CEFET/RJ
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO

MINERAÇÃO DE PROCESSOS: UMA ANÁLISE DO IMPACTO ANTES E DEPOIS DA PANDEMIA NOS PROCESSOS ADMINISTRATIVOS, SOB A PERSPECTIVA ORGANIZACIONAL

Mirley Bitencourt Ferreira

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de Mestre.

Orientador: Rafael Barbastefano
Coorientador: Diego Carvalho

Rio de Janeiro
03/2023

MINERAÇÃO DE PROCESSOS: UMA ANÁLISE DO IMPACTO ANTES E DEPOIS DA PANDEMIA NOS PROCESSOS ADMINISTRATIVOS, SOB A PERSPECTIVA ORGANIZACIONAL

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de Mestre.

Mirley Bitencourt Ferreira

Banca Examinadora:

Presidente, Professor Dr. Rafael Garcia Barbastefano (CEFET/RJ) (orientador)

Professor Dr. Diego Carvalho (CEFET/RJ) (Coorientador)

Professor Dr. Eduardo Ogasawara (CEFET/RJ)

Professora Dra. Cristina Gomes De Souza (CEFET/RJ)

Professor Dr. Eduardo Gonçalves Serra (UFRJ)

SUPLENTES

Professor Dr. Igor Leão dos Santos (CEFET/RJ)

Rio de Janeiro
03/2023

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

F383 Ferreira, Mirley Bitencourt
Mineração de processos: uma análise do impacto antes e depois da pandemia nos processos administrativos, sob a perspectiva organizacional / Mirley Bitencourt Ferreira. — 2023.
162f. : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2023.
Bibliografia : f. 157-162
Orientador: Rafael Barbastefano
Coorientador: Diego Carvalho

1. Processo administrativo. 2. Mineração de dados (Computação). 3. COVID-19, Pandemia de, 2020-. 4. Eficiência organizacional. 5. Redes sociais. I. Barbastefano, Rafael (Orient.). II. Carvalho, Diego (Coorient.). III. Ttulo.

CDD 658

Elaborada pela bibliotecária Tania Mello – CRB/7 nº 5507/04

RESUMO

Mineração de processos: uma análise do impacto antes e depois da pandemia nos processos administrativos, sob a perspectiva organizacional

A pandemia do COVID-19 afetou muitos processos e modelos de negócios nos últimos anos. Neste período, as organizações tiveram que se adaptar ao novo formato de trabalho em casa. A mineração de processo (MP) é uma área que liga os processos de negócios e a mineração de dados onde extrai conhecimento de registros de eventos reais registrados em sistemas de informação. Esta pesquisa faz um estudo bibliométrico sobre MP e analisa os processos administrativos no período antes e depois da pandemia de uma instituição pública de ensino no Brasil. Aplicam-se técnicas de MP sob a perspectiva organizacional para entender como o referido conjunto de processos convencionais registrados em papel foi executado durante a pandemia. Este trabalho mostra que houve mudanças nos processos e que ocorreu uma simplificação na maioria dos processos administrativos, não apenas pela queda significativa em torno de 50% do volume dos eventos, mas na reestruturação dos departamentos. Encontrou-se nos dados minerados vários indícios de redução na duração do tempo e na transferência de trabalho entre departamentos e, também, que vários departamentos centrais foram modificados entre os períodos pandêmicos. Os resultados apresentam que os processos demandados são basicamente os mesmos, porém, há indicadores concretos de que os trabalhadores implementaram uma nova estrutura de processos durante a pandemia o que mostrou maior eficiência. Prevê-se que essa análise seja um ponto de partida para discussões onde pode-se indicar se o “novo normal” imposto pela COVID-19 se tornou a regra ou foram apenas mecanismos adotados no momento em que foram aplicadas medidas restritivas mundialmente. E consequentemente, contribuir para novas estratégias da diretoria desta instituição e na evolução de seus processos administrativos.

Palavras-chave: Mineração de Processos; Covid-19; Redes Sociais; Perspectiva Organizacional.

ABSTRACT

Process mining: An analysis of the impact before and after the pandemic on the administrative processes, from the organizational perspective

The COVID-19 pandemic has affected many processes and business models in the last years. In this period, organizations had to adapt to the new format of working from home. Process mining (PM) is an area that links business processes and data mining where it extracts knowledge from logs of real events recorded in information systems. This research makes a bibliometric study on PM and analyzes the administrative processes in the period before and after the pandemic of a public educational institution in Brazil. PM techniques are applied from an organizational perspective to understand how the aforementioned set of conventional processes recorded on paper was executed during the pandemic. This work shows that there were changes in the processes and that there was a simplification in most administrative processes, not only due to the significant drop of around 50% in the volume of events, but also in the restructuring of the departments. It was found in the mined data several indications of a reduction in the duration of time and in the transfer of work between departments and, also, that several central departments were modified between the pandemic periods. The results show that the processes demanded are basically the same, however, there are concrete indicators that workers implemented a new process structure during the pandemic, which showed greater efficiency. This analysis is expected to be a starting point for discussions where it can be indicated whether the “new normal” imposed by COVID-19 has become the rule or were just mechanisms adopted at the time restrictive measures were applied worldwide. And consequently, contribute to new strategies of the board of this institution and in the evolution of its administrative processes.

Keywords: Process Mining; Social Networks; Covid-19; Organizational Perspective

LISTA DE FIGURAS

Figura 1 – Rede de amizades da turma de Redes Sociais – 2020.1 24
Figura 2 - Grafo conexo e desconexo com 2 componentes 28
Figura 3 - Conjuntos de vértices (sombreados) com muitas arestas internas e poucas arestas externas a esta área..... 30
Figura 4 - Tela do plugin de conversão do arquivo CSV em XES 35
Figura 5 - Diferentes tipos de mineração de processo e sua relação com modelo e registro de eventos 35
Figura 6 - Plugin do DFvM no PROM – suas entradas, processamento e saída..... 39
Figura 7 - Exemplo de processo tipo “Lasanha”..... 42
Figura 8 - Exemplo de processo tipo “Espaguete” 43
Figura 9 - O processo administrativo 45
Figura 10 - Estrutura do levantamento bibliométrico com análise de rede social e agrupamento de temas dos artigos..... 49
Figura 11 - Fluxo de agrupamento dos temas dos clusteres menores 53
Figura 12- Metodologia PM^2 para Projeto de Mineração de Processo 55
Figura 13 - Arquivo de definição da rede no Pajek..... 59
Figura 14 - Modelo de arquivo CSV para ser inserido no PROM 60
Figura 15 - Menu de Configuração no PROM 61
Figura 16 - Rede bipartida de autores e textos científicos..... 66
Figura 17 - Componente Gigante da rede bipartida 66
Figura 18 - Rede de coautoria dos pesquisadores..... 68
Figura 19 - Componente Gigante da rede de coautoria de pesquisadores..... 69
Figura 20 - Rede de artigos 71
Figura 21 - Componente gigante de artigos 71
Figura 22 - Comunidades Louvain 72
Figura 23 - Agrupamentos (MDS) identificados para os artigos priorizados da componente gigante 74
Figura 24 - Dendrograma das classificações dos temas 77
Figura 25 - Primeiro Arquivo das Instâncias dos Processos..... 89
Figura 26 - Segundo Arquivo das Informações dos Processos 90
Figura 27 – Fases para Agrupamento dos Processos..... 92
Figura 28 - Tratamento no OpenRefine com métodos fingerprint..... 93
Figura 29 - Tratamento no OpenRefine com métodos ngram-fingerprint..... 94
Figura 30 - Fases no Orange para agrupar os Processos..... 95
Figura 31 - Períodos da análise comparativa antes e depois da pandemia..... 102
Figura 32 - Histograma de Distribuição de duração média dos processos antes e depois da pandemia 104
Figura 33 - Histograma de distribuição do número médio de etapas por processos . 104
Figura 34 – Rede Social do macroprocesso Gestão de Carreira TAE Antes da Pandemia 113
Figura 35 - Rede Social do macroprocesso Gestão de Carreira TAE depois da pandemia..... 113
Figura 36- Lista de departamentos do macroprocesso de gestão da carreira TAE e suas métricas de centralidade antes e depois da pandemia 115
Figura 37 -Lista de departamentos do macroprocesso de gestão da carreira TAE e suas frequências relativas antes e depois da pandemia 117
Figura 38 - Fluxo do processo de Gestão de Carreira TAE antes da pandemia..... 118
Figura 39 - Fluxo do processo de Gestão de Carreira TAE depois da pandemia..... 118
Figura 40 - Rede Social do macroprocesso Gestão de Carreira Docente antes da pandemia..... 120
Figura 41 - Rede Social do macroprocesso Gestão de Carreira Docente depois da

pandemia 121
Figura 42 - Lista de departamentos do macroprocesso de Gestão de Carreira Docente e suas métricas de centralidade antes e depois da pandemia 124
Figura 43 - Lista de departamentos do macroprocesso de Gestão de Carreira Docente e suas frequências relativas antes e depois da pandemia 125
Figura 44 - Fluxo do processo de Gestão Carreira Docente 126
Figura 45 - Fluxo do processo de Gestão Carreira Docente 127
Figura 46 – Rede Social do macroprocesso Folha de Pessoal Antes da Pandemia . 129
Figura 47 - Rede Social do macroprocesso Folha de Pessoal depois da pandemia . 129
Figura 48 - Lista de departamentos do macroprocesso de Folha de Pessoal e suas métricas de centralidade antes e depois da pandemia..... 132
Figura 49 - Lista de departamentos do macroprocesso Folha de Pessoal e suas frequências relativas antes e depois da pandemia..... 133
Figura 50 - Fluxo do processo de Folha de Pessoal antes da pandemia 134
Figura 51 - Fluxo do processo de Folha de Pessoal depois da pandemia..... 135
Figura 52 – Rede Social do macroprocesso Reembolso e Pagamentos antes da Pandemia 137
Figura 53 - Rede Social do macroprocesso Reembolso e Pagamentos depois da pandemia..... 137
Figura 54 - Lista de departamentos do macroprocesso de Reembolso e Pagamentos e suas métricas de centralidade antes e depois da pandemia 140
Figura 55 - Lista de departamentos do macroprocesso de Reembolso e Pagamentos e suas frequências relativas antes e depois da pandemia 141
Figura 56 - Fluxo do processo de Reembolso antes da pandemia..... 142
Figura 57 - Fluxo do processo de Reembolso depois da pandemia 142
Figura 58 – Rede Social do macroprocesso Compra de Material Antes da Pandemia 144
Figura 59 - Rede Social do macroprocesso Compra de Material depois da pandemia 145
Figura 60 - Lista de departamentos do macroprocesso de Compra de Material e suas métricas de centralidade antes e depois da pandemia..... 147
Figura 61 - Lista de departamentos do macroprocesso de Compra de Material e suas frequências relativas antes e depois da pandemia..... 149
Figura 62 - Fluxo do processo de Compra de Material antes da pandemia..... 150
Figura 63 - Fluxo do processo de Compra de Material depois da pandemia..... 151

LISTA DE TABELAS

Tabela 1 - Lacuna na literatura versus contribuição de preenchimento da lacuna 17
Tabela 2 - Exemplo de Registros de Eventos de Sistemas de Informação 33
Tabela 3 - Ambiguidade de Nome dos autores 50
Tabela 4 – Consultas para busca específica..... 53
Tabela 5 - Quantidade de autores por artigo..... 68
Tabela 6 - Métricas de centralidade dos autores que se destacaram..... 69
Tabela 7 - Distribuição de artigos em cada agrupamento da componente gigante 73
Tabela 8 - Palavras frequentes de cada cluster da componente gigante 74
Tabela 9 - Distribuição dos temas por comunidade da componente gigante..... 75
Tabela 10 - Quantidade de artigos por agrupamento dos 7 componentes isoladas com mais de 10 vértices..... 76
Tabela 11 - Palavras frequentes de cada cluster 77
Tabela 12 - Distribuição dos temas por comunidade das redes de artigos desconexas de até 10 vértices 78
Tabela 13 - Estratos Qualis da busca restrita 80
Tabela 14 - Autores com maior número de artigos publicados e seu índice "h" da Scopus 81
Tabela 15 - Artigos Relevantes pelo Número de Citações..... 82
Tabela 16 - Artigos Relevantes pelo número de vezes que apareceu na busca 83
Tabela 17 - Ambiguidade no Nome dos Processos 91
Tabela 18 - Exemplos de macroprocessos padronizados após a limpeza e agrupamento dos nomes dos processos despadronizados. 96
Tabela 19 - Exemplo de inconsistência de datas 97
Tabela 20 - Ranking dos Macroprocessos Gerais com o corte seguindo a Lei de Pareto 98
Tabela 21 -Ranking dos macroprocessos principais antes da pandemia com o corte seguindo a Lei de Pareto 105
Tabela 22 - Ranking dos macroprocessos principais pós pandemia com o corte seguindo a Lei de Pareto 106
Tabela 23 - Visão Geral do macroprocesso de Gestão de Carreira TAE 112
Tabela 24 - Métricas Gerais das redes do macroprocesso de Gestão de Carreira TAE 112
Tabela 25 – Visão Geral do macroprocesso de Gestão de Carreira Docente 119
Tabela 26 - Métricas gerais das redes do macroprocesso de Gestão de Carreira Docente 120
Tabela 27 – Visão Geral do macroprocesso de Folha de Pessoal 128
Tabela 28 - Métricas Gerais das redes do macroprocesso de Folha de Pessoal 128
Tabela 29 – Visão Geral do macroprocesso de Reembolso e Pagamentos..... 136
Tabela 30 - Métricas Gerais das redes do macroprocesso de Reembolso e Pagamentos 136
Tabela 31 – Visão Geral do macroprocesso de Compra de Material 143
Tabela 32 - Métricas Gerais das redes do macroprocesso de Compra de Material .. 144

LISTA DE SIGLAS

AP - *ANTES DA PANDEMIA*

ARQUIVO GABIN - *ARQUIVO GABINETE*

ARS- *ANÁLISE DE REDES SOCIAIS*

BI - *BUSINESS INTELIGENCE*

BPM- *BUSINESS PROCESS MANAGEMENT*

CAPES- *COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR*

CCONC - *COORDENADORIA DE CONCURSOS*

CEFET/RJ - *CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA CELSO SUCKOW DA FONSECA DO RIO DE JANEIRO*

CEPE - *CONSELHO DE ENSINO, PESQUISA E EXTENSÃO*

COVID - *CORONA VIRUS DISEASE*

CPPD - *COMISSÃO PERMANENTE DE PESSOAL DOCENTE*

CRM - *CUSTOMER RELATIONSHIP MANAGEMENT*

CSV - *COMMA-SEPARATED VALUES*

DASPE - *DIVISÃO DE ATENÇÃO À SAÚDE E PERÍCIA*

DECOF - *DEPARTAMENTO DE CONTABILIDADE E FINANÇAS*

DEFIS - *DEPARTAMENTO DE FÍSICA*

DEMEC - *DEPARTAMENTO DE ENGENHARIA MECÂNICA*

DEMET - *DEPARTAMENTO DE ENSINO MÉDIO E TÉCNICO*

DEPAD - *DEPARTAMENTO DE ADMINISTRAÇÃO*

DEPAF - *DEPARTAMENTO DE PLANEJAMENTO E FINANÇAS*

DEPEC - *DEPARTAMENTO DE ENGENHARIA CIVIL*

DEPEQ - *DEPARTAMENTO DE PESQUISA*

DEPES - *DEPARTAMENTO DE EDUCAÇÃO SUPERIOR*

DERAC - *DEPARTAMENTO DE ADMINISTRAÇÃO E REGISTROS ACADÊMICOS*

DFM - *DIRECTLY-FOLLOWS MULTIGRAPH*

DFvM - *VISUAL DIRECTLY FOLLOWS*

DGORC - *DEPARTAMENTO DE GESTÃO ORÇAMENTÁRIA*

DGP - *DEPARTAMENTO DE GESTÃO DE PESSOAS*

DGP/DAPEN - *DIVISÃO DE APOSENTADORIA E PENSÃO DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/DICAD - *DIVISÃO DE CADASTRO DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/DICAP - *DIVISÃO DE CAPACITAÇÃO E DESENVOLVIMENTO DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/DILEN - *DIVISÃO DE LEGISLAÇÃO E NORMAS DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/DIMOV - *DIVISÃO DE MOVIMENTAÇÃO, ADMISSÃO E DIMENSIONAMENTO DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/DIPAG - *DIVISÃO DE PAGAMENTO DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/GABIN - *GABINETE DA DIRETORIA DE GESTÃO DE PESSOAS*

DGP/RECEPÇÃO - *RECEPÇÃO DA DIRETORIA DE GESTÃO DE PESSOAS*

DICOD - *DIVISÃO DE COMPRAS DIRETAS*

DICOS - *DIVISÃO DE COMPRAS SISTÊMICAS*

DIEMP - *DIVISÃO DE INTEGRAÇÃO EMPRESARIAL*

DILIC - *DIRETORIA DE LICENCIAMENTO AMBIENTAL*

DIMAT - *DIVISÃO DE MATERIAL*

DIPAT - *DIVISÃO DE PATRIMÔNIO*
DIPPG - *DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO*
DIRAP - *DIRETORIA DE ADMINISTRAÇÃO E PLANEJAMENTO*
DIREG - *DIREÇÃO GERAL*
DIREN - *DIRETORIA DE ENSINO*
DIREX - *DIRETORIA DE EXTENSÃO*
DISCO - *DIVISÃO DE SERVIÇOS CONTÍNUOS*
DTINF - *DEPARTAMENTO DE TECNOLOGIA DA INFORMAÇÃO*
ERP - *ENTERPRISE RESOURCE PLANNING*
GECOF - *GERÊNCIA CONTÁBIL E FINANCEIRA*
ID - *IDENTICATION*
LMS - *LEARNING MANAGEMENT SYSTEM*
MP - *MINERAÇÃO DE PROCESSOS*
OMS - *ORGANIZAÇÃO MUNDIAL DE SAÚDE*
PP - *PÓS PANDEMIA*
PROJU - *PROCURADORIA JURÍDICA*
PROM - *PROCESS MINING FRAMEWORK*
RSC - *RECONHECIMENTO DE SABERES E COMPETÊNCIAS*
SCM - *SUPPLY CHAIN MANAGEMENT*
SECAD - *SECRETARIA ADMINISTRATIVA*
SEORC - *SEÇÃO DE EXECUÇÃO ORÇAMENTÁRIA*
SEPG - *SISTEMA ELETRÔNICO DE PROTOCOLO GERAL*
SGRAF - *SEÇÃO DE SERVIÇOS GRÁFICOS*
SIG - *SISTEMA DE GESTÃO INTEGRADA*
SPED - *SISTEMA PÚBLICO DE ESCRITURAÇÃO DIGITAL*
SPROT - *SEÇÃO DE PROTOCOLO*
TAE - *TÉCNICO-ADMINISTRATIVOS EM EDUCAÇÃO*
UNED ANGRA DOS REIS - *UNIDADE DE ENSINO DESCENTRALIZADA ANGRA DOS REIS*
UNED ITAGUAÍ - *UNIDADE DE ENSINO DESCENTRALIZADA ITAGUAÍ*
UNED MARIA DA GRAÇA - *UNIDADE DE ENSINO DESCENTRALIZADA MARIA DA GRAÇA*
UNED NOVA FRIBURGO - *UNIDADE DE ENSINO DESCENTRALIZADA NOVA FRIBURGO*
UNED PETRÓPOLIS - *UNIDADE DE ENSINO DESCENTRALIZADA PETRÓPOLIS*
UNED VALENÇA - *UNIDADE DE ENSINO DESCENTRALIZADA VALENÇA*
WOS - *WEB OF SCIENCE*
XES - *EXTENSIBLE EVENT STREAM*
XML - *EXTENSIBLE MARKUP LANGUAGE*

SUMÁRIO

1- Introdução 14

1.1. Motivação e Justificativa 16

1.2. Perguntas de Pesquisa e Objetivos 18

1.3. Contribuições 19

1.4. Metodologia 20

1.5. Estrutura do Trabalho 22

2- Rede Social 23

2.1. Centralidade 26

2.2. Componentes 27

2.3. Comunidades 29

2.3.1. Modularidade 29

3- Mineração de Processos 31

3.1. *Registros* de Eventos na Mineração de Processos 32

3.2. Ferramenta PROM 34

3.2.1. O formato XES 34

3.3. Tipos de Mineração de Processos 35

3.4. Diferentes Perspectivas de Mineração de Processo 36

3.5. Algoritmos de descoberta de processos 37

3.6. Mineração Organizacional 39

3.7. Classificação e desafios dos Processos de Negócio 41

3.8. Processos na Administração Pública 43

3.8.1. Administração Pública no Brasil 45

4- Metodologia 48

4.1. Desenvolvimento do Estudo Bibliométrico 48

4.1.1. Desenvolvimento do Estudo Bibliométrico – Geral 48

4.1.2. Desenvolvimento do Estudo Bibliométrico – Busca Específica 53

4.1.3. Desenvolvimento do Estudo Bibliométrico – Busca Expandida 54

4.2. Desenvolvimento do Estudo de Caso 55

4.2.1. Metodologia para Projeto de Mineração de Processo 55

4.3. Ferramentas metodológicas 58

5- Estudo Bibliométrico sobre Mineração de Processos 62

5.1. Busca Geral – Resultados do estudo bibliométrico 62

5.1.1. Rede bipartida de autores e artigos em mineração de processos 65

5.1.2. Rede de coautoria em mineração de processos 67
5.1.3. Rede Social de artigos publicados sobre mineração de processos 70
5.1.4. Encontrando temas estudados das redes de artigos 72
5.1.4.1. Encontrando temas estudados da componente gigante 72
5.1.4.2. Encontrando temas estudados pelas comunidades das redes desconexas 76
5.2. Busca Específica – Resultados do estudo bibliométrico 79
5.3. Busca Expandida – Resultados do estudo bibliométrico 84
5.4. Discussão do estudo bibliométrico 85

6- Estudo de Caso da Análise dos Processos antes e depois da pandemia 87
6.1. A instituição do estudo de caso: CEFET/RJ 87
6.2. Planejamento Estudo de Caso 88
6.3. Extração dos dados 88
6.4. Primeira iteração de Análises 90
6.4.1. Processamento de dados 91
6.4.2. Mineração e Análises 97
6.4.3. Avaliação 101
6.5. Segunda Iteração de Análises 102
6.5.1. Processamento dos dados 102
6.5.2. Mineração de Processo e Análises 103
6.5.3. Avaliação 109
6.6. Terceira Iteração de Análises 110
6.6.1. Processamento dos dados 111
6.6.2. Mineração de Processo e Análises – Gestão Carreira TAE 111
6.6.3. Avaliação – Gestão Carreira TAE 119
6.6.4. Mineração de Processo e Análises – Gestão Carreira Docente 119
6.6.5. Avaliação – Gestão Carreira Docente 127
6.6.6. Mineração de Processo e Análises – Folha de Pessoal 128
6.6.7. Avaliação – Folha de Pessoal 135
6.6.8. Mineração de Processo e Análises – Reembolso 136
6.6.9. Avaliação – Reembolso 142
6.6.10. Mineração de Processo e Análises – Compra de Material 143
6.6.11. Avaliação – Compra de Material 151
6.6.12. Discussão sobre o Estudo de Caso 152

Conclusões 155

Referências 158

1- Introdução

A pandemia do COVID-19 está sendo um marco histórico mundial, conforme alerta-nos Pinzaru, F. *et al.*, (2020) e Bailey e Breslin (2021). Sendo um vírus altamente transmissível, as primeiras recomendações da Organização Mundial de Saúde - OMS foram voltadas para manter o distanciamento social (OMS, 2021). Esse procedimento causou impacto, tanto na vida social e na economia, quanto nas organizações que tiveram que readaptar suas rotinas, gestão, modelo de negócio e processos, conforme análise de Pinzaru, F. *et al*., (2020), Deloitte, (2020) e Accenture, (2020). O primeiro setor a sentir o impacto do surto pandêmico foi a saúde, com o aumento de casos nas emergências, como é exposto no trabalho de Chang *et al.* (2020). Houve setores que sofreram uma queda na demanda, como o de turismo, enquanto que outros tiveram alta demanda, como a tecnologia. O primeiro teve queda, pois com as medidas de distanciamento social, cidades foram fechadas, voos foram cancelados, entre outras. Já o segundo tem tido um grande crescimento, pois para contornar o distanciamento social, tanto a comunidade quanto as organizações têm-se readaptado para uso da tecnologia, visando a transformação digital, segundo as informações encontradas em Accenture, (2020); Pinzaru, F. *et al.*, (2020); Bailey; Breslin, (2021); Mckinsey, (2020) e BBC (2021).

Um movimento rápido, muito desafiador e necessário na pandemia foi o de digitalizar processos, para que estes suportem os negócios diante dessa nova realidade na economia, o distanciamento social e a nova forma de trabalho em casa, na percepção de Pinzaru, F. *et al.* (2020) e Bartsch et al. (2021). Muitas empresas estavam na posição mais forte para manter suas operações, pois já faziam uso da tecnologia para seu crescimento e até mesmo como seu modelo de negócio. Pode-se citar empresas já criadas totalmente digital como *Uber, Nubank, iFood,* entre outras. Assim como, empresas privadas e instituições públicas, que mesmo não tendo o seu *core business* digital, já adotavam processos digitais. Estas últimas, ainda assim, tiveram que extrapolar em suas tecnologias e, por exemplo, fazer uso de inteligência artificial, como parte de sua gestão. Já aquelas empresas que tinham pouco ou nada de processos digitais, o surto da COVID-19 criou um desafio imediato para a transformação digital. Forçando, assim, uma disrupção nos processos de negócios e na operação das empresas, tornando-as cada vez mais digitais e *on-line,* onde são conectadas internamente e com demais partes interessadas*,* de acordo com Deloitte, (2020) e Kamal, (2020).

A medida em que os processos de negócios se tornam digitais, simultaneamente há necessidade de sistemas de informação que os suportem. Esses sistemas de informações, que atuam como parte do avanço tecnológico ocorrido nas empresas, são por exemplo ERP (*Enterprise Resource Planning* (ou planejamento de recursos da empresa)), CRM (*Customer Relationship Management* (ou gestão do relacionamento com os clientes)), SCM (*Supply Chain Management* (ou gestão da cadeia de fornecimento)), SIG (Sistema de Gestão Integrada), LMS (*Learning Management System (*ou Sistema de Gestão de Aprendizagem), até sistemas desenvolvidos internamente para suprir suas necessidades específicas dos processos de negócio. Todos esses têm como objetivo a coleta dos dados e controlar as interações que ocorrem entre processos, partes interessadas e tecnologias, que trabalham em conjunto. E isto gera o elemento principal que é a informação. O qual vai embasar as tomadas de decisões dos gestores para alcançar os objetivos definidos pela empresa, conforme orienta-nos Van Der Aalst, W. M. P. (2011).

Os avanços provocados pelo uso dos Sistemas de Informações, além de apoiar a gestão de processos de negócios (BPM - *Business Processe Management*), geram um grande volume de dados, e consequentemente novas práticas para melhoria da eficiência dos processos operacionais e gerenciais. BPM que é a prática tradicional de gerir os processos, citam Van der Aalst e Rosa (2016), e com a mineração de dados fez surgiu um novo campo de estudo que é a área de mineração de processos, nos termos de Van Der Aalst (2016). O diferencial da mineração de processos é ter o mapeamento detalhado dos processos de negócios com dados reais. Vand Der Aalst (2009) faz uma analogia de mineração de processos como um sistema de navegação. Sendo assim, com esta tecnologia é possível direcionar a gestão nas organizações promovendo melhorias e transformações não somente em seus processos, mas em outras partes de todo seu ecossistema, afirma o autor Van Der Aalst (2011). Ela vai além dos simples mapas de processo, pois atua em diferentes perspectivas como: caso (o que?), processo (como?) e organizacional (quem?). A mineração de processos sob a perspectiva organizacional é denominada pelos autores por Song; Van der Aalst (2008) como mineração organizacional. Esta é focada para descobrir estruturas organizacionais e redes sociais e mostrar como esses modelos podem ajudar a compreender e melhorar os processos. Para Pika *et al.* (2017) a identificação de tais padrões comportamentais de recursos pode ajudar a melhorar o desempenho individual do funcionário e da equipe, impactando diretamente o processo geral e o desempenho organizacional.

Esta pesquisa examina como a pandemia do COVID-19 interferiu no funcionamento das empresas olhando para o seu núcleo: os processos de negócios.

Investiga-se uma instituição pública de ensino profissionalizante e superior no Brasil. Durante a pandemia, a organização forçou a execução dos processos registrados em papel em uma implementação virtual, uma vez que um subconjunto de seus processos internos é realizado com processos administrativos convencionais registrados em papel devido a restrições legais. Este referido subconjunto foi prejudicado diretamente pela pandemia pois o trabalho em *home office* era a única forma viável de operar devido às intervenções não farmacológicas.

Dentro deste contexto, este trabalho é um estudo de caso para analisar os impactos, no período antes e depois da pandemia, nos processos administrativos registrados do CEFET/RJ. Este estudo é pautado na mineração organizacional com foco nas análises de redes sociais avaliando assim como tais impactos interferiram nos relacionamentos entre os recursos.

1.1. Motivação e Justificativa

A Mineração de Processos (MP) tem como base os *registros* de eventos dos sistemas que registram toda instância de cada processo em tempo real. Uma sequência de eventos registra principalmente: o código, a data do evento, a atividade executada, o recurso responsável e entre outros atributos. E então, ao aplicar sua metodologia e técnicas, é possível fazer diversas análises como descoberta de processos através de algoritmos, comparações com a conformidade do processo real versus o modelo, previsão de comportamento dos processos e até estudos sobre a forma como as pessoas estão relacionando-se, como indica-nos Van Der Aalst (2011). Vale destacar que para esta última prática, a literatura apresenta fortes conceitos em análise de redes sociais para realizar a mineração organizacional e, assim, obter entendimento dessas relações entre os recursos das empresas, como pode ser visto em Appice *et al.* (2016); Ferreira (2012); Van Der Aalst; Reijers; Song, (2005) e em Van Der Aalst; Song, (2004).

Segundo Van Der Aalst *et al.* (2012), através das métricas obtidas a partir da análise de redes sociais é possível tratar a mineração de processos na perspectiva organizacional. E assim, aprofundar sobre a colaboração entre departamento para obter mais informações das trocas entre os recursos, como por exemplo, subcontratações, transferência de trabalho e a estrutura organizacional, entre outros. Porém, pesquisas tradicionais de mineração de processos se concentram nos processos de negócios da perspectiva da atividade (fluxo de controle), enquanto menos atenção foi dedicada à perspectiva organizacional, de acordo com Zhao (2014).

Pesquisadores como Tiwari *et al.* (2008) e Thiede *et al.* (2018) apontam que é observada uma lacuna na aplicação das técnicas e uso da mineração de processos nas organizações. Thiede *et al.* (2018) afirmam que a grande maioria das pesquisas estão voltadas para desenvolvimento de ferramentas e algoritmos. E em muitas destas pesquisas, os dados utilizados para testes das ferramentas não são reais, o que implica na falta de aplicação prática, não gerando novos conhecimentos e melhorias para a gestão. A maioria das vezes em que foram feitos estudos de casos, esses foram concentrados em certos setores como finanças, saúde e indústria. Além disso, há uma forte representação destes determinados setores reutilizando os mesmos arquivos de *registros* em vários estudos, como observa-se em Thiede *et al.* (2018).

Até o momento do presente trabalho, apenas uma pesquisa foi encontrada avaliando o impacto, antes e depois da pandemia, com as técnicas de mineração de processo. A referida pesquisa faz análise de tempo de permanência de pacientes no departamento de emergência, como pode ser visto em Chang *et al.* (2020). O mesmo afirma que o impacto imediato da pandemia ocorre na área da saúde em geral. E que em seu estudo houve aumento da demanda de recursos médicos afetando a duração nos atendimentos. No entanto, uma análise comparativa do impacto da pandemia para o processo administrativo de um setor diferente da saúde não foi estabelecida.

A motivação sobre a escolha dos processos administrativos do protocolo geral deu-se pela possibilidade de aprofundar no entendimento de como estes executados em formato tradicional de papel e regulados por um arcabouço jurídico robusto transacionaram entre os departamentos durante a pandemia. Outras motivações foram a facilidade de acesso à base de dados, a possibilidade de trazer contribuições para uma instituição pública sobre o momento atual com os impactos da pandemia e pela inexistência de restrições de sigilo quanto à discussão dos resultados. Este estudo justifica-se também para contribuir no preenchimento de parte daquelas lacunas identificadas na literatura, conforme apresentados na Tabela 1.

Ao adotar tais procedimentos, deseja-se que o estudo seja referência em termos práticos para a interpretação dos impactos da pandemia a fim de promover ações de melhoria ou manutenção dos processos administrativos da referida instituição.

Tabela 1 - Lacuna na literatura *versus* contribuição de preenchimento da lacuna

Lacuna identificada na literatura	**Contribuição de preenchimento da Lacuna**
Pouca aplicação das técnicas nas organizações	Estudo de caso nos processos administrativos do CEFET/RJ
Bases de dados fictícias ou reutilizadas	Base de dados com *registros* reais e não reutilizada

Falta de diversificação de setores com aplicação	Setor administrativo de uma instituição educacional
Pouca atenção dedicada à perspectiva organizacional	Estudo sobre mineração de processo com foco na perspectiva organizacional

Fonte: Própria (2020)

1.2. Perguntas de Pesquisa e Objetivos

Esta pesquisa tem como objetivo geral analisar o impacto antes e depois da pandemia nos processos de protocolo geral do CEFET/RJ, com mineração de processos sob a perspectiva organizacional. Sendo assim, o presente estudo propõe responder às seguintes perguntas (P) de pesquisa:

P1: Qual é o panorama da produção científica sobre Mineração de Processos, em termos gerais e da perspectiva organizacional, segundo parâmetros da bibliometria?

P2: “Houve alteração nos processos administrativos do Centro Tecnológico Universitário no período anterior e posterior a pandemia?”

Diante do exposto, para responder a P1, este trabalho aplicar-se-á do método de bibliometria. Para resolver a P1, definiu-se os objetivos específicos a serem analisados sobre a contribuição do conhecimento científico:

- Há crescimento da produção científica sobre MP em termos gerais?
- Existe uma comunidade científica pesquisando sobre MP em termos gerais?
- Há surgimento de novos autores que publicam trabalhos de MP em termos gerais?
- Quais a categorizações e tipos dos textos científicos de MP em termos gerais? E MP sob perspectiva organizacional?
- Quais os temas publicados nas pesquisas de MP em termos gerais?
- Quem são os principais autores de MP em termos gerais? E MP sob perspectiva organizacional?
- Quais países mais se destacam a pesquisa de MP em termos gerais? E MP sob perspectiva organizacional?
- Há expressivo crescimento da produção científica brasileira em termos gerais? E MP sob perspectiva organizacional?
- Quais os periódicos com mais publicações de MP sob perspectiva organizacional? E quais artigos relevantes de MP sob perspectiva organizacional?

Para responder a P2, este trabalho compreende também em um estudo de caso e utilizar-se-á das técnicas de mineração de processos e redes sociais (que é a análise sob a perspectiva organizacional). A fim de responder a P2, definiu-se os objetivos específicos a serem estudados:

- P2.1 - Identificar e analisar os processos principais da base geral e seus comportamentos através da frequência dos processos e duração de tempo;
- P2.2 - Identificar e analisar os processos principais do recorte da base nos períodos antes e depois da pandemia e seus comportamentos através da frequência dos processos, duração do tempo e frequência das etapas;
- P2.3 - Descobrir e analisar a rede social dos departamentos envolvidos nos processos administrativos antes e depois da pandemia através das métricas de centralidade baseadas na transferência de trabalho.

1.3. Contribuições

Este trabalho contribui para tratamento de parte das lacunas identificadas na literatura, conforme apresentado na Tabela 1. E serve como modelo para futuras aplicações da metodologia ao demonstrar a viabilidade e a relevância da utilização de ferramentas de MP para o mapeamento e análise dos processos, com foco na perspectiva organizacional. Este estudo de caso detalha como um projeto com características similares possa ser realizado, apoia na formação de ideias, resolução de problemas e interpretação dos resultados para tomada de decisões. É esperado que os analistas e pesquisadores de MP possam ganhar maior entendimento nos típicos desafios e dificuldades que enfrentarão ao longo de um projeto de MP. Pois, é dada ênfase na aplicação das técnicas, descrevendo cada atividade envolvida, considerações e ferramentas utilizadas.

Adicionalmente, deseja-se que o estudo contribua na definição de novas estratégias não apenas da diretoria do CEFET/RJ como demais organizações que enfrentam desafios semelhantes. E assim, sejam tomadas decisões e ações para o gerenciamento desses processos administrativo, tanto na perspectiva de eventos, quanto organizacional, especialmente, para aqueles que foram impactados na era pandêmica. Logo, espera-se que seja possível maximizar os ganhos dos resultados para uma posterior revisão, otimização e evolução dos processos administrativos analisados para todos os envolvidos.

1.4. Metodologia

Os objetivos da pesquisa são obtidos seguindo uma metodologia desenvolvida em dois estágios. No primeiro estágio efetua-se uma busca na literatura com base no método de bibliometria e no segundo estágio é um estudo de caso aplicando-se técnicas de mineração organizacional.

No primeiro estágio tem-se o objetivo de responder a primeira pergunta de pesquisa onde para obter um panorama e compreensão sobre MP através da análise bibliométrica das contribuições do conhecimento científico derivado das publicações utiliza-se de pesquisas feitas na WOS - *Web Of Science*. Este estágio foi dividido em três partes: busca geral, busca específica e bus expandida. A primeira está relacionada ao levantamento de artigos sobre a mineração de processos em termos gerais bem como entender as redes sociais envolvendo os autores e os artigos. A segunda parte é com foco na mineração de processos sob a perspectiva organizacional, ou seja, mineração organizacional que objetiva situar o autor sobre os autores, periódicos, países e artigos relevantes sobre mineração organizacional. E por fim, na busca expandida entende-se a pesquisa de alguns trabalhos trazendo uma atualização na busca e também pelo método *Snowballing*. Neste estágio, a exploração feita pode ser definida como um estudo quantitativo.

