página 53.

ERDŐS, P.; RÉNYI, A. On random graphs. **Publicationes Mathematicae Debrecen**, v. 6, p. 290–297, 1959. Citado na página 35.

_____. On the evolution of random graphs. **Publ. Math. Inst. Hung. Acad. Sci.**, v. 5, n. 1, p. 17–60, 1960. Citado nas páginas 35 e 36.

ERIKSEN, K. A.; SIMONSEN, I.; MASLOV, S.; SNEPPEN, K. Modularity and extreme edges of the internet. **Physical review letters**, APS, v. 90, n. 14, p. 148701, 2003. Citado na página 48.

FIEDLER, M. Algebraic connectivity of graphs. **Czechoslovak mathematical journal**, Institute of Mathematics, Academy of Sciences of the Czech Republic, v. 23, n. 2, p. 298–305, 1973. Citado na página 48.

FORTUNATO, S. Community detection in graphs. **Physics reports**, Elsevier, v. 486, n. 3-5, p. 75–174, 2010. Citado na página 49.

FORTUNATO, S.; HRIC, D. Community detection in networks: A user guide. **Physics reports**, Elsevier, v. 659, p. 1–44, 2016. Citado nas páginas 47 e 50.

GALLARDO, P. F. Google's secret and linear algebra. **EMS Newsletter**, v. 63, p. 10–15, 2007. Citado na página 45.

GARAS, A.; ARGYRAKIS, P.; ROZENBLAT, C.; TOMASSINI, M.; HAVLIN, S. Worldwide spreading of economic crisis. **New journal of Physics**, IOP Publishing, v. 12, n. 11, p. 113043, 2010. Citado na página 21.

GERLACH, M.; ALTMANN, E. G. Testing statistical laws in complex systems. **Physical review letters**, APS, v. 122, n. 16, p. 168301, 2019. Citado na página 66.

GILBERT, E. N. Random graphs. **The Annals of Mathematical Statistics**, JSTOR, v. 30, n. 4, p. 1141–1144, 1959. Citado na página 35.

GIRVAN, M.; NEWMAN, M. E. Community structure in social and biological networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 99, n. 12, p. 7821–7826, 2002. Citado nas páginas 47 e 48.

GOYAL, S. **Connections: an introduction to the economics of networks**. [S.l.]: Princeton University Press, 2012. Citado na página 53.

GUIMERA, R.; AMARAL, L. A. N. Functional cartography of complex metabolic networks. **nature**, Nature Publishing Group, v. 433, n. 7028, p. 895, 2005. Citado na página 50.

HAGBERG, A.; SWART, P.; CHULT, D. S. **Exploring network structure, dynamics, and function using NetworkX**. [S.l.], 2008. Citado na página 57.

HAMMOND, A. L. Digitally empowered development. **Foreign Aff.**, HeinOnline, v. 80, p. 96, 2001. Citado na página 24.

HUMPHRIES, M. D.; GURNEY, K. Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence. **PloS one**, Public Library of Science, v. 3, n. 4, p. e0002051, 2008. Citado na página 63.

JOHNSON, N. F. **Two’s company, three is complexity: A simple guide to the science of all sciences**. [S.l.]: Oneworld Pubns Ltd, 2007. Citado na página 21.

KANG, U.; PAPADIMITRIOU, S.; SUN, J.; TONG, H. Centralities in large networks: Algorithms and observations. In: SIAM. **Proceedings of the 2011 SIAM international conference on data mining**. [S.l.], 2011. p. 119–130. Citado nas páginas 58, 66 e 72.

KARONSKI, M. I.; RUCINSKI, A. The origins of the theory of random graphs. **The Mathematics of Paul Erdős I**, Springer, p. 371, 2013. Citado na página 36.

KHAN, B. S.; NIAZI, M. A. Network community detection: A review and visual survey. **arXiv preprint arXiv:1708.00977**, 2017. Citado nas páginas 49 e 58.

KLUYVER, T.; RAGAN-KELLEY, B.; PÉREZ, F.; GRANGER, B. E.; BUSSONNIER, M.; FREDERIC, J.; KELLEY, K.; HAMRICK, J. B.; GROUT, J.; CORLAY, S. *et al.* Jupyter notebooks-a publishing format for reproducible computational workflows. In: **ELPUB**. [S.l.: s.n.], 2016. p. 87–90. Citado na página 57.

KRANTON, R. E.; MINEHART, D. F. A theory of buyer-seller networks. **American economic review**, v. 91, n. 3, p. 485–508, 2001. Citado na página 53.

LANGVILLE, A. N.; MEYER, C. D.; FERNÁNDEZ, P. Google’s pagerank and beyond: The science of search engine rankings. **The Mathematical Intelligencer**, Springer, v. 30, n. 1, p. 68–69, 2008. Citado na página 45.