O segundo estágio desta pesquisa é caracterizada como um estudo de caso, com enfoque quantitativo e fundamentação teórica. O estudo de caso é uma forma popular de pesquisa qualitativa, e apresenta um exame aprofundado de um determinado tema, de acordo com a conceituação de Marczyk; Dematteo; Festinger (2005). Neste estudo de caso utiliza-se dados reais de *registros* de eventos dos processos administrativos do CEFET/RJ no período entre 2013 a junho de 2021. Esta base reúne informações sobre solicitações gerais tanto de colaboradores internos quanto do público externo. Como por exemplo, solicitação de férias de docentes, reembolso de curso de técnico administrativo, reingresso de alunos, solicitações de compra de material entre outros. Estes dados serão a base para extrair o conhecimento necessário para responder a segunda pergunta de pesquisa.

Para desenvolver o segundo estágio, este estudo de caso implementa o *framework* PM^2 (*Process Mining Project Methodology*) para Projetos de Mineração de Processo de Van Eck *et al.* (2015). O *framework* é subdivido em três grandes tópicos que são:

1. **Início:** Planejar o projeto definindo os objetivos do projeto e a base de dados.

E extrair os dados dos *registros* de eventos dos processos administrativos do Sistema Eletrônico de Protocolo Geral do CEFET/RJ em arquivos.

2. **Interações de Análises:**

   2.1. Primeira iteração: analisar os processos da base de dados extraindo estatísticas básicas (como por exemplo, frequência de processos, histograma de etapas e da duração média) para compreender os processos da base geral;

   2.2. Segunda iteração: analisar os processos da base de dados extraindo estatísticas básicas (como por exemplo, frequência de processos, histograma de etapas e da duração média) para compreender os processos do recorte da base nos períodos antes e depois da pandemia

   2.3. Terceira iteração: realizar a mineração organizacional extraindo métricas de redes sociais (como por exemplo, grau de centralidade) do recorte da base nos períodos antes e depois da pandemia.

3. **Melhorias de Processo e Apoio**: Deseja-se as conclusões deste estudo de caso possa promover novas ações e estratégias para os processos administrativos, para o sistema e o setor de Protocolo Geral.

Na primeira iteração, identificou-se que os nomes dos processos não eram padronizados. No levantamento inicial haviam em torno de 6 mil nomes de processos distintos e essas variedades de nomes em muitos casos tratavam do mesmo processo (como por exemplo “Compra de Material”, “Compra de Equipamento”, “Aquisição de Materiais”). Com o objetivo de padronizar o nome dos processos utilizou-se algoritmos mineração de textos para agrupamento dos nomes dos processos.

Esses dois estágios estão inseridos no cenário descrito por Marconi; Lakatos (2003), que afirmam que o instrumental metodológico da pesquisa está relacionado com o problema de estudo. Os problemas estudados, muitas vezes, necessitam mais de uma técnica. Neste trabalho, os problemas estão fortemente inseridos no contexto da análise quantitativa, mas não excluído da análise qualitativa.

E por fim, as ferramentas metodológicas utilizadas foram: *Microsoft Excel* para manipulação e primeiro levantamentos dos dados extraídos do sistema eletrônico de protocolo geral, o *Open Refine* para tratamento dos nomes dos processos e dos nomes dos autores na fase do levantamento bibliométrico; o *software* de redes sociais *Pajek* com o objetivo de identificar os principais autores e artigos durante o levantamento bibliométrico e também das redes sociais antes e depois da pandemia, o *software Orange* para agrupamento e geração das nuvens de palavras dos temas durante o levantamento bibliométrico e o PROM software de mineração de processos para descoberta e simulação dos processos com foco na perspectiva organizacional.

1.5. Estrutura do Trabalho

A pesquisa está estruturada por esta Introdução, mais seis capítulos e a conclusão. O segundo capítulo apresenta os principais conceitos de redes sociais, o terceiro mostra a revisão teórica sobre Mineração de Processos incluindo Mineração Organizacional e sobre Processos na Administração Pública, o quarto detalha a metodologia aplicada à pesquisa adotada tanto para o levantamento bibliométrico quanto para o estudo de caso, o quinto apresenta o levantamento bibliométrico sobre mineração de processos, o sexto apresenta o estudo de caso e por fim, são apresentadas as conclusões desta dissertação.

2- Rede Social

Uma rede social, segundo Wasserman e Faust (1994), é formada por pontos, chamados de atores ou vértices, que se ligam por linhas chamadas de arestas as quais representam alguma relação existente entre esses vértices. Esses pontos podem representar pessoas, empresas, países; e as arestas representam o tipo de ligação existente entre eles, que pode ser de afinidade, troca de informação, materiais, documentos, entre outras.

Sua relevância dá-se pelos padrões de conexões ou interações entre as partes de um sistema. Beneficia-se enormemente da ampla gama de pontos de vista trazidos por tantas disciplinas diferentes, como aponta-nos Newman (2010). É uma estrutura abstrata e sua representação é conhecida como grafo. Um Grafo G (V,E), pode ser definido como um conjunto finito de pontos denominados vértice (representados por “V”), e alguns pares destes que são ligados entre si, denominados arestas (representadas por “E”), como orienta-nos Merris (2000). O poder dos grafos é descrever com clareza as características gráficas e topológicas das redes facilitando sua compreensão.

Já as Redes Sociais, é um tipo específico de rede, embora não seja recente, esse conceito tem ganhado força e visibilidade com as novas tecnologias e massa de dados. Moreno (1934), um médico e psicólogo que se dedicou a estudos de terapias em grupo, foi o pioneiro neste campo introduzindo o conceito de Sociometria. Segundo ele, o átomo social é composto pelo indivíduo e seus laços econômicos, culturais e sociais. Esta é uma rede de pessoas, embora às vezes possa ser uma rede de grupos, de outras entidades, como empresas. Ou seja, é uma estruturação estática para melhor entendimento das relações sociais em um determinado grupo em estudo, na concepção de Newman (2010). Numa rede social, os vértices V, são atores, e as arestas, E, representam alguma relação existente entre esses pontos. Esses atores podem representar pessoas, animais, instituições, departamentos, países entre outros; e as arestas podem representar afinidade, amizade, troca, documentos. Essas arestas podem conter peso que represente numericamente esta relação. Sendo assim, tem aplicabilidade em diversas áreas: sociologia, biotecnologia, comunicação, logística, entre outras, como afirma Scott (2000).

Há muitos trabalhos de redes sociais aplicados em diversas áreas de estudo. Pode-se citar o trabalho de Albuquerque; Monteiro (2019) sobre redes empresariais. Nele cada vértice representa um conselheiro e, há arestas se ambos forem membros

de um mesmo conselho, simultaneamente. Há redes com foco em análise de coautoria de artigos conforme indica-nos os estudos de Barbastefano *et al.* (2013), Souza; Barbastefano (2011), Bordin *et al.* (2014) e Souza *et al.* (2012). A pesquisa de Cerqueira *et al.* (2014) utiliza-se de redes sociais aplicada em uma cadeia de suprimentos em que os vértices representam como atores os fornecedores e os centros de distribuição e as arestas estabelecem a relação contratual entre os atores envolvidos. E também em estudos de processos de negócios onde utiliza-se dos vértices sendo os atores como os recursos que executam as atividades, como pode ser visto no trabalho de Van Der Aalst *et al.* (2012). O qual utiliza-se das técnicas e ferramentas de redes sociais para identificar melhorias nos processos. E mais recentemente, trabalho de redes sobre a temática da pandemia do corona vírus, onde o estudo de Belli *et al.* (2020) apresentou redes para identificar os países e organizações que possuem ligações mais fortes. Avaliando a capacidade das pesquisas de gerar conhecimento científico sobre a crise de saúde mundial.

Para demonstrar um exemplo de rede social é apresentado um grafo, na Figura 1 com o relacionamento de amizade entre alunos na turma da disciplina de Rede Social. O levantamento das informações foi feito pela autora conversando com os colegas de turma e entendendo quantas disciplinas cada um cursou um com o outro. Então, os vértices da rede são os alunos e o professor e as arestas ligam cada um que tenha cursado pelo menos uma disciplina. O peso dessas ligações são quantas disciplinas já cursaram juntos até o momento. Foi definido que seriam arestas pois a relação é bidirecional, ou seja, o aluno 1 cursou 1 disciplina com aluno 2 e vice-versa.

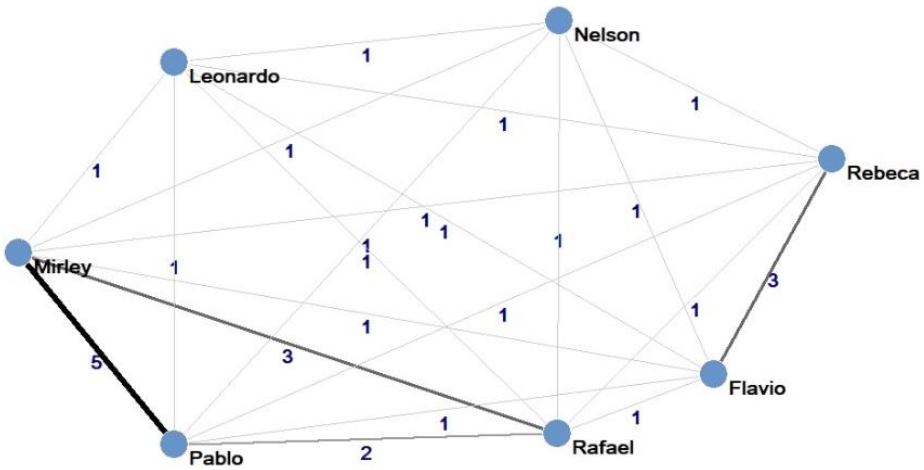


Figura 1 – Rede de amizades da turma de Redes Sociais – 2020.1
Fonte: Própria (2020)

A rede social consta com 7 vértices, 6 alunos e 1 professor, 17 arestas com peso 1, 2 arestas com peso 3, 1 aresta com peso 2 e também 1 aresta com peso 5. Percebe-se que a grande maioria dos colegas se encontra pela primeira vez nesta turma.

Destacam-se com as arestas mais grossas Pablo e Mirley cursando 5 disciplinas, seguido de Mirley e Rafael (professor), Flávio e Rebeca com 3 disciplinas e por fim Pablo e Rafael com 2 disciplinas.

Essa visualização gráfica, como Freeman (2000) afirma sempre teve um papel fundamental na pesquisa sobre rede social. As primeiras redes foram criadas manualmente e posteriormente foram desenvolvidos algoritmos de computadores para que pudessem exibi-los. A visualização gráfica auxilia tanto o pesquisador a entender a estrutura da rede quanto a comunicá-la. As redes sociais criadas virtualmente, cada vez mais, permitem imagens poderosas as quais facilitam e permitem a iteração destas com os analistas. Este autor também afirma que há uma nova era iniciando-se nos arranjos das estrutura de redes social com grafos tridimensionais, cores e animações. Newman (2010) complementa que, nos últimos anos, tem-se criado redes cada vez maiores e mais complexas, e que houve um grande esforço de pesquisadores para que serviços como lista de amigos do *Facebook* pudessem fornecer dados que antes eram inacessíveis.

Além da visualização, os analistas em redes sociais verificam os resultados, através da quantificação de medidas gerando conhecimento mais aprofundados do modelo desenhado. Essas medidas permitem modelar os relacionamentos para descrever a estruturas da rede geral, dos indivíduos ou de grupos. Pode-se então estudar o impacto dessa estrutura no funcionamento do grupo e/ou a influência dessa estrutura nos indivíduos do grupo de acordo com a observação de Wasserman; Faust, (1994).

A Análise de Redes Sociais (ARS) avalia o desempenho dos seus “atores”, a importância e a influência no comportamento deles na rede. Ela detém uma série de ferramentas que podem ser utilizadas para estudar a interação entre os indivíduos. Isso é possível através de métricas principalmente de centralidade, segundo informa-nos Scott (2000). Para realizar uma análise de rede social há diversas métricas que permitem mensurar a rede como um todo, como por exemplo, densidade e grau médio. E métricas que permitem medir os atores, ou seja, o quão um vértice tem ligações com os demais, e este é o grau do vértice. Uma outra análise possível é a abordagem de comunidade e subgrupos nas redes sociais – grupos de nós possuindo mais conectividade, como divisão de componentes fracas ou fortes. E também análises como as partições e vetores. Partições de uma rede que é uma classificação de vértices na rede por algum tipo de informação atribuído a exatamente uma classe. E os vetores são valores numéricos reais atribuído a cada vértice. Assim, as partições representam propriedades discretas enquanto os vetores traçam informações contínuas nos vértices, referenciando-se em Nooy; MRVAR; Batagelj (2005).

A seguir são detalhadas outras técnicas e métricas sobre análise de rede social importantes para este trabalho.

2.1. Centralidade

Já partindo para algumas métricas o conceito de centralidade é muito importante, principalmente no que diz respeito a análise do fluxo de informação. Muitas vezes, existem vértices que possuem maior influência sobre os outros, e que possa estar influenciando em suas ações em certo grau. Há também aqueles servindo de único intermédio entre dois grupos e que possuem um papel importante na comunicação entre grupos. Assim, centralidade é o número de relações que um nó tem com os demais nós de uma rede, é um estudo empírico que tenta identificar os atores (vértices) mais importantes e ou influentes na rede, conforme orienta-nos Scoot (2000).

Há diversas métricas que permitem quantificar a importância de um determinado ator ou um grupo de nós numa rede. Todas essas medidas são calculadas para cada ator da rede. Assim, o maior nó é aquele que tem a maior influência da rede, porém é importante ressaltar que essas métricas de centralidade devem ser avaliadas em conjunto para apontar o nó de mais poder na rede. A seguir, serão detalhadas as medidas de centralidade de grau, centralidade de proximidade e centralidade de intermediação.

A centralidade de grau é o número de relações que um nó tem com os demais nós de uma rede. Aponta-se a quantidade de conexões no vértice, sejam elas entradas e saídas. Com uma rede onde as arestas são sem direções ela é simplesmente a quantidade de conexão do nó. Essa propriedade varia de um valor mínimo de zero, caso não exista nenhum vértice adjacente a um dado vértice (este chama-se vértice isolado). E possui o valor máximo igual ao número total de vértices menos um, caso o vértice em questão esteja conectado com todos os demais, na orientação de Wasserman; Faust (1994). Na Figura 1 o grau de todos os vértices é igual a 6 devido a rede criada.

Uma outra métrica que auxilia na análise de influência dos nós centrais é a medida de distância. A distância é o caminho mais curto entre os nós *u* e *v*, também conhecido como comprimento geodésico. Esta distância permite medir o número total de passos entre uma ramificação extrema e outro da rede, ou entre dois nós numa mesma rede. E quanto menor esta distância mais influente pode-se dizer que este nó é entre os demais, como indica-nos Nooy; MRVAR; Batagelj (2005).

Já a medida de centralidade de proximidade (ou *closeness*) é oriunda do

conceito de distância, que é baseada na distância total entre um vértice e todos os outros vértices, onde distâncias maiores resultam em pontuações de centralidade de proximidade mais baixa. Logo, essa métrica indica a proximidade de um ator em relação aos demais. Caso uma rede não esteja fortemente conectada, ou seja, se há muitos nós que não são possíveis de ser conectados, não é recomendável o cálculo desta métrica, como adverte-nos Nooy; MRVAR; Batagelj (2005). Essa medida pode ser definida, matematicamente, como:

$$C_C(v) = \frac{n-1}{\sum_{t \in V/v} d_G(v,t)}$$

onde corresponde à distância geodésica entre os vértices *v* e *t,* segundo Wasserman; Faust (1994).

E por fim, a medida de centralidade de intermediação (ou *betweeness*) indica os nós que está servindo de ponte para facilitar o fluxo de informação na rede. Esta medida indica até que ponto o autor é indispensável como uma ligação entre as demais cadeias de contatos facilitando a disseminação de informações dentro da rede. Quanto mais uma pessoa é um intermediário, mais central é sua posição na rede e por conseguinte, um ator que está situado na geodésica entre a maiorias dos pares, ou seja, combinando com outras métricas, ele é um grande central, na concepção de Nooy; MRVAR; Batagelj (2005). Essa medida foi definida, matematicamente, por Freeman (1977), como:

$$Bet(v) = \sum_{i \neq j \neq v, i \neq v} \frac{\sigma_{i,j}(v)}{\sigma_{i,j}}$$

onde $\sigma_{i,j}(v)$ representa o número de caminhos geodésicos que ligam os vértices i e j passando por v; já $\sigma_{i,j}$ é o número de caminhos geodésicos que ligam os vértices i e j. Ou seja, tem um alto valor de centralidade de intermediação um vértice, se v ocorre um mais vezes entre os menores caminhos entre todos os pares de vértices do grafo que os demais vértices $u \in V$.

De acordo com Freeman (1979), essas métricas de centralidade também possuem as medições referente a rede como um todo. É importante ressaltar que há uma diferença entre centralidade e centralização. Basicamente centralidade mede os nós individualmente e centralização calcula a rede toda. Uma rede com a clareza entre o centro e as ramificações mais distantes é uma rede altamente centralizada pois o centro é responsável para espalhar as informações, esclarece-nos Nooy; MRVAR; Batagelj (2005) e Freeman (1979).

2.2. Componentes

As redes podem ser compostas por partes individuais ou por grupos. Segundo Wasserman, Faust (1994) uma propriedade é analisar se em uma rede há grafos conectados ou não. Uma definição simples para um grafo conexo é que haja um caminho, ou seja uma aresta, entre cada par de vértices. Se um grafo não for conexo, ele será desconexo. Considere a Figura 2 como exemplo onde tem-se do lado esquerdo o grafo único, conexo e o grafo à direita desconexo com dois subgrafos conexos. Estes subgrafos conexos são chamados de componentes como em Wasserman, Faust (1994).

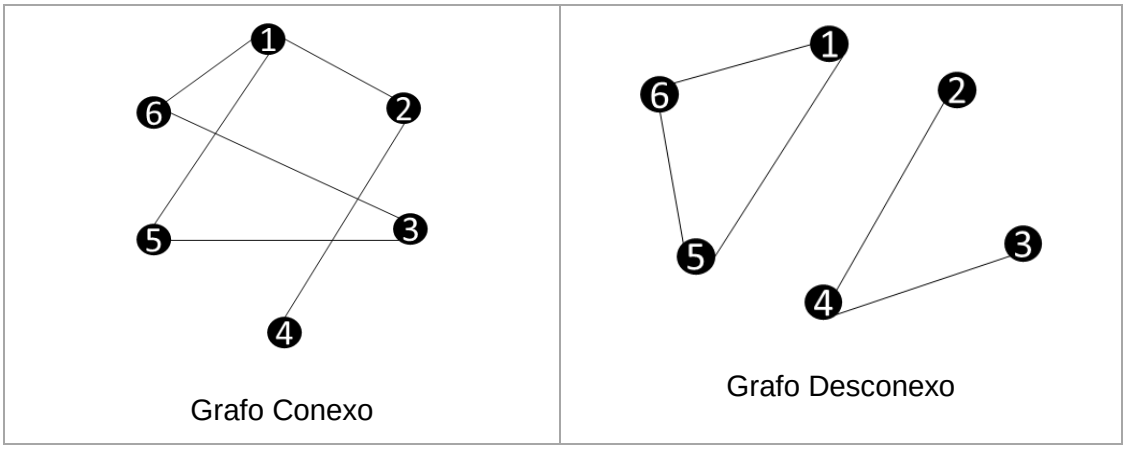


Figura 2 - Grafo conexo e desconexo com 2 componentes
Fonte: Adaptado de Wasserman; Faust (1994)

Uma outra definição muito importante para uma rede social é a componente gigante. Ela possui o maior subgrafo conexo da rede, ou seja, o maior subconjunto de vértices e arestas. No trabalho de Souza *et al.* (2012) esta componente gigante representa mais de 50% dos vértices indicando o total de autores que possuem ligação com os demais integrantes da rede. Essa componente gigante significa ser possível alcançar um grande número de outros autores do mesmo grafo começando por um dos vértices e movendo-se ao longo das bordas e permitindo assim a comunicação fluida entre os membros.

Esses vértices do componente gigante também formam conexões entre si e novas conexões ligando os vértices já existentes. Isso ocorre na maioria das redes sociais, o que corrobora com a hipótese do mundo pequeno, criada por Milgram em 1967. Neste estudo verificou-se que em várias redes sociais a distância média entre dois vértices tomados ao acaso era muito pequena se encaixa na hipótese do mundo pequeno. Em algumas áreas, esta conexão abrange mais de 90% das vértices, como

afirma Newman (2010).

2.3. Comunidades

Redes complexas estão cada vez mais comum e, cada vez mais, necessita-se de técnicas de análise de redes sociais para facilitar a identificação de características e novas descobertas de conhecimentos. E muitas vezes, é possível notar alguns grupos que possuem mais conexão. Em ARS, esses grupos são chamados de subgrupos coesos. Neles, são agrupados nós que possuem algum tipo de entrosamento, seja implícito ou explícito, como exposto por Nooy; MRVAR; Batagelj (2005).

A literatura apresenta várias técnicas para detectar subgrupos coesos com base na densidade e conectividade, como componentes, k-núcleos e cliques ou sub-redes completas, como pode ser visto em Nooy; MRVAR; Batagelj (2005), Newman (2010) e Wasserman; Faust (1994). Uma dessas técnicas defini que um grupo coeso é aquele que possui a densidade relativamente alta dentro do subgrupo, ou seja, há conexão forte neste subgrupo e laços relativamente fracos fora deste subgrupo.

Diante do exposto, é possível encontrar grupos denominado estrutura de comunidades, a qual fornece uma grande compreensão e visualização da estrutura da rede complexa em análise. Newman e Girvan (2004) definem essa estrutura como segregação de nós da rede completa em grupos dentro dos quais são densamente conectados, mas entre grupos são menos densos.

2.3.1. Modularidade

A modularidade é uma métrica proposta por Newman e Girvan (2004) que mede a qualidade de uma possível grupo de redes, ou seja, se uma divisão específica de uma rede em comunidades é boa ou não. É um valor escalar entre -1 e 1 que mede a densidade das ligações internas das comunidades em comparação com as ligações entre comunidades. Assume valores positivos se houver mais arestas entre vértices do mesmo tipo do que esperaríamos ao acaso, e valores negativos se houver menos. A modularidade Q é definida como:

$$Q = \frac{1}{2m}\sum_{ij}\left[\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(C_i, C_j)\right]$$

Onde m é o número de arestas, A_{ij} é o elemento *ij* da matriz de incidência A, k_i é o grau do vértice *i*, k_j é o grau do vértice j, C_i, é o calor do componente do vértice *i*, C_j é o calor do componente do vértice *j*, A função δ retorna 1 caso *i* e *j* pertençam ao mesmo componente e 0 caso contrário. O somatório percorre todos os pares de vértices *ij* da rede, como delimitado por Newman (2006).

Conforme a **Error! Reference source not found.**, há três comunidades nas quais a modularidade alta significa conexão mais densa na mesma comunidade e mais esparsa entre comunidades diferentes. A comunidade é considerada a mais ideal quando sua modularidade atinge o máximo. Han *et al.* (2012) afirma que a modularidade pode ser usada para medir a qualidade do agrupamento.

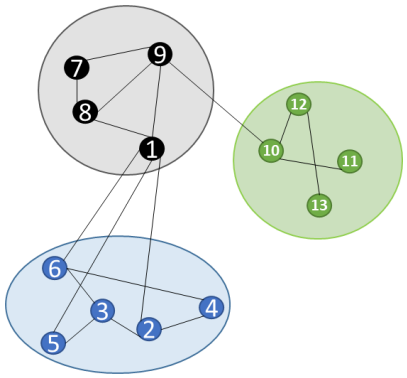


Figura 3 - Conjuntos de vértices (sombreados) com muitas arestas internas e poucas arestas externas a esta área
Fonte: Adaptado de Newman, M. E. J. (2006)

Há implementação de um algoritmo para detectar comunidades com base em métrica de modularidade chamado método de *Louvain* e foi criado por Blondel *et al.* (2008). Ele é um algoritmo progressivo o qual primeiro detecta pequenas comunidades otimizando a modularidade localmente em todos os nós. Então cada pequena comunidade é agrupada em um nó e assim, a primeira etapa é repetida novamente. Encontra-se disponível no *software* como o *Pajek* para que possa aplicar os estudos de caso. Como indica-nos Ferreira e Alves (2012); Appice *et al.* (2016) e Blondel; Guillaume; Lefebvre (2008), a modularidade tem tido sucesso em vários tipos de redes para descobrir diferentes comunidades.

3- Mineração de Processos

A mineração de processo é uma disciplina que apresenta uma variedade de técnicas que ajudam as organizações a descobrir seus processos de negócios reais. As primeiras publicações na literatura foram desenvolvidas pelos precursores Wil van der Aalst e Ton Weijters em projetos de pesquisas acadêmicos iniciando por volta dos anos 2000 conforme apresenta-nos Van Der Aalst (2011).

Davenport e Spanyi (2019) citam que a mineração de processos é inovadora por ser capaz de resolver dois grandes problemas das organizações: um envolve o pouco interesses dos gestores no entendimento de como está seus processos sendo executados no estado atual e o outro é a falta de conexão entre processos de negócios e os sistemas de informações presentes nas organizações.

Estes autores citam ainda sobre um primeiro problema que, mesmo aquelas organizações, que fazem um levantamento do estado atual, são muitas vezes baseadas em entrevistas, são demorados e os gestores julgam subjetivos e tratam com ceticismo considerável. E, sobre um segundo problema, eles relatam que sistemas como sistema de gestão integrado (*Enterprise Resource Planning* **–** ERP) que são orientados a processos é no sentido de se ter uma sequência de atividades e que não há facilidades para entendimento total do processo em tempo real. Adicionalmente, relatam outros *softwares* que apoiam a modelagem de processos exigem muitas etapas manuais e poucos recursos para sintetizar o andamento dos processos no dia a dia.

Por outro lado, a mineração de dados, como uma área multidisciplinar, que está contemplada na Ciência da Computação, possui técnicas centradas em dados e não extrapolam a compreensão mais abrangente para os processos. Já as ferramentas de Inteligência de Negócio (ou *Business Inteligence* – *BI*), consomem dados dos sistemas de informações e se concentram em gerar painéis e relatórios simples e não apresentam percepções bem definidas dos processos. Já os *softwares* de BPM visam muito desenhos futuros de processos mas, são limitados para entender os processos no estado atual como adverte-nos Van Der Aalst (2011).

Expostos estes problemas a Mineração de Processos é um elo entre duas grandes áreas que são a ciência de dados e a gestão de processos de negócios. Ela preenche uma lacuna entre a análise de processos baseado no modelo tradicional (como por exemplo, mapeamento para entendimento de seu estado atual (“*as is”*) e simulações) e as análises de dados através das técnicas da ciência de dados, na

observação de Van Der Aalst (2016).

A mineração de processos é uma metodologia para descrever uma estrutura de processo a partir de um conjunto de dados de execuções reais oriundos dos sistemas de informação das organizações. Esta metodologia vai além, pois não se refere apenas à descoberta desta estrutura, mas também verifica a conformidade, gera indicadores para monitoramento de desempenho e aprimoramento de processos reais. Ou seja, as técnicas coletam o que realmente está acontecendo, quem está fazendo, quanto tempo está gastando e os *gaps* e gargalos dos processos, como alerta-nos Van Der Aalst (2011).

Aplica-se também a inteligência artificial com as técnicas de ciência de dados para detectar causas raízes, prever gargalos e atrasos e assim, apoiar a tomada de decisões e recomendar projetos de processos. Esse elo abordado por esta metodologia torna os modelos de processos dinâmicos e traz os grandes volumes de dados de hoje em um contexto de processo. Ou seja, um contexto de ponta a ponta sobre a organização e as análises de dados através das técnicas de mineração de dados, como exposto por Van Der Aalst (2016). A mineração de processos fornece não apenas uma ponte entre a mineração de dados e o gerenciamento de processos de negócios mas também ajuda a resolver a divisão clássica entre "negócios" e "TI", na compreensão de Van Der Aalst (2011).

Durante a primeira década dos anos 2000 ela desenvolveu-se e foi expandindo-se por novos estudos e pesquisadores no meio acadêmico como pode ser visto no capítulo 4, com o levantamento bibliográfico (os conceitos de estrutura organizacional e informal). Porém, esta metodologia vem ganhando visibilidade e aplicabilidade nas empresas e já apresenta muitos fornecedores de *softwares* para manipulação de suas técnicas como mostra o guia de mercado para mineração de processos do *Gartner,* informado por Kerremans (2019).

3.1. *Registros* de Eventos na Mineração de Processos

A mineração de processos tem como ponto de partida as informações dos *registros* de eventos de sistemas de informação. Estes *registros* são insumos primordiais para aplicar as técnicas e algoritmos desenvolvidos para obter as melhorias na gestão dos processos. E os componentes para aplicação desta prática são: o caso,

a atividade, o evento, o tempo, o recurso e outras propriedades complementares. O caso (instância de processo) é objeto sobre o qual ocorrem todos os demais componentes e cada caso tem uma identidade única, por exemplo, um pedido de compra ou uma solicitação de viagem. Atividades, ações ou tarefas são operações em um caso, como registro, verificação ou aprovação. Evento é uma atividade executada no processo e registrada no registro do sistema. O tempo refere-se a data e hora de ocorrência, que pode ser gravado como um período, ou apenas como um único momento. O recurso é responsável pela execução da atividade podendo ser também o departamento ou máquinas onde esta foi realizada. Entre outras, também podem fazer parte dos registros de eventos, permitindo complementar o conhecimento extraído através da mineração de processos, como esclarece-nos Van Der Aalst (2011).

Na Tabela 2 cada linha corresponde a um evento extraído dos *registros* dos sistemas. Cada evento tem várias propriedades, algumas básicas para a mineração de processo e outras para análises mais detalhadas e complementares. A tabela refere-se a processamento de pedidos onde a primeira coluna é o número do pedido, ou seja, é o ID de um caso de processo. A segunda é o ID único do evento. A terceira é o carimbo de data e hora da ocorrência. Em seguida, a descrição da ação realizada. Logo após é o recurso responsável por realizá-la. E por fim há o valor da ordem de compra.

Tabela 2 - Exemplo de Registros de Eventos de Sistemas de Informação

Identificação do Processo		Propriedades			
ID Compra	**ID do evento**	**Data e hora**	**Atividade**	**Usuário**	**Custo**
1	35654423	30/12/2010 11:02	Registrar solicitação	Pete	50
1	35654424	31/12/2010 10:06	Analisar solicitação	Sue	400
1	35654425	05/01/2011 15:12	Verificar tíquete	Mike	100
1	35654426	06/01/2011 11:18	Decidir	Sara	200
1	35654427	07/01/2011 14:24	Rejeitar Solicitação	Pete	200
2	35654483	30/12/2010 11:32	Registrar Solicitação	Mike	50
2	35654485	30/12/2010 12:12	Verificar Tíquete	Mike	100
2	35654487	30/12/2010 14:16	Analisar Solicitação	Pete	400
2	35654488	05/01/2011 11:22	Decidir	Sara	200
2	35654489	08/01/2011 12:05	Pagar Compensação	Ellen	200

Fonte: Adaptado de Van Der Aalst (2011)

Para que seja possível realizar atividades de mineração de processos, é necessário que cada registro de evento contenha informações sobre qual instância este evento ocorreu, à qual atividade ele corresponde e em que momento ele foi registrado.

Informações como o responsável pela execução da atividade, o departamento ou unidade organizacional onde esta foi realizada, entre outras, também podem fazer parte dos registros de eventos, permitindo complementar o conhecimento extraído através da mineração de processos, conforme orienta-nos Van Der Aalst *et al*. (2005).

3.2. Ferramenta PROM

O PROM, conforme citam Van der Aalst *et al.* (2013) é um *software* que possui um conjunto de ferramentas e algoritmos de mineração de processos para análise de processos a partir do *registro* de eventos. Ela é gratuita e pode-se ter acesso a ela através do site promtools.org, logo recebe contribuição de diversos pesquisadores da área que, a partir dos pacotes existentes na ferramenta, criam novos modelos ou melhoram os existentes.

O PROM é uma das primeiras ferramentas e altamente utilizada pela academia, porém, como a metodologia de MP vem ganhando aplicabilidade nas empresas foi apresentado por Kerremans (2019) dezenove fornecedores distintos com ferramentas comerciais. Neste trabalho é utilizado o PROM para as técnicas de MP.

3.2.1. O formato XES

O formato do arquivo de entrada dos *registros* de eventos adotado pela IEEE *Task Force*[8] em mineração de processos é o XES (eXtensible Event Stream) conforme cita Gunther e Verbeek (2014). Ele está padronizado desde 2010 nas ferramentas de mineração de processos como o PROM e tem a sua biblioteca base OpenXES. É possível converter vários formatos, como por exemplo MS Acess e CSV em XES, que nada mais do que um arquivo baseado em XML formatado para a mineração de processos. Assim, no PROM, através do *plugin* “Convert CSV to XES”, conforme Figura 4, é possível transformar um arquivo CSV numa extensão XES. E a partir de então utilizar algoritmos e técnicas para a mineração de processos (Van der Aalst, 2011).

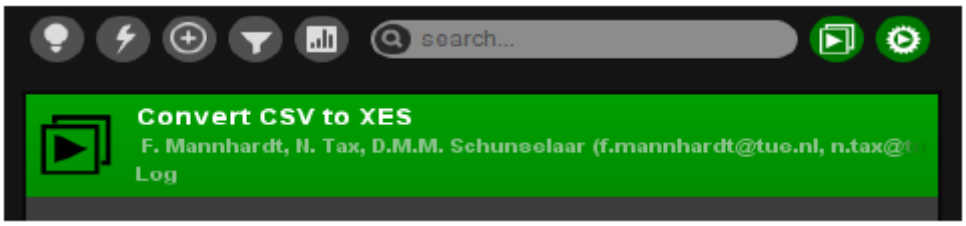


Figura 4 - Tela do *plugin* de conversão do arquivo CSV em XES
Fonte: recorte da tela do PROM

3.3. Tipos de Mineração de Processos

A literatura menciona diferentes tipos de mineração de processos conforme exposto por Hakvoort; Sluiter (2006); Rojas et al. (2016); Van Der Aalst (2011, 2018). Esses tipos podem ser descobertos, conformidade, aprimoramento e suporte operacional. Na Figura 5 apresenta-se estes tipos e como se relacionam com as entradas de dados e modelo de processo.

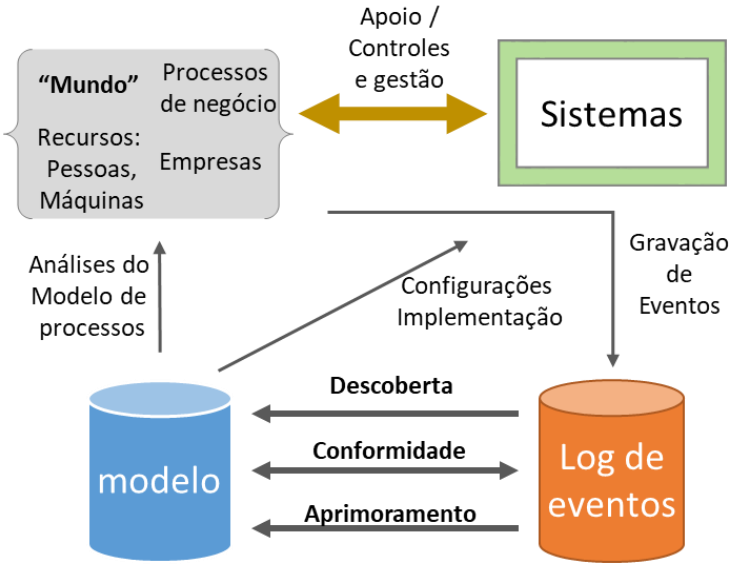


Figura 5 - Diferentes tipos de mineração de processo e sua relação com modelo e *registro* de eventos
Fonte: Adaptado de Van Der Aalst, W. M. P., (2011)

A descoberta de processos (*discovery*) cria um modelo do *registro* de eventos sem usar nenhuma informação adicional. É usado para aprender como o processo flui. Esta é uma técnica onde utiliza-se a base de *registros* de eventos e produz-se um modelo de processo, um exemplo desta técnica é o algoritmo *Alfa*, citado por Van Der

Aalst; Weijters; Maruster (2004), que utiliza um registro de eventos e produz um modelo de processo (uma rede de Petri) explicando o comportamento registrado no registro.

A conformidade do processo (*conformance*) compara um modelo de processo existente com o *registro* de eventos desse processo. Com essa técnica pode-se verificar se a realidade, conforme apresentada no registro, está de acordo com o modelo e vice-versa.

O aprimoramento do processo (*enhancement*) tenta estender ou melhorar um processo já existente usando informações sobre o processo real registrado em algum *registro* de eventos. Embora a conformidade do processo verifique apenas o modelo existente, a melhoria do processo altera ou estende este modelo. Van der Aalst (2018) chama esse processo de reengenharia. Um exemplo é a extensão de um modelo de processo com informações de desempenho, por exemplo, mostrando gargalos.

O quarto tipo mencionado por Van Der Aalst (2018) é o suporte operacional (*operational support*) quando as técnicas são usadas de modo *online*. Um caso é a detecção de não conformidade no momento em que o desvio realmente ocorre. Outro é a previsão de tempo para casos em execução, ou seja, dado um caso parcialmente executado, o tempo de processamento restante é estimado com base nas informações históricas de casos semelhantes. E por fim, outro caso é de recomendações, o sistema recomendar com 85% de certeza uma ação futura, e assim, reduzir o custo ou tempo total de processamento, por exemplo.