MANZI, V. A. **Compliance no Brasil: consolidação e perspectivas**. [S.l.]: Saint Paul Editora, 2008. 15 p. Citado na página 23.

MELNIK, S.; GUBAREV, A.; LONG, J. J.; ROMER, G.; SHIVAKUMAR, S.; TOLTON, M.; VASSILAKIS, T. Dremel: Interactive analysis of web-scale datasets. In: **Proc. of the 36th Int’l Conf on Very Large Data Bases**. [s.n.], 2010. p. 330–339. Disponível em: <<http://www.vldb2010.org/accept.htm>>. Citado na página 54.

NATIONS, U. United nations e-government survey 2014: E-government for the future we want. **United Nations Department of economic and social affairs**, 2014. Citado na página 24.

_____. United nations e-government survey 2018: E-government for the future we want. **United Nations Department of economic and social affairs**, 2018. Citado na página 24.

NEWMAN, M. Networks: An introduction. Oxford University Press, Inc., 2010. Citado nas páginas 36, 37, 38, 39, 40, 43, 47 e 58.

NEWMAN, M. E. Random graphs as models of networks. **arXiv preprint cond-mat/0202208**, 2002. Citado na página 38.

_____. Analysis of weighted networks. **Physical review E**, APS, v. 70, n. 5, p. 056131, 2004. Citado na página 59.

_____. Fast algorithm for detecting community structure in networks. **Physical review E**, APS, v. 69, n. 6, p. 066133, 2004. Citado nas páginas 49, 50 e 72.

_____. Modularity and community structure in networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 103, n. 23, p. 8577–8582, 2006. Citado na página 48.

NEWMAN, M. E.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical review E**, APS, v. 69, n. 2, p. 026113, 2004. Citado na página 49.

NG, A. Y.; JORDAN, M. I.; WEISS, Y. On spectral clustering: Analysis and an algorithm. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2002. p. 849–856. Citado na página 48.

OOBJ. 2007. Publicado por Amanda Gouvêa em 26 de julho de 2017. Disponível em: <<https://blog.oobj.com.br/importancia-guardar-notas-fiscais-5-anos/>>. Acesso em: 1 set. 2019. Citado na página 27.

PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. **The PageRank citation ranking: Bringing order to the web**. [S.l.], 1999. Citado nas páginas 44 e 45.

RODRIGUES, F. A.; COSTA, L. d. F. Caracterização, classificação e análise de redes complexas. 2007. Citado na página 72.

SCHWEITZER, F.; FAGIOLO, G.; SORNETTE, D.; VEGA-REDONDO, F.; WHITE, D. R. Economic networks: What do we know and what do we need to know? **Advances in Complex Systems**, World Scientific, v. 12, n. 04n05, p. 407–422, 2009. Citado na página 53.

SECRETARIA DA FAZENDA. 2019. Estatísticas em Arquivo. Quantidade de NF-e, por mês, de 2016 a 2018. Disponível em: <<http://www.nfe.fazenda.gov.br/portal/infoEstatisticas.aspx>>. Acesso em: 13 novembro de 2019. Citado na página 59.

SEYED-ALLAEI, H.; BIANCONI, G.; MARSILI, M. Scale-free networks with an exponent less than two. **Physical Review E**, APS, v. 73, n. 4, p. 046113, 2006. Citado na página 63.

SOUMA, W.; FUJIWARA, Y.; AOYAMA, H. Complex networks and economics. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 324, n. 1-2, p. 396–401, 2003. Citado na página 53.

SUGIYAMA, K.; HONDA, O.; OHSAKI, H.; IMASE, M. Application of network analysis techniques for japanese corporate transaction network. In: IEEE. **6th Asia-Pacific Symposium on Information and Telecommunication Technologies**. [S.l.], 2005. p. 387–392. Citado nas páginas 58, 61, 62 e 66.

TRAVERS, J.; MILGRAM, S. The small world problem. **Psychology Today**, JSTOR, v. 1, n. 1, p. 61–67, 1967. Citado na página 38.

UMBHAUER, G. Introduction. In: STAHN, H.; LLERENA, P.; COHENDET, P.; UMBHAUER, G. (Ed.). **The economics of networks: interaction and behaviours**. [S.l.]: Springer, 1998. p. 1. Citado na página 53.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **nature**, Nature Publishing Group, v. 393, n. 6684, p. 440, 1998. Citado nas páginas 39, 40, 46 e 62.

WELLMAN, B. The development of social network analysis: A study in the sociology of science. **Contemporary Sociology**, American Sociological Association, v. 37, n. 3, p. 221, 2008. Citado na página [21](#).

WHITE, D. Taxes in the ancient world. university of pennsylvania. **Almanac**, v. 48, n. 28, 2002. Citado na página [32](#).

LISTA DE INTERESSE DE CFOPS

Tabela 5 – Lista de interesse de CFOPs separada por um especialista fiscal com objetivo de restringir as operações que qualificassem transação comercial envolvendo recebíveis. Em outras palavras, exclui-se todas as operações referentes a manipulação fiscal sem rendimentos diretos (e.g. realocação de estoque, bonificação, doação, e etc.), restando apenas operações com natureza rentável (compras, vendas, industrialização, e afins).

Descrição	CFOP
ENTRADAS OU AQUISIÇÕES DE SERVIÇOS DO ESTADO	1000
Compra para industrialização ou produção rural	1101
Compra para comercialização	1102
Compra para industrialização ou produção rural originada de encomenda para recebimento futuro	1116
Compra para comercialização originada de encomenda para recebimento futuro	1117
Compra para industrialização em que a mercadoria foi remetida pelo fornecedor ao industrializador sem transitar pelo estabelecimento adquirente	1122
Industrialização efetuada por outra empresa	1124
Industrialização efetuada por outra empresa quando a mercadoria remetida para utilização no processo de industrialização não transitou pelo estabelecimento adquirente da mercadoria	1125
COMPRAS PARA INDUSTRIALIZAÇÃO, COMERCIALIZAÇÃO OU PRESTAÇÃO DE SERVIÇOS	1100
Compra de mercadoria para comercialização pelo adquirente originário, entregue pelo vendedor remetente ao destinatário, em venda à ordem	1118
Compra para industrialização, em venda à ordem, já recebida do vendedor remetente	1120
Compra para comercialização, em venda à ordem, já recebida do vendedor remetente	1121
COMPRAS DE ENERGIA ELÉTRICA	1250
Compra de energia elétrica para distribuição ou comercialização	1251
Compra de energia elétrica por estabelecimento industrial	1252
Compra de energia elétrica por estabelecimento comercial	1253
Compra de energia elétrica por estabelecimento prestador de serviço de transporte	1254
Compra de energia elétrica por estabelecimento prestador de serviço de comunicação	1255

Compra de energia elétrica por estabelecimento de produtor rural	1256
Compra de energia elétrica para consumo por demanda contratada	1257
AQUISIÇÕES DE SERVIÇOS DE COMUNICAÇÃO	1300
Aquisição de serviço de comunicação para execução de serviço da mesma natureza	1301
Aquisição de serviço de comunicação por estabelecimento industrial	1302
Aquisição de serviço de comunicação por estabelecimento comercial	1303
Aquisição de serviço de comunicação por estabelecimento de prestador de serviço de transporte	1304
Aquisição de serviço de comunicação por estabelecimento de geradora ou de distribuidora de energia elétrica	1305
Aquisição de serviço de comunicação por estabelecimento de produtor rural	1306
AQUISIÇÕES DE SERVIÇOS DE TRANSPORTE	1350
Aquisição de serviço de transporte para execução de serviço da mesma natureza	1351
Aquisição de serviço de transporte por estabelecimento industrial	1352
Aquisição de serviço de transporte por estabelecimento comercial	1353
Aquisição de serviço de transporte por estabelecimento de prestador de serviço de comunicação	1354
Aquisição de serviço de transporte por estabelecimento de geradora ou de distribuidora de energia elétrica	1355
Aquisição de serviço de transporte por estabelecimento de produtor rural	1356
Aquisição de serviço de transporte por contribuinte substituto em relação ao serviço de transporte	1360
ENTRADAS DE MERCADORIAS SUJEITAS AO REGIME DE SUBSTITUIÇÃO TRIBUTÁRIA	1400
Compra para industrialização ou produção rural em operação com mercadoria sujeita ao regime de substituição tributária	1401
Compra para comercialização em operação com mercadoria sujeita ao regime de substituição tributária	1403
Compra de bem para o ativo imobilizado cuja mercadoria está sujeita ao regime de substituição tributária	1406
Compra de mercadoria para uso ou consumo cuja mercadoria está sujeita ao regime de substituição tributária	1407
OPERAÇÕES COM BENS DE ATIVO IMOBILIZADO E MATERIAIS PARA USO OU CONSUMO	1550
Compra de bem para o ativo imobilizado	1551
Compra de material para uso ou consumo	1556
OUTRAS ENTRADAS DE MERCADORIAS OU AQUISIÇÕES DE SERVIÇOS	1900
Entrada para industrialização por encomenda	1901
Entrada de mercadoria remetida para industrialização e não aplicada no referido processo	1903
Entrada de mercadoria recebida para depósito em depósito fechado ou armazém geral	1905
Lançamento efetuado a título de simples faturamento decorrente de compra para recebimento futuro	1922
Lançamento efetuado a título de reclassificação de mercadoria decorrente de formação de kit ou de sua desagregação	1926