3.4. Diferentes Perspectivas de Mineração de Processo

É reconhecido pelos pesquisadores da área que a maioria das publicações são voltadas para descoberta do fluxo e conformidade dos processos, como encontramos em Zaoh (2014); Appice *et al.* (2016) e Ferreira; Alves (2012). Entretanto, dentro desta grande área há diferentes perspectivas que podem ser combinadas em modelos mais completos, conforme aponta-nos Jans; Alles; Vasarhelyi (2013); Van Der Aalst; Weijters; Song (2020) e Van Dongen *et al.* (2005). Então, analisar modelos de processo, não é apenas controle de fluxo, é também sobre dados, tempo, recursos, custos e risco. É preciso combinar todas essas perspectivas diferentes em um único modelo para ser capaz de realmente aplicar a mineração de processo de uma forma significativa. As perspectivas para a Mineração de Processos são: Fluxo de Atividades, Organizacional,

Casos e Tempo, como orienta-nos Van Der Aalst (2011).

A perspectiva do processo (*process perspective*) concentra-se no fluxo de atividades focada no encadeamento e ordenação de atividades realizadas. Segundo Van Der Aalst; Weijters; Maruster (2004) e Weijters; Van Der Aalst; de Medeiros (2006), o objetivo nesta perspectiva é encontrar boas descrições de todos os caminhos possíveis, geralmente resultando em um modelo de processo. A perspectiva organizacional (*organizational perspective*) concentra-se em como a estrutura organizacional apresenta-se, como representa o relacionamento entre os atores dentro de uma instituição e os eventos. O objetivo dessa perspectiva é determinar quais funções estão presentes e como elas interagem, conforme indica-nos Van Der Aalst; Reijers; Song (2005) e Song; Van Der Aalst (2008). A perspectiva do caso (*case perspective)* concentra-se nas características de um caso, como seu caminho e atores. Van Der Aalst, *et al.* (2012) menciona uma quarta perspectiva no manifesto Mineração de Processo, a perspectiva de tempo (*time perspective*). Essa perspectiva concentra-se no tempo dos eventos, o que pode levar à descoberta de gargalos, monitorar recursos, prever o tempo restante e medir os níveis de serviço.

Essas perspectivas ajudam a determinar como um processo foi executado, quem estava envolvido e o que aconteceu, de acordo com Jans; Alles; Vasarhrlyi (2013). Portanto, as perspectivas indicam como um usuário vê um modelo de processo. Enquanto os tipos de mineração de processo, por outro lado, determinam como a mineração de processo pode ser usada.

3.5. Algoritmos de descoberta de processos

Esclarece Van der Aalst (2011) que uma das análises típicas em um estudo de mineração de processos é descobrir automaticamente um modelo de processo a partir de um registro de eventos, que apresenta as etapas do processo (em diferentes perspectivas) que foram executadas e sua ordem. Espera-se que esse modelo de processo reflita a realidade baseado nos dados tanto quanto possível. Através dele é possível calcular medidas de desempenho e identificar gargalos, carga de trabalho da equipe e desvios. Em muitos casos, as técnicas de descoberta de processos fazem concessões entre as quatro dimensões de qualidade de adequação, precisão, simplicidade e generalização.

Tiwari et al. (2008), Van der Aalst (2016), Zaoh (2014); Appice et al. (2016) e Ferreira; Alves (2012) reforçam que muitos esforços nas pesquisas de mineração de processos são focados na descoberta. Logo, criou-se muitos algoritmos que são apontados por Tiwari et al. (2008): do tipo Genéticos: Algoritmos projetados em torno do processo de seleção natural de Darwin como *Genetic Miner*; Abordagens de algoritmos gerais: Algoritmos customizados, projetados para MP por autores individuais *Alfa Miner* e *Beta Miner, Fuzzy Miner* e *Heuristics Miner*; Abordagem markoviana: Algoritmos que examinam o comportamento passado e futuro para definir um potencial estado corrente como *Expectation-Maximization Miner*; Rede Neural: Modelam baseado na forma de aprender do raciocínio humano para identificar os padrões nos dados como *RNet*; Análise de Cluster: Divide grupos de soluções em subgrupos homogêneos.

Outros algoritmos foram criados e novas abordagens ganharam visibilidade tanto em ferramentas comerciais quanto acadêmicas. Segundo Leemans *et al.* (2019) um dos algoritmos mais simples para criação do modelo do processo é o *Directly-Follows Multigraph* (DFM). DFM é um grafo que mostra a relação entre as atividades em um processo de negócio incorporando vários casos. As relações no DFM mostram como o controle no processo pode passar de uma atividade para outra. Leemans *et al.* (2019) diz que, sintaticamente, um modelo de seguimento direto (DFM) é um grafo direcionado no qual os nós são anotados com uma atividade, início ou fim.

Definição DFM sintaticamente, segundo Leemans *et al.* (2019b): Dado um alfabeto Σ tal que início ∉ Σ e fim ∉ Σ, um *Directly-Follows* é um grafo direcionado (N, E), tal que N: Σ ∪ {início, fim} é um conjunto de nós e E: N × N é um conjunto de arestas. O início não tem arestas entrando e o fim não tem arestas saindo: $\neg\exists_n(\mathrm{n}, \mathrm{início}) \in \mathrm{E}$ e $\neg\exists_n(\mathrm{fim}, \mathrm{n}) \in \mathrm{E}$.

Leemans *et al.* (2019b) complementa que, semanticamente, a linguagem de um DFM consiste em todos os traços que podem ser obtidos iniciando no nó inicial e caminhando pelas bordas, adicionando eventos para cada atividade encontrada, ao nó final. Semelhante aos *registros* de eventos, um evento de início é opcional, enquanto um evento de fim é obrigatório para cada atividade para garantir que a atividade seja realmente executada.

No PROM, Leemans et al. (2019a) apresenta que o DFM está implementado através do *plugin* Minerador *Visual Directly Follows* (DFvM). Conforme apresentado na Figura 6, para a entrada é necessário o arquivo XES do registro de eventos e aplica automaticamente uma série de etapas, o que permite ao usuário realizar análises baseadas em processos no log. Assim, primeiramente um modelo DFM é descoberto

por um algoritmo de descoberta DFM. Em depois, o DFM é alinhado com o registro de eventos e os resultados são mostrados. Este *plugin* é prático e possui outros recursos de forma a facilitar a manipulação de configurações para o fluxo do processo como nos apresenta (LEEMANS, 2017) e Leemans et al. (2019a). Como por exemplo, os desvios entre o registro de eventos e o DFM podem ser mostrados. Além disso, é possível configurar o classificador de exibição das etapas (pode ser atividades, recursos, tempos entre outros), várias medidas de desempenho, como a frequência com que partes do modelo são executadas e o tempo gasto esperando ou executando essas partes, são computadas e visualizadas. Interessantes citar também que há simulação de cada instância de processo por todo o fluxo através de bolas amarelas fluindo pelo modelo, o que permite a observação de gargalos, comportamento de lotes e sazonalidade. O DFvM facilita essas consultas filtrando rapidamente de várias maneiras os percentuais de atividades e caminhos possível, de tal forma que se chegue num modelo ideal.



Figura 6 - Plugin do DFvM no PROM – suas entradas, processamento e saída
Fonte: PROM

Neste trabalho é apresentado fluxos de processos baseado no algoritmo DFM devido a sua simplicidade de implementação e interpretação, pois será como uma análise complementar a fim de exemplificar a aplicação de descoberta do processo, não sobre as atividades mas na perspectiva de recursos utilizando o classificador organizacional.

3.6. Mineração Organizacional

Os autores Song; Van der Aalst (2008) descrevem que a mineração organizacional é a mineração de processos sob a perspectiva organizacional. Esta perspectiva visa unir os conhecimentos dos *registros* dos eventos dos processos com o foco no atributo de recursos (este podendo ser o colaborador, departamentos, funções), análise de redes sociais (como por exemplo, quais atores estão envolvidos na realização

das atividades e como eles se relacionam) e até técnicas de mineração de dados.

Segundo Song; Van der Aalst (2008) a mineração organizacional pode ser explorada de acordo com os três tipos de MP mencionados anteriormente (ou seja, descoberta, conformidade e aprimoramento). Estes autores ressaltam ainda que há diferentes tipos de mineração organizacional: descoberta de estrutura organizacional, análise de redes sociais, alocação de recursos e mineração de papéis.

A descoberta da estrutura organizacional visa derivar o modelo organizacional real que representa a situação atual dos processos de negócios a partir de *registros* de eventos. Ela não pretende derivar uma hierarquia organizacional explícita, mas um grupo de originadores que possui características semelhantes na execução do processo. Um grupo pode ser uma unidade organizacional ou uma função.

Foram desenvolvidos diversos algoritmos para descobrir o processo real, analisar conformidades e identificar causas para os modelos dos processos. Também tem-se avançado nas pesquisas para além de apenas mapeamento das atividades mas também sobre as estruturas organizacionais, como atores, papéis e suas relações, como apresenta-nos Zhao (2014). Nos próximos parágrafos são apresentadas algumas pesquisas como foco em mineração organizacional.

A primeira ideia publicada para a junção dos conhecimentos entre mineração de processos e análise de rede social está na trabalho de Van Der Aalst; Song (2004), onde foi apresentado o termo de Mineração Organizacional. E neste mesmo artigo foram apresentados o modelo de extração dos recursos que fazem as atividades em cada *registro* de evento, métricas para serem utilizadas como pesos dos arcos das redes sociais e até uma ferramenta para obter uma rede social. As métricas se destacaram neste artigo por envolver tanto a relação entre os indivíduos quanto com as atividades dos processos. Elas foram divididas em quatro tipos: Métricas baseadas em causalidade, métricas baseadas em casos conjuntos, métricas baseadas em atividades e métricas baseadas em tipos de eventos especiais.

Essas métricas foram detalhadas e apresentadas em um estudo de caso real nas pesquisa de Van der Aalst; Reijers; Song (2005). E como resultado deste trabalho foram apresentados vários sociogramas aplicando as métricas desenvolvidas pelos autores e, também técnicas de análise de rede social (como centralidades). Estes foram validados pelo corpo de gestão do escritório estudado que tanto apoiaram a correção das análises quanto obtiveram surpreendentes informações adicionais, como por exemplo, equipe de *sêniors* deixando de executar suas análises e transferindo o trabalho para colaboradores *júniores*. Validando assim a relevância da mineração de

processos no contexto organizacional. Em Van der Aalst; Song (2004) e Van der Aalst; Reijers; Song (2005) os autores baseiam-se principalmente na conversão de um evento entrar em uma rede social de recursos e gerar métricas de rede social, a fim de determinar padrões organizacionais relevantes. Por exemplo, entre (a razão com base no número de caminhos geodésicos que visitam um determinado nó) pode ser usada para encontrar possíveis gargalos.

A abordagem de Song; Van der Aalst (2008) foca na análise de rede social e fluxos de informações entre entidades organizacionais. A qual apresentou o tema da descoberta do modelo organizacional, ou seja, a estrutura organizacional baseada nos recursos dos *registros* de eventos dos processos. Com base na ideia de identificar comunidades aplicando vários algoritmos de agrupamento oriundas das técnicas de análise de redes sociais. Diferente de Song; Van der Aalst (2008) que usa apenas dados estáticos, Appice *et al.* (2016) propôs uma técnica para descobrir e atualizar continuamente uma rede social descoberta de *registros* de eventos. Nesta abordagem, é possível rastrear mudanças na estrutura organizacional ao longo de janelas de tempos e obter conhecimento históricos e reais. Ainda, em Song; Van der Aalst (2008) é sugerido a busca pela organização hierárquica mais próxima da realidade e trabalhos futuros que possam desenvolver recursos para verificação e auditoria do modelo descoberto quanto ao real. Já a busca do presente trabalho é identificar justamente um modelo de organização informal.

Os autores Ferreira e Alves (2012) identificam que a aplicação das técnicas pioneiras de Van der Aalst; Reijers; Song (2005) e Song; Van der Aalst (2008) resultam em redes complexas e propõe a combinação de *cluster* hierárquico juntamente com a métrica de modularidade. Essa métrica tem como objetivo medir a qualidade da divisão dos grupos e, assim, identificar vértices que melhor se agrupam. O trabalho discorreu sobre a rede social considerando a métrica de trabalho conjunto e não similaridade de tarefas.

Já a pesquisa feita por Yang; Ouyang; Van Der Aalst (2020) apresenta novos modelos de descoberta de modelo organizacional a partir de registros de eventos sem utilizar conceitos de redes sociais para criação de grupos. Há também uma abordagem para aplicação e ligação entre os clusteres e também uma forma de avaliar a conformidade do modelo organizacional proposto com relação ao modelo real.

3.7. Classificação e desafios dos Processos de Negócio

A classificação dos processos pode do tipo estruturado e o outro o não estruturado aponta Van Der Aalst (2020). Na prática, a maioria dos processos está situada em algum ponto entre esses dois extremos de classificação. O "processo lasanha", pode ser visto na Figura 7, é do tipo estruturado pois é regular, repetitivo e pode ser controlado. Portanto, é previamente entendido que as instâncias do processo ocorrem numa frequência relativamente constante. Há aproximadamente a mesma duração de tempo e seguem um ritmo contínuo. No geral, o caminho principal do fluxo do processo é representado por um modelo relativamente simples, e os usuários chaves do processo sabem validar sua conformidade. Van der Aalst (2011) recomenda que este modelo de processo acordado tenha uma adequação de pelo menos 0,8, ou seja, mais de 80% dos eventos acontecem conforme planejado e as partes interessadas confirmam a validade do modelo.

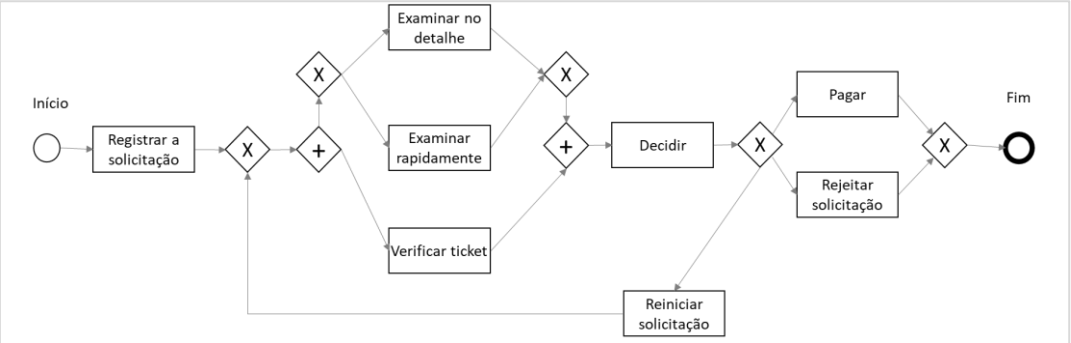


Figura 7 - Exemplo de processo tipo “Lasanha”
Fonte: Adaptado de Van der Aalst (2011)

Por outro lado os processos não estruturados são denominados de "processo espaguete", representado na Figura 8. Ele é irregular, flexível e muito variável. O formato que o modelo do processo normalmente toma lembra um espaguete por isto este nome, onde há tantas conexões que fica muito difícil entender e validar qual é a estrutura do fluxo do processo. Não identificando assim um caminho claro. Nesta Figura 1Figura 8 o fluxo representa o diagnóstico e tratamento de 2.765 pacientes em um hospital holandês. O modelo do processo foi construído com base 114.592 eventos. Existem 619 atividades diferentes executadas por 266 pessoas diferentes.

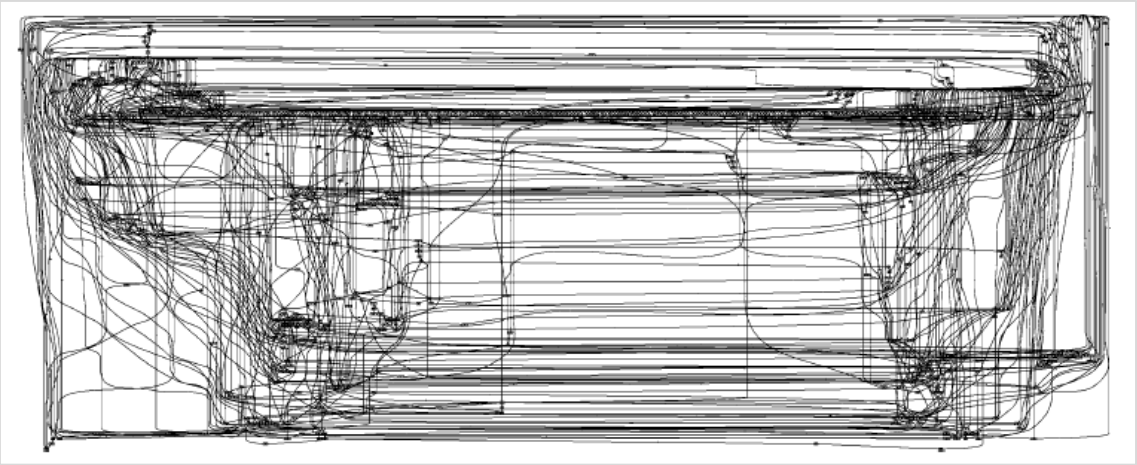


Figura 8 - Exemplo de processo tipo “Espaguete”
Fonte: Van der Aalst (2020)

Na Mineração de Processos ainda há muitos desafios e obstáculos principalmente na etapa de processamento e pré análise dos dados. Tiwari *et al.* (2008) compartilhas os mais comuns: Ruído: é quando os dados estão incorretos ou incompletos, trazendo problemas quando esses dados são minerados e interpretados. . Análises local/global: estratégias locais restringem o espaço de busca, mas são menos complexas e mais focadas. Já estratégias globais são complicadas, mas tem uma melhor chance de encontrar a solução ótima. Processos concorrentes: quando os processos ocorrem ao mesmo tempo. Loops: quando um ou mais eventos são executados várias vezes. Perspectivas diferenciadas: quando há informação adicional nos eventos dos *registros* que permitem gerar diferentes perspectivas de mineração e análise dos dados. Tarefas escondidas e duplicadas: tarefas que existem, porém não podem ser encontradas nos dados por diversos motivos, e tarefas que ocorrem quando dois casos de processos se referem ao mesmo modelo de processo.

3.8. Processos na Administração Pública

De acordo com Chiavenatto (2004) as empresas são organizações constituídas por associações de pessoas que trabalham em conjunto nos processos buscando um objetivo em comum explorando algum negócio. Negócios estes que podem ser do tipo públicos, privados, não-governamentais com fins lucrativos ou não. As empresas são uma das mais sofisticadas e complexas invenções do homem na era moderna. Segundo

este mesmo autor há duas classificações para as empresas que são privadas ou públicas. As empresas privadas são aquelas de propriedade particular voltada para produzir produtos ou serviços com o foco de obter retorno para os proprietários, acionistas, investidores e cada vez mais para a sociedade em geral. As empresas públicas são propriedades do Estado ou com participação do Estado e seu objetivo é prestar serviços à sociedade em geral.

A palavra administração vem do latim *ad* (junto de) e *minister* (prestação de serviço, obediência), ou seja, ação de dirige obedecendo à vontade de quem o contratou. Segundo Chiavenato (2008) a administração é fazer as coisas por meio do trabalho em conjunto das pessoas de maneira eficiente e eficaz para conseguir objetivos comuns. A administração pública representa o funcionamento do Estado e estabelece um importante segmento da ciência da Administração. O governo tem como instrumento as atividades dos processos relacionadas com o planejamento, organização, direção e controle de todas as ações administrativas, no sentido de dar plena satisfação das necessidades básicas à sociedade. O progresso e o desenvolvimento de uma nação passam obrigatoriamente pela Administração. E por fim, este mesmo autor lembra Peter Drucker que não existem países subdesenvolvidos ou desenvolvidos, porém países bem-administrados e mal administrados.

Chiavenato (2004) define as funções administrativas como as seguintes: Planejar, Organizar, Dirigir e Controlar. Planejar: é a primeira função administrativa onde são definidos os objetivos para desempenho organizacional atual e futuro e também os recursos e atividades necessárias para alcançá-los adequadamente. Organizar: é estabelecer meios e integrar recursos e órgãos estabelecendo as atribuições e relações entre eles, no sentido de possibilitar a realização do planejamento. A organização é a função administrativa relacionada com distribuição, agrupamento atribuição de tarefas nas equipes ou departamentos. Dirigir: consiste em interpretar o plano, influenciar e instruir as pessoas da organização sobre as atividades relacionadas com as diversas tarefas para garantir o mesmo objetivo da organização. Controlar é a função administrativa que monitora, acompanha, avalia, mede e assegura que a empresa esteja funcionando de maneira certa e no tempo certo. Garante que os resultados esperados sejam produzidos e alcançados conforme os objetivos propostos. O controle consiste basicamente em um processo que de fim a fim que guia a atividade exercida com foco no que foi previamente estabelecido. Essa sequência de funções administrativas formam um processo administrativo, apresentado na Figura 9, que se integram e se relacionar numa ação dinâmica na organização.



Figura 9 - O processo administrativo
Fonte: Adaptado de Chiavenato (2008)

As organizações públicas ou privadas ao longo das últimas décadas estão se modernizando juntamente com as transformações tecnológicas, tornando seus processos administrativos mais integrados e automatizados, a fim de alcançar assertividade na tomada de decisão, agilidade, compartilhamento de informações em tempo real e facilitar o controle de seus resultados.

3.8.1. Administração Pública no Brasil

Chiavenato (2008) apresenta que no âmbito das organizações públicas tem-se três modelos básicos: a administração pública patrimonialista, a burocrática e a gerencial. Este último é o modelo atual e, Pinho (2008) discorre que há um movimento, desde a década de 90, chamado Nova Gestão Pública e tem impulsionado as novas Tecnologias de Informação e Comunicação, abreviada por “TICs”. Destes surgem iniciativas como e-governo, governo eletrônico, que priorizam o desenvolvimento de tecnologias de informação para trocas de informação e prestação de serviço à distância com os cidadãos e empresas tornando uma gestão mais participativa e transparente para a sociedade em geral Governo (2021). Parece soar estranho sobre modernização do setor público pois sua administração sempre foi vista como ineficiente e burocrática.

No Brasil, ações deste tipo estão sendo aplicadas através do Decreto Presidencial de 03 de abril de 2000 com o objetivo de revisar e propor políticas e diretrizes com base nas interações eletrônicas e incorporando frequentemente o uso das tecnologias da informação Marcondes e Jardim (2003). Como dois grandes exemplos podemos citar o Portal Previdência Social onde a Declaração de Imposto de Renda pela Internet e SPED/e-Social – Sistema Público de Escrituração Digital/e-Social. Os desafios com administração pública ficaram ainda maiores na pandemia da Covid-19 e muitas pesquisas estão explorando as tecnologias aplicadas e as adaptações necessárias nesse novo cenário. O trabalho de Moura; Brauner; Janissek-Muniz (2020) investigou as estratégias digitais identificadas nas cidades brasileiras sob uma ótica de acompanhamento ou de antecipação governamental à situação pandêmica, baseada no conceito de *smart city*. Foi conduzida em abril de 2020 e identificou que as cidades brasileiras pesquisadas estão seguindo, em algum nível, as tendências de otimização e de enfrentamento adotadas mundialmente. Eles ressaltaram que há tendência de adoção das estratégias nos mesmos locais de contexto historicamente favorável à inovação ou maior PIB per capita. A pesquisa de Ferreira, I. D. V. *et al*. (2021) é um relato sobre da experiência vivenciada pela Coordenadoria Administrativo-Financeira da Secretaria da Saúde de Sobral, com a visão de gestão relativo às ações no que concerne ao enfrentamento da pandemia. Importante ressaltar que as discussões dos relatos de experiências ocorrem a partir das informações registradas em sistemas de informação, planilhas e anotações gerais disponíveis nas gerências Financeira, Gestão de Pessoas, Logística e Patrimônio, Infraestrutura e Manutenção de Equipamentos e Transportes, considerando o período de março a novembro de 2020. Com a finalidade de garantir uma memória e um guia minimamente numa gestão pública sobre como agir, em cenários semelhantes ao vivido atualmente. Uma revisão sistemática pautou o estudo de Moura et al. (2020) apresentar as potenciais aplicações e consequências do uso da *Blockchain* para a administração pública. Os impactos de seu uso são as melhorias na gestão dos dados, diminuição da burocracia e necessidade de afinar a relação entre Estado, sociedade e mercado. Reforçando que a tecnologia de *Blockchain* apresenta-se capaz de renovar os processos de gestão, mas criando novos desafios à administração pública.

Essas grandes mudanças, provocadas pelo uso dos Sistemas de Informações além de apoiar os processos de negócios geram um grande volume de dados. Muitos destes dados são os registros dos *registros* de eventos como data e hora, atividade executada, recurso responsável entre outros. E através destes dados, conforme é

explanado na sessão 3- Mineração de Processos, é possível identificar a maneira como os processos estão estruturados e de que forma as pessoas estão se relacionando. Portanto, é possível realizar análises nos processos afim aprimorar a administração pública visando a eficiência, eficácia e efetividade pelo bem e interesse da sociedade em geral. Este destaque é importante pois este trabalho iremos tratar de um estudo de caso em um processo de uma Universidade Pública no Brasil o qual se utilizar de meios eletrônicos para acompanhar as solicitações feitas por qualquer cidadão.

Os estudos apresentam a aproximação da Mineração de Processos com a Gestão Pública no Brasil com aplicações práticas para melhorias dos processos administrativos. Da Costa e Rodrigues (2020) demonstram a viabilidade de MP ao processos orçamentário na Marinha do Brasil e tem como objetivos específicos, a identificação de modelos de processos reais, gargalos e desvios neste processo. Apoiaram-se no software Disco com base em *registros* de eventos do novo Sistema de Acompanhamento do Plano Diretor. E como desejo final é servir como modelo para futuras aplicações da metodologia aos demais macroprocessos da Marinha. A pesquisa de Gandulfo et al. (2016) também aplica-se algumas técnicas de MP numa empresa pública jurídica e realizar a análise do processo de controle de férias de servidores públicos. Foram apresentados alguns indicativos, como por exemplo, há um volume considerável de retrabalho em duas tarefas existentes no processo, que precisam ser revisados no processo para uma execução mais eficiente nessa organização.

Neste contexto, nos deparamos com a gestão dos processos administrativos do protocolo geral do CEFET/RJ elemento essencial de relacionamento e solicitações gerais dos alunos, professores, funcionários e cidadãos em geral.

4- Metodologia

Inicialmente, para o desenvolvimento deste trabalho, realiza-se um estudo bibliométrico sobre a mineração de processos com o objetivo de identificar sua abrangência, principais autores e técnicas em termos gerais sobre o estado da arte em MP. Posteriormente, restringe-se aos trabalhos voltados à estrutura organizacional através da mineração de processos. E por fim, na busca expandida atualiza-se a pesquisa com alguns trabalhos. Nesta parte da presente pesquisa de dissertação, a exploração feita pode ser definida como um estudo quantitativo, a fim de responder a primeira pergunta de pesquisa P.1 detalhada na sessão 1.2 - Perguntas de Pesquisa e Objetivos.

No segundo momento deste trabalho realiza-se um estudo de caso nos processos administrativos de uma instituição de pública de ensino profissionalizante e de graduação tecnológica. Aplica-se técnicas de mineração de processos para responder a pergunta de pesquisa P.2 detalhada na sessão 1.2 - Perguntas de Pesquisa e Objetivos baseado em um framework de implantação de projetos de mineração de processos.

4.1. Desenvolvimento do Estudo Bibliométrico

A seguir são apresentadas as três partes do estudo bibliométrico: busca geral, busca específica e bus expandida.

4.1.1. Desenvolvimento do Estudo Bibliométrico – Geral

Primeiramente, efetuou-se uma busca na literatura com foco da análise bibliométrica da mineração de processos de com a *string* de busca de forma geral. O levantamento das informações foi feito através de uma pesquisa na base *Web Of Science* (WOS), acessada através do Portal de Periódicos da CAPES e teve como termo de pesquisa “*process mining*” no campo tópico. Esta consulta foi realizada em 31 de

outubro de 2020 e obteve-se retorno de 1.969 documentos científicos abrangendo textos de vários países e diversos anos de publicação a datar desde o ano de 1999. Na representação esquemática da Figura 10, é possível verificar de forma resumida os passos aplicados para obtenção e discussões das análises. A partir desta base foi possível recuperar informações sobre a categoria dos artigos, países núcleos, evolução do tema ao longo do tema entre outras. Nesta análise bibliométrica contempla-se também análise de redes sociais para entendimento dos autores influenciadores e dos artigos de maiores destaques e seus temas.

No primeiro passo, após o levantamento da base geral do *Web Of Science (WOS),* analisa-se os dados a fim de obter os indicadores mais gerais da bibliometria. Após essa etapa obteve-se os registros de dados dos artigos, como por exemplo, nome dos autores (que posteriormente serão tratados como os vértices) e os resumos. Para obter as informações necessárias para criação das redes sociais, foi-se necessário transformar esta tabela da WOS numa correspondente, onde para cada linha havia o título do texto e apenas um autor, e assim, na linha seguinte o título poderia repetir-se, porém, com o outro autor, caso houvesse. Nesta nova tabela, com um autor por linha, foi possível identificar uma ambiguidade entre seus nomes devidas as abreviações nas publicações. Barbastefano *et al.* (2015) ressalta a importância de se adotar procedimentos de tratamento para estes casos, bem como apresentar detalhamento das fases escolhidas para mitigar erros nas métricas das redes, retrabalho e na duplicação de vértices tratando-se do mesmo autor. Como o nome dos autores são apenas textos, utilizou-se técnicas de mineração de texto para recuperar e tratar as informações necessárias. Segundo Han (2012) a mineração de texto tem sido muito ativa e tem como objetivo obter informações de alta qualidade do texto. Isso geralmente é feito por meio da descoberta de padrões, avaliação e interpretação da saída. E este mesmo autor cita que uma das tarefas típicas de mineração de texto é o agrupamento de texto e que há vários tipos de software e ferramentas de mineração e análise de texto disponíveis em instituições acadêmicas, fóruns de código aberto e indústria.

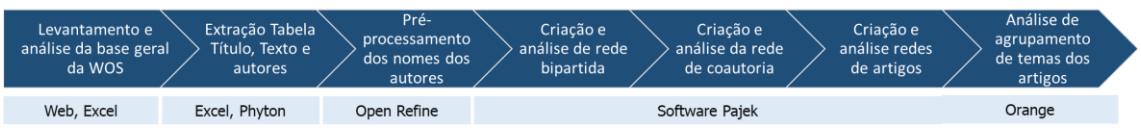


Figura 10 - Estrutura do levantamento bibliométrico com análise de rede social e agrupamento de temas dos artigos
Fonte: Própria (2020)

O pré-processamento dos nomes dos autores deu-se através do uso da ferramenta *Open Refine* para mineração de texto*. Software* de código aberto para limpeza e tratamento de dados textuais, no qual foi possível tratar vários nomes de autores através de algoritmos de agrupamento selecionados. Por padrão, o primeiro algoritmo de agrupamento é o mais restrito e mais rígido, o método de colisão de chave (*key collision*) denominado *fingerprint*, o qual apresentou os nomes identificados como praticamente os mesmos autores. E então, as sugestões foram analisadas e mescladas. Porém, após esta primeira limpeza ainda haviam várias ambiguidades, então, partiu-se para o segundo método de vizinhos mais próximos (*nearest neighbor*) chamado *levenshtein*, que pode encontrar tipos de variações que não são simples. Neste método, teve-se que inserir as variáveis necessárias para os agrupamentos (*clusteres*) que foram: o raio de distância (*Radius*) e cadeia de caracteres (*Block chars*). O parâmetro configurado para o raio foi igual a 1 e variou-se a cadeia de caracteres de 15 até 6 (padrão do Refine). Para cada agrupamento gerado foram analisadas as combinações apresentadas para definir quais seriam mesclados ou não. Entre as cadeias de 15 a 12 foram praticamente quase todos mesclados, a partir de 11 até 9 foram em torno de 50% das sugestões e entre 9 e 6 quase nenhuma foi mesclada. Na Tabela 3, apresentam-se exemplos de agrupamentos encontrados.

Tabela 3 - Ambiguidade de Nome dos autores

Exemplos de Nome com Ambiguidade nas publicações
van der Aalst, WMP (215) ;van der Aalst, W (29) ;van der Aalst, AMP (6) ;Aalst, WMPD (1) ;van der Aalst, WMR (1) ;
van Dongen, B (8) ;**van Dongen, BF** (36)
Maggi, FM (43) ;Maggi, F (1)
Rosa, NS (1) ;**Rosa, N** (1)
ter Hofstede, AHM (28) ;ter Hofstede, A (3)
Reijers, H (3) ;**Reijers, HA** (22)
Wynn, M (2) ;**Wynn, MT (21)**
De Weerdt, J (21) ;Deweerdt, J (1)
Di Ciccio, C (18) ;Ciccio, C (1)
Buijs, JCAM (18) ;Buijs, J (2)

Fonte: Própria (2020)

Como parte ainda da análise do tratamento dos nomes dos autores, vários artigos foram visitados em seu conteúdo original na *web* para certificar se o

entendimento para mesclar estava correto. Assim, foi identificado um padrão de que há um maior número de publicações em certa abreviação por autor, e este foi o nome preservado para mesclar (na Tabela 3 identificados em negrito). Além disso, os agrupamentos geralmente apareciam em 2 nomes, o primeiro com o maior número de registros e o segundo, frequentemente, com o seguinte padrão: sobrenome vírgula primeira letra de seu nome. E então, foram nestes casos que foram feitos estes tratamentos.

Além dos autores foram identificados também 19 textos publicados com o mesmo título. Foram analisados na base do *WOS* que para cada registro ele continha um ID diferente (*Unique WOS* ID e também IDS *Number*). Além desses, ainda continham outras informações que se distinguia como o tipo de documento, a conferência de publicação ou o título da fonte. Para fins deste levantamento que terá uma relação entre textos e autores foram mantidos estes registros e tratados com identificadores únicos. Han (2012) reforça que a limpeza de dados é uma etapa essencial para ajudar a evitar redundâncias, remove ruídos, corrigir inconsistências nos dados, ou seja, prepara dados para as análises.

As relações entre os vértices foram estabelecidas através da rede original bipartida a qual foi dividida em outras duas redes: a rede de coautoria e a rede de artigos. Na primeira rede os vértices os autores são ligados por uma aresta quando dois autores publicaram um mesmo artigo. E na segunda rede os vértices representam os artigos e suas arestas ligam dois artigos quando publicados com o mesmo autor. As arestas na rede bipartida são todas valoradas como um, porém, nas duas outras redes foram valoradas pelo número de artigos verificados em coautoria. Para as análises dessas três redes foram extraídas informações como número de vértices, número de linhas, grau da rede e também a componente gigante. Especificamente, na rede de coautoria foram calculadas as métricas de centralidade (grau, proximidade e intermediação) para todos os vértices envolvidos a fim de avaliar a classificação dos principais autores. Nooy; Mrvar; Batageli (2005) descrevem que se deve ter cuidado na interpretação e cálculos das medidas em rede de dois modos, pois elas têm significados diferentes entre as classes analisadas (autores e títulos de textos). Como o grau por exemplo, os vértices da classe de autores representam quantos textos eles publicaram, enquanto a classe dos títulos dos textos, representam quantos autores participaram na realização e publicação da pesquisa. Nooy *et al.* (2005), ressaltam ainda que há outras restrições para as medidas em um grafo bipartido, como por exemplo, a densidade, que é o número real de linhas dividido pelo número máximo possível de linhas, deve ser

calculada de forma diferente para redes de um e dois modos. Por isso, aquelas métricas que foram adotadas nas redes de coautoria serão diferentes na rede bipartida. Mas, métricas especiais para redes de dois modos (bipartidos) estão fora do escopo deste estudo.

Na rede de artigos, teve uma análise complementar onde foram avaliadas a presença de comunidades através do algoritmo de Louvain e o agrupamento de temas através de algoritmos de mineração de texto. Conforme aponta-nos Han *et al.* (2012) o agrupamento (clustering) é uma técnica que pode ser usada para facilmente resumir ou caracterizar os dados. Utilizou-se desta técnica para encontrar os temas dos artigo de maneira que cada artigo fez parte de apenas um dos grupos, ou seja é necessário organizar um conjunto de objetos em vários grupos exclusivos. Os autores também indicam que um método efetivo com este requisito é o agrupamento hierárquico ou "árvore" de agrupamentos. O Dendrograma é uma forma visual de representar estes agrupamentos nele é possível reprocessars recursivamente os agrupamentos até que uma granularidade desejada seja alcançada. Para a criação desta visualização de agrupamento hierárquido onde a entrada são textos, a distância de cosseno dever ser usada para comparar os documentos juntamente com vetores de termo-frequência.

Para realizar o agrupamento em temas foram extraídos alguns dados textuais dos referidos artigos pertencente as redes selecionadas através do *Software Orange*. Como mostrado na Figura 11, transformou-se o arquivo dos artigos em um dado do tipo documento ("*corpus")* e selecionou os seus respectivos: título do artigo, palavras chaves e resumo. Destes, resultaram no total de 16.292 palavras (*tokens*) dos 133 artigos científicos. o 1º passo foi tratar as palavras e permitir o n-grama de 2 palavras para melhor entendimento e contextualização dos termos. Depois criou-se a *bag of words* com a frequência de termos com a contagem e frequência de documentos igual "IDF" e em seguida preparou-se o vetor para cálculo das distancias de coseno. E através das distancias obteve-se a hierarquia de palavras para a separação dos agrupamentos dos temas (*clusters*).