Aquisição de serviço de transporte iniciado em unidade da Federação diversa daquela onde inscrito o prestador	1932
Aquisição de serviço tributado pelo ISSQN	1933
Entrada de mercadoria recebida do vendedor remetente, em venda à ordem	1923
Entrada para industrialização por conta e ordem do adquirente da mercadoria, quando esta não transitar pelo estabelecimento do adquirente	1924
Lançamento efetuado pelo tomador do serviço de transporte quando a responsabilidade de retenção do imposto for atribuída ao remetente ou alienante da mercadoria, pelo serviço de transporte realizado por transportador autônomo ou por transportador não inscrito na unidade da Federação onde iniciado o serviço	1931
ENTRADAS OU AQUISIÇÕES DE SERVIÇOS DE OUTROS ESTADOS	2000
Compra para industrialização ou produção rural	2101
Compra para comercialização	2102
Compra para industrialização ou produção rural originada de encomenda para recebimento futuro	2116
Compra para comercialização originada de encomenda para recebimento futuro	2117
Compra para industrialização em que a mercadoria foi remetida pelo fornecedor ao industrializador sem transitar pelo estabelecimento adquirente	2122
Industrialização efetuada por outra empresa	2124
Industrialização efetuada por outra empresa quando a mercadoria remetida para utilização no processo de industrialização não transitou pelo estabelecimento adquirente da mercadoria	2125
Compra para utilização na prestação de serviço	2126
COMPRAS PARA INDUSTRIALIZAÇÃO, COMERCIALIZAÇÃO OU PRESTAÇÃO DE SERVIÇOS	2100
Compra de mercadoria para comercialização pelo adquirente originário, entregue pelo vendedor remetente ao destinatário, em venda à ordem	2118
Compra para industrialização, em venda à ordem, já recebida do vendedor remetente	2120
Compra para comercialização, em venda à ordem, já recebida do vendedor remetente	2121
COMPRAS DE ENERGIA ELÉTRICA	2250
Compra de energia elétrica para distribuição ou comercialização	2251
Compra de energia elétrica por estabelecimento industrial	2252
Compra de energia elétrica por estabelecimento comercial	2253
Compra de energia elétrica por estabelecimento prestador de serviço de transporte	2254
Compra de energia elétrica por estabelecimento prestador de serviço de comunicação	2255
Compra de energia elétrica por estabelecimento de produtor rural	2256
Compra de energia elétrica para consumo por demanda contratada	2257
AQUISIÇÕES DE SERVIÇOS DE COMUNICAÇÃO	2300
Aquisição de serviço de comunicação para execução de serviço da mesma natureza	2301
Aquisição de serviço de comunicação por estabelecimento industrial	2302
Aquisição de serviço de comunicação por estabelecimento comercial	2303
Aquisição de serviço de comunicação por estabelecimento de prestador de serviço de transporte	2304
Aquisição de serviço de comunicação por estabelecimento de geradora ou de distribuidora de energia elétrica	2305

Aquisição de serviço de comunicação por estabelecimento de produtor rural	2306
Aquisição de serviço de transporte para execução de serviço da mesma natureza	2351
Aquisição de serviço de transporte por estabelecimento industrial	2352
Aquisição de serviço de transporte por estabelecimento comercial	2353
Aquisição de serviço de transporte por estabelecimento de prestador de serviço de comunicação	2354
Aquisição de serviço de transporte por estabelecimento de geradora ou de distribuidora de energia elétrica	2355
Aquisição de serviço de transporte por estabelecimento de produtor rural	2356
ENTRADAS DE MERCADORIAS SUJEITAS AO REGIME DE SUBSTITUIÇÃO TRIBUTÁRIA	2400
Compra para industrialização ou produção rural em operação com mercadoria sujeita ao regime de substituição tributária	2401
Compra para comercialização em operação com mercadoria sujeita ao regime de substituição tributária	2403
Compra de bem para o ativo imobilizado cuja mercadoria está sujeita ao regime de substituição tributária	2406
Compra de mercadoria para uso ou consumo cuja mercadoria está sujeita ao regime de substituição tributária	2407
Entrada de mercadoria recebida com fim específico de exportação	2501
OPERAÇÕES COM BENS DE ATIVO IMOBILIZADO E MATERIAIS PARA USO OU CONSUMO	2550
Compra de bem para o ativo imobilizado	2551
Compra de material para uso ou consumo	2556
Entrada de bem do ativo imobilizado de terceiro, remetido para uso no estabelecimento	2555
CRÉDITOS E RESSARCIMENTOS DE ICMS	2600
Ressarcimento de ICMS retido por substituição tributária	2603
Compra de combustível ou lubrificante para industrialização subsequente	2651
Compra de combustível ou lubrificante para comercialização	2652
Compra de combustível ou lubrificante por consumidor ou usuário final	2653
Entrada de combustível ou lubrificante para armazenagem	2663
OUTRAS ENTRADAS DE MERCADORIAS OU AQUISIÇÕES DE SERVIÇOS	2900
Entrada para industrialização por encomenda	2901
Entrada de mercadoria remetida para industrialização e não aplicada no referido processo	2903
Entrada de mercadoria recebida para depósito em depósito fechado ou armazém geral	2905
Lançamento efetuado a título de simples faturamento decorrente de compra para recebimento futuro	2922
Aquisição de serviço de transporte iniciado em unidade da Federação diversa daquela onde inscrito o prestador	2932
Aquisição de serviço tributado pelo ISSQN	2933
Entrada de mercadoria recebida do vendedor remetente, em venda à ordem	2923
Entrada para industrialização por conta e ordem do adquirente da mercadoria, quando esta não transitar pelo estabelecimento do adquirente	2924

Lançamento efetuado pelo tomador do serviço de transporte quando a responsabilidade de retenção do imposto for atribuída ao remetente ou alienante da mercadoria, pelo serviço de transporte realizado por transportador autônomo ou por transportador não inscrito na unidade da Federação onde iniciado o serviço	2931
ENTRADAS OU AQUISIÇÕES DE SERVIÇOS DO EXTERIOR	3000
Compra para industrialização ou produção rural	3101
Compra para comercialização	3102
Compra para utilização na prestação de serviço	3126
COMPRAS PARA INDUSTRIALIZAÇÃO, COMERCIALIZAÇÃO OU PRESTAÇÃO DE SERVIÇOS	3100
Compra para industrialização sob o regime de "drawback"	3127
COMPRAS DE ENERGIA ELÉTRICA	3250
Compra de energia elétrica para distribuição ou comercialização	3251
AQUISIÇÕES DE SERVIÇOS DE COMUNICAÇÃO	3300
Aquisição de serviço de comunicação para execução de serviço da mesma natureza	3301
AQUISIÇÕES DE SERVIÇOS DE TRANSPORTE	3350
Aquisição de serviço de transporte para execução de serviço da mesma natureza	3351
Aquisição de serviço de transporte por estabelecimento industrial	3352
Aquisição de serviço de transporte por estabelecimento comercial	3353
Aquisição de serviço de transporte por estabelecimento de prestador de serviço de comunicação	3354
Aquisição de serviço de transporte por estabelecimento de geradora ou de distribuidora de energia elétrica	3355
Aquisição de serviço de transporte por estabelecimento de produtor rural	3356
OPERAÇÕES COM BENS DE ATIVO IMOBILIZADO E MATERIAIS PARA USO OU CONSUMO	3550
Compra de bem para o ativo imobilizado	3551
Compra de material para uso ou consumo	3556
Compra de combustível ou lubrificante para industrialização subsequente	3651
Compra de combustível ou lubrificante para comercialização	3652
Compra de combustível ou lubrificante por consumidor ou usuário final	3653
OUTRAS ENTRADAS DE MERCADORIAS OU AQUISIÇÕES DE SERVIÇOS	3900
Lançamento efetuado a título de entrada de bem sob amparo de regime especial aduaneiro de admissão temporária	3930
SAÍDAS OU PRESTAÇÕES DE SERVIÇOS PARA O ESTADO	5000
VENDAS DE PRODUÇÃO PRÓPRIA OU DE TERCEIROS	5100
Venda de produção do estabelecimento	5101
Venda de mercadoria adquirida ou recebida de terceiros	5102
Venda de produção do estabelecimento que não deva por ele transitar	5105
Venda de produção do estabelecimento originada de encomenda para entrega futura	5116
Industrialização efetuada para outra empresa	5124
Industrialização efetuada para outra empresa quando a mercadoria recebida para utilização no processo de industrialização não transitar pelo estabelecimento adquirente da mercadoria	5125
Venda de produção do estabelecimento, efetuada fora do estabelecimento	5103
Venda de mercadoria adquirida ou recebida de terceiros, efetuada fora do estabelecimento	5104