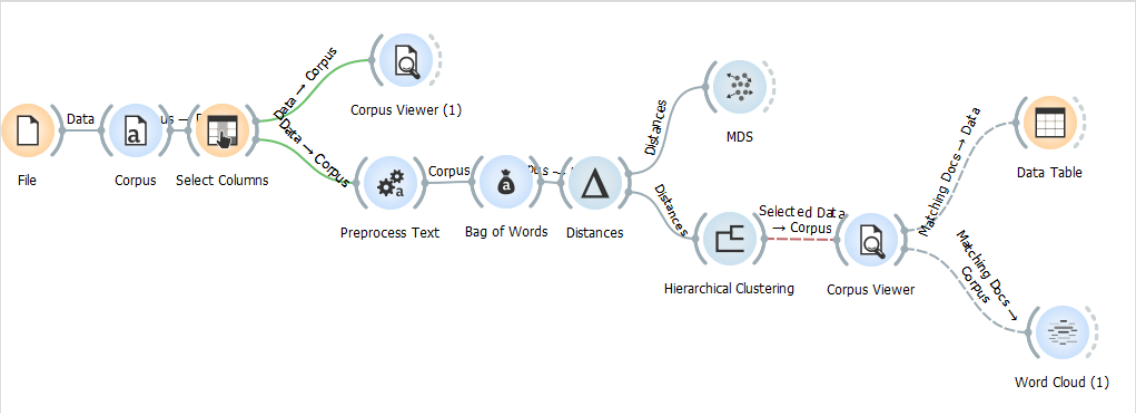


Figura 11 - Fluxo de agrupamento dos temas dos clusteres menores
Fonte: Própria (2020)

4.1.2. Desenvolvimento do Estudo Bibliométrico – Busca Específica

Entende-se que, após a análise geral dos artigos publicados na área de mineração de processo, foi preciso reduzir a abrangência da busca. Para o levantamento bibliométrico, com a busca específica sobre a mineração de processos com foco na perspectiva organizacional, restringiu-se a consulta na base do *Web Of Science* (considerando o mesmo perído da coleta do item anterior) conforme as "*strings*" da Tabela 4:

Tabela 4 – Consultas para busca específica

String da consulta	Qnt.
"process mining" and resource	128
"process mining" and Social Networks	**41**
"process mining" and Social Network Analysis	**34**
"process mining" and "Social Networks"	**18**
"process mining" and Organizational Perspective	**14**
organizational mining	**14**
"process mining" and resource and Organizational Perspective	**7**
"process mining" and Organisational Perspective	**2**
Grand Total	258

Fonte: Própria (2020)

Para a construção desta tabela uniu-se as extrações de todas as consultas para fins das análises, ou seja um artigo pode aparecer em mais de uma busca. O que retornou mais artigos foi com a busca *"process mining" and resource* apresentando 128 publicações. Ao verificar os títulos e resumos entendeu-se que haviam artigos tratando

recursos (*resource*) não apenas como recursos humanos o que não eram interessante para este trabalho que tem foco em pessoas, pois trata-se da perspectiva organizacional. Logo essa busca com 128 artigos foi descartada, e então seguiu-se para as demais consultas utilizando palavras como redes sociais (*social networks*) e perspectiva organizacional (*Organizational Perspective*).

Entendeu-se que os artigos que retornaram nas novas consultas (em negrito na Tabela 4) estavam mais aderente ao assunto deste presente trabalho. Analisou-se em seguida os artigos que apareceram em mais de uma consulta. Sendo assim, foi encontrado um artigo que apareceu em seis consultas. Houveram seis artigos que apareceram em quatro consultas. Outros treze artigos que apareceram em três consultas. Tiveram vinte e cinco artigos que retornaram em duas consultas. E por fim, vinte e nove apareceram em apenas uma consulta. Totalizou-se então 74 artigos únicos da base geral que apareceram em pelo menos uma dessas consultas restritivas, exceto aquela que possui *"process mining" and resource*. Para a análise bibliométrica desta busca restrita selecionou-se esses 74 artigos.

4.1.3. Desenvolvimento do Estudo Bibliométrico – Busca Expandida

A busca expandida para este levantamento deu-se de três formas diferentes. A primeira foi a atualização da busca restrita para os anos entre 2020 e 2021, pois na consulta anterior foi com o ano 2020 incompleto e havia apenas um artigo pré-publicado em 2021. A segunda foi a revisão de *string* de busca acrescentando as *strings " process mining" and covid* e também *"process mining" and "pandemic"*. A terceira foram as indicações de referências pelo orientador e também a técnica *Snowballing*, buscando por referências citadas nos artigos relevantes levantados na busca específica.

Seguindo essas etapas discriminadas na sessão 4.1 foi possível identificar a abrangência, principais autores e o conjunto de publicações como estado da arte e uma melhor compreensão sobre em MP.

4.2. Desenvolvimento do Estudo de Caso

Este estágio da pesquisa é caracterizada como um estudo de caso com enfoque quantitativo com fundamentação teórica e, seguido da aplicação de mineração de processos com a extração e tratamento do *registro* de eventos, conversão do formato dos dados e avaliação dos dados. O estudo de caso é um método muito popular usado em estudos qualitativos que visa coletar dados ou observar uma unidade social, como este trabalho realiza sobre uma instituição. Este procedimento metodológico leva-nos a estudar os fenômenos em profundidade ao invés de amplamente com a intenção de obter uma descrição verdadeira e completa do caso que o pesquisador selecionou para estudar, como adverte-nos Marczyk; Dematteo; Festinger (2005). É uma investigação empírica que procura por um fenômeno contemporâneo dentro de seu contexto de vida real, especialmente quando as fronteiras entre fenômeno e contexto não são claramente evidentes, segundo os termos de Yin, (2002).

4.2.1. Metodologia para Projeto de Mineração de Processo

Na representação esquemática da Figura 12, é possível ver de forma resumida os passos aplicados para o desenvolvimento desse trabalho.

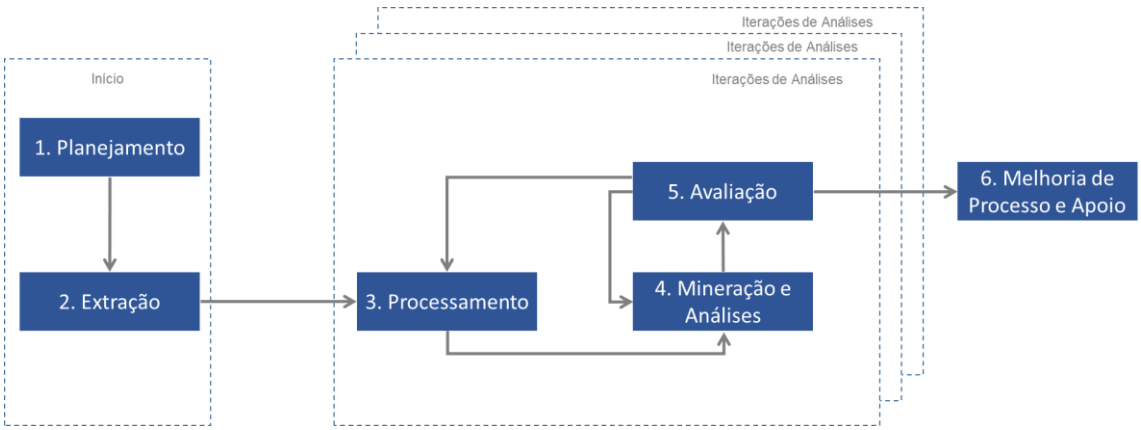


Figura 12- Metodologia PM^2 para Projeto de Mineração de Processo
Fonte: Adaptado de Van Eck *et al.* (2015)

A PM^2 - *Process Mining Project Methodology* é uma metodologia para projetos de mineração de processos baseada em três grandes tópicos: início, iterações de análises e as melhorias nos processos. Com as questões de pesquisas levantadas, em 1.2 Perguntas de Pesquisa e Objetivos, as iterações permitem tanto o refinamento quanto as respostas do projeto. As duas primeiras fases do fluxo é (1) Planejamento e (2) extração dos dados, nas quais são definidas as perguntas iniciais e os dados a ser levantados para respondê-las. Após essas atividades, é previsível ter várias iterações entre as fases (3) Processamento, (4) Mineração e Análise e (5) Avaliação e não necessariamente na mesma ordem. Uma iteração busca diagnosticar e responder cada um dos objetivos específicos desta pesquisa. E então a última fase (6) Melhoria e Suporte do Processo que, muitas vezes, pode ser caracterizada como um projeto apartado.

A Fase 3, Processamento, é muito importante para preparar os dados para a fase analítica e de mineração. Há quatro elementos importantes que podem ser considerados nessa fase: criação de visualizações, enriquecimento de *registros*, agregação de eventos e filtragem de *registros*. Nesta fase são explorados todos estes elementos: criação de visualizações dos dados, enriquecimento de atributos para auxiliar no aprofundamento das análises, agregação nos nomes de processos, e filtragem dos *registros* de eventos. A agregação tem o objetivo de padronizar o nome dos processos portanto, utilizou-se algoritmos mineração de textos para agrupamento dos nomes dos processos. Segundo Bose; Aalst (2009) e Han (2012) a agregação ajuda a reduzir a complexidade e melhorar a estrutura dos resultados da mineração, portanto ela ocorre na primeira iteração e é usada como base para as demais iterações e análises. O enriquecimento dos dados ocorre para facilitar e aprofundar algumas análises, como por exemplo, uma nova coluna para identificar se o evento ocorreu dentro do período antes da pandemia ou depois da pandemia. Já as filtragens dos *registros* de eventos ocorrem em várias iterações distintas dependendo da análise que se pretende fazer e do objetivo específico que se pretende responder. Exemplo de filtragens são: eventos filtrados com datas inconsistentes para limpeza dos dados, são filtradas instâncias de processos que ocorrem mais vezes na base e também são filtrados na última iteração instâncias de processos que foram configurados no período antes e depois da pandemia. E a criação de visualização modela novos formatos de base para serem inseridos nos softwares PROM e Pajek os quais permitem a geração das redes sociais quanto a descoberta e simulação do processo.

Na Fase 4, Mineração e Análise, é esperado que numa fase inicial do estudo

podem ser aplicadas em *registros* de eventos análises exploratórias para obter uma visão geral do processo de negócios e posteriormente focar nas análises mais específicas. Nela possui dois elementos que são: as técnicas de mineração de processos, apresentadas por Van Der Aalst, W. M. P. (2011), que são descoberta, conformidade e aprimoramento. Neste presente trabalho o foco é a descoberta das redes sociais dos departamentos envolvidos nos processos administrativos do Protocolo Geral e também a descoberta do processo através do fluxo onde as etapas dos processos são os departamentos; o segundo elemento é a análise de processos que são outras técnicas complementares de análise, por exemplo, análise visual (por exemplo, histogramas de eventos por caso) e mineração de dados e, que podem ser aplicadas no contexto de processos de negócios de acordo com Leoni *et al.* (2014). Esta análise de processos é repetida várias vezes para aumentar o conhecimento sobre a base de dados e sobre os processos tanto para uma visão geral quanto para analisar os comportamentos antes e depois da pandemia.

A fase 5, Avaliação, recebe as informações dos resultados das métricas e modelo com objetivo de relacionar os achados da análise às ideias de melhoria. Ela tem dois elementos: diagnosticar que interpreta os resultados, distingue resultados interessantes ou incomuns dos esperados e identificar ou refinar questões de pesquisa para possíveis iterações posteriores; Verificar e Validar que compara os resultados obtidos e também valida com as alegações das partes interessadas no processo. Para se fazer essas alegações e interpretações dos resultados, Van Eck et al. (2015) reforça que um dos desafios em projetos de MP é muitas vezes os analistas e pesquisadores não são especialistas no domínio do processo que estão analisando. Implicando em dificuldades em diagnosticar, determinar as causas dos resultados da análise e propor melhoria. Para mitigar essas dificuldades são abordados colaboradores do CEFET/RJ conhecedores do processo administrativo para apoiar nas análises e interpretação dos resultados.

E por fim, a fase 6, Melhoria e Suporte do Processo, tem como entrada as ideias de melhoria da fase de avaliação e as saídas desta etapa são modificações do processo. No entanto, Van Eck et al. (2015) aponta que a implementação real de modificações de processo é geralmente um projeto separado e uma área de especialização diferente. E este é o cenário deste estudo onde serão descritos os resultados encontrados, porém não é extensivo a implantá-los.

De forma geral para responder as perguntas e alcançar os objetivos deste estudo de caso resumidamente são necessários: a preparação e processamento da base nas

várias interações, aplicação de modelos analíticos, conforme discorrido por de Van Eck et al. (2015) para extrair métricas como, por exemplo, tempo total de duração dos macroprocessos e a distribuição de etapas do processo; aplicação da mineração organizacional conforme apresentado em Song; Van Der Aalst (2008) para descobrir as redes sociais, extrair os conhecimentos e descrever os resultados. A aplicação desses procedimentos podem ser vistos em Chang *et al. (*2020), Leoni *et al.* (2014), Van der Aalst, W. M.P. *et al.* (2007) e Van der Aalst, Wil M. P.; REIJERS; SONG, (2005).

4.3. Ferramentas metodológicas

As ferramentas utilizadas para desenvolvimento deste estudo de caso e do estudo bibliométrico são: *Microsoft Excel* para manipulação e primeiro levantamentos dos dados extraídos, o *Open Refine* para tratamento dos nomes dos autores (no estudo bibliométrico) e nomes dos processos (no estudo de caso); o *software* de redes sociais *Pajek* com o objetivo de analisar as redes sociais antes e depois da pandemia, o *software Orange* para análise de coautoria (do estudo bibliométrico) e para o agrupamento dos processos (no estudo de caso) e o PROM, software de mineração de processos para descoberta e simulação dos processos com foco na perspectiva organizacional. Tanto o Pajek quanto o PROM exigem uma entrada de dados diferente do arquivo padrão de *registros* de eventos. Portanto, em ambos os casos os dados são adequados para o formato necessário.

No Pajek é necessário criar um arquivo que receba os vértices, as arestas e, caso houver os pesos das arestas. Para tal, Batagelj; Mrvar (2008) orienta que é necessário construir o arquivo conforme modelo da Figura 13 e sua extensão é “.net”.

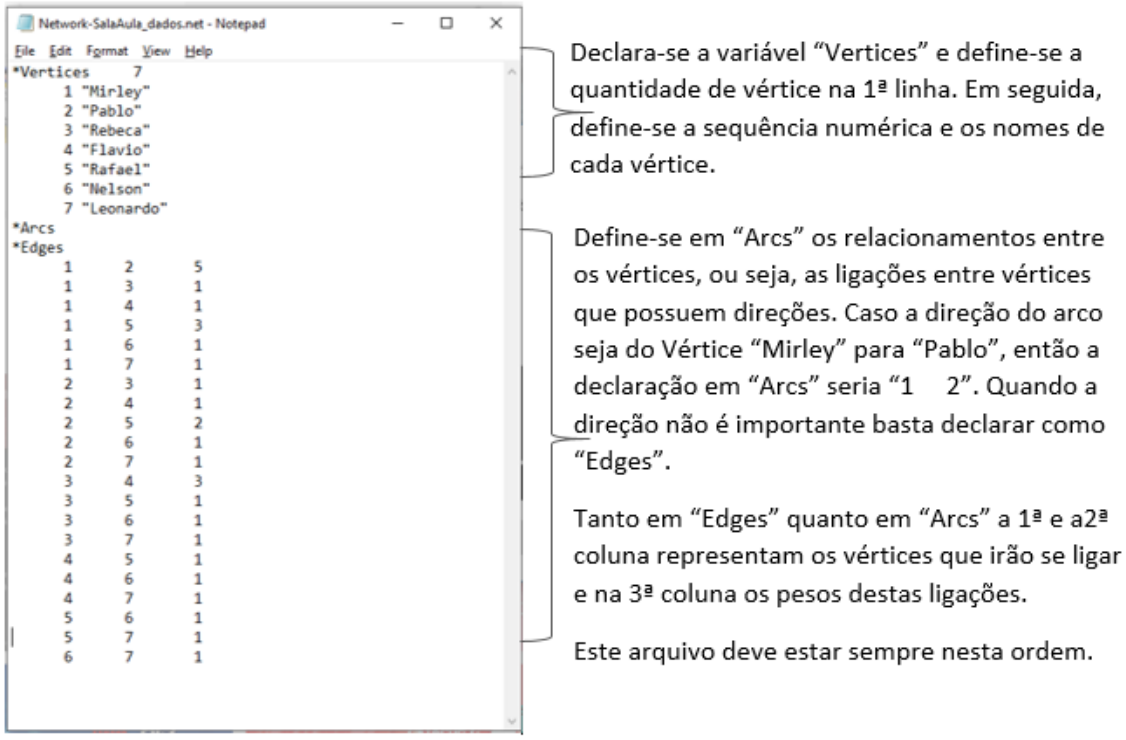


Figura 13 - Arquivo de definição da rede no Pajek
Fonte: Própria (2021)

Por exemplo, no caso das redes sociais aplicadas na mineração de processos, os vértices são os departamentos que tramitam a solicitação aberta no Sistema de Protocolo Geral, estes são ligados por arcos que indicam a movimentação da origem ao destino e os pesos dos arcos é o número da transferência de trabalho, ou seja, o número de vezes que o processo tramitou entre os dois departamentos.

Para a descoberta e simulação dos fluxos do processo é preciso fazer duas conversões para ser utilizado pelo PROM: a primeira é criar uma extensão ".csv", conforme a Figura 14. Neste arquivo contém o "Número do Processo", "Data" e o "Recurso" (esta coluna precisa ser formatada com *org:Resource*) para o algoritmo de mineração "*Directly-Follows*" identificar que o classificador será recursos. Ou seja, as caixas apresentadas na descoberta do fluxo são os departamentos e não as atividades. Uma vez formatado este arquivo, o mesmo é inserido no PROM, e posteriormente é transformado na extensão XES através do *plugin* "*Convert CSV to XES*". Assim, este arquivo. XES é a base de dados de entrada para o algoritmo de "*Directly-Follows*"

NUMERO	DATA	org:resource
23041	13/12/2018 00:00	CPPD
23041	13/12/2018 00:00	PROTOCOLO GERAL
23041	19/12/2018 00:00	DIREG
23041	03/01/2019 00:00	DGP
23041	04/01/2019 00:00	DGP/DICAD
23041	25/01/2019 00:00	DGP/DIPAG
23042	13/12/2018 00:00	CPPD
23042	13/12/2018 00:00	PROTOCOLO GERAL
23042	07/05/2019 00:00	DGP
23042	09/05/2019 00:00	DGP/DICAD
23052	17/12/2018 00:00	CPPD
23052	17/12/2018 00:00	PROTOCOLO GERAL
23052	26/04/2019 00:00	DIREG
23052	09/05/2019 00:00	DGP
23052	13/05/2019 00:00	DGP/DICAD

Figura 14 - Modelo de arquivo CSV para ser inserido no PROM
Fonte: Própria (2022)

Segundo Leemans (2017) é possível configurar diversas visualizações que auxiliam na análise dos fluxos dos processos. Neste trabalho, a descoberta do fluxo é feita através do software PROM utilizando o *plugin* para descoberta do processo chamado *Mine with Directly Follows visual Miner* (este *plugin* foi desenvolvido por Leemans (2017)). Todos os fluxos são gerados com 80% dos principais caminhos. E o classificador utilizado foram os departamentos. Esta é uma configuração do classificador não é padrão deste *plugin*. Esta configuração onde há a classificação dos departamentos é recomendada pro Leemans (2017) para analisar as transferências de trabalho entre os recursos. Logo, a configuração completa está de acordo com Figura 15.

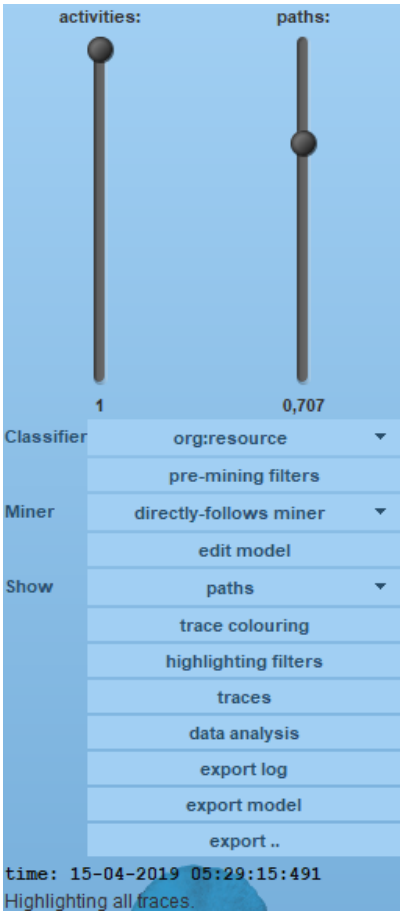


Figura 15 - Menu de Configuração no PROM
Fonte: Software PROM

5- Estudo Bibliométrico sobre Mineração de Processos

O estudo bibliométrico ocorreu em três fases: busca geral, busca específica e busca expandida. Na primeira fase buscou-se entender as pesquisas científicas sobre a área de mineração de processo de forma mais geral. Já na segunda fase aprofundou-se na busca restritiva sobre mineração de processos na perspectiva organizacional. E por fim, a busca expandida para completar a coleção de estudos deste trabalho.

5.1. Busca Geral – Resultados do estudo bibliométrico

Na consulta realizada na Web Of Science, conforme apresentado no capítulo de metodologia, em 31 de outubro de 2020 e obteve-se retorno de 1.969 documentos científicos abrangendo textos de vários países, desde o ano de 1999. Haviam 19 artigos sem data de publicação, os quais foram pesquisados na *web* para recuperar esta informação. Há um artigo inclusive dito como publicado em 2021 para o periódico *Inform Sciences*. Quanto aos tipos de documentos recuperados, retornaram-se 1.220 artigos publicados em anais de conferências/eventos (62%), seguido de 710 (36%) em periódicos. Ele também registra outros tipos como *Review* (revisões de artigos e pesquisas da literatura publicados anteriormente); *Early Access* (artigos que foram publicados eletronicamente por um periódico antes de ser atribuído a um volume), Editorial Material (artigos sobre opinião de uma pessoa, grupo ou organização como por exemplo entrevistas e mesas redondas), *Meeting Abstract* (resumo geral dos trabalhos concluídos que foram ou serão apresentados em um simpósio ou conferência) e por último *Correction* (correção de erros encontrados em artigos publicados anteriormente e que foram divulgados após a sua publicação).

Pode-se perceber uma grande evolução da quantidade de artigos, autores na disseminação do tema *"process mining*" com seus trabalhos publicados no Gráfico 1. O ano com maior número de documentos publicados foi em 2018 com 1.000 autores (um autor pode estar em mais de uma publicação) em 263 publicações. Tornando-se possível o acumulado de 6.843 (ou seja o número de arestas na rede bipartida) autores publicando 1.969 textos científicos, ou seja, esta última é o número de vértices da classe de instituições na rede bipartida. E deste número de publicações, apenas 135 (6,9%)

foram escritos por apenas um autor, ou seja, todas os vértices da classe de artigos de grau igual a 1. Os demais envolveram relação de coautoria (1.834 publicações representando 93,1%). É importante ainda ressaltar que no ano de 2020 teve uma expressiva queda de artigos e acredita-se que é devido a pandemia do vírus Covid-19, o qual teve um impacto mundial, onde prevaleceram fechadas muitas instituições e eventos não foram realizados.

Gráfico 1 - Publicações ano a ano



Fonte: Própria (2020)

Foi interessante excluir também a multiplicidade dos autores ao longo do tempo o qual permitiu analisar, conforme o Gráfico 2, a quantidade de novos autores que surgem. O acumulado de autores que tem tratado do assunto de *process mining* somam-se em 3.324, ou seja número de vértice da classe de autores do grafo bipartido. E que a média entre os anos de 2015 a 2019 (últimos 5 anos completos) é de 387,4 novos autores a cada ano.

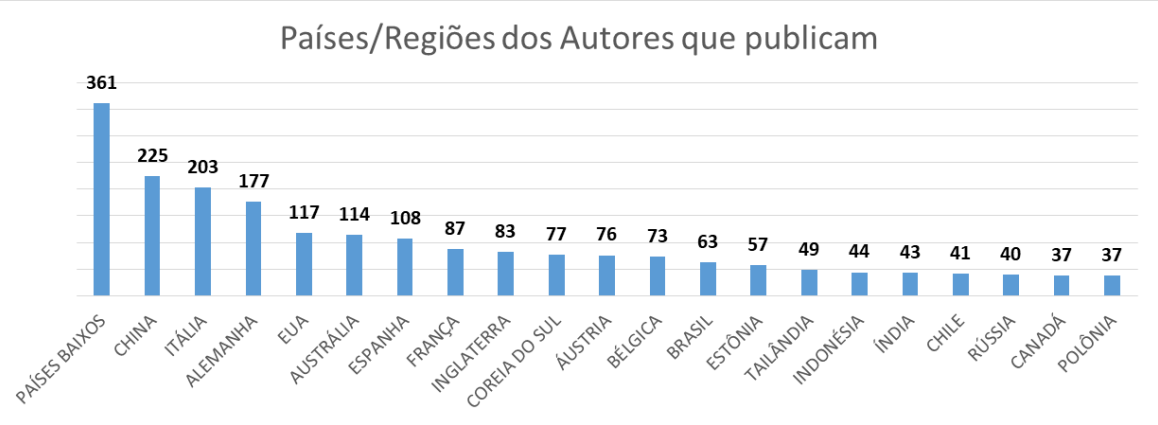
Gráfico 2 - Surgimento de novos autores anualmente



Fonte: Própria (2020)

Para apresentar como este tema está sendo trabalhado nas diversas instituições no mundo, verificou-se que há participação dos autores de 80 países diferentes. Como pode-se verificar no Gráfico 3, o principal núcleo de desenvolvimento neste tema é da região dos Países Baixo com 361 registros, seguidos da China (225) e Itália (203).

Gráfico 3 - Países e regiões dos autores



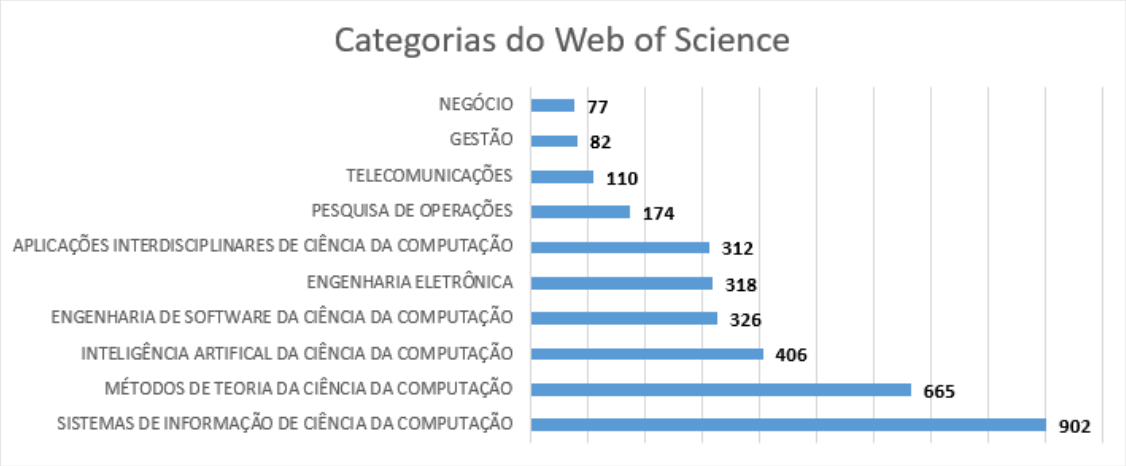
Fonte: Própria (2020)

O recorte acima foi com base na distribuição de Pareto ao qual os 21 países apresentados no gráfico representam 81% do total de publicações. E que nesta seleção o Brasil está contemplado com 63 publicações.

Para entender de quais categorias são os textos publicados sobre a mineração de processos, foi retirado, diretamente do *site do WOS* na sessão de análise, a informação de que ele abrange 96 categorias. Um texto científico pode ser classificado em uma ou mais categorias diferentes. De acordo com o Gráfico 4 o que mais se destaca

é Sistemas de Informação da Ciência da Computação com 902, seguida de Métodos de Teoria da Ciência da Computação com 665 e Inteligência Artificial da Ciência da Computação com 406 registros. Entende-se que essa categorização deve-se por tratar a mineração de processo, do levantamento, análises e teorias sobre os *registros* nos sistemas da informação.

Gráfico 4 - Categorização dos textos científicos



Fonte: Própria (2020)

O recorte acima foi com base na distribuição de Pareto no qual as 10 categorias apresentadas no gráfico representam 81% do total de publicações.

5.1.1. Rede bipartida de autores e artigos em mineração de processos

A Figura 16 abaixo representa a rede bipartida entre autores e os textos científicos publicados o qual o somatório total de vértices são 5293 (sendo 1.959 de textos científicos e os outros 3.324 de autores). Além, disso há um grau médio de 2,58 e densidade para este tipo de grafo de dois modos igual a 0,001.

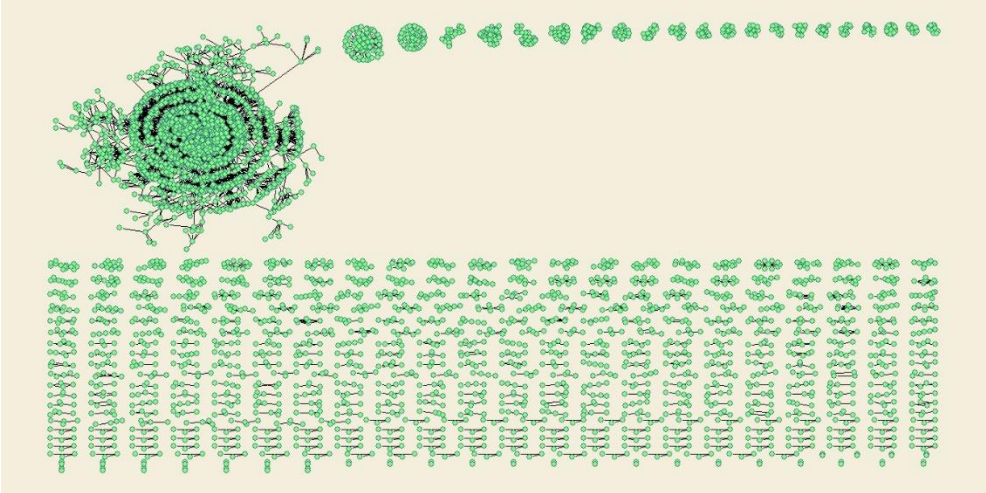


Figura 16 - Rede bipartida de autores e textos científicos
Fonte: Própria (2020)

Verificou-se que há 469 componentes isoladas, onde a componente gigante, Figura 17, possui 2.537 vértices correspondendo a aproximadamente 50% de toda rede.

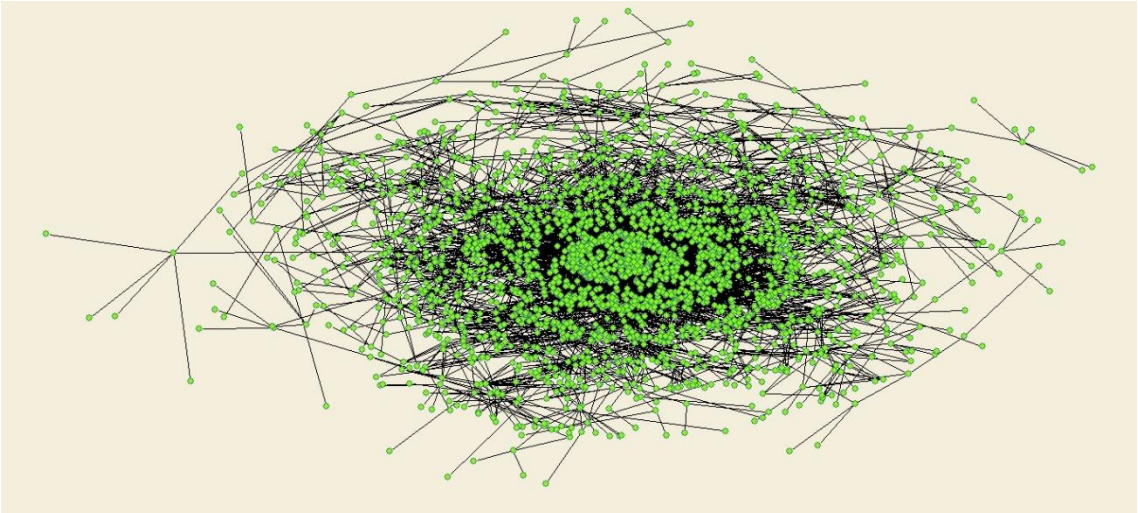
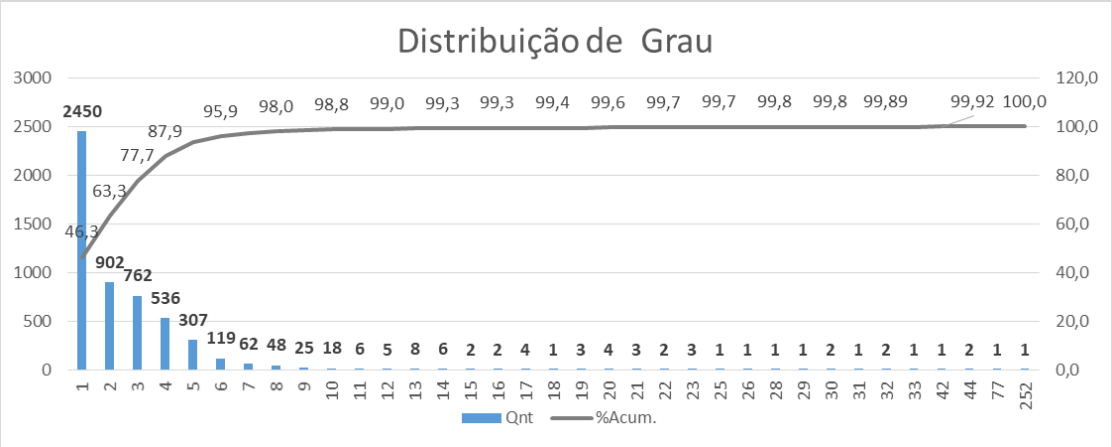


Figura 17 - Componente Gigante da rede bipartida
Fonte: Própria (2020)

Para conhecer outras informações foi feita a distribuição de grau no Gráfico 5. No qual identificou-se que há 2.450 vértices de grau 1 (46% da rede). E com até 4 graus obtém-se as ligações da rede em 87,9%. Importante destacar também que há um vértice da classe de autor com grau igual a 252, ou seja, 252 publicações distintas, e este é Van Der Aalst, o grande autor desta área de mineração de processos. Outro destaque é o artigo "*Process Mining Manifesto*" com grau 77, ou seja, 77 autores participaram da publicação deste.

Gráfico 5 - Distribuição de graus da rede bipartida



Fonte: Própria (2020)

5.1.2. Rede de coautoria em mineração de processos

Barbastefano *et al.* (2013), Souza; Barbastefano (2011), Bordin *et al.* (2014) e Souza *et al.* (2012) destaca-nos que é possível conhecer os autores influenciadores e de maiores destaques através de redes sociais. Para conhecer a rede de coautoria dos autores, foi-se necessário criar a rede, Figura 18, um modo a qual possui então 3.324 vértices (autores), 10.418 arestas onde 8.924 (85%) tem valores iguais a 1, ou seja, os autores destas ligações tiveram apenas um artigo publicado em conjunto. E os demais 15% (1.494) mais do que apenas uma publicação. A densidade desta rede é 0,001 e o número de grau médio é de 6,26.

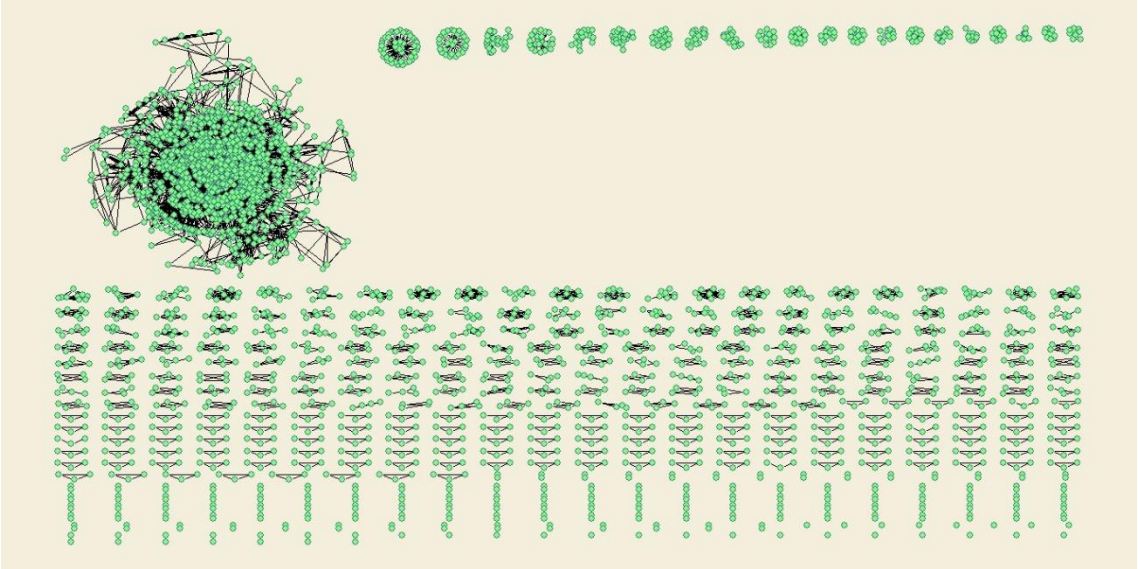


Figura 18 - Rede de coautoria dos pesquisadores
Fonte: Própria (2020)

Já na Tabela 5 a seguir está a distribuição de autores por artigo apresentando a extensão da colaboração científica existente nestes textos científicos deste presente estudo.

Tabela 5 - Quantidade de autores por artigo

Qnt autor	Nº artigo	%
1	142	7,2%
2	408	20,7%
3	590	30,0%
4	426	21,6%
5	244	12,4%
6	77	3,9%
7	37	1,9%
8	26	1,3%
9	7	0,4%
10	12	0,6%
Total	1969	100,0%

Fonte: Própria (2020)

Também há o aparecimento de uma Componente Gigante, Figura 19, nesta rede de autores, representando mais de 43% (1.442 dos 3.324 totais) dos autores, entendendo-se que este tema de mineração de processo possui características típicas de uma forte comunidade científica. E há ainda estudos isolados entre 468 outros componentes apresentados.