Venda de mercadoria adquirida ou recebida de terceiros, que não deva por ele transitar	5106
Venda de produção do estabelecimento, destinada à Zona Franca de Manaus ou Áreas de Livre Comércio	5109
Venda de mercadoria adquirida ou recebida de terceiros, destinada à Zona Franca de Manaus ou Áreas de Livre Comércio	5110
Venda de mercadoria adquirida ou recebida de terceiros, originada de encomenda para entrega futura	5117
Venda de produção do estabelecimento entregue ao destinatário por conta e ordem do adquirente originário, em venda à ordem	5118
Venda de mercadoria adquirida ou recebida de terceiros entregue ao destinatário por conta e ordem do adquirente originário, em venda à ordem	5119
Venda de mercadoria adquirida ou recebida de terceiros entregue ao destinatário pelo vendedor remetente, em venda à ordem	5120
Venda de produção do estabelecimento remetida para industrialização, por conta e ordem do adquirente, sem transitar pelo estabelecimento do adquirente	5122
Venda de mercadoria adquirida ou recebida de terceiros remetida para industrialização, por conta e ordem do adquirente, sem transitar pelo estabelecimento do adquirente	5123
VENDAS DE ENERGIA ELÉTRICA	5250
Venda de energia elétrica para distribuição ou comercialização	5251
Venda de energia elétrica para estabelecimento industrial	5252
Venda de energia elétrica para estabelecimento comercial	5253
Venda de energia elétrica para estabelecimento prestador de serviço de transporte	5254
Venda de energia elétrica para estabelecimento prestador de serviço de comunicação	5255
Venda de energia elétrica para estabelecimento de produtor rural	5256
Venda de energia elétrica para consumo por demanda contratada	5257
Venda de energia elétrica a não contribuinte	5258
PRESTAÇÕES DE SERVIÇOS DE COMUNICAÇÃO	5300
Prestação de serviço de comunicação para execução de serviço da mesma natureza	5301
Prestação de serviço de comunicação a estabelecimento industrial	5302
Prestação de serviço de comunicação a estabelecimento comercial	5303
Prestação de serviço de comunicação a estabelecimento de prestador de serviço de transporte	5304
Prestação de serviço de comunicação a estabelecimento de geradora ou de distribuidora de energia elétrica	5305
Prestação de serviço de comunicação a estabelecimento de produtor rural	5306
Prestação de serviço de comunicação a não contribuinte	5307
PRESTAÇÕES DE SERVIÇOS DE TRANSPORTE	5350
Prestação de serviço de transporte para execução de serviço da mesma natureza	5351
Prestação de serviço de transporte a estabelecimento industrial	5352
Prestação de serviço de transporte a estabelecimento comercial	5353
Prestação de serviço de transporte a estabelecimento de prestador de serviço de comunicação	5354
Prestação de serviço de transporte a estabelecimento de geradora ou de distribuidora de energia elétrica	5355
Prestação de serviço de transporte a estabelecimento de produtor rural	5356

Prestação de serviço de transporte a não contribuinte	5357
Prestação de serviço de transporte a contribuinte ou a não contribuinte quando a mercadoria transportada está dispensada de emissão de nota fiscal	5359
Prestação de serviço de transporte a contribuinte substituto em relação ao serviço de transporte	5360
SAÍDAS DE MERCADORIAS SUJEITAS AO REGIME DE SUBSTITUIÇÃO TRIBUTÁRIA	5400
Venda de produção do estabelecimento em operação com produto sujeito ao regime de substituição tributária, na condição de contribuinte substituto	5401
Venda de produção do estabelecimento de produto sujeito ao regime de substituição tributária, em operação entre contribuintes substitutos do mesmo produto	5402
Venda de mercadoria adquirida ou recebida de terceiros em operação com mercadoria sujeita ao regime de substituição tributária, na condição de contribuinte substituto	5403
Venda de mercadoria adquirida ou recebida de terceiros em operação com mercadoria sujeita ao regime de substituição tributária, na condição de contribuinte substituído	5405
SISTEMAS DE INTEGRAÇÃO	5450
OPERAÇÕES COM BENS DE ATIVO IMOBILIZADO E MATERIAIS PARA USO OU CONSUMO	5550
Venda de bem do ativo imobilizado	5551
CRÉDITOS E RESSARCIMENTOS DE ICMS	5600
Ressarcimento de ICMS retido por substituição tributária	5603
Utilização de saldo credor de ICMS para extinção por compensação de débitos fiscais	5606
Venda de combustível ou lubrificante de produção do estabelecimento destinado à industrialização subsequente	5651
Venda de combustível ou lubrificante de produção do estabelecimento destinado à comercialização	5652
Venda de combustível ou lubrificante de produção do estabelecimento destinado a consumidor ou usuário final	5653
Venda de combustível ou lubrificante adquirido ou recebido de terceiros destinado à industrialização subsequente	5654
Venda de combustível ou lubrificante adquirido ou recebido de terceiros destinado à comercialização	5655
Venda de combustível ou lubrificante adquirido ou recebido de terceiros destinado a consumidor ou usuário final	5656
OUTRAS SAÍDAS DE MERCADORIAS OU PRESTAÇÕES DE SERVIÇOS	5900
Lançamento efetuado a título de simples faturamento decorrente de venda para entrega futura	5922
Lançamento efetuado a título de reclassificação de mercadoria decorrente de formação de kit ou de sua desagregação	5926
Lançamento efetuado a título de baixa de estoque decorrente do encerramento da atividade da empresa	5928
Prestação de serviço de transporte iniciada em unidade da Federação diversa daquela onde inscrito o prestador	5932
Prestação de serviço tributado pelo ISSQN	5933
Lançamento efetuado a título de baixa de estoque decorrente de perda, roubo ou deterioração	5927