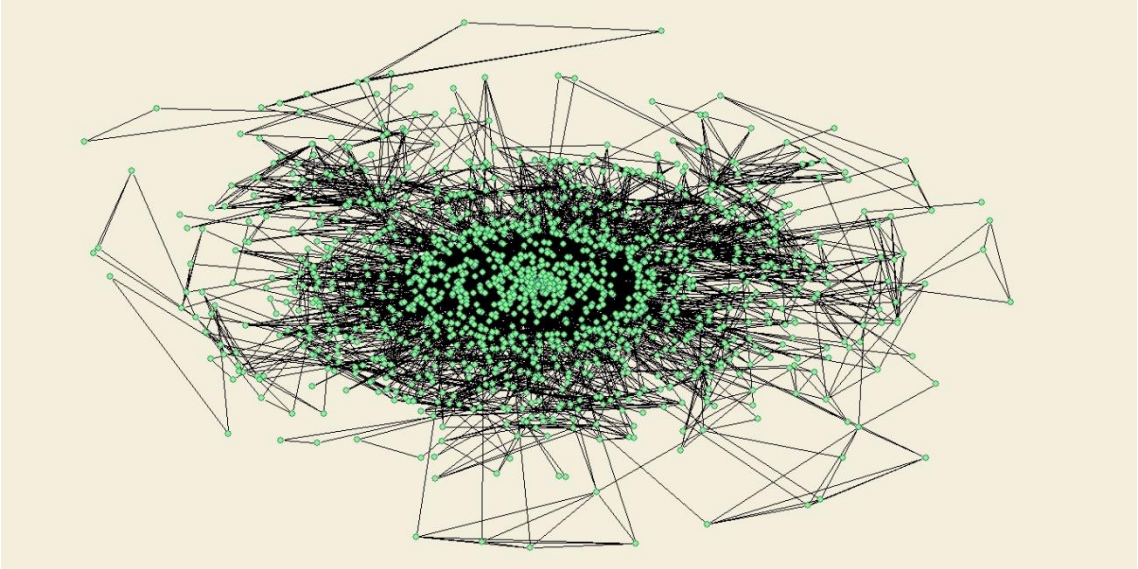


Figura 19 - Componente Gigante da rede de coautoria de pesquisadores
Fonte: Própria (2020)

De acordo com a Tabela 6, abaixo, foi-se possível classificar os pesquisadores com as maiores medidas de centralidade de intermediação em decorrência das demais, pois entende-se que este possui um caráter mais global e que destaca o processo de disseminação do conhecimento em se tratando de rede de coautoria.

Tabela 6 - Métricas de centralidade dos autores que se destacaram

Autor	Grau	Proximidade	Intermediação	Rank Grau	Rank Prox.	Rank Inter.
Van Der Aalst,	248	0,176	0,062	1	1	1
Song, M	120	0,163	0,029	4	2	2
Wen, LJ	114	0,162	0,025	8	4	3
Munoz-Gama, J	137	0,162	0,020	2	3	4
Mendling, J	116	0,162	0,016	6	5	5
Sepulveda, M	122	0,161	0,015	3	7	6
Van Dongen, BF	104	0,161	0,008	11	6	7
Maggi, FM	115	0,159	0,008	7	10	8
ter Hofstede, AHM	117	0,159	0,005	5	8	9
Wynn, MT	108	0,159	0,003	9	9	10
Dumas, M	105	0,158	0,003	10	11	11

Fonte: Própria (2020)

Tem-se então o autor Van Der Aalst com as maiores métricas de centralidade (grau, proximidade e intermediação), sendo o pesquisador com a maior influência no tema de mineração de processo. A segunda melhor métrica de intermediação é o autor Song, M, que também teve a segunda melhor métrica de proximidade. Porém, é possível

verificar que mesmo ele estando em segundo lugar há uma diferença bem significativa entre ele e o Van Der Aalst nessas métricas.

5.1.3. Rede Social de artigos publicados sobre mineração de processos

Para conhecer a rede de artigo, foi-se necessário criar a rede, Figura 20, um modo a qual possui então 1.969 vértices (artigos), 42.321 arestas onde 36.851 (87%) tem valores iguais a 1, ou seja nesta ligação um autor publicou ambos os artigos desta aresta (ligação). A densidade desta rede é 0,022, mais densa do que a rede de autores. E o número de grau médio é de 42,98, onde há 323 vértices com grau zero, ou seja não há autores em comum entre eles. Entre os graus “1-25” há 1.139 vértices, entre os graus “26-250” há 255 vértices, entre os graus “251-325” há 251 artigos e, por fim, há um vértice, ou seja, um artigo que tem grau 594. Este último é o artigo “*Process Mining Manifesto*” composto por 77 autores, o qual é uma referência em MP pois, lista os desafios e princípios para guiar as partes interessadas como cientistas, desenvolvedores e gerentes de negócios que precisem lidar com a Mineração de Processos.

Por esta grande quantidade de artigos com graus autores, levou a média da rede para este valor de 42,98. Para entender do que tratam os temas destes artigos buscou-se entender os componentes da rede.

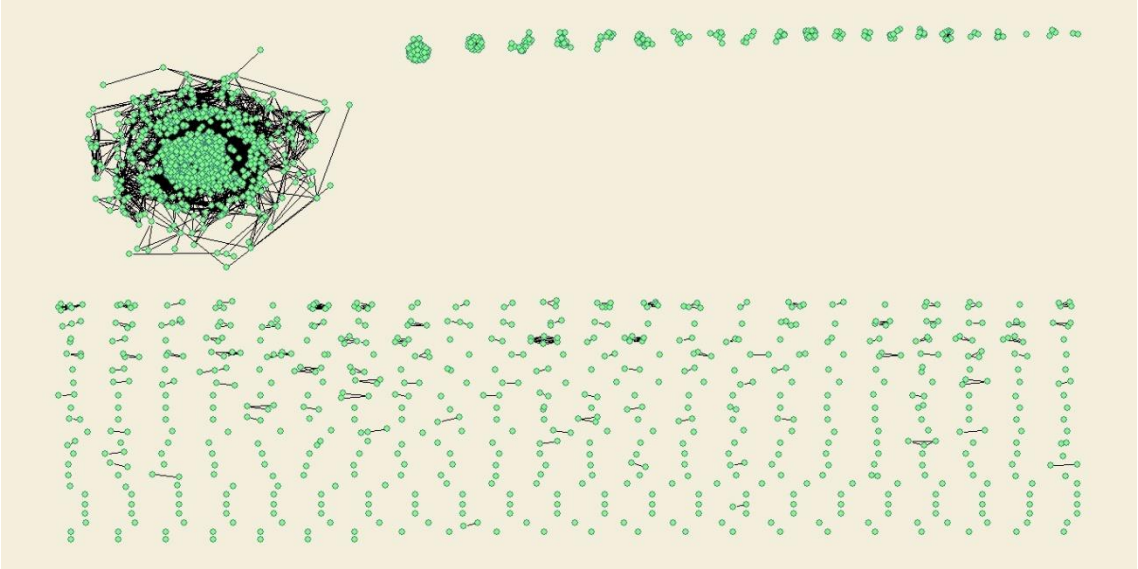


Figura 20 - Rede de artigos
Fonte: Própria (2020)

A Componente Gigante, Figura 21, é o destaque desta rede de artigos, representando mais de 55,6% (1.095 dos 1.969 totais) dos artigos. E há outras redes desconexas de interesse deste estudo que são aquelas com mais de dez vértices, estas somam-se outros sete componentes isoladas. Tanto para a componente gigante quanto estes sete componentes serão apresentados os temas dos artigos que encontram-se neles.

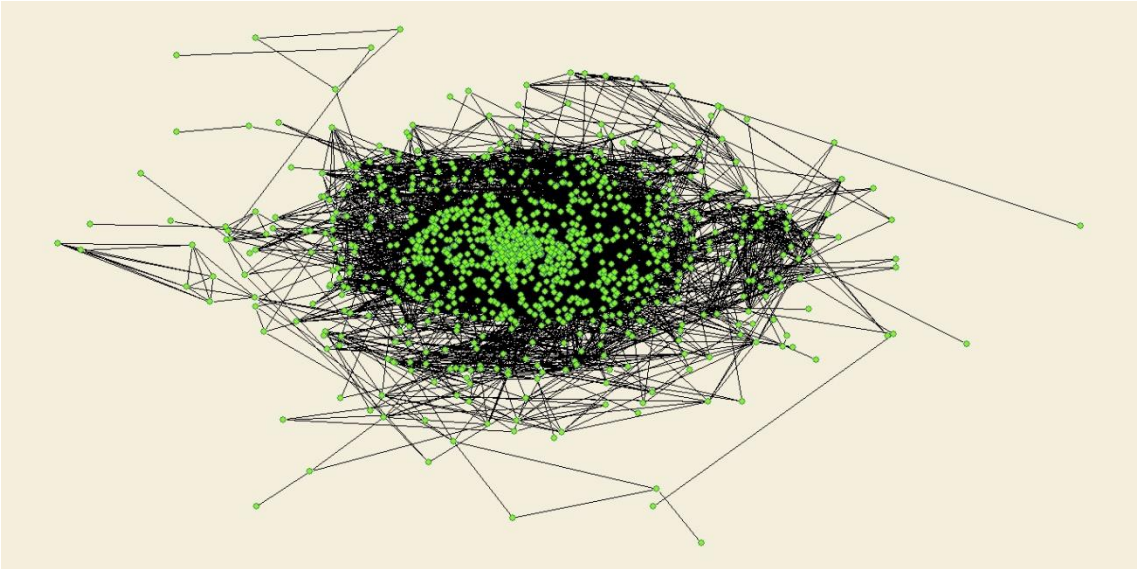


Figura 21 - Componente gigante de artigos
Fonte: Própria (2020)

Para aprofundar nos temas que os artigos da competente gigante foi feito

levantamento do agrupamento pelo método de Louvain, com o valor de modularidade (Q) igual a 0,346 conforme apresentado na Figura 22. Identificou-se que há vários 32 núcleos/agrupamentos, aonde não há nenhum agrupamento com apenas um artigo, há no mínimo com dois artigos e o maior dele passou a ter 261 artigos.

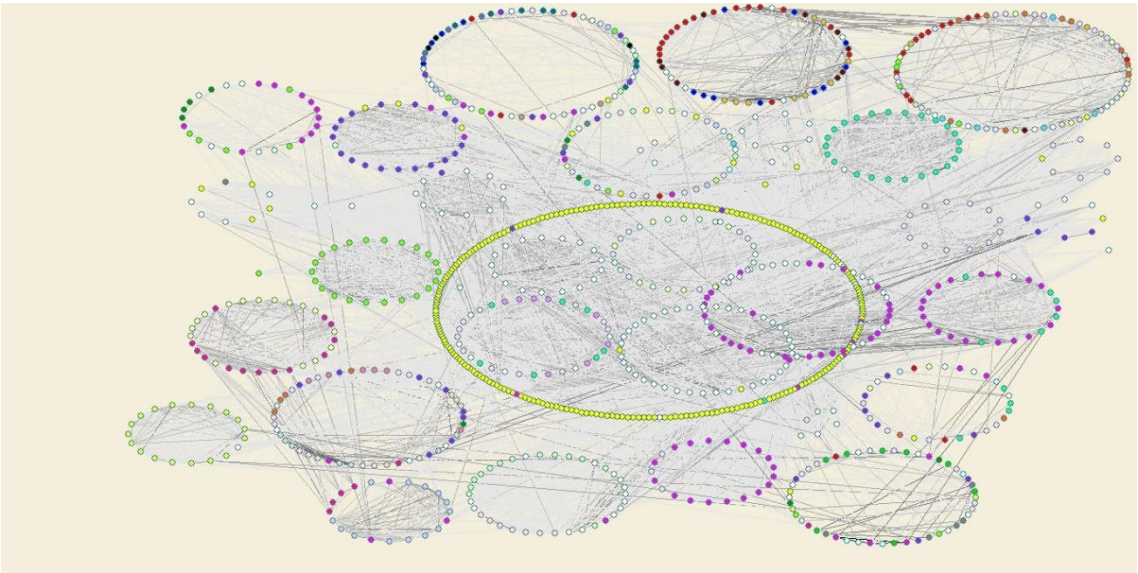


Figura 22 - Comunidades *Louvain*
Fonte: Própria (2020)

5.1.4. Encontrando temas estudados das redes de artigos

Para entender quais temas tem-se discutido nos artigos cientificos coletados, volta-se na rede principal de artigos, Figura 20, onde se tem 1.969 artigos. São isoladas os artigos que fazem parte componente gigante e as redes de componentes desconexas com mais de dez vértices, totalizando oito redes que representam 62,4% da rede. Primeiramente serão apresentados os temas identificados na componente gigante e após desta são apresentados os temas nas componentes isoladas.

5.1.4.1. Encontrando temas estudados da componente gigante

Para encontrar os temas estudados nos artigos que fazem parte a componente gigante foi utilizado o processo de clusterização apresentado na Figura 11 priorizando

os artigos pertencentes as comunidades com mais de 30 artigos, conforme a distribuição destes na

Tabela 7 abaixo:

Tabela 7 - Distribuição de artigos em cada agrupamento da componente gigante

Nº da comunidade	Frequência
101	252
8	75
3	63
5	50
9	50
31	48
4	48
43	41
1	41
13	34
7	32
2	30
Total	**764**

Fonte: Própria (2020)

Nota-se que esta seleção corresponde a aproximadamente 70% da componente gigante (que possui 1.095 artigos) e que há uma maior concentração de artigo no núcleo 101 com 252 artigos e o segundo apresenta-se com apenas 75 artigos. Como são muitos artigos ficou inviável a apresentação destes no Dendrograma (recurso de visualização de agrupamento hierárquico), porém no *software Orange* também foi possível verificar estes *clusteres* e buscar pelos seus temas através desta figura. Para demonstrar os *clusteres* encontrados para este caso, será apresentado o MDS, disponibilizado no *Orange* também, numa visualização cartesiana, conforme Figura 23 (atente-se, que nela os pontos representam os artigos e as cores o *cluster*).

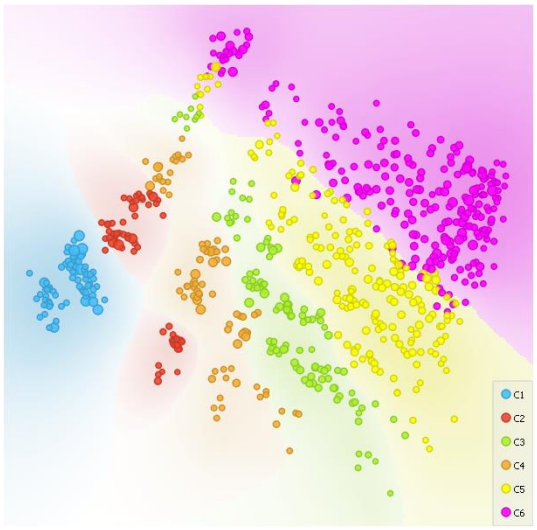


Figura 23 - Agrupamentos (MDS) identificados para os artigos priorizados da componente gigante
Fonte: Própria (2020)

Através desta Figura 23, nota-se as faixas de delimitação dos grupos o qual se destaca o grupo 6, em rosa, com a maior parte dos artigos, e na outra ponta, em azul, já um agrupamento menos e mais específico e mais próximo um do outro. Então segue-se para o processo de identificar os temas de cada um destes agrupamentos através da nuvem de palavras também. Na Tabela 8 abaixo, onde na coluna “Palavras mais frequentes” está respeitando a ordem de maior frequência da esquerda para direita que apareceram na nuvem de palavras:

Tabela 8 - Palavras frequentes de cada *cluster* da componente gigante

Cluster	Palavras mais frequentes
Cluster 1	*business model, conformance checking, machine learning, concept drift, learning strategy, self regulated, regulated learning*
Cluster 2	*constrained trace, machine learning, software development, predictive monitoring, business management, transition system, performance analysis*
Cluster 3	*business management, conformance checking, legacy system, business model, human resource, resource allocation, learning analytics*
Cluster 4	*business management, performance analysis, sequential pattern, trace clustering, visual analytics, risk assessment, electronic health*
Cluster 5	*conformance checking, business management, petri net, predictive monitoring, declarative model, multi perspective, case study*
Cluster 6	*petri net, business management, conformance checking, workflow management, information system, business intelligence, business analysis*

Fonte: Própria (2020)

Ao analisar os textos agrupados no *cluster* 1 percebe-se este grupo tratando de

estudos de casos no desenvolvimento de técnicas de verificação de conformidade que é uma subárea da mineração de processos. No *cluster* 2 monitoramento de processo baseado em predição. O *cluster* 3 trata-se aspectos comportamentais dos recursos humanos nas interações com os processos. O *cluster* 4 trata-se de rastreamento de *registro* e sequência de padrões através da mineração de processos. O *cluster* 5 e 6 estão mais gerais conforme apresentado as 3 primeiras palavras mais agrupadas tratam do mesmo tema.

Neste agrupamento houve maior dificuldade de clusterizar os temas fazendo diversas abordagens e mudanças de parâmetros no processo de mineração de texto. Percebeu-se que para agrupar nos temas acima houve uma melhorar clareza dos temas quando manteve-se apenas o título e a palavras chave como atributos. Quando estava-se analisando também com o resumo termos frequentes de mineração de processos pareciam com frequência nos 4 primeiros lugares como por exemplo, *"event log, information system, conformance checking, business management, petri net".* Mesmo acrescentando estas palavras como *stop words* no pre-processamento a dificuldade mantinha-se. Na Tabela 9 há distribuição das comunidades (com pelo menos de 30 artigos) da componente gigante com relação aos temas abordados por eles.

Tabela 9 - Distribuição dos temas por comunidade da componente gigante

Comunidade	C1	C2	C3	C4	C5	C6	Total
1	6	1	9	9	8	8	41
2	3	7	1	5	9	5	30
3	4	3	6	12	19	19	63
4	5	3	13	5	11	11	48
5	16	2	6	5	10	11	50
7	3	2	6	8	7	6	32
8	10	9	10	9	23	14	75
9	3	5	10	5	13	14	50
13	3	5	2	6	9	9	34
31	9	7	7	3	19	3	48
43	4	6	7	2	20	2	41
101	9	10	31	16	43	143	252
Total	**75**	**60**	**108**	**85**	**191**	**245**	**764**

Fonte: Própria (2020)

Pode-se notar que as comunidades mapeadas tem desenvolvido diversos temas não havendo uma concentração específica e vice-versa. Entendeu-se que o cluster 6 é o maior, logo o mais genérico. Com menos concentração é o cluster 2.

5.1.4.2. Encontrando temas estudados pelas comunidades das redes desconexas

Nas redes desconexas de até 10 vértices da rede de artigos (Figura 20) foram encontradas 7 comunidades, do mesmo modo desenvolido na componente gigante. Na Tabela 10, está apresentada a frequência de artigos em cada comunidade.

Tabela 10 - Quantidade de artigos por agrupamento dos 7 componentes isoladas com mais de 10 vértices

Nº da comunidade	Frequência
45	11
94	13
107	44
109	20
148	20
160	12
173	13
Total	**133**

Fonte: Própria (2020)

Para estes agrupamentos de temas aplica-se as etapas apresentas na Figura 11. Os resultados apontam que a melhor separação foi com agrupamentos em 5 grupos conforme o Dendrograma na Figura 24.

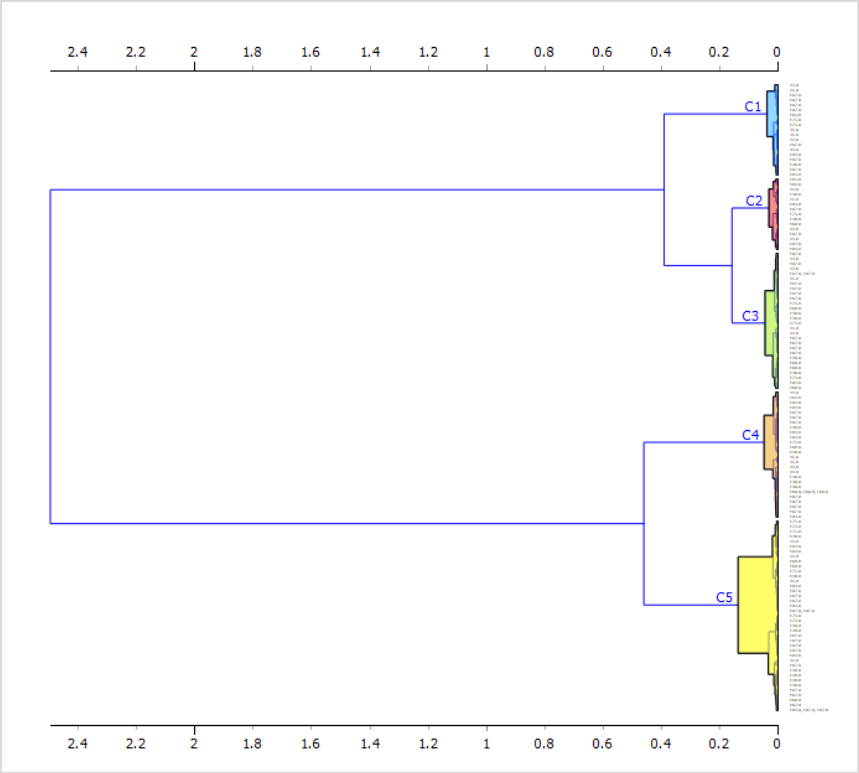


Figura 24 - Dendrograma das classificações dos temas
Fonte: Própria (2020)

Então seguiu-se para uma análise da nuvem de palavras para identificar os temas específicos de cada grupo. Pode-se avaliar temas bem demarcados para os *clusteres* 1, 2, 3 e 4 e sendo o 5 de outros assuntos conforme a Tabela 11 abaixo:

Tabela 11 - Palavras frequentes de cada *cluster*

Cluster	Palavras mais frequentes
Cluster 1	*event log, miner algorithm, alpha miner, information system, information event, business event, case study*
Cluster 2	*social network, network miner, network analysis, event log, work metric, using social, analysis social*
Cluster 3	*fuzzy miner, event log, customer service, miner algorithm, result study, using fuzzy*
Cluster 4	*using framework, provide model, model paper, paper proposes, framework workflow, workflow management, activity using*
Cluster 5	*event log, petri net, business model, alpha algorithm, information system, case study, data data*

Fonte: Própria (2020)

Ao analisar os textos agrupados no *cluster* 1 percebe-se este grupo tratando de estudos de casos no desenvolvimento de melhoria no Algoritmo *Alpha* (este algoritmo

inicial de mineração de processos proposto por Van Der Aalst) e comparação entre o desempenho deles. Já no *cluster* 2 há uma dedicada a mineração de processo focada nos recursos e seu relacionamento na perspectiva organizacional dos processos de negócios. No *cluster* 3 há um direcionamento de aplicações do algoritmo de mineração *Fuzzy* no *registro* de eventos. No *cluster* 4, há artigos focados na previsão de atividades usando o *framework WoMan* para gerenciamento de *workflows*. E por fim, no *cluster* 5 são de assunto mais gerais não convertidos para um tema específico. Logo, há no cluster 2 um pequenos volume de artigos voltados para mineração de processos com perspectiva organizacional que é o foco do estudo de caso posteriormente.

Após este levantamento, criou-se uma matriz de distribuição das comunidades da rede de artigos com relação aos temas abordados por eles. Foram priorizadas as comunidades com pelo menos 8 artigos para esta análise. Conforme a Tabela 12, há diversas comunidades participando no desenvolvimento do tema do *cluster* 1, com maior concentração na comunidade 148. Já no *cluster* 2 há uma concentração maior em 107. Nos *clusteres* 3 e 4 a concentração são nos 107 e 160 respectivamente.

Tabela 12 - Distribuição dos temas por comunidade das redes de artigos desconexas de até 10 vértices

Nº da comunidade	C1	C2	C3	C4	C5	Total
45		1			10	11
94	5				8	13
107	5	6	19		14	44
109	3				17	20
148	6				14	20
160				4	8	12
173					13	13
Total	19	7	19	4	84	133

Fonte: Própria (2020)

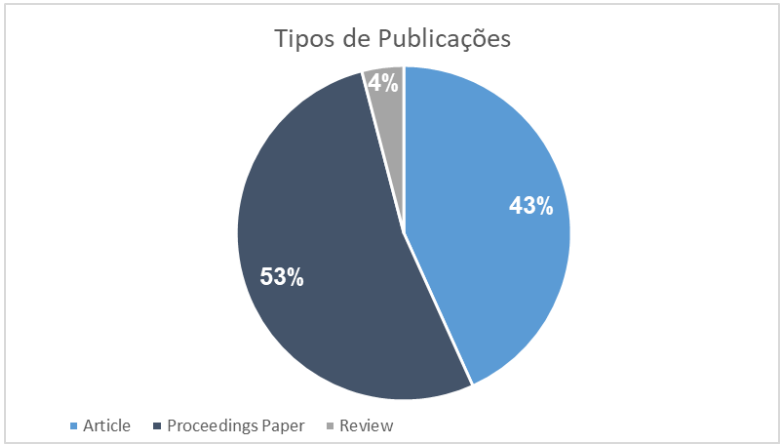
Percebe-se que há uma distribuição maior dos núcleos no *cluste*r 5, que é mais geral e então, entendeu-se que as comunidades não necessariamente tratam de temas específicos em suas pesquisas científicas, mas sim de vários assuntos distintos. Pode-se notar um melhor agrupamento e definição de temas quando a seleção tratou de uma

quantidade menor de artigos (com 133 da rede desconexa) do que com um número maior (764 da componente gigante).

5.2. Busca Específica – Resultados do estudo bibliométrico

Entende-se que, após a análise geral dos artigos publicados na área de mineração de processo, foi preciso reduzir a abrangência da busca. Com base nos 74 artigos selecionados, conforme apresentado na sessão 4.1.2 - Desenvolvimento do Estudo Bibliométrico – Busca Específica, o Gráfico 6 apresenta os percentuais dos tipos de publicações destes artigos.

Gráfico 6 - Tipos de publicações



Fonte: Própria (2020)

Percebe-se que a maioria são artigos apresentados em conferências. E ainda que há 35 artigos em periódicos científicas dos quais 3 são artigos de revisões. Selecionou-se esses 35 artigos e discriminou-se os periódicos os quais foram

publicados. Com este filtro retornaram 28 periódicos científicas diferentes. Ordenou-se pelo número médio de citações dos artigos de periódicos e na Tabela 13 abaixo são apresentados os 10 periódicos mais relevantes.

Tabela 13 - Estratos *Qualis* da busca restrita

Periódicos	Média Citações por artigo
INFORMATION SYSTEMS	323
DECISION SUPPORT SYSTEMS	75
BUSINESS PROCESS MANAGEMENT	38,3
DATA & KNOWLEDGE ENGINEERING	28
INTERNATIONAL JOURNAL OF ONLINE PEDAGOGY AND COURSE	16
COMPUTERS IN HUMAN BEHAVIOR	15
SAFETY SCIENCE	14
COMPUTERS IN BIOLOGY AND MEDICINE	13
ADVANCED ENGINEERING INFORMATICS	13
JOURNAL OF BIOMEDICAL INFORMATICS	12

Fonte: Própria (2020)

Observou-se que os três primeiros periódicos também estão presentes na lista da busca geral. Nesta seleção também há dois periódicos que tratam especificamente de temas na área biomédica, dois nas áreas humanas, 3 na área de engenharia e sistemas de informação outros 2 na área de gestão. Esses quatro temas parecem apresentar a abrangência do assunto para a área de sistemas, engenharia e gestão. Como há um foco em mineração de processos na perspectiva organizacional, que tem como principal elemento o recuso humano, é interessante notar que há um periódico sobre comportamento humano nesta seleção. Outros periódicos não apareceram nesta lista, por não fazerem parte das 10 mais citadas, embora possuam temas de publicações nestas áreas e também em redes sociais e simulações.

Levantou-se também os 10 autores que possuem mais artigos nesta busca restrita na Tabela 14. É possível notar que o autor Premchaiswadi possui dez artigos e seu índice “h” da *Scopus* é igual a 13. Em seguida é o autor Porouhan com 7 artigos e com índice “h” igual 11. Na leitura da base entendeu-se que estes dois primeiros autores são parceiros em suas pesquisas com diversas publicações em conjunto e também notou-se que eles não estão na lista da busca geral (Tabela 6). Já o autor Van Der Aalst está na terceira posição com o índice “h” igual a 107, muito alto em relação a todos os demais. Isto reafirma a importância deste autor para a temática do presente trabalho e que há seis artigos de sua autoria focado também na perspectiva organizacional. Assim como os autores Song, Mendling, Munoz-Gama e Sepulveda que também aparecem na

lista dos principais autores de mineração de processo da Tabela 6. E os autores Jung, Rojas e Appice também possuem estudos que podem auxiliar no aprofundamento dessa pesquisa.

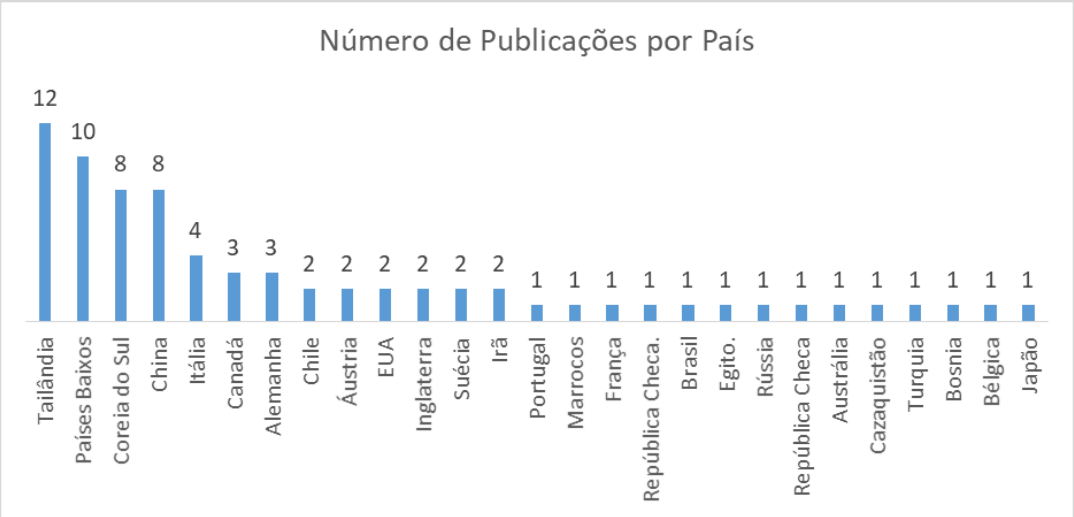
Tabela 14 - Autores com maior número de artigos publicados e seu índice "h" da *Scopus*

Autor	Quant. Artigos	Índice h *Scopus*
Premchaiswadi, W	10	13
Porouhan, P	7	11
Van Der Aalst, WMP	6	107
Jung, JY	5	77
Song, M	5	64
Rojas, E	3	47
Mendling, J	2	53
Appice, A	2	20
Munoz-Gama, J	2	17
Sepulveda, M	2	14

Fonte: Própria (2020)

Levantou-se também o número de publicações por países desta busca restrita, conforme apresentado no Gráfico 7. Verificou-se que o maior número são de autores com vínculos em instituições da Tailândia, o que parece indicar que há um núcleo de pesquisa no assunto de mineração de processos sobre a perspectiva organizacional. Comparando este resultado com o Gráfico 3 da busca geral, é possível notar que os três primeiros países daquele gráfico, Países Baixos, China e Itáia permanecem entre os cinco principais países desta busca restrita. Entende-se então que há um fomento nestes locais para o assunto não apenas geral sobre mineração de processo mas também na perspectiva organizacional. O Brasil também aparece no Gráfico 7 - Número de Publicações por País com apenas uma publicação, concluindo-se assim que este assunto da mineração de processo na perspectiva organizacional ainda é pouco explorado em nosso país. Logo, isto é um bom indicador que o assunto da presente pesquisa é relevante para ampliar as pesquisas e estudos de casos aqui no Brasil.

Gráfico 7 - Número de Publicações por País



Fonte: Própria (2020)

Para a seleção dos trabalhos mais relevantes levou-se em consideração o número de citações dos artigos e o número de vezes que o artigo retornou naquelas consultas apresentadas na

Tabela 15. Este último justifica-se pelo fato de quanto mais vezes o artigo retornou em algumas daquelas strings de busca mais próximo do assunto este artigo encontra-se. Os artigos com mais de 100 citações estão listados na tabela abaixo:

Tabela 15 - Artigos Relevantes pelo Número de Citações

Artigos Relevantes pelo número de citações	Nº de Citações	Quant vezes retornou
Título: Business process mining: An industrial application **Autor**: Van Der Aalst, WMP; Reijers, HA; Weijters, AJMM; van Dongen, BF; de Medeiros, AKA; Song, M; Verbeek, HMW **Tipo**: Artigo de Revista (estudo de caso) **Publicado em**: *INFORMATION SYSTEMS* - **Ano**: 2007	323	1
Título*: Application of Process Mining in Healthcare - A Case Study in a Dutch Hospital* **Autor**: Mans, RS; Schonenberg, MH; Song, A; Van Der Aalst, WMP; Bakker, PJM **Tipo**: Artigo de Anais de Congresso **Publicado em**: *BIOMEDICAL ENGINEERING SYSTEMS AND TECHNOLOGIES* - **Ano**: 2008	138	1
Título: *Towards comprehensive support for organizational mining* **Autor**: Song, M; Van Der Aalst, WMP **Tipo**: Artigo de Revista **Publicado em**: *DECISION SUPPORT SYSTEMS* - **Ano**: 2008	134	4

Título: *Mining social networks: Uncovering interaction patterns in business processes* **Autor**: Van Der Aalst, WMP; Song, M **Tipo**: Artigo de Revista **Publicado em**: *BUSINESS PROCESS MANAGEMENT* **Ano**: 2004	109	3

Fonte: Própria (2020)

Notou-se que em todos os artigos desta listagem há o autor Van der Aalst e não há publicações de Premchaiswadi e Porouhan (primeiros autores da Tabela 6). Os dois primeiros artigos mais citados apareceu apenas uma vez em uma das consultas. Estes são artigos de aplicação geral de mineração de processos, são estudos de caso, um aplicado em uma indústria e outro aplicado a área da saúde. Nele há prática e interpretação dos resultados das técnicas de mineração de processo dos diferentes tipos e perspectivas. Já os dois últimos, retornaram mais de uma vez nas consultas e são completamente voltados ao assunto de mineração de processo na persperctiva organizacional. O artigo "*Mining social networks: Uncovering interaction patterns in business processes*" é a primeira publicação nesse assunto e é abordado pelos principais autores mapeados no levantamento bibliométrico. Nele é apresentado as justificativas e as primeiras técnicas de como se utilizar os *registros* de eventos para mapear a iteração entre os recursos humanos nos processos. É importante notar que, apesar do grande número de citação de cada artigo, esses dois últimos foram publicados entre 2004 e 2008. Também foram selecionados como relevantes, os artigos que retornaram mais vezes nas *string* da busca restrita e estes estão apresentados na Tabela 16.

Tabela 16 - Artigos Relevantes pelo número de vezes que apareceu na busca

Artigos Relevantes pelo número de vezes que apareceu na busca	Qtd vezes retornou	Nº de Citações
Título: *Process Mining from the Organizational Perspective* **Autor**: Zhao, WD; Zhao, XD **Tipo**: Artigos de Anais de Congresso **Publicado em**: *FOUNDATIONS OF INTELLIGENT SYSTEMS* (ISKE 2013) - **Ano**: 2014	6	15
Título: *OrgMiner: A Framework for Discovering User-Related Process Intelligence from Event Logs* **Autor**: Deokar, AV; Tao, J **Tipo**: Artigo de Revista **Publicado em***: INFORMATION SYSTEMS FRONTIERS* - **Ano**:2020	4	0

Título: *Discovering and Tracking Organizational Structures in Event Logs* **Autor**: Appice, A; Di Pietro, M; Greco, C; Malerba, D **Tipo**: Artigo de Anais de Congresso **Publicado em***: NEW FRONTIERS IN MINING COMPLEX PATTERNS* **Ano**:2016	4	2
Título: *Discovering User Communities in Large Event Logs* **Autor**: Ferreira, DR; Alves, C **Tipo**: Artigo de Anais de Congresso **Publicado em**: *BUSINESS PROCESS MANAGEMENT WORKSHOPS* - **Ano**: 2012	4	5

Fonte: Própria (2020)

Através da última coluna desta tabela percebe-se que estes artigos foram menos citados que a busca anterior. Porém, o período é mais recente e está entre 2012 e 2020. Pode-se perceber também que a maioria foram publicados em congressos internacionais. O primeiro é uma referência importante sobre um levantamento de pesquisas e oportunidades de avanços da mineração de processos sobre a perspectiva organizacional. O segundo é uma proposta de *framework* sobre mineração organizacional. O terceiro e o quarto são propostas de melhorias técnicas sobre redes sociais aplicadas à mineração de processos sob a perspectiva organizacional.

5.3. Busca Expandida – Resultados do estudo bibliométrico

Para as duas primeiras buscas expandidas encontrou-se mais 18 referências, onde 4 são do ano de 2020 e outras 14 são do ano de 2021. Destes, 12 são artigos de periódicos e os outros são de anais de congressos. Foi observado apenas o autor Wynn, M.T. que foi destaque nos levantamentos bibliométricos anteriores.

Da seleção de artigos enviados pelo orientador identificou-se alguns destes como pré-selecionados da listagem do subitem 4.2. Identificou-se um artigo que não havia sido encontrado em nenhuma pesquisa anterior, sendo ele dos autores, Van Der Aalst; Reijers; Song de 2005 . Ele tem o título "*Discovering Social Networks from Event Logs"*. Este artigo é muito importante pois é um estudo de caso de um instituição holandesa, no qual utiliza-se das técnicas de mineração de processos sob a perspectiva organizacional descrista no artigo "*Mining social networks: Uncovering interaction patterns in business processes"* (selecionado como relevante no item 4.2) para compor

o conjunto de artigos. Outro artigo selecionado, também da coleção do orientador foi *"How is process mining technology used by organizations? A systematic literature review of empirical studies".* Esse artigo traz uma revisão empírica de 2008 desenvolvida pelos autores Thiede, D. F. e Barquet, Ana Paula B. Por fim, mais um artigo selecionado através da técnica *slowballing* foi o "*OrgMining 2.0: A Novel Framework for Organizational Model Mining from Event Logs*" de 2020 dos autores Yang, Jing; Ouyang, Chun ; Van Der Aalst, Wil M. P; ter Hofstede, Arthur H. M ; Yu, Yang de 2020 com uma nova proposta de descoberta de estrutura organizacional através de *registros* de eventos.

5.4. Discussão do estudo bibliométrico

Obteve-se um panorama sobre as contribuições do conhecimento científico derivado das publicações afim de responder a primeira pergunta de pesquisa utilizando-se de estudos bibliométricos, redes sociais e técnicas de mineração de texto. Dividiu-se essas análises em busca geral, busca específica e busca expandida sobre mineração de processos. Para entender a MP em termos gerais coletou-se 1.969 artigos científicos através de pesquisas feitas na WOS e identificou-se que há um crescimento tanto em termos de publicações quanto de novos autores estudando sobre este tema. E que a maioria dos documentos gerados são em anais de conferências e eventos. No ano de 2020 houve uma expressiva queda de artigos e acredita-se que foi reflexo devido a pandemia, o qual teve um impacto mundial, onde prevaleceram fechadas muitas instituições e eventos não foram realizados. Os principais paises núcleos do desenvolvimento neste tema são: a região dos Países Baixo, seguidos da China e Itália. No Brasil existem estudos sobre MP e está entre os 21 paises com maiores publicações. Aprofundou-se neste levantamento bibliométrico, utilizando a análise de rede social e verificou-se na rede de cooautoria (rede de autores) há características típicas de uma forte comunidade científica para os estudos de MP. E através das métricas de centralidade (grau, proximidade e intermediação) observou-se que o autor Van Der Aalst é o pesquisador mais influente seguidos de seus parceiros Song, M e Wen, LJ. Este resultado já era esperado pois trata-se do "pai" de todo o desenvolvimento deste assunto. Durante o levantamento dos nomes dos autores foi essencial a sua padronização no pré-processamento para evitar ambiguidade e retrabalho ao longo das etapas seguintes, pois percebeu-se que o mesmo autor tinham várias publicações com formatos de nomes distintos. Identificou-se também, na rede artigos, que a publicação

“*Process Mining Manifesto*” é uma referência com 77 autores a qual é um guia de MP tanto para profissionais acadêmicos quanto de negócio. Ainda nesta rede de artigos pode-se segregar as redes em uma rede pertencente a componente gigante (conexa) e em outras as redes desconexas. Nessas redes aplicou-se técnicas de mineração de textos agrupando os artigos por temas. Porém, por mais que tenha sido detectado comunidades através da ARS os temas abordados por elas foram bem mesclados não havendo especificidades de temas por cada comunidade.

No estudo bibliométrico focado na busca específica de mineração sob a perspectiva organizacional reduziu-se para 74 artigos e percebeu-se uma grande similaridade com a busca geral sobre as principais local de publicação, categorias, periódicos e países de publicação assim como alguns autores. Foi possível identificar novos autores não destacados na busca geral, porém, Van der Aalst é referência e possui os artigos com maiores citações e pioneiros em mineração organizacional. Importante ressaltar que ficou claro que mineração organizacional é menos explorado, principalmente em nosso país com apenas 1 publicação. Entende-se que éum indicativos que esta presente pesquisa é relevante para ampliar os estudos aqui no Brasil. Entendeu-se que foi muito importante para aprofundar numa coleção de artigos fundamentais de mineração organizacioanl, assim como, conhecer os principais autores, países de fomento, os assuntos que estão sendo abordados e também os principais períodicos.

A busca expandida contribuiu com artigos recentes contemplando a mineração de processos na era pandêmica e reuniu também novas pesquisas que não haviam sido captadas pelas buscas anteriores. Foram agregados 18 novas publicações importantes em uma pesquisa atualizada até 2021. Percebeu-se que este processo de levantamento de estudos científicos é longo e um grande desafio é delimitar o periodo de tempo para a busca. Entendeu-se que com essas três buscas obteve-se um conjunto fundamental sobre o estado da arte que serviu como base para o desenvolvimento desta dissertação e a primeira pergunta de pesquisa foi respondida.

6- Estudo de Caso da Análise dos Processos antes e depois da pandemia

Neste capítulo, será apresentada a abordagem para este estudo de caso envolvendo todas as fases do framework PM^2. Vale destacar que a análise dos processos administrativos acontece em três iterações: a primeira é o entendimento dos processos gerais em todo os períodos de tempo; a segunda, é a análise dos processos no período de tempo antes da pandemia e, por fim a descoberta e análise das redes sociais antes e depois da pandemia.

6.1. A instituição do estudo de caso: CEFET/RJ

Este estudo visa analisar os processos de solicitação de Protocolo Geral do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ), localizado na cidade do Rio de Janeiro, mais especificamente na unidade Maracanã. A qual é a sede de outras unidades do estado do Rio de janeiro. Esta é uma instituição federal de ensino que oferece vários cursos técnicos, no nível médio, subsequentes (pós-médio), tecnológicos, graduação e pós-graduação, tanto *lato sensu* e quanto *stricto sensu* (mestrado e doutorado), nas modalidades presencial e a distância.

Os processos de negócios desta instituição têm duas categorias de processos: A primeira representa as práticas acadêmicas que são fiscalizadas por sistemas de informação, e a segunda categoria retrata os procedimentos administrativos regulados por um arcabouço jurídico rigoroso que visa à prestação de contas. Todos os processos desta segunda categoria são executados em formato tradicional de papel, e o Sistema Eletrônico de Protocolo Geral (SEPG) registra a identificação, descrição, proprietário, datas, movimentações, localização, unidade organizacional em qualquer momento da vida do processo.

No SEPG há o registro de eventos que é possível coletar um conjunto de dados (identificação, descrição, proprietário, datas, movimentações, localização, unidade organizacional) que são registrados no contexto dos processos. O Setor de Protocolo é o responsável pelas atividades de recebimento e expedição de processos internos e externos. O número anual de processos é de aproximadamente 3,4 mil. E incluiu

atendimento a outras unidades, para além de cidadãos, alunos, professores e funcionários.

6.2. Planejamento Estudo de Caso

O planejamento é a primeira fase aplicada conforme a Figura 12. Esta sessão descreve os objetivos e a organização deste estudo de caso que consistirá em:

- **Objetivos**: As questões iniciais da pesquisa a serem respondidas e objetivos a serem atingidos estão definidos em 1.2 Perguntas de Pesquisa e Objetivos. E os objetivos específicos também devem ser alcançados pelos passos seguintes do PM^2.
- **Extração**: Como o foco deste estudo é a mineração organizacional, a base de dados a ser extraída (conforme fase 2 da Figura 12) deve conter, minimamente, os *registros* de eventos do Sistema de Protocolo Geral com as seguintes informações: a data do evento, o nome e número da instância de processo e o recurso responsável.
- **Iterações de Análises**: Cada iteração passa nas Fases 3, 4 e 5 da Figura 12 e cada uma dessas iterações objetiva responder um objetivo específico desta pesquisa, logo, são três iterações que são: na primeira, processamento, análise e avaliação dos principais processos da base geral e seus comportamentos; na segunda, processamento, análise e avaliação dos principais processos do recorte da base nos períodos antes e depois da pandemia e seus comportamentos; na terceira, processamento, mineração organizacional com foco na descoberta das redes sociais e avaliação dos processos selecionados (baseados nas análises anteriores) .
- **Melhoria e Suporte do Processo**: o cenário deste estudo não é extensivo a implantar as melhorias. E posteriormente, servirá como apoio para a instituição avaliar e evoluir seus processos.

6.3. Extração dos dados

As informações são extraídas do Sistema de Protocolo Geral do período entre 2013 a junho de 2021 em dois arquivos de Excel distintos.

O primeiro arquivo (Arquivo 1), Figura 25, possui as informações gerais dos processos. Nele contém 30.871 instâncias de processos. Os atributos disponíveis são:

Número (código sequencial do processo), Processo (Número Interno do Protocolo que é entregue para o solicitante acompanhar o andamento do processo e está oculto na para proteger as informações), Texto 1 (Informação Complementar da Solicitação), Texto 2 (Nome do Processo), Interessado (refere-se ao nome do solicitante e este foi anonimizado por “Interessado 1”, “Interessado 2” e *etc,* quando preenchido, para proteger as informações) e *Status.* Porém, estas duas últimas estão com a maioria das linhas sem informação. Ou seja, nesse arquivo tem todas as instâncias de processos e seus respectivos nomes a que se destina a solicitação.

COD	TXT1	TXT2	INTERESSADO	STATUS
4133		PROGRESSAO FUNCIONAL	Interessado 1	ABERTO
4134		AFASTAMENTO DO PAÍS	Interessado 2	ABERTO
4135		PROGRESSÃO FUNCIONAL POR AVALIAÇÃO	Interessado 3	ABERTO
4136		Auxílio saúde indenizatório	Interessado 4	ABERTO
4137		Aposentadoria		
4138		Convênio		
4139		Convênio		
4140		Convênio		
4141		Convênio		
4145		Prog. Funcional		
4146		Prog. Funcional		
4147		Prog. Funcional		
4148		Prog. Funcional		
4149		Prog. Funcional		
4150		Prog. Funcional		
4151		Prog. Funcional		
4152		Auxílios		
4153		Aux. Nat. e e aux. pré-escolar		

Figura 25 - Primeiro Arquivo das Instâncias dos Processos
Fonte: Própria (2021)

O segundo arquivo (Arquivo 2), Figura 26, possui as informações dos *registros* eventos dos processos. Nele contém 193.227 registros de eventos. Os atributos disponíveis são: Data (data do evento), Seção (refere-se aos recursos, neste caso o departamento de atendimento), Número (código sequencial do processo), Posição (número sequencial por processo), De (departamento de origem) e Para (departamento de destino). Ou seja, aqui o número do processo (que é o mesmo que a coluna “COD” no Arquivo 1) neste arquivo se replica, pois para cada evento é acrescido um novo registro.

DATA	SECAO	NUMERO	POSICAO	DE	PARA
17/01/2013	DGP	37	1	DGP	PROTOCOLO GERAL
17/01/2013	PROTOCOLO GERAL	37	2	PROTOCOLO GERAL	DGP/DICAD
22/01/2013	DGP/DICAD	37	3	DGP/DICAD	DGP
07/03/2013	DGP	37	4	DGP	DGP
12/03/2013	DGP	37	5	DGP	DGP/DIPAG
28/06/2013	DGP/DIPAG	37	6	DGP/DIPAG	DGP
17/01/2013	CPPD	38	1	CPPD	PROTOCOLO GERAL
17/01/2013	PROTOCOLO GERAL	38	2	PROTOCOLO GERAL	DIREG
15/02/2013	DIREG	38	3	DIREG	DGP/DILEN
20/02/2013	DGP/DILEN	38	4	DGP/DILEN	DGP/DILEN
21/02/2013	DGP/DILEN	38	5	DGP/DILEN	DGP
26/06/2013	DGP	38	6	DGP	COORD. CURSO SUP. ADM. - CSADM
26/06/2013	COORD. CURSO SUP. ADM. - CSADM	38	7	COORD. CURSO SUP. ADM. - CSADM	PROJU
12/07/2013	PROJU	38	8	PROJU	DIREG
18/07/2013	DIREG	38	9	DIREG	DIREG
29/10/2013	DIREG	38	10	DIREG	DGP/DICAD
19/11/2013	DGP/DICAD	38	11	DGP/DICAD	DGP/DIPAG
17/01/2013	PROTOCOLO GERAL	39	1	PROTOCOLO GERAL	DERAC
17/01/2013	PROTOCOLO GERAL	40	1	PROTOCOLO GERAL	COORD. ENG. ELÉTRICA - CEELE
17/01/2013	COORD. ENG. ELÉTRICA - CEELE	40	2	COORD. ENG. ELÉTRICA - CEELE	COORD. ENG. ELÉTRICA - CEELE

Figura 26 - Segundo Arquivo das Informações dos Processos
Fonte: Própria (2021)

Com estes dados brutos oriundos do sistema de Protocolo Geral é possível verificar a qualidade dos dados de cada coluna e identificar aqueles dados que estão inconsistentes, faltantes ou precisão ser enriquecidos ou agregados para serem tratados na fase seguinte de processamento de dados.

6.4. Primeira iteração de Análises

Nesta primeira iteração de análises pretende-se responder o objetivo específico P2.1 apresentado em 1.2 Perguntas de Pesquisa e Objetivos, que é: Identificar e analisar os processos principais da base geral e seus comportamentos através da frequência dos processos e duração de tempo. Esta iteração se faz necessária para conhecimento geral da base e entender quais os limites de recorte da base para o período antes da pandemia e depois da pandemia. Nela serão utilizados os *registros* de eventos no período de 2013 a junho de 2021.

Conforme o framework PM^2, Figura 12, as iterações possuem três fases que são: processamento, mineração e análises e avaliação. E em cada uma há diferentes elementos que podem ser usados a depender do objetivo da iteração. Todos estes especificados no capítulo 4.2 - Desenvolvimento do Estudo de Caso. Sendo assim, na fase de processamento aplica-se os elementos de agregação e a filtragem. Na fase de mineração e análises implementa-se o elemento de análise de processos para criar uma análise visual. E por último, avalia os resultados para relacionar os achados da análise

e, então responder o objetivo específico P2.1.

6.4.1. Processamento de dados

Ao verificar os dados de cada coluna dos arquivos extraído identificou-se a falta de padronização, principalmente, da coluna “Texto 2” do primeiro arquivo. Para tratar esta despadronização implementa-se o elemento de agregação. E também algumas inconsistências na coluna “Data” no segundo arquivo. Para tratar estas inconsistências implementa-se o elemento de filtragem. A Tabela 17 apresenta alguns exemplos desta falta de padronização referente a coluna “Texto 2 “, na primeira coluna. E na segunda coluna apresenta-se um possível nome padronizado do processo.

Tabela 17 - Ambiguidade no Nome dos Processos

Exemplos do Nome do Processo no “Texto 2”	Processo Referente ao processo
SOLICITAÇÃO DE RSC, RSC, SOLICITAÇÃO DO RSC, solicitação de rsc, SOLICITAÇÃO RSC, RECONHECIMENTO DE SABERES E COMPETÊNCIAS(RSC), Reconhecimento de saberes, Reconhecimento de Saberes e Competências	Reconhecimento de Saberes e Competências (RSC)
CONVENIO, Convênio, Covênio, Convênio - C. Educ. Realengo, Convênio - Campanhia, Convênio - Est -1 elem. sol., Convênio - Hotel Urbano, Convênio - Multisuprimentos	Convênio
COMPRA DE MATERIAL, COMPRA DE EQUIPAMENTO, AQUISIÇÃO DE MATERIAIS, AQUISIÇÃO DE BENS MATERIAIS, Aquisição de Equipamento, AQUISIÇÃO DE EQUIPAMENTOS,	Compra de Material e equipamentos
ADIANTAMENTO DA METADE DO 13º SALÁRIO, adiantamento do 13 salário, Adiantamento do 13º salário, ANTECIPAÇÃO DO DÉCIMO TERCEIRO	Décimo terceiro salário

Fonte: Própria (2021)

Com estes dados brutos, os números de processos distintos eram no total de 6.470. Isso se deve à falta de padronização na abertura do processo no sistema de informação do protocolo geral. Foi-se necessário tratar esses nomes dos processos para então agrupá-los. Este agrupamento, segundo Bose; Van der Aast (2009), reduz a complexidade, melhora a estrutura dos resultados da mineração e análises. Portanto,

este procedimento unifica os processos que são os mesmos, porém estavam descritos com nomes diferentes.

Primeiramente, tratou-se o atributo “Texto 2”, pois um mesmo processo contém mais de um nome pela falta de padronização, conforme alguns exemplos apresentados na Tabela 17. Para saneamento da base os dados foram submetidos às técnicas de mineração de texto. As 4 fases deste processamento estão apresentadas na Figura 27 a seguir:

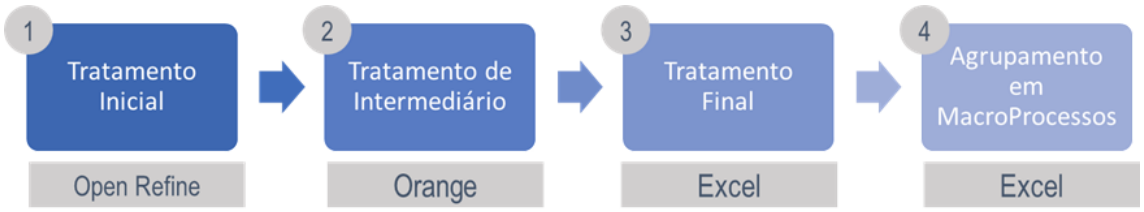


Figura 27 – Fases para Agrupamento dos Processos
Fonte: Própria (2021)

Na fase 1, para o tratamento inicial e uma limpeza mais geral, utilizou-se o *Open Refine*. Este é um *software* gratuito do *Google*. Ele é *desktop* e de código aberto destinado a limpeza e transformação de dados. Foi utilizada sua versão 3.4.1, disponibilizada em 24 de setembro de 2020. Após instalar este *software,* foi importado o Arquivo 1 e migrado para as transformações apenas o atributo “Texto 2”. Como este dado trata-se de texto (*strings*) este então foi configurado para o tipo “*Facet*” para que o *OpenRefine* possa disponibilizar os algoritmos de tratamento e agrupamento de textos, em consonância com OPENREFINE (2021). Feito essa configuração foi possível os métodos de agrupamento de texto. Os métodos disponíveis são: colisão de chaves (*key collision*) e vizinho mais próximo (*nearest neighbor*). A colisão de chaves possui os algoritmos *fingerprint*, *ngram-fingerprint*, *cologne-phonetic, Daitch-Mokotoff e Beider-Morse*. E para o segundo método estão disponíveis os seguintes algoritmos: *levenshtein* e *ppm,* como propõe OPENREFINE (2021).

Todos estes algoritmos foram exaustivamente processados, a fim de encontrar os agrupamentos necessários. Pode-se dizer que esta fase de pré-processamento é semi-automática, pois, o aplicativo, em cada método ao ser processado apresenta o agrupamento encontrado e então, é preciso escolher quais deles são válidos. Sendo assim, é preciso uma intervenção da autora para finalizar o processo.

Na primeira rodada do método colisão de chaves no processamento dos algoritmos *fingerprint* foram encontrados 694 agrupamentos. É possível verificar parte

deste tratam na Figura 28. Foram dez grupos (*clusters*) com um total de 649 linhas, para as quais sugere-se agrupá-las e renomeá-las todas como “INCENTIVO A QUALIFICAÇAO”. E também, outro agrupamento com 10 agrupamentos envolvendo 42 linhas e com a sugestão de renomeá-lo como “Auxílio Saúde Suplementar”.



Figura 28 - Tratamento no *OpenRefine* com métodos *fingerprint*
Fonte: Própria (2021)

A partir de então, estes foram escolhidos para que sejam em linhas agrupadas. O *Open Refine,* então, renomeia o atributo “Texto 2” para o novo nome sugerido em todas as linhas encontradas. Quando não há mais agrupamentos encontrados ele apresenta uma mensagem informando que nenhum *cluster* foi encontrado para o método selecionado. Houve algum agrupamento sugerido em que, pela descrição do processo, não foi aceito o agrupamento. Um exemplo para este caso foi ao utilizar o método *ngram-fingerprint* com o tamanho de *Ngram* igual a 4. Ele agrupou 42 linhas de processos com descrição bem diferentes conforme a Figura 29 abaixo.



Figura 29 - Tratamento no *OpenRefine* com métodos *ngram-fingerprint*
Fonte: Própria (2021)

Após passar pelos dois métodos do *Open Refine* e seus devidos funções, foram aceitos aqueles agrupamentos com melhor correspondência, então as linhas foram renomeadas. Para aqueles que não havia uma correspondência ideal, não houve mudança nas linhas. Assim, seguiu-se para a fase de Tratamento Intermediário (fase 2, conforme Figura 27).

Na fase 2, foi feita uma análise de agrupamento de processos através de mineração de texto no *Software Orange*. O *Orange* é um programa de código aberto, gratuito e está disponível para o *desktop*. Foi utilizada a versão *Orange* 3, versão 3.23.1. Ele possui vários *plugins* de fácil manipulação e visual para o usuário. Estes *plugins* são tanto para visualização de dados quanto algoritmos de aprendizado de máquina. Pela sua facilidade de uso, este foi escolhido para aplicar um agrupamento hierárquico passando por todas as fases de mineração de texto sobre o atributo “Texto 2”. A Figura 30 apresenta o processo para agrupamento.

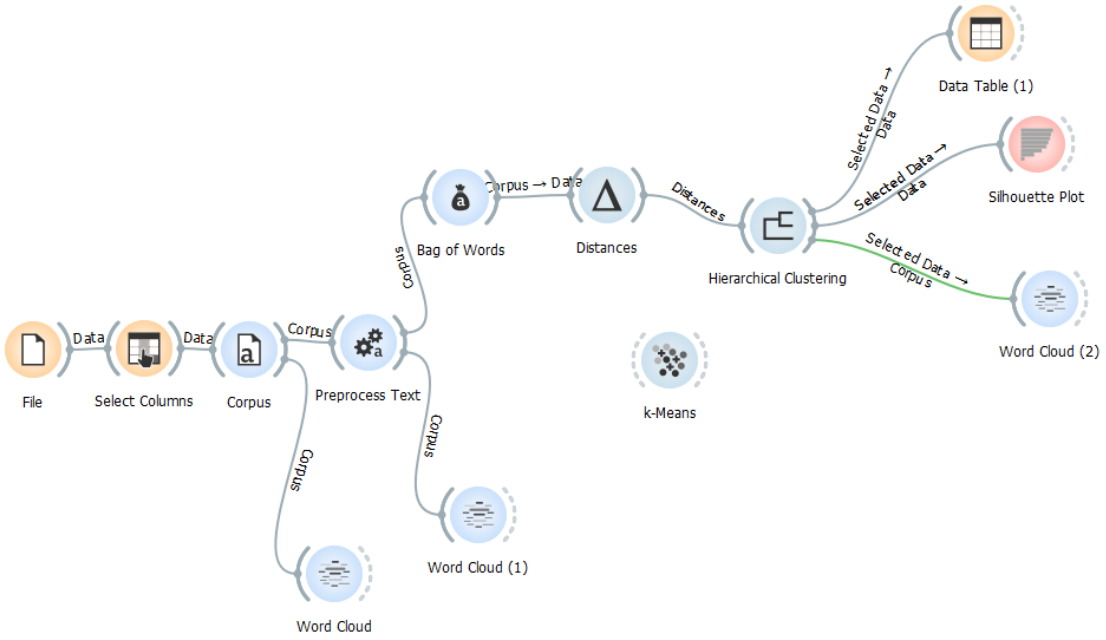


Figura 30 - Fases no *Orange* para agrupar os Processos
Fonte: Própria (2021)

Ao final deste processo foram agrupados linhas que ainda não tinham sido tratadas, conforme os exemplos na Tabela 18, Solicitação de Retribuição por Titulação ficou no mesmo grupo de Retribuição por Titulação. Assim como, Averbação de Tempo Especial ficou no grupo de Averbação de Tempo de Serviço. Analisando esses agrupamentos propostos foi possível tratar mais o atributo “Texto 2”. Porém, diferente do *OpenRefine,* o *Orange* não altera as linhas e isto teve que ser feito manualmente.

Ao tratar manualmente esses dados no *Excel*, percebeu-se outros tratamentos mais refinados que não foi possível reconhecer por nenhuma das duas técnicas. Então, partiu-se para a fase 3, do pré-processamento, que foi totalmente manual. Pode-se citar o exemplo do processo nomeado no “Texto 2” como “Reconhecimento de Saberes e Competências” e este ficou isolado do grupo RSC. Porém, RSC é apenas a abreviação. Assim como, um processo com o nome de “Adiantamento do 13º salário” e outro com “ANTECIPAÇÃO DO DÉCIMO TERCEIRO”.

Após estes tratamentos, executou-se a fase 4 que requer o agrupamento dos processos afins. Este processo está referenciado em Bose; Aalst (2009) e Van Eck *et al.* (2015), os quais defendem que este procedimento é uma agregação “is-a” onde diferentes tipos de instância de processos pertencentes a uma classe de processos equivalente, porém mais geral, enquanto o número de eventos e instância de processos permanecem o mesmo. Por exemplo, dois processos rotulados com “COLAÇÃO DE

GRAU” “COLAÇÃO DE GRAU ESPECIAL DE URGÊNCIA” são considerados instâncias da classe de “COLAÇÃO DE GRAU”, mas permanecem como dois processos. A partir daqui estes agrupamentos de processos é denominado como “macroprocessos”.

A Tabela 18 apresenta alguns exemplos de macroprocessos na primeira coluna. E na segunda coluna estão os nomes tratados (após o tratamento completo nas 3 fases de processamento) do atributo “Texto 2”:

Tabela 18 - Exemplos de macroprocessos padronizados após a limpeza e agrupamento dos nomes dos processos despadronizados.

Macroprocessos	Atributo “Texto 2” tratado
CONVÊNIO	CONVÊNIO; TERMO DE CONVÊNIO/ ESTÁGIO SUPERVISIONADO; CONVÊNIO/ESTÁGIO; MINUTA DE CONVÊNIO; CONVENIO; PROPOSTA DE MINUTA DE CONVÊNIO DE ESTÁGIO; JUSTIFICATIVA PARA ASSINATURA DE CONVÊNIO; CONVÊNIO COM O CREA; RENOVAÇÃO DO CONVÊNIO CBPF; SOLICITAÇÃO DE ENCAMINHAMENTO DE CONVÊNIO; SOLICITAÇÃO DE ASSINATURA DE CONVÊNIO; TERMO ADITIVO AO CONVÊNIO; ANÁLISE DE CONVÊNIO; TERMO DE CONVÊNIO;EDITAL DE CONVÊNIO - SANTANDER; ADITIVO AO CONVÊNIO;EXTENSÃO DO PRAZO DE CONVÊNIOS
ESTÁGIO PROBATÓRIO	ESTÁGIO PROBATÓRIO; AVALIAÇÃO DE ESTÁGIO PROBATÓRIO; HOMOLOGAÇÃO DE ESTÁGIO PROBATÓRIO; PRIMEIRA AVALIAÇÃO NO ESTÁGIO PROBATÓRIO; SEGUNDA AVALIAÇÃO NO ESTÁGIO PROBATÓRIO; SOLICITAÇÃO DE DOCUMENTO COMPROBATÓRIO; SOLICITAÇÃO DE DECLARAÇÃO COMPROBATÓRIA; CONCLUSÃO DE ESTÁGIO PROBATÓRIO; ATA DO ESTÁGIO PROBATÓRIO; DECLARAÇÃO COMPROBATÓRIA
TEMPO DE SERVIÇO	AVERBAÇÃO DE TEMPO DE SERVIÇO; CONTAGEM DE TEMPO DE SERVIÇO; CERTIDÃO DE TEMPO DE SERVIÇO; DECLARAÇÃO DE TEMPO DE CONTRIBUIÇÃO; AVERBAÇÃO DE TEMPO ESPECIAL; CERTIDÃO DE TEMPO EXCEDENTE; DESAVERBAÇÃO DE TEMPO EXCEDENTE; EMISSÃO DE CERTIDÃO DE TEMPO DE CONTRIBUIÇÃO; SOLICITAÇÃO DECLARAÇÃO TEMPO SERVIÇO; CERTIDÃO DO TEMPO NO CEFET -RJ; CONTAGE DE TEMPO
COLAÇÃO DE GRAU	COLAÇÃO DE GRAU; COLAÇÃO DE GRAU ESPECIAL DE URGÊNCIA; ANTECIPAÇÃO DE COLAÇÃO DE GRAU; COLAÇÃO DE GRAU / DIPLOMA; ADIANTAMENTO DE COLAÇAO

Fonte: Própria (2021)

Após todas essas fases o número inicial de nomenclaturas distintas para os processos que eram de 6.470, passou a ser 1.762. Isso corresponde a uma redução de aproximadamente de 73% das variações nominais iniciais. Com esses macroprocessos é possível verificar a distribuição de suas ocorrências, a duração média e a distribuição

do número de etapas de cada um.

Nesta fase de processamento, o segundo elemento utilizado é a filtragem. A coluna de “Data” do Arquivo 2 foi tratada por inconsistência de datas, conforme apresentados alguns exemplos na Tabela 19. Foram eliminados 63 linhas que continham esses erros, pois não havia uma forma de recuperá-los. Isso corresponde apenas a 0,03% do total de linhas do arquivo 2. A não retirada dessas datas na base impactaria no cálculo do tempo dos processos.

Tabela 19 - Exemplo de inconsistência de datas

18/03/3013; 18/03/3013; 28/2/0138; 10/04/2107; 18/03/3013; 28/2/0138; 07/08/2012 26/09/2113; 8/8/1013; 0/04/2107; 26/2/0142; 27/6/0214; 4/12/0014; 30/11/0201; 4/12/0014; 5/2/0152

Fonte: Própria (2021)

Importante destacar que este processamento da primeira iteração é a base para todas as outras iterações, pois nela foram preparados os macroprocessos e o tratamento das datas inconsistentes.

6.4.2. Mineração e Análises

Esta fase de Mineração e Análises na primeira iteração implementa-se o elemento de análise de processos e utiliza os *registros* de eventos no período de 2013 a junho de 2021. Ela é focada numa análise exploratória inicial da base completa e para responder o objetivo específico P2.1 que é voltado para uma análise visual. Este elemento é referenciado no framework PM^2 e amplamente utilizado na literatura CHANG *et al.* (2020), Leoni; Van der Aalst; Dees (2014) e Van der Aalst. *et al.* (2007). Esta iteração se faz necessária também para conhecer e entender quais os limites de recorte da base para o período antes da pandemia e depois da pandemia.

Nesta fase utilizou-se o Arquivo 2 dos *registros* de evento destes processos do Protocolo Geral para extrair conhecimento. Para este trabalho existem quatro informações essenciais para a execução na análise do processo e disponíveis nos arquivos: ID do processo, nome do grupo do processo (já tratado), data do evento (também já tratada) e departamento (recurso organizacional). O conjunto de cada sequência, de cada etapa é o que consiste em um *registro* de eventos

Para se ter uma visão geral dos processos, primeiramente foram definidos os

principais macroprocessos. Lembrando que em um macroprocesso pode ter mais de uma instância de processo. Através da Lei de Pareto, foram identificados 24 macroprocessos que possuem maior frequência de instâncias de processos totalizando 24.869. Estes representam 80,6% do total de 30.871 instâncias de processos. O primeiro macroprocesso está relacionado à “Gestão de Carreira do Docente”. Só este grupo representa 14% das ocorrências da base total. Ele envolve os protocolos referentes a: Promoções, progressões, redistribuição de titulação, reconhecimento de saberes e competências. Todos os 24 macroprocessos serão denominados a partir daqui como Macroprocessos Principais. Quando for utilizado a base geral, e não somente destes 24 macroprocessos, são denominados Macroprocessos Gerais. Pela mesma forma, quando as instâncias de processos presentes nestes 24 macroprocessos são denominadas a partir daqui como Processos Principais. Quando for utilizado a base geral, e não somente instâncias destes 24 macroprocessos, são denominados Processos Gerais. A lista dos macroprocessos principais estão apresentados na Tabela 20.

Tabela 20 - *Ranking* dos Macroprocessos Gerais com o corte seguindo a Lei de Pareto

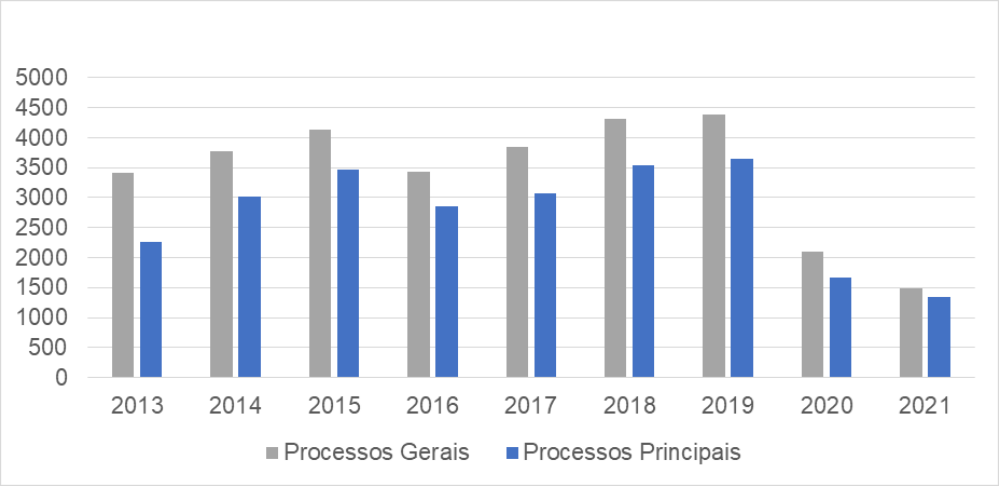
Posição	Macroprocessos Principais	Número de Instâncias de processos	Frequência Relativa Acumulada
1	GESTÃO DE CARREIRA DOCENTE	4.334	14,0%
2	CONVÊNIO	2.554	22,3%
3	COMPRA DE MATERIAL	2.369	29,9%
4	DIPLOMA	2.313	37,5%
5	FOLHA DE PESSOAL	1.754	43,2%
6	GESTÃO DE CARREIRA TAE	1.451	47,9%
7	REEMBOLSO E PAGAMENTOS GERAIS	1.288	52,0%
8	AFASTAMENTO E LICENÇA	1.277	56,2%
9	ESTÁGIO PROBATÓRIO	768	58,7%
10	ADMISSÃO EM CARGO	741	61,0%
11	TEMPO DE SERVIÇO	725	63,4%
12	INSCRIÇÃO EM CURSO E EVENTOS	701	65,7%
13	CONTRATAÇÃO DE SERVIÇO	683	67,9%
14	AVALIAÇÃO DE DESEMPENHO	676	70,1%
15	COLAÇÃO DE GRAU	493	71,7%
16	BOLSAS AUXÍLIO	489	73,3%
17	MATRÍCULA	438	74,7%
18	APOSENTADORIA	319	75,7%
19	ALTERAÇÃO DE DADOS CADASTRAIS	303	76,7%

20	CONTRATAÇÃO DE DOCENTE	258	77,5%
21	JORNADA DE TRABALHO	257	78,4%
22	REDISTRIBUIÇÃO	237	79,1%
23	GESTÃO PATRIMONIAL	229	79,9%
24	FICHA FINANCEIRA	212	80,6%

Fonte: Própria (2020)

Em seguida analisou-se a frequência de instâncias de processos por ano, tanto dos processos gerais quanto dos processos principais. Através do Gráfico 8 é possível verificar que nos três primeiros anos houve um crescimento linear. Porém, no ano de 2016 houve uma queda de 25% referente a 2015. E que desde então havia um crescimento do número de processos até 2019. Em 2019, foi o ano que houve maior número de processos, com 4.392 novos processos. Em 2020, ano da pandemia, houve uma queda significativa em torno de 50% de redução comparando-se com o ano anterior. Já no ano de 2021, com coleta até junho, há aproximadamente 1.500 processos. Percebe-se também que tanto Processos Gerais quanto Processos Principais possuem o mesmo comportamento.

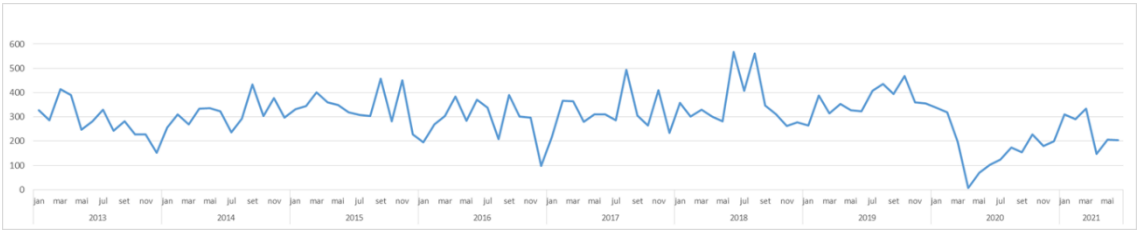
Gráfico 8 – Frequência de Processos anualmente



Fonte: Própria (2020)

O Gráfico 9 representa a série histórica das novas aberturas de solicitações no protocolo geral. Através dele identificou-se algumas sazonalidades no período de 2013 a 2019. Percebe-se que entre os meses de dezembro e janeiro a quantidade de novos processos caem. E que entre os quartos e quintos bimestres ocorrem os picos de aberturas de novos processos. Destaca-se também que no mês de abril de 2020 o número de novos processos abertos caiu drasticamente, foi muito próximo de zero.