Lançamento efetuado em decorrência da responsabilidade de retenção do imposto por substituição tributária, atribuída ao remetente ou alienante da mercadoria, pelo serviço de transporte realizado por transportador autônomo ou por transportador não inscrito na unidade da Federação onde iniciado o serviço	5931
Lançamento efetuado em decorrência de emissão de documento fiscal relativo a operação ou prestação também registrada em equipamento Emissor de Cupom Fiscal	5929
SAÍDAS OU PRESTAÇÕES DE SERVIÇOS PARA OUTROS ESTADOS	6000
VENDAS DE PRODUÇÃO PRÓPRIA OU DE TERCEIROS	6100
Venda de produção do estabelecimento	6101
Venda de mercadoria adquirida ou recebida de terceiros	6102
Venda de produção do estabelecimento que não deva por ele transitar	6105
Venda de produção do estabelecimento originada de encomenda para entrega futura	6116
Industrialização efetuada para outra empresa	6124
Industrialização efetuada para outra empresa quando a mercadoria recebida para utilização no processo de industrialização não transitar pelo estabelecimento adquirente da mercadoria	6125
Venda de produção do estabelecimento, efetuada fora do estabelecimento	6103
Venda de mercadoria adquirida ou recebida de terceiros, efetuada fora do estabelecimento	6104
Venda de mercadoria adquirida ou recebida de terceiros, que não deva por ele transitar	6106
Venda de produção do estabelecimento, destinada a não contribuinte	6107
Venda de mercadoria adquirida ou recebida de terceiros, destinada a não contribuinte	6108
Venda de produção do estabelecimento, destinada à Zona Franca de Manaus ou Áreas de Livre Comércio	6109
Venda de mercadoria adquirida ou recebida de terceiros, destinada à Zona Franca de Manaus ou Áreas de Livre Comércio	6110
Venda de mercadoria adquirida ou recebida de terceiros, originada de encomenda para entrega futura	6117
Venda de produção do estabelecimento entregue ao destinatário por conta e ordem do adquirente originário, em venda à ordem	6118
Venda de mercadoria adquirida ou recebida de terceiros entregue ao destinatário por conta e ordem do adquirente originário, em venda à ordem	6119
Venda de mercadoria adquirida ou recebida de terceiros entregue ao destinatário pelo vendedor remetente, em venda à ordem	6120
Venda de produção do estabelecimento remetida para industrialização, por conta e ordem do adquirente, sem transitar pelo estabelecimento do adquirente	6122
Venda de mercadoria adquirida ou recebida de terceiros remetida para industrialização, por conta e ordem do adquirente, sem transitar pelo estabelecimento do adquirente	6123
VENDAS DE ENERGIA ELÉTRICA	6250
Venda de energia elétrica para distribuição ou comercialização	6251
Venda de energia elétrica para estabelecimento industrial	6252
Venda de energia elétrica para estabelecimento comercial	6253
Venda de energia elétrica para estabelecimento prestador de serviço de transporte	6254
Venda de energia elétrica para estabelecimento prestador de serviço de comunicação	6255
Venda de energia elétrica para estabelecimento de produtor rural	6256
Venda de energia elétrica para consumo por demanda contratada	6257
Venda de energia elétrica a não contribuinte	6258