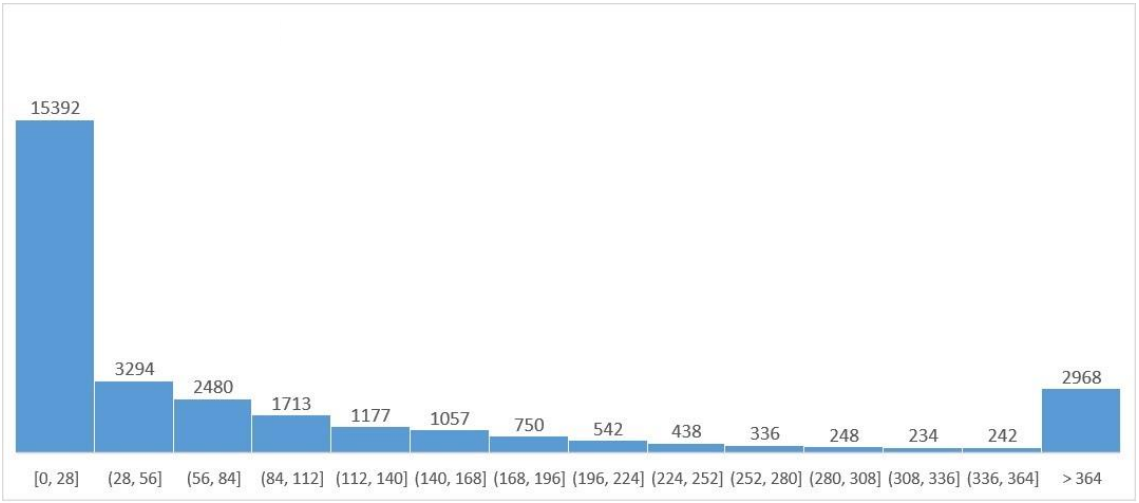
Gráfico 9 - Série histórica dos novos processos desde 2013 até 2021



Fonte: Própria (2020)

Para selecionar o recorte de instâncias de processos e eventos do período antes e pós pandemia foi necessário verificar a distribuição da duração das instâncias processos em dias. O Gráfico 10 apresenta este histograma e observa-se que em torno de 50% dos processos tem a duração de até 28 dias. E também que existem 2.968 processos com duração maior que um ano.

Gráfico 10 - Histograma da Duração dos Processos Geral



Fonte: Própria (2020)

Nesse histograma nota-se que utilizando a Lei de Pareto, aproximadamente 83% dos processos tem duração até de zero até a faixa entre 168 e 196 dias. Ou seja, a grande maioria dos processos são tramitados em torno do período de 6 meses. Este corte de 6 meses de duração é utilizado para reduzir os vieses do recorte da base para os períodos antes e depois da pandemia.

6.4.3. Avaliação

Esta é a fase de avaliar os resultados para relacionar os achados da análise e, então responder o objetivo específico P2.1: Identificar e analisar os processos principais da base geral e seus comportamentos através da frequência dos processos e duração de tempo.

Foram identificados os 24 macroprocessos principais representando 80,6% das instâncias de processos. Destaca-se como os cinco primeiros macroprocessos principais como: Gestão de Carreira Docente (referente Promoções, progressões, redistribuição de titulação, reconhecimento de saberes e competências dos docentes), Convênio (refere-se a convênios entre a faculdade e instituições para estágios dos alunos), Compra de Material (refere-se a vários tipos de compras feitas, como por exemplo, extintores e softwares de computadores), Diploma (emissão ou remissão de diploma para os alunos) e Folha de Pessoal (refere-se aos processos atrelado a Folha de Pessoal como plano de saúde, auxílio funeral e auxílio transporte).

Percebeu-se que havia um crescimento no número de novos processos desde 2016 e que em 2019 foi o ano com maior quantidade. Porém, com a pandemia e todas as medias indicadas pela OMS (2021) reduziu-se consideravelmente o número de novos processos abertos. E consequentemente, as sazonalidades observadas na série história antes de 2020 não ocorrem nos anos pandêmicos, entre 2020 e 2021. Conclui-se que houve um forte impacto reduzindo os processos administrativo de solicitações do protocolo geral.

A partir desta análise uma nova iteração de processamento de dados deve ser gerada para retirar da base os processos que ocorreram há menos de 6 meses antes da coleta da base. E a partir daí, pode-se comparar os processos principais antes e depois da pandemia, assim como a duração, o número de etapas e a realizar a mineração organizacional destas duas bases.

Identificou-se como pontos de melhoria a padronização dos nomes dos processos no Sistema de Protocolo Geral. Isso irá facilitar tanto futuras análises sobre essa base de registro de eventos quanto para o colaborador que abre e tramita a solicitação apenas selecionando o macroprocesso desejado. E outro ponto de melhoria também no sistema é ter uma validação no campo de data para evitar inconsistências.

6.5. Segunda Iteração de Análises

Nesta segunda iteração de análises pretende-se responder o objetivo específico P2.2 apresentado em 1.2 Perguntas de Pesquisa e Objetivos, que é: P2.2 - Identificar e analisar os processos principais do recorte da base nos períodos antes e depois da pandemia e seus comportamentos através da frequência dos processos, duração do tempo e frequência das etapas.

Na fase de processamento desta nova iteração aplica-se os elementos filtragem e enriquecimento. Na fase de mineração e análises implementa-se, novamente, o elemento de análise de processos para criar uma análise visual. E por último, avalia-se os resultados.

6.5.1. Processamento dos dados

Para a análise comparativa dos processos antes da pandemia (AP) e pós pandemia (PP), foi necessário reprocessar os dados e realizar o filtro para coletar seis meses na base para estes dois períodos, conforme a análise do Gráfico 10. O período utilizado para a seleção dos dados dos processos é mostrado na Figura 31. Vale ressaltar que o corte da base de dados foi até dia 16 de junho de 2021, conforme a letra "E" desta figura.

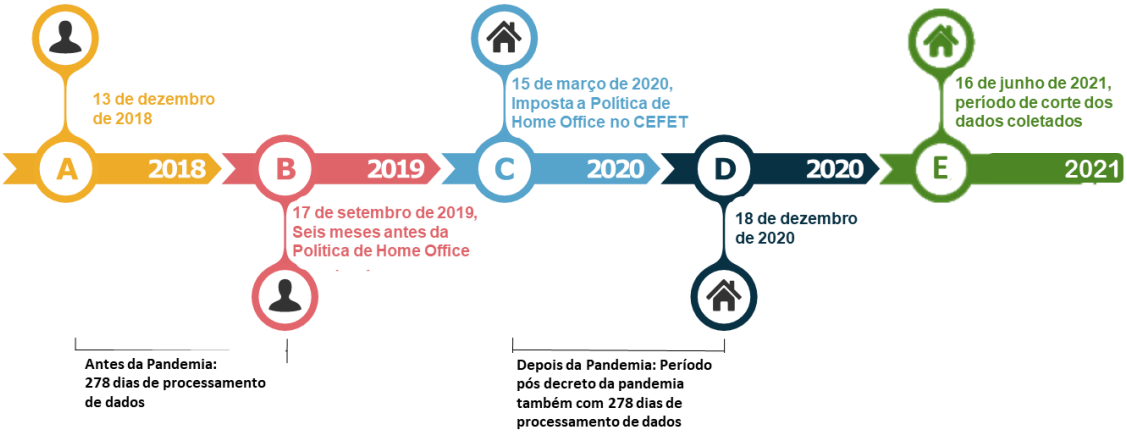


Figura 31 - Períodos da análise comparativa antes e depois da pandemia
Fonte: Própria (2021)

A instituição impôs uma política de home office em 15 de março de 2020, representado pela letra C na Figura 31. Essa data foi utilizada para criar a divisão entre os períodos AP e PP. São apresentados esses dois períodos AP de 278 dias compreende do dia 13 dezembro de 2018 até 17 de setembro de 2019, representado pelas letras A e B da Figura 31. E, o período PP compreende do dia 15 de março de 2020 a 18 de dezembro de 2020 representado pelas letras C e D da Figura 31. Portanto, os dados da base são enriquecidos inserindo uma nova coluna “Pandemia”. Esta é preenchida com os valores “Antes”, “Depois” ou “Desconsiderar” dependendo da data do evento do registro.

Vale ressaltar que todo o tratamento dos dados da primeira iteração foi também utilizado nesta iteração. Adicionalmente, filtrou-se, eliminando da base, aquelas instâncias de processos sem informação de recurso (departamento) que atuou. E por fim, também, excluiu-se também etapas de processos com mais de 180 dias.

Com este novo processamento dos dados onde houve o filtro dos processos para os períodos conforme a Figura 31, o número de novos processos antes da pandemia são 3.144 e após a pandemia são 1.177 processos. Para o mesmo período de observação (278 dias) teve uma queda de aproximadamente de 63% no número de processos antes e depois da pandemia.

6.5.2. Mineração de Processo e Análises

Nesta segunda iteração na fase de Mineração e Análises implementa-se, novamente, o elemento de análise de processos para fazer análises antes e depois da pandemia. Utiliza-se os *registros* de eventos somente dos períodos antes e depois da pandemia. Ela é focada para responder o objetivo específico P2.2.

A

Figura *32* exibe os histogramas da duração das instâncias de processos, onde o processo tende a ser mais curto, em média, após a pandemia. De forma geral, o processo de negócios mudou drasticamente.

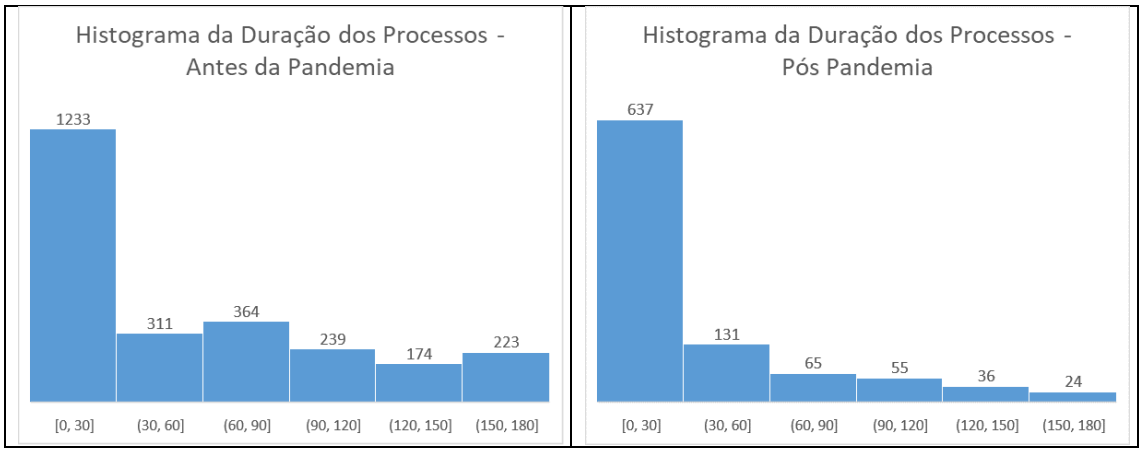


Figura 32 - Histograma de Distribuição de duração média dos processos antes e depois da pandemia
Fonte: Própria (2021)

Além disso, há indícios concretos de foi implementada uma nova estrutura de processos durante a pandemia. O processo apresenta um número menor de etapas depois da pandemia conforme mostra a Figura 33.

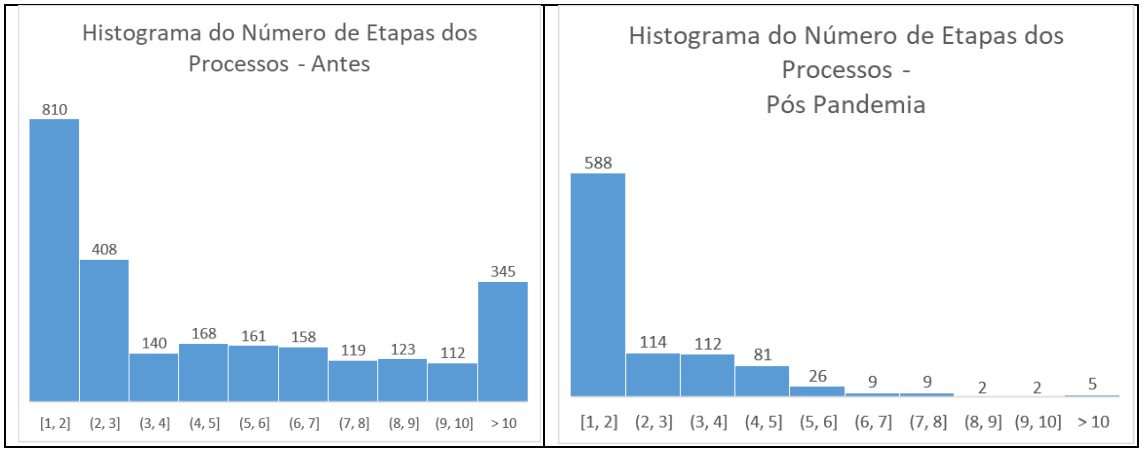


Figura 33 - Histograma de distribuição do número médio de etapas por processos
Fonte: Própria (2021)

Antes da pandemia, uma unidade organizacional supervisionava fisicamente a movimentação e o registro de processos no Sistema Eletrônico de Protocolo Geral. Durante a política de home office, os trabalhadores movimentam eletronicamente o processo anterior registrado em papel. Uma hipótese para a forte mudança na duração e número de tarefas por processo é a suposição de subnotificação, onde o trabalhador evita o registro de processos no sistema.

Os macroprocessos principais foram definidos primeiro para obter uma visão geral dos processos. Na Tabela 21 são listados esses macroprocessos antes da

pandemia obtidos através da aplicação da Lei de Pareto. Onde foram identificados 20 macroprocessos principais que possuem maior ocorrência antes da pandemia totalizando 2.650 instâncias de processos. Estes representam 80,9% do total de 3.144 instância de processos.

Tabela 21 -Ranking dos macroprocessos principais antes da pandemia com o corte seguindo a Lei de Pareto

Posição	Macroprocessos Principais	Número de Instâncias de processos	Frequência Relativa Acumulada
1	DIPLOMA	552	17,6%
2	GESTÃO DE CARREIRA DOCENTE	361	29,0%
3	GESTÃO DE CARREIRA TAE	213	35,8%
4	CONVÊNIO	213	42,6%
5	AFASTAMENTO E LICENÇA	167	47,9%
6	REEMBOLSO E PAGAMENTOS GERAIS	128	52,0%
7	BOLSAS AUXÍLIO	105	55,3%
8	FOLHA DE PESSOAL	103	58,6%
9	JORNADA DE TRABALHO	91	61,5%
10	COMPRA DE MATERIAL	84	64,6%
11	MATRÍCULA	83	66,8%
12	ALTERAÇÃO DE DADOS CADASTRAIS	73	69,1%
13	TEMPO DE SERVIÇO	61	71,0%
14	ESTÁGIO PROBATÓRIO	58	72,9%
15	CONTRATAÇÃO DE DOCENTE	51	74,5%
16	INSCRIÇÃO EM CURSO E EVENTOS	48	76,0%
17	REDISTRIBUIÇÃO	41	77,3%
18	APOSENTADORIA	39	78,6%
19	CONTRATAÇÃO DE SERVIÇO	37	79,8%
20	COLAÇÃO DE GRAU	36	80,9%

Fonte: Própria (2020)

A coluna “Posição” é o novo *ranking* e em primeiro lugar está o macroprocesso Diploma representando 17,6% do total. Em seguida tem-se a Gestão de Carreira Docente e Gestão de Carreira TAE. Estes dois últimos tratam-se das evoluções de carreira tanto dos docente quanto dos funcionários técnicos administrativos.

Para ordenar macroprocessos principais no período pós pandemia, utilizou-se da Lei de Pareto que apresentou 18 macroprocessos com maior ocorrência. Estes totalizam 924 instâncias de processos. Logo, representam 80,54% do total de 1.177 instâncias de processos. Na Tabela 22 é apresentada a lista destes macroprocessos principais e seus respectivos valores.

Tabela 22 - Ranking dos macroprocessos principais pós pandemia com o corte seguindo a Lei de Pareto

Posição	Macroprocessos Principais	Número de Instâncias de processos	Frequência Relativa Acumulada
1	GESTÃO DE CARREIRA DOCENTE	175	14,9%
2	CONVÊNIO	153	27,9%
3	DIPLOMA	118	37,9%
4	GESTÃO DE CARREIRA TAE	80	44,7%
5	COMPRA DE MATERIAL	67	50,4%
6	COLAÇÃO DE GRAU	45	54,2%
7	ESTÁGIO PROBATÓRIO	41	57,7%
8	FOLHA DE PESSOAL	36	60,8%
9	INSCRIÇÃO EM CURSO E EVENTOS	32	63,5%
10	EDITAL	31	66,1%
11	REDISTRIBUIÇÃO	30	68,6%
12	CONTRATAÇÃO DE SERVIÇO	25	70,8%
13	AFASTAMENTO E LICENÇA	24	72,8%
14	ADMISSÃO EM CARGO	23	74,8%
15	REEMBOLSO E PAGAMENTOS GERAIS	19	76,4%
16	TEMPO DE SERVIÇO	19	78,0%
17	OFÍCIO	16	79,4%
18	MATRÍCULA	14	80,5%

Fonte: Própria (2020)

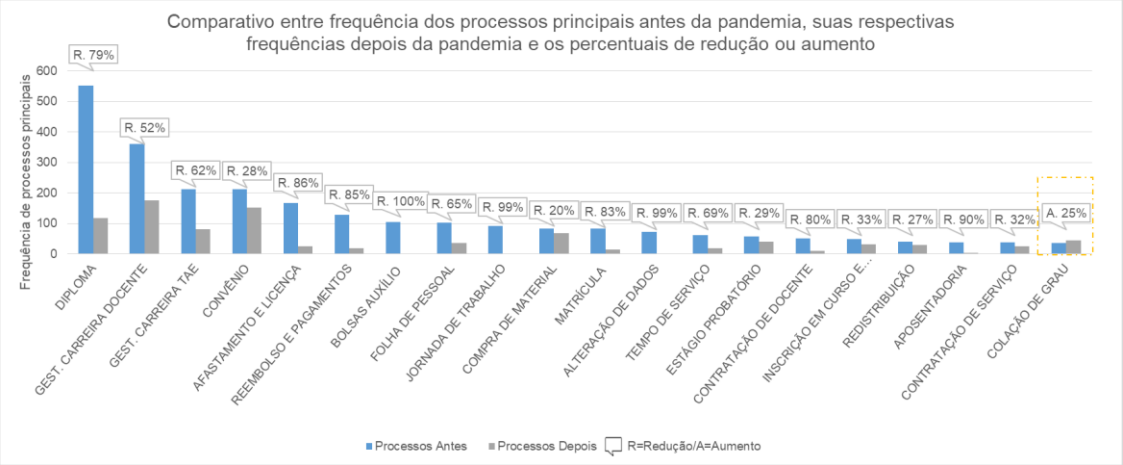
A coluna “Posição” é o *ranking* e em primeiro lugar está o macroprocesso “Gestão de Carreira Docente” representando 14,8% do total. Em seguida tem-se o “Convênio” e o “Diploma”.

Percebe-se que o “Diploma” que estava em primeiro lugar antes da pandemia, este macroprocesso segue em terceiro lugar no recorte de pós pandemia. O “Gestão de Carreira Docente” que estava em segundo lugar antes da pandemia, liderou os processos no período de pós pandemia. O macroprocesso “Gestão de Carreira TAE” que antes da pandemia que ficou em terceiro lugar no período pós pandemia ficou em quarto lugar. E “Convênio que estava em quarto lugar antes da pandemia ocupou o segundo lugar no período pós pandemia. Nota-se que os quatro macroprocessos mais comuns são os mesmos no período antes e pós pandemia. Estes diferem-se entre a posição na ordenação, na quantidade de processos e também o seu percentual de cada um deles em relação ao total.

Para avaliar um comparativo entre os processos principais que ocorriam antes da pandemia e como ficaram depois da pandemia foi desenvolvido o Gráfico 11. Nele são fixados e ordenados pela frequência de processos principais antes da pandemia na primeira coluna, em azul, e são apresentados estes mesmos processos depois da

pandemia na segunda coluna, em cinza. E em destaque estão os seus respectivos percentuais de aumento ou redução depois da pandemia. Analisa-se que a grande maioria dos macroprocessos tiveram redução com exceção de “Colação de Grau” (aumento de 25%). Este estava em último no recorte, Tabela 21, antes da pandemia com 36 instâncias de processos. E no período pós pandemia ficou em sexto, Tabela 22, com 45 instâncias de processos.

Gráfico 11 – Comparativo entre frequência dos processos principais antes da pandemia, suas respectivas frequências depois da pandemia e os percentuais de redução ou aumento

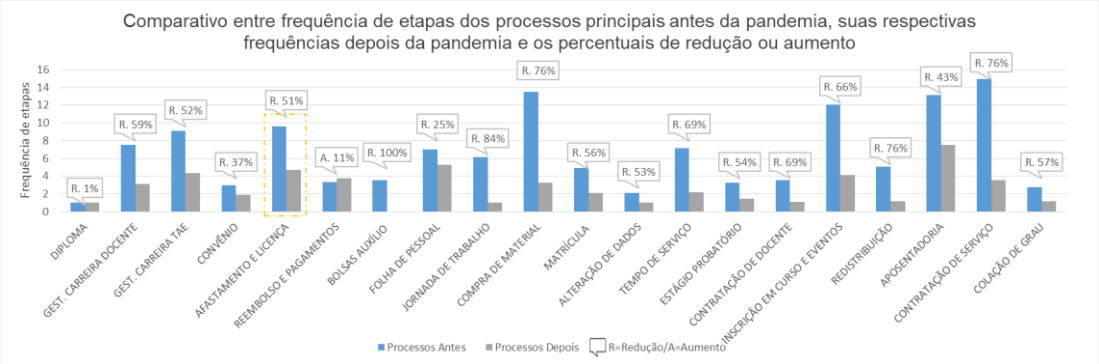


Fonte: Própria (2020)

O macroprocesso de “Bolsa auxílio” que antes da pandemia estava sétimo lugar com 105 instâncias de processos mas no período pós pandemia não foi encontrado nenhum novo processo. Outros macroprocessos que tiveram redução maior ou igual a 90% foram: “Jornada de Trabalho”, “Alteração de Dados Cadastrais”, “Aposentadoria”. Os macroprocessos que tiveram uma redução de até 30% foram “Convênio”, “Compra de Material”, “Estágio Probatório” e “Redistribuição”.

No Gráfico 12 mantem-se a ordenação pela frequência de processos principais antes da pandemia na primeira coluna, em azul. Adicionalmente, são apresentadas as médias das frequências de etapas, em dias, e suas respectivas médias depois da pandemia na segunda coluna, em cinza. E em destaque estão os percentuais de aumento ou redução depois da pandemia. Essa análise é importante pois auxilia na verificação de simplificação do processo, ou seja, o número de etapas reduzindo houve menos trocas dos processos entre os recursos (departamentos).

Gráfico 12 - Comparativo entre as médias de frequência de etapas dos processos principais antes da pandemia, suas respectivas médias depois da pandemia e os percentuais de redução ou aumento

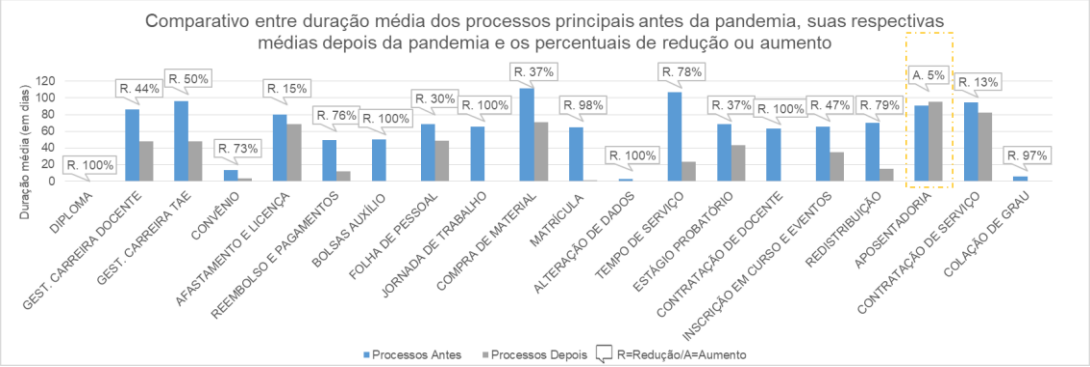


Fonte: Própria (2022)

Percebe-se que apenas o macroprocesso “Reembolso e Pagamento” teve um aumento de 11%. O que ajuda a descartar a hipótese de subnotificação, uma vez que os processos têm mais tarefas após a política de home office do que antes. Os macroprocessos “Gestão de Carreira Docente” e “Gestão de Carreira TAE”, como segundo e terceiro lugares de frequência de instâncias de processos antes da pandemia, teve redução de 59% e 52%, respectivamente. O macroprocesso de “Diploma” tem apenas uma etapa. Com exceção deste “Diploma”, o macroprocesso “Folha de Pessoal” com a menor redução foi “Folha de Pessoal” com uma redução de 25%.

No Gráfico 13 mantem-se a ordenação pela frequência de processos principais antes da pandemia na primeira coluna, em azul. Adicionalmente, são apresentadas as durações médias, em dias, e suas respectivas durações médias depois da pandemia na segunda coluna, em cinza. E em destaque estão os percentuais de aumento ou redução depois da pandemia. Analisa-se que apenas o macroprocesso “Aposentadoria” teve um aumento de 5% e que todos os demais tiveram o tempo reduzido. O macroprocesso “Diploma”, mesmo sendo o mais frequente se manteve com duração em 0 dias, configurando-se numa entrega instantânea do Diploma ao solicitante. O “Convênio” teve uma significativa redução de duração de tempo, 73%, e a sua redução de frequência foi apenas de 28%. Por outro lado, “Afastamento e licenças” que teve uma redução significativa de frequência de processo teve apenas 15% de redução na duração média.

Gráfico 13 - Comparativo entre duração média dos processos principais antes da pandemia, suas respectivas médias depois da pandemia e os percentuais de redução ou aumento



Fonte: Própria (2022)

Acredita-se que as reduções relacionadas aos tempos de duração são insuficientes para justificar a considerável discrepância observada, indicando mudanças estruturais nos processos. Para avançar com as análises na próxima iteração é implementado a mineração organizacional com a técnica de rede social em alguns macroprocessos específicos.

6.5.3. Avaliação

Esta é a fase de avaliar os resultados do objetivo específico P2.2: “Identificar e analisar os processos principais do recorte da base nos períodos antes e depois da pandemia e seus comportamentos através da frequência dos processos, duração do tempo e frequência das etapas”.

Verificou-se através dos histogramas de duração e etapas que os processos tendem a ser mais curto, em média, e ter menos etapa após a pandemia. E observou-se que nos períodos selecionados houve também a redução da frequência dos processos depois da pandemia o que é um resultado esperado após a análise da primeira iteração e dado que a instituição do estudo de caso (CEFET/RJ) ficou um período sem funcionamento presencial, conforme orientações da (OMS, 2021). Os processos demandados antes e depois da pandemia são os mesmos, mas eles se alternam entre as posições à medida que se tornam menos ou mais demandados. Os macroprocessos “Gestão de Carreira Docente” e “Gestão de Carreira TAE”, que

preenchem a posição de segundo e terceiro lugar dos processos mais frequentes antes da pandemia, tiveram reduções nas três métricas apresentadas (frequência de processos, duração médias e média das frequências de etapas) no período após a pandemia.

Diagnosticou-se que o macroprocesso de “Colação de Grau” foi o único que teve aumento na frequência dos macroprocessos principais em 25% pós pandemia. Assim como houve um aumento na duração média do macroprocesso de “Aposentaria”, mesmo com a sua redução da sua frequência de instâncias de processos em 90%. Avaliou-se que outros macroprocessos poderiam ter aumentado suas durações também. Portanto, supõe-se que neste cenário pandêmico outros macroprocessos poderiam ter aumentado suas durações e frequências também. Afirmando-se que não são tão óbvios os resultados de reduções significativas. Entretanto, é possível avaliar que houve mudança nos processos para que se adaptassem a nova realidade da pandemia e estas mudanças indicam ser positivas por terem reduzido a duração média em dias. .

Sendo assim, não são conclusivas os impactos e alterações ocorridos depois da pandemia. Para tanto é preciso ser feito um aprofundamento maior sobre estes casos na mineração organizacional. Logo, serão avaliadas nas seções posteriores a relação entre os departamentos com as transferências de trabalhos para identificar as mudanças nos processos destaques conhecidos: “Gestão de Carreira Docente”, “Gestão de Carreira TAE”, “Folha de Pessoal”, “Reembolso de Pagamentos”, e “Compra de Material”. Apesar do macroprocesso “Diploma” ser destaque também, ele é bem simples e composto basicamente por uma etapa e um departamento, por isso não foi feita uma análise mais aprofundada.

6.6. Terceira Iteração de Análises

Nesta última iteração de análises pretende-se responder o objetivo específico P2.3 apresentado em 1.2 Perguntas de Pesquisa e Objetivos. Na fase de processamento aplicam-se os elementos criação de visualizações e filtragem de *registros*. A criação de visualização se faz necessária pois é preciso adaptar os arquivos de registro de eventos para serem processados na ferramenta de mineração de processos, PROM, e na ferramenta para análise de rede social, Pajek. Na fase de

mineração e análises implementa-se a mineração de processos do tipo descoberta sob a perspectiva organizacional. Ou seja, a mineração organizacional será utilizada com técnicas de redes sociais conforme apresentado por Song; Van der Aalst (2008). Complementarmente, é apresentada a visão do fluxo do processo com base nos departamentos. E por último, avalia-se os resultados.

6.6.1. Processamento dos dados

Para a mineração organizacional são reprocessados os dados e criados novos arquivos nos formatos esperados pelo Pajek, inicialmente, e em seguida para o PROM. A mineração organizacional é realizada nos macroprocessos principais conhecidos: “Gestão de Carreira Docente”, “Gestão de Carreira TAE”, “Folha de Pessoal”, “Reembolso de Pagamentos”, e “Compra de Material”. Logo, em cada um destes macroprocessos são criados dois arquivos do Pajek (no formato .net) onde um tem os dados antes da pandemia e o outro tem os dados depois da pandemia para a descoberta das redes sociais. Nestes casos, os vértices são os departamentos que tramitam a solicitação aberta no Sistema de Protocolo Geral, estes são ligados por arcos que indicam a movimentação da origem ao destino e os pesos dos arcos é o número da transferência de trabalho, ou seja, o número de vezes que o processo tramitou entre os dois departamentos. Na visão complementar para descoberta dos fluxos de processos com bases nos recursos, também são criados outros novos arquivos (extensão .CSV e .XES) para esta mesma seleção dos macroprocesso..

6.6.2. Mineração de Processo e Análises – Gestão Carreira TAE

Na Tabela 23 é apresentada alguns dados gerais do macroprocesso de Gestão de Carreira TAE antes e depois da pandemia. Percebe-se que houve uma redução aproximadamente de 62% no número de processos. E também houve redução em torno de 50% na duração média e no número médio de etapas dos processos.

Tabela 23 - Visão Geral do macroprocesso de Gestão de Carreira TAE

Gestão Carreira TAE	Antes	Depois
Número de processos	213	80
Média de duração (dias)	96,1	48,1
Média do nº de etapas	9,1	4,4

Fonte: Própria (2022)

Para a análise da rede de Gestão de Carreira TAE foram analisadas as métricas apresentadas na Tabela 24. Após a pandemia observa-se que houve menos dez departamentos envolvidos, 50 menos arcos e cinco menos departamentos com loop. A frequência dos pesos dos arcos também reduziu. O grau médio indica o número de departamentos em média que um departamento se relaciona, ou seja, antes era em torno de sete departamentos e passou a ser em torno de quatro. A densidade e a medida de centralidade intermediação da rede entre os períodos ficaram próximos.

Tabela 24 - Métricas Gerais das redes do macroprocesso de Gestão de Carreira TAE

Métricas gerais das redes	Antes	Depois
Número de departamento (vértices)	18	8
Número de arcos	67	17
Total dos pesos dos arcos	1940	350
Número de departamento com loops	8	3
Densidade da Rede	0,21	0,26
Grau médio da rede	7,44	4,25
Intermediação Rede	0,28	0,24

Fonte: Própria

As Figura 34 e Figura 35 apresentam as redes sociais derivadas dos registros de eventos do macroprocesso Gestão de Carreira TAE. As métricas destas redes são baseadas em transferência de trabalho e o tamanho do vértice é o grau de intermediação. Verifica-se visualmente, por exemplo, a redução do número de departamento, número de arcos e o grau médio dos departamentos.

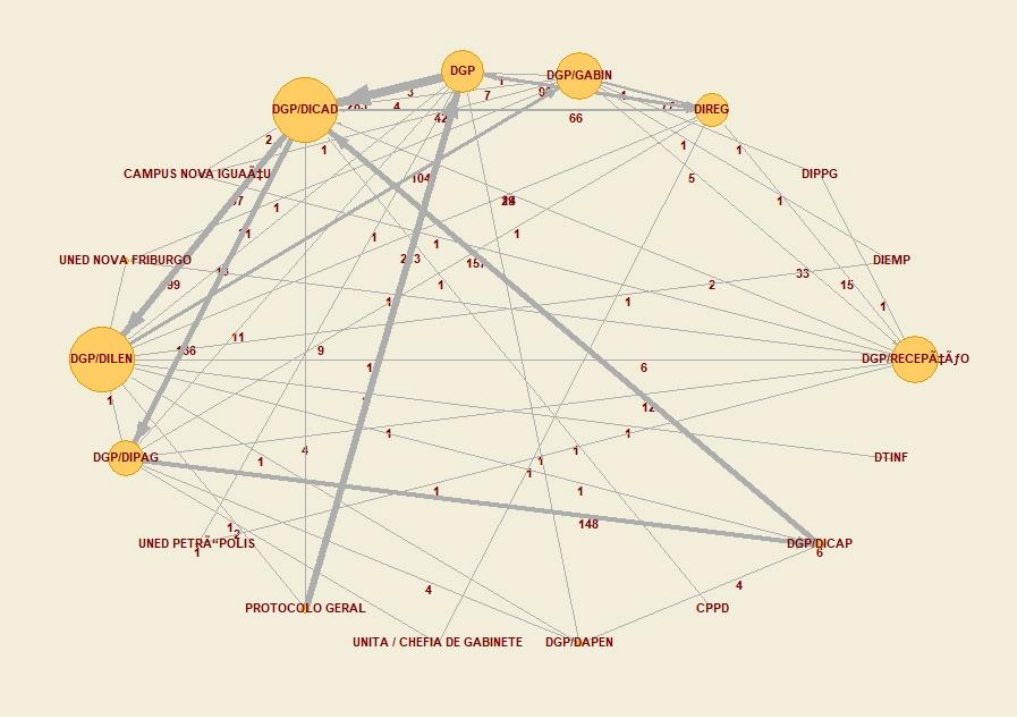


Figura 34 – Rede Social do macroprocesso Gestão de Carreira TAE Antes da Pandemia

Fonte: Própria (2022)

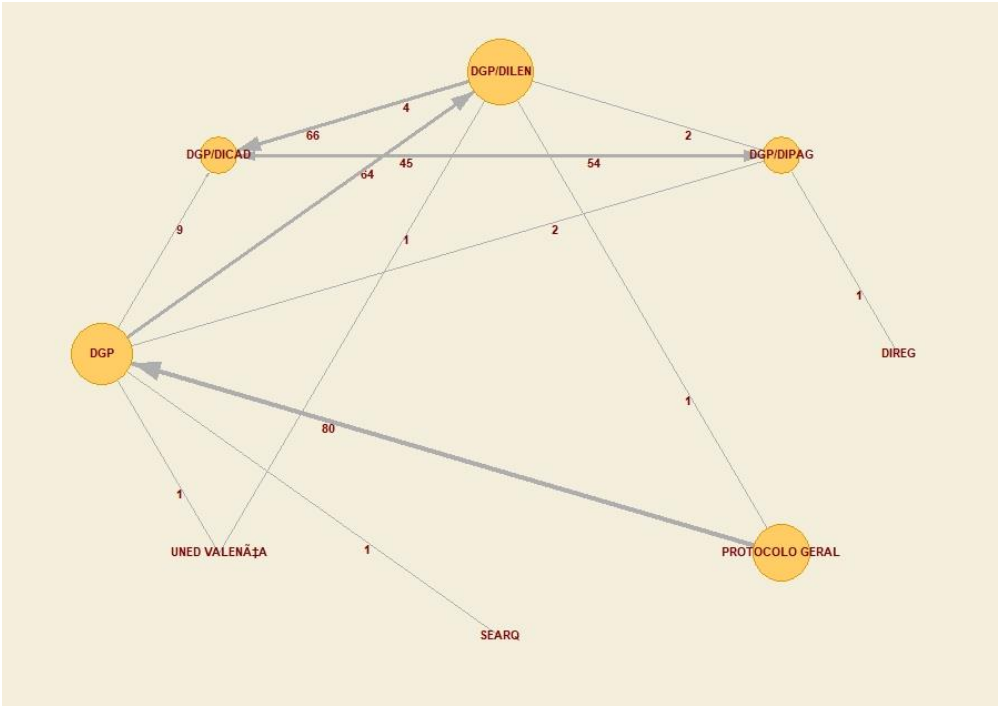


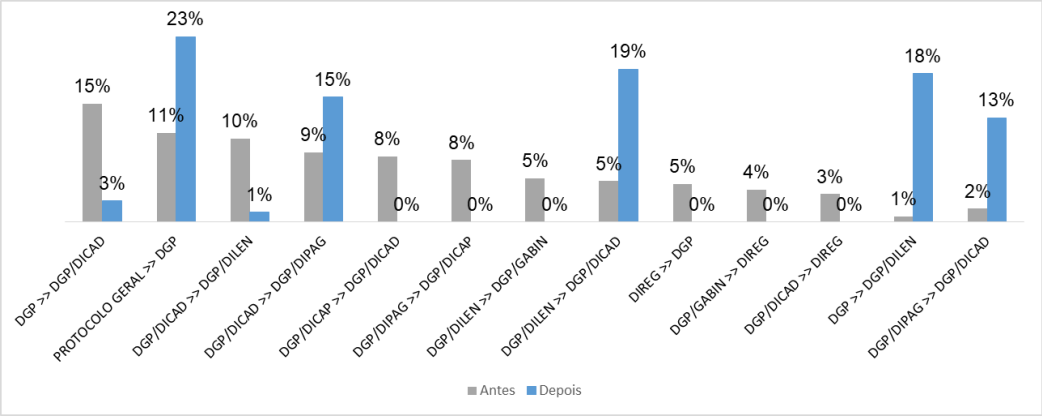
Figura 35 - Rede Social do macroprocesso Gestão de Carreira TAE depois da pandemia

Fonte: Própria (2022)

Através destas redes observou-se que houve alteração no arco de maior peso. Antes da pandemia era “DGP > DGP/DICAD“ com 283 transferências de trabalho, representando 15% do total. Este arco “DGP > DGP/DICAD“ existe depois da pandemia, porém, ele representa apenas 3% do total de transferências de trabalho. E no período pós pandemia o arco de maior peso é o “PROTOCOLO GERAL > DGP” com 80 transferências de trabalho representando 23% do total. Este arco existia antes da pandemia e era o segundo arco mais relevante representando 11% do total.

Para um melhor entendimento desta análise e um detalhamento mais profundo dos arcos, foi feita a frequência relativa dos arcos em cada período, ou seja, calculou-se o peso (a frequência) de cada arco sobre o peso total. E então, foi feita uma extração dos arcos mais relevantes, utilizando a lei de Pareto, antes da pandemia e depois da pandemia. Esta extração está representada no Gráfico 14:

Gráfico 14 - Arcos mais relevantes do macroprocesso de Gestão de Carreira TAE antes e depois da pandemia



Fonte: Própria

Percebeu-se nestas relações os seguintes pontos importantes:

- seis arcos relevantes antes da pandemia que desapareceram. São eles: DGP/DICAP >> DGP/DICAD, DGP/DIPAG >> DGP/DICAP, DGP/DILEN >> DGP/GABIN, DIREG >> DGP, DGP/GABIN >> DIREG e DGP/DICAD >> DIREG.
- dois arcos relevantes antes da pandemia que deixou de ser relevante reduzindo-se significativamente sua frequência relativa. Que são: DGP >> DGP/DICAD e DGP/DICAD >> DGP/DILEN.
- três arcos que já eram importantes antes da pandemia e aumentaram ainda mais sua frequência relativa depois da pandemia que são: PROTOCOLO GERAL >> DGP, DGP/DICAD >> DGP/DIPAG e DGP/DILEN >> DGP/DICAD

- E dois arcos que não estavam dentro dos arcos mais importantes da lei de Pareto antes da pandemia, tornaram-se relevante no período pós pandemia. Que são: DGP >> DGP/DILEN e DGP/DIPAG >> DGP/DICAD.

De modo geral antes da pandemia eram necessários onze arcos (representados pelas onze primeiras barras cinzas no Gráfico 14) para cobrir 83% do total dos pesos dos arcos. E após a pandemia são necessários apenas 5 arcos (representados pelas cinco barras azuis com a frequência relativa maior que 10% no Gráfico 14) para cobrir 88% do total dos pesos dos arcos. Assim, apenas 3 arcos relevantes mantêm-se relevantes após a pandemia. Houve a simplificação de trâmite entre os departamentos pois foram eliminados 6 arcos relevantes. E por fim, com essas alterações houve uma necessidade de adaptação tornando-se relevantes 2 arcos no período pós da pandemia.

Para analisar no detalhe a relação entre os departamentos antes e depois da pandemia e na identificação do departamento mais central das redes calculou-se as métricas de centralidade de cada vértice: grau de entrada, grau de saída e intermediação. A Figura 36 apresenta essas métricas de cada departamento envolvido antes e depois da pandemia. Para facilitar a visualização, nesta figura o tamanho das barras representa o número do grau com relação a coluna correspondente, logo, uma barra totalmente preenchida representa que aquele departamento tem maior grau.

DEPARTAMENTO	Grau de Entrada		Grau de Saída		Intermediação	
	Antes	Depois	Antes	Depois	Antes	Depois
DGP/DICAD	10	4	10	3	0,35	0,12
DGP/DILEN	8	4	9	4	0,34	0,36
DGP/DIPAG	6	4	6	3	0,10	0,12
DIREG	6	1	7	0	0,09	0,00
DGP/RECEPCAO	6	0	9	0	0,18	0,00
DGP/GABIN	6	0	8	0	0,17	0,00
DGP	5	1	4	5	0,15	0,31
DGP/DICAP	5	0	2	0	0,01	0,00
PROTOCOLO GERAL	2	1	1	1	0,01	0,26
UNED NOVA FRIBURGO	2	0	1	0	0,00	0,00
DGP/DAPEN	2	0	2	0	0,00	0,00
UNED PETROPOLIS	2	0	2	0	0,00	0,00
CAMPUS NOVA IGUAÇU	2	0	1	0	0,00	0,00
DIPPG	1	0	1	0	0,00	0,00
UNITA / CHEFIA DE GABINETE	1	0	1	0	0,00	0,00
DTINF	1	0	1	0	0,00	0,00
DIEMP	1	0	1	0	0,00	0,00
CPPD	1	0	1	0	0,00	0,00
UNED VALENÇA	0	1	0	1	0,00	0,00
SEARQ	0	1	0	0	0,00	0,00

Figura 36- Lista de departamentos do macroprocesso de gestão da carreira TAE e suas métricas de centralidade antes e depois da pandemia
Fonte: Própria

Percebe-se que houve mudança no departamento central. O departamento “DGP/DICAD” antes da pandemia possuiu os maiores valores em todas as métricas de centralidade analisadas. Logo, ele é o departamento central antes da pandemia. O segundo e o terceiro mais centrais são “DGP/DILEN” e o “DGP/Recepção”, respectivamente.

Depois da pandemia não houve um departamento que estivesse os maiores valores de todas as métricas centrais. Porém, destaca-se o DGP/DILEN como o mais central depois da pandemia, pois, ele teve as maiores métricas no grau de entrada e de intermediação. Em seguida, o “DGP” com a maior métrica do grau de saída e a segunda maior métrica do grau de intermediação. E no terceiro lugar, se destacam dois departamentos “DGP/DICAD” e “DGP/DILEN”, ambos com os maiores valores de grau de entrada e terceiros e quartos maiores valores dos graus de saída e de intermediação.

Nota-se que os nós mais centrais fazem parte da DGP (Diretoria de Gestão de Pessoas) a qual é responsável pelos casos dos processos de Gestão de Carreira TAE. Consequentemente, nas redes Figura 34 e Figura *35* é possível verificar que os arcos que os envolvem tem maior peso, ou seja, tem maior frequência da transferência de trabalho. Já os departamentos com valores menores de grau são de unidades externas ou de diretorias variadas que desempenha um papel de forma mais isolada e não estão envolvidos em muitas transferências de trabalho.

Para detalhar e aprofundar no entendimento da participação de cada departamento, a Figura 37 apresenta dois cálculos de % de frequência. A primeira “Frequência Relativa - Casos” a qual é o % de casos (instâncias de processos) contendo uma única vez o determinado departamento (vértice). Logo, é o percentual de presença do departamento em todos os processos sem repetição (mesmo que tenha loop). Nesta análise é possível perceber, por exemplo, se tem algum departamento que participa 100% dos processos. E a segunda é “Frequência Relativa - Eventos” é o percentual de vezes que um departamento atua em eventos (atividades) sobre a quantidade de casos (instâncias de processos). Ou seja, é o número de vezes que este departamento participou de alguma ação, mesmo tendo repetição, sobre o total de processos. Uma análise possível identificar se há algum departamento altamente requisitado, mais de uma vez por processo. Assim, se houve um departamento com número alto de loops, este poderá ter o percentual maior do que 100%.

DEPARTAMENTO	Frequência Relativa - Casos		Frequência Relativa - Eventos	
	Antes	Depois	Antes	Depois
DGP	100%	96%	**145%**	96%
PROTOCOLO GERAL	100%	100%	100%	100%
DGP/DICAD	99%	88%	**228%**	93%
DGP/DILEN	95%	83%	**108%**	88%
DGP/DIPAG	86%	58%	88%	60%
DGP/DICAP	74%	0%	77%	0%
DIREG	73%	0%	75%	0%
DGP/GABIN	50%	0%	59%	0%
DGP/RECEPCAO	24%	0%	26%	0%
DGP/DAPEN	2%	0%	2%	0%
UNED NOVA FRIBURGO	1%	0%	1%	0%
CAMPUS NOVA IGUAÇU	1%	0%	1%	0%
UNED PETROPOLIS	0%	0%	1%	0%
DIPPG	0%	0%	0%	0%
UNITA / CHEFIA DE GABINETE	0%	0%	0%	0%
DTINF	0%	0%	0%	0%
DIEMP	0%	0%	0%	0%
CPPD	0%	0%	0%	0%
UNED VALENÇA	0%	0%	0%	0%
SEARQ	0%	0%	0%	0%

Figura 37 -Lista de departamentos do macroprocesso de gestão da carreira TAE e suas frequências relativas antes e depois da pandemia

Fonte: Própria (2022)

Analisou-se que apesar do departamento “Protocolo Geral” não ser destaque nas métricas de centralidades de grau ele é o único que aparece em 100% das instâncias de processos nos períodos antes e depois da pandemia. Na “Frequência Relativa - Casos” DGP/DIPAG antes da pandemia ocorreu em 86% das instâncias de processos e apenas 58% no período da pandemia. Importante destacar também que na “Frequência Relativa - Eventos”, no período antes da pandemia, os departamentos “DGP/DICAD”, “DGP” e “DGP/DILEN” tem um alto percentual. Isso mostra um número de eventos muito maior que o número de instâncias de processos, podendo indicar possíveis gargalos e retrabalhos. Já para os processos depois da pandemia este comportamento não ocorreu, mesmo nos nós mais centrais.

A visão do fluxo do processo de Gestão de Carreira TAE conforme a Figura 38 complementa as análises antes e depois da pandemia. Este fluxo é baseado na técnica de descoberta de fluxo através do algoritmo “*Directly-Follows*” e as configurações do mesmo estão explanadas na sessão de Metodologia e na subsessão 4.2.1. Percebe-se que a grande maioria dos processos inicia-se pelo protocolo, indicados pelo círculo verde. E finaliza-se pelo DGP/DICAP, indicado pelo círculo vermelho. Além disso, o departamento DGP/DICAD é um departamento decisório, pois através dele, ou o fluxo

segue mais à direita entre os departamentos DGP/GABIN, DIREG, DGP/RECEPÇÂO e DGP/DILEN, ou mais à esquerda passando apenas por mais 2 departamentos DGP/DIPG e DGP/DICAP para finalizar.

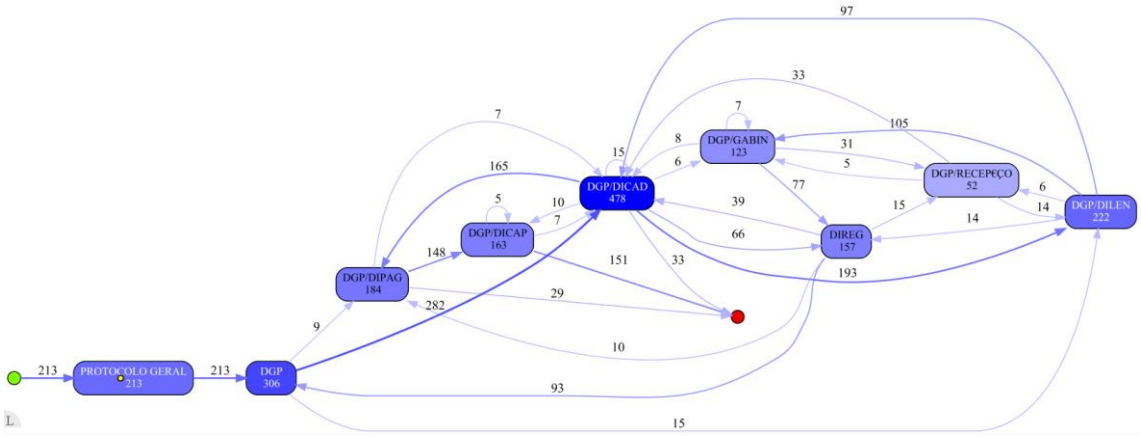


Figura 38 - Fluxo do processo de Gestão de Carreira TAE antes da pandemia

Fonte: Própria(2022)

No fluxo pós pandemia, apresentado Figura 39, percebe-se uma simplificação do processo que segue um processo mais direto. A maioria dos processos (em torno de 85% dos processos) são tramitados entre três ou quatro departamentos após a entrada no Protocolo Geral. Observou-se também que em qualquer um dos eventos o processo pode seguir para o fim. Os caminhos possíveis são:

- Protocolo Geral> DGP > DGP/DILEN > DGP/DICAD. Representados por 26 processos.
- Protocolo Geral> DGP > DGP/DILEN > DGP/DICAD > DGP/DIPAG. Representados por 42 processos.

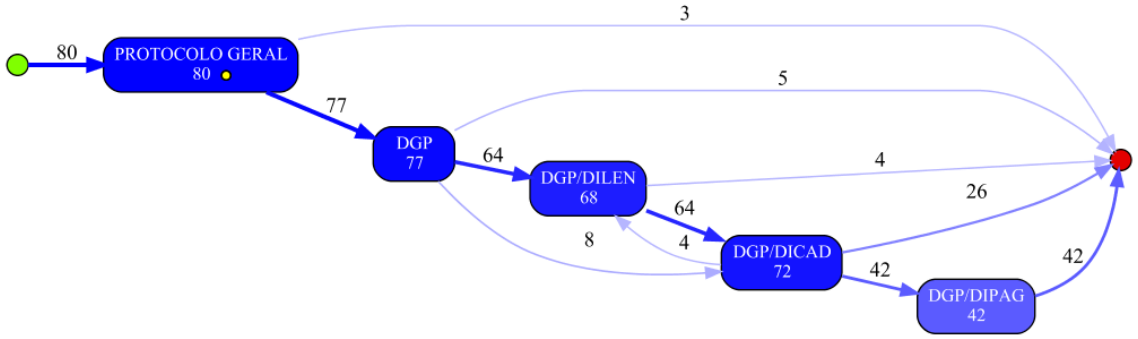


Figura 39 - Fluxo do processo de Gestão de Carreira TAE depois da pandemia

Fonte: Própria(2022)

6.6.3. Avaliação – Gestão Carreira TAE

Através das análises anteriores fica claro que houve simplificação no processo e que as melhorias não ocorreram eliminando apenas departamentos não centrais ou reduzindo arcos irrelevantes. Observou-se que houve uma simplificação estruturada nos processos no período pós pandemia, onde mesmo arcos e departamentos importantes antes da pandemia foram revisados e modificados.

E isso contribuiu para redução no trâmite entre departamentos (desburocratizando as transferências de trabalho), possíveis revisões de gargalos e redução de retrabalho nos processos de Gestão de Carreira TAE. Resultando assim, na diminuição da duração em dias e no número de etapas. Uma das suposições é a criação de listas de verificação no site da instituição contemplando, principalmente a Progressão da Carreira TAE. Assim, o funcionário sabe com antecedência o que é preciso para seguir com maior fluidez à sua solicitação. E uma segunda suposição é a evolução do processo por digitalização através do e-mail. Onde os registros feitos no sistema de Protocolo Geral são marcados apenas nas etapas principais e as demais etapas ficam registradas apenas na caixa do e-mail.

6.6.4. Mineração de Processo e Análises – Gestão Carreira Docente

Na Tabela 25 é apresentada alguns dados gerais do macroprocesso de Gestão de Carreira Docente antes e depois da pandemia. Percebe-se que houve uma redução aproximadamente de 52% no número de processos. E também houve redução em torno de 44% na duração média e 59% de redução no número médio de etapas dos processos.

Tabela 25 – Visão Geral do macroprocesso de Gestão de Carreira Docente

Gestão Carreira Docente	Antes	Depois
Número de processos	361	175
Média de duração (dias)	86,1	48,0
Média do nº de etapas	7,5	3,1

Fonte: Própria

Conforme apresentado na Tabela 26 após a pandemia é possível notar que houve menos 13 departamentos envolvidos, 88 menos arcos e 8 menos departamentos com loop. O grau de intermediação da rede e a frequência dos pesos dos arcos também reduziram consideravelmente. O grau médio indica o número de departamentos em média que um departamento se relaciona, ou seja, antes era em torno de 8,86 departamentos e passou a ser apenas 4,80. Porém, a densidade da rede se manteve igual nos dois períodos.

Tabela 26 - Métricas gerais das redes do macroprocesso de Gestão de Carreira Docente

Métricas gerais das redes	Antes	Depois
Número de departamentos (vértices)	28	15
Número de arcos	124	36
Total dos pesos dos arcos	2706	544
Número de departamentos com loops	11	3
Densidade da Rede	0,16	0,16
Grau médio da rede	8,86	4,80
Intermediação Rede	0,48	0,29

Fonte: Própria

As Figuras 40Figura 41 - Rede Social do macroprocesso Gestão de Carreira Docente depois da pandemia e Figura 41 apresentam as redes sociais derivadas dos *registros* de eventos do macroprocesso Gestão de Carreira Docente e com as métricas baseadas em transferência de trabalho. Verifica-se visualmente, por exemplo, a redução do número de departamento, número de arcos e o grau médio dos departamentos.

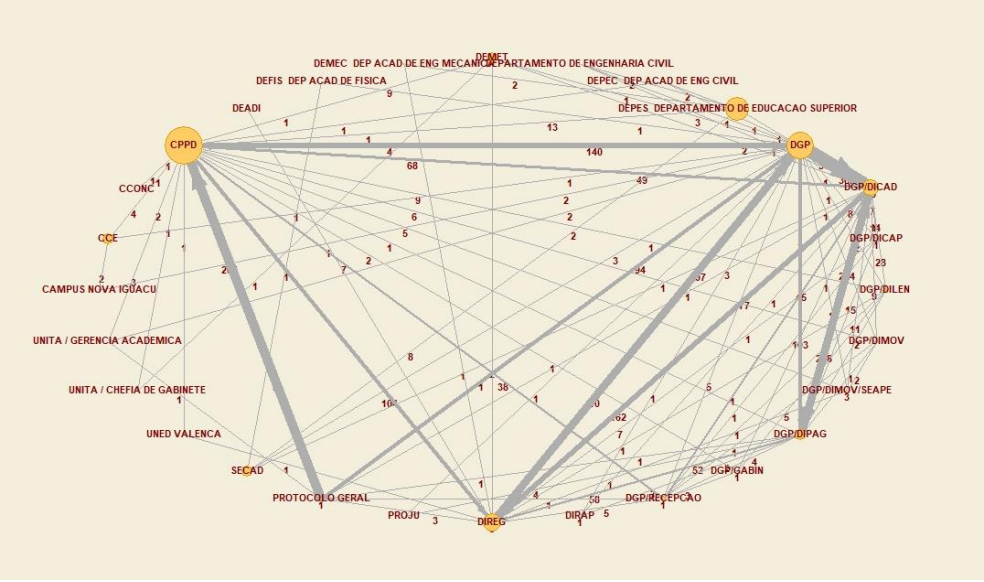


Figura 40 - Rede Social do macroprocesso Gestão de Carreira Docente antes da pandemia

Fonte: Própria (2022)

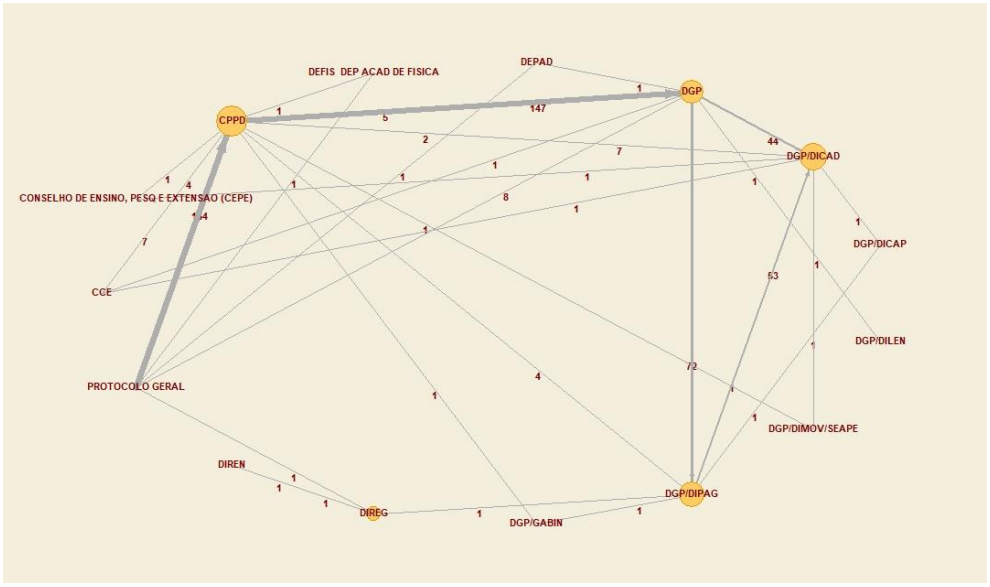
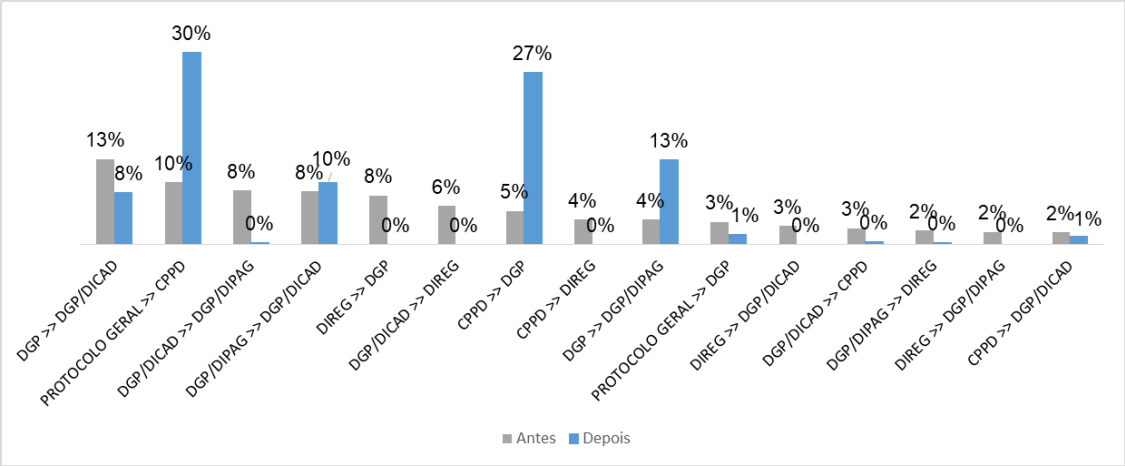


Figura 41 - Rede Social do macroprocesso Gestão de Carreira Docente depois da pandemia
Fonte: Própria (2022)

Através destas redes observou-se que houve alteração no arco de maior peso. Antes da pandemia era “DGP > DGP/DICAD“ com 362 transferências de trabalho, representando 13% do total. Este arco “DGP > DGP/DICAD“ existe depois da pandemia, porém, ele representa 8% do total de transferências de trabalho. E no período pós pandemia o arco de maior peso é o “PROTOCOLO GERAL > CPPD” com 164 transferências de trabalho representando 30% do total. Este arco existia antes da pandemia e era o segundo arco mais relevante representando 10% do total.

Para um melhor entendimento desta análise e um detalhamento mais profundo dos arcos, foi feita a frequência relativa dos arcos em cada período, ou seja, calculou-se o peso (a frequência) de cada arco sobre o peso total. E então, foi feita uma extração dos arcos mais relevantes, utilizando a lei de Pareto, antes da pandemia e depois da pandemia. Esta extração está representada no Gráfico 15:

Gráfico 15 - Arcos mais relevantes do macroprocesso de Gestão de Carreira Docente antes e depois da pandemia



Fonte: Própria (2022)

Percebeu-se nestas relações os seguintes pontos importantes:

- 10 arcos selecionados antes da pandemia que desapareceram ou reduziu-se para menos que 1%. São eles: DGP/DICAD >> DGP/DIPAG, DIREG >> DGP, DGP/DICAD >> DIREG, CPPD >> DIREG, PROTOCOLO GERAL >> DGP, DIREG >> DGP/DICAD, DGP/DICAD >> CPPD, DGP/DIPAG >> DIREG, DIREG >> DGP/DIPAG e CPPD >> DGP/DICAD.
- 4 arcos que foram selecionados antes da pandemia e tiveram sua frequência relativa aumentada depois da pandemia são: PROTOCOLO GERAL >> CPPD, DGP/DIPAG >> DGP/DICAD, CPPD >> DGP e DGP >> DGP/DIPAG

De modo geral antes da pandemia eram necessários 15 arcos (representados pelas quinze barras cinzas no Gráfico 15 para cobrir 81% do total dos pesos dos arcos. E após a pandemia são necessários apenas 4 arcos (representados pelas cinco barras azuis com a frequência relativa maior ou igual a 10% no Gráfico 15) para cobrir 80% do total dos pesos dos arcos. Assim, apenas 4 arcos mantêm-se relevantes após a pandemia. Houve a simplificação de trâmite entre os departamentos pois foram eliminados 10 arcos relevantes.

Para analisar no detalhe a relação entre os departamentos antes e depois da pandemia e na identificação do departamento mais central das redes calculou-se as métricas de centralidade de cada vértice: grau de entrada, grau de saída e intermediação. A Figura 42 apresenta essas métricas de cada departamento envolvido antes e depois da pandemia e ela foi construída conforme a Figura 36.

Percebe-se que o departamento principal para estas redes se manteve o mesmo que é o “CPPD”. Logo, ele é o departamento central antes e depois da pandemia. O

segundo e o terceiro mais centrais antes da pandemia são “DGP” e o “DGP/DICAD”, respectivamente. No período pós pandemia o segundo departamento mais central “DGP/DICAD”, pois, ele tem as segundas maiores de todas as métricas. O terceiro departamentos mais central é “DGP/DIPAG”. Diferentemente da Gestão de Carreira TAE, este macroprocesso de Gestão de Carreira Docente o departamento CPPD é um departamento de entrada das solicitações diferentes dos demais por não ser o departamento padrão de “Protocolo Geral”.

De modo geral, assim como ocorre com a Gestão de Carreira TAE, nota-se que os nós mais centrais fazem parte da DGP (Diretoria de Gestão de Pessoas) a qual é responsável pelos casos dos processos de gestão de carreira. Consequentemente, na rede das Figuras 40Figura 41 - Rede Social do macroprocesso Gestão de Carreira Docente depois da pandemia e 41 é possível verificar que os arcos que os envolvem tem maior peso, ou seja, tem maior frequência da transferência de trabalho. Já os departamentos com valores menores de grau são de unidades externas, departamentos acadêmicos ou de diretorias variadas que desempenha um papel de forma mais isolada e não estão envolvidos em muitas transferências de trabalho.

DEPARTAMENTO	Grau de Entrada		Grau de Saída		Intermediação	
	Antes	Depois	Antes	Depois	Antes	Depois
CPPD	19	7	13	7	0,51	0,34
DGP	19	5	12	4	0,23	0,19
DGP/DICAD	14	6	10	5	0,09	0,26
DIREG	11	3	9	1	0,11	0,07
DGP/RECEPCAO	10		4	0	0,01	0,00
DGP/DIPAG	6	6	10	3	0,04	0,21
DGP/DICAP	5	1	5	1	0,00	0,00
DGP/DIMOV/SEAPE	4	1	3	1	0,00	0,00
DGP/DILEN	4	1	6	1	0,01	0,00
PROTOCOLO GERAL	4	0	8	5	0,01	0,00
DEPES DEPARTAMENTO DE EDUCACAO SUPERIOR	4		7	0	0,18	0,00
DGP/GABIN	3	1	4	1	0,00	0,00
DIRAP	3		4	0	0,00	0,00
DGP/DIMOV	2		4	0	0,00	0,00
UNITA / GERENCIA ACADEMICA	2		2	0	0,00	0,00
PROJU	2		0	0	0,00	0,00
CCE	1	1	3	3	0,03	0,00
DEFIS DEP ACAD DE FISICA	1	1	1	1	0,00	0,00
DEMET	1		5	0	0,03	0,00
SECAD	1		4	0	0,03	0,00
DEMEC DEP ACAD DE ENG MECANICA	1		2	0	0,00	0,00
DEPEC DEP ACAD DE ENG CIVIL	1		2	0	0,00	0,00
UNITA / CHEFIA DE GABINETE	1		1	0	0,00	0,00
UNED VALENCA	1		1	0	0,00	0,00
CCONC	1		1	0	0,00	0,00
CAMPUS NOVA IGUACU	1		1	0	0,00	0,00
DEPARTAMENTO DE ENGENHARIA CIVIL	1		2	0	0,00	0,00
DEADI	1		0	0	0,00	0,00
DEPAD	0	1	0	1	0,00	0,00
CONSELHO DE ENSINO, PESQ E EXTENSAO (CEPE)	0	1	0	1	0,00	0,00
DIREN	0	1	0	1	0,00	0,00

Figura 42 - Lista de departamentos do macroprocesso de Gestão de Carreira Docente e suas métricas de centralidade antes e depois da pandemia
Fonte: Própria (2022)

Para detalhar e aprofundar no entendimento da participação de cada departamento, a Figura 43 apresenta-se dois cálculos de % de frequência e ela foi construída conforme a Figura 37.

DEPARTAMENTO	Frequência Relativa - Casos		Frequência Relativa - Eventos	
	Antes	Depois	Antes	Depois
CPPD	91%	91%	106%	97%
DGP	87%	67%	143%	70%
DGP/DICAD	88%	4%	161%	4%
DIREG	71%	1%	98%	1%
DGP/RECEPCAO	15%	0%	16%	0%
DGP/DIPAG	80%	31%	89%	31%
DGP/DICAP	2%	1%	5%	1%
DGP/DIMOV/SEAPE	2%	1%	2%	1%
DGP/DILEN	7%	1%	9%	1%
PROTOCOLO GERAL	100%	100%	101%	100%
DEPES DEPARTAMENTO DE EDUCACAO SUPERIOR	4%	0%	5%	0%
DGP/GABIN	2%	1%	2%	1%
DIRAP	1%	0%	2%	0%
DGP/DIMOV	2%	0%	3%	0%
UNITA / GERENCIA ACADEMICA	1%	0%	1%	0%
PROJU	0%	0%	1%	0%
CCE	1%	3%	1%	3%
DEFIS DEP ACAD DE FISICA	0%	1%	0%	1%
DEMET	2%	0%	2%	0%
SECAD	2%	0%	2%	0%
DEMEC DEP ACAD DE ENG MECANICA	0%	0%	1%	0%
DEPEC DEP ACAD DE ENG CIVIL	1%	0%	1%	0%
UNITA / CHEFIA DE GABINETE	0%	0%	0%	0%
UNED VALENCA	0%	0%	0%	0%
CCONC	0%	0%	0%	0%
CAMPUS NOVA IGUACU	1%	0%	1%	0%
DEPARTAMENTO DE ENGENHARIA CIVIL	1%	0%	1%	0%
DEADI	0%	0%	0%	0%
DEPAD	0%	1%	0%	1%
CONSELHO DE ENSINO, PESQ E EXTENSAO (CEPE)	0%	1%	0%	1%
DIREN	0%	1%	0%	1%

Figura 43 - Lista de departamentos do macroprocesso de Gestão de Carreira Docente e suas frequências relativas antes e depois da pandemia
Fonte: Própria (2022)

Identificou-se que apesar do departamento “Protocolo Geral” não ser destaque nas métricas de centralidades de grau ele é o único que aparece em 100% das instâncias de processos nos períodos antes e depois da pandemia. Na “Frequência Relativa - Casos” DGP/DIPAG antes da pandemia ocorreu em 80% das instâncias de processos e apenas 31% no período da pandemia. Importante destacar também que na “Frequência Relativa - Eventos”, no período antes da pandemia, os departamentos “DGP/DICAD”, “DGP” e “CPPD” tem um alto percentual. Isso mostra um número de

eventos muito maior que o número de instâncias de processos, podendo indicar possíveis gargalos e retrabalhos. Já para os processos depois da pandemia este comportamento não ocorreu, mesmo nos nós mais centrais.

Complementarmente, o fluxo do processo de Gestão Carreira Docente entre os departamentos antes da pandemia é apresentado na Figura 44. Percebe-se que conforme visto através das análises de redes sociais a maioria das instâncias dos processos chegam pelo CPPD e que DGP/DICAD é o departamento que passa o maior número de instâncias de processos. Nota-se também que há um fluxo grande chegando no “Protocolo Geral”, mesmo ele não sendo o principal departamento de entrada deste processo. Porém dos 257 processos chegando pelo CPPD esta mesma quantidade é enviada até o “Protocolo Geral”. Observa-se vários caminhos possíveis para resolução das solicitações de Gestão de Carreira Docente

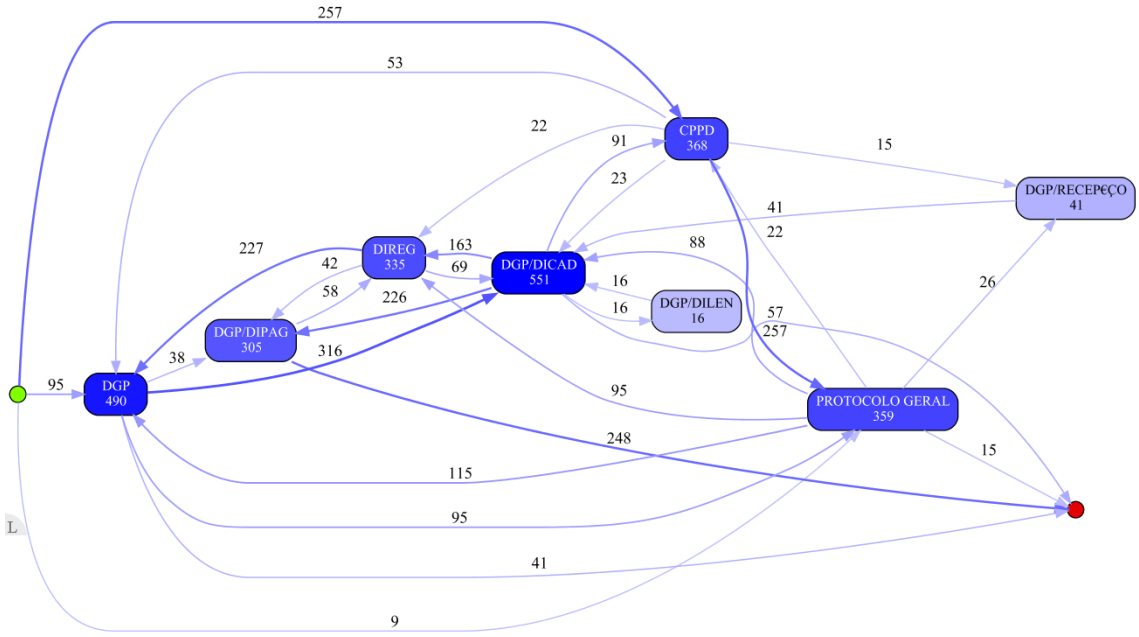


Figura 44 - Fluxo do processo de Gestão Carreira Docente
entre os departamentos antes da pandemia
Fonte: Própria (2022)

O fluxo de processo da Gestão Carreira Docente pós pandemia é mais simples e direto, conforme a Figura 45. Percebe-se que permanece o CPPD como o departamento de entrada para estas solicitações. Este por sua vez encaminha para o Protocolo Geral e a partir daí tem 3 caminhos possíveis: fechar, encaminhar para DGP e fechar ou encaminhar para DGP, DGP/DIPAG e fechar.

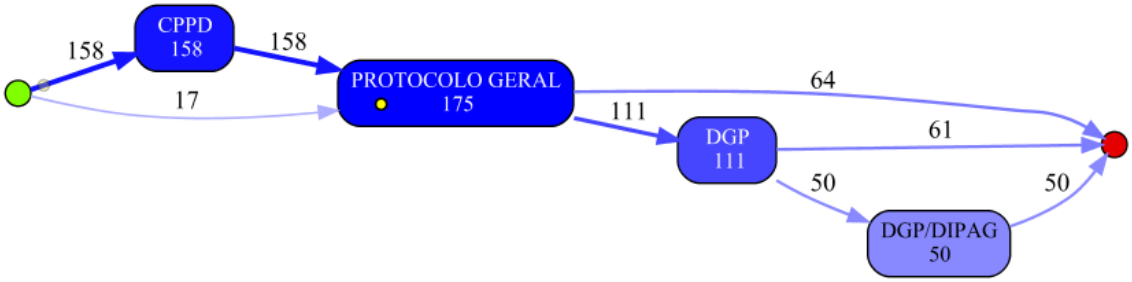


Figura 45 - Fluxo do processo de Gestão Carreira Docente
entre os departamentos depois da pandemia
Fonte: Própria (2022)

6.6.5. Avaliação – Gestão Carreira Docente

Pela mesma forma da Gestão de Carreira TAE, ficou evidente que houve simplificação no onde mesmo arcos e departamentos importantes antes da pandemia foram impactados. Neste processo a maioria dos vértices centrais permaneceram os mesmos entre os dois períodos analisados. Redução significante de possíveis gargalos no departamento DGP e DGP/DIPAG. Este departamento DGP/DIPAG, é responsável pelos pagamentos do CEFET, nele tinham 80% dos processos antes da pandemia e depois da pandemia em torno de 30%. Avaliou-se juntamente com o especialista, e estes são fortes indícios de que não estavam registrando oficialmente todas as etapas no sistema de Protocolo Geral e também tramitadas por subnotificação por e-mail. Ainda assim, algumas tarefas foram movidas e relatadas as suas atividades. E que as listas de verificações podem ter sido criadas para facilitar ao solicitantes dos documentos prévio necessários para dar continuidade ao processo com maior fluidez.

E outra suposição é que alguns departamento pode ter sido incorporado para dentro do DGP onde os lançamentos não estão sendo feitos a nível de departamento mas sim apenas a nível da diretoria. Em conversas com o especialista entendeu-se que na criação do sistema de protocolo geral, foram criados os nomes das diretorias para localizar prontamente o processo. A única diretoria que solicitou que fossem abertos mais níveis de seus departamentos foi a DGP (Diretoria de Gestão de Pessoas). Entendeu-se que esta é uma boa prática para identificar com mais precisão