PRESTAÇÕES DE SERVIÇOS DE COMUNICAÇÃO	6300
Prestação de serviço de comunicação para execução de serviço da mesma natureza	6301
Prestação de serviço de comunicação a estabelecimento industrial	6302
Prestação de serviço de comunicação a estabelecimento comercial	6303
Prestação de serviço de comunicação a estabelecimento de prestador de serviço de transporte	6304
Prestação de serviço de comunicação a estabelecimento de geradora ou de distribuidora de energia elétrica	6305
Prestação de serviço de comunicação a estabelecimento de produtor rural	6306
Prestação de serviço de comunicação a não contribuinte	6307
PRESTAÇÕES DE SERVIÇOS DE TRANSPORTE	6350
Prestação de serviço de transporte para execução de serviço da mesma natureza	6351
Prestação de serviço de transporte a estabelecimento industrial	6352
Prestação de serviço de transporte a estabelecimento comercial	6353
Prestação de serviço de transporte a estabelecimento de prestador de serviço de comunicação	6354
Prestação de serviço de transporte a estabelecimento de geradora ou de distribuidora de energia elétrica	6355
Prestação de serviço de transporte a estabelecimento de produtor rural	6356
Prestação de serviço de transporte a não contribuinte	6357
Prestação de serviço de transporte a contribuinte ou a não contribuinte quando a mercadoria transportada está dispensada de emissão de nota fiscal -	6359
SAÍDAS DE MERCADORIAS SUJEITAS AO REGIME DE SUBSTITUIÇÃO TRIBUTÁRIA	6400
Venda de produção do estabelecimento em operação com produto sujeito ao regime de substituição tributária, na condição de contribuinte substituto	6401
Venda de produção do estabelecimento de produto sujeito ao regime de substituição tributária, em operação entre contribuintes substitutos do mesmo produto	6402
Venda de mercadoria adquirida ou recebida de terceiros em operação com mercadoria sujeita ao regime de substituição tributária, na condição de contribuinte substituto	6403
Venda de mercadoria sujeita ao regime de substituição tributária, cujo imposto já tenha sido retido anteriormente	6404
OPERAÇÕES COM BENS DE ATIVO IMOBILIZADO E MATERIAIS PARA USO OU CONSUMO	6550
Venda de bem do ativo imobilizado	6551
CRÉDITOS E RESSARCIMENTOS DE ICMS	6600
Ressarcimento de ICMS retido por substituição tributária	6603
Venda de combustível ou lubrificante de produção do estabelecimento destinado à industrialização subsequente	6651
Venda de combustível ou lubrificante de produção do estabelecimento destinado à comercialização	6652
Venda de combustível ou lubrificante de produção do estabelecimento destinado a consumidor ou usuário final	6653
Venda de combustível ou lubrificante adquirido ou recebido de terceiros destinado à industrialização subsequente	6654

Venda de combustível ou lubrificante adquirido ou recebido de terceiros destinado à comercialização	6655
Venda de combustível ou lubrificante adquirido ou recebido de terceiros destinado a consumidor ou usuário final	6656
OUTRAS SAÍDAS DE MERCADORIAS OU PRESTAÇÕES DE SERVIÇOS	6900
Lançamento efetuado a título de simples faturamento decorrente de venda para entrega futura	6922
Prestação de serviço de transporte iniciada em unidade da Federação diversa daquela onde inscrito o prestador	6932
Prestação de serviço tributado pelo ISSQN	6933
Lançamento efetuado em decorrência da responsabilidade de retenção do imposto por substituição tributária, atribuída ao remetente ou alienante da mercadoria, pelo serviço de transporte realizado por transportador autônomo ou por transportador não inscrito na unidade da Federação onde iniciado o serviço	6931
Lançamento efetuado em decorrência de emissão de documento fiscal relativo a operação ou prestação também registrada em equipamento Emissor de Cupom Fiscal	6929
SAÍDAS OU PRESTAÇÕES DE SERVIÇOS PARA O EXTERIOR	7000
VENDAS DE PRODUÇÃO PRÓPRIA OU DE TERCEIROS	7100
Venda de produção do estabelecimento	7101
Venda de mercadoria adquirida ou recebida de terceiros	7102
Venda de produção do estabelecimento, que não deva por ele transitar	7105
Venda de mercadoria adquirida ou recebida de terceiros, que não deva por ele transitar	7106
Venda de produção do estabelecimento sob o regime de "drawback"	7127
VENDAS DE ENERGIA ELÉTRICA	7250
Venda de energia elétrica para o exterior	7251
PRESTAÇÕES DE SERVIÇOS DE COMUNICAÇÃO	7300
Prestação de serviço de comunicação para execução de serviço da mesma natureza	7301
PRESTAÇÕES DE SERVIÇO DE TRANSPORTE	7350
Prestação de serviço de transporte	7358
EXPORTAÇÃO DE MERCADORIAS RECEBIDAS COM FIM ESPECÍFICO DE EXPORTAÇÃO	7500
Exportação de mercadorias recebidas com fim específico de exportação	7501
OPERAÇÕES COM BENS DE ATIVO IMOBILIZADO E MATERIAIS PARA USO OU CONSUMO	7550
Venda de bem do ativo imobilizado	7551
Venda de combustível ou lubrificante de produção do estabelecimento	7651
Venda de combustível ou lubrificante adquirido ou recebido de terceiros	7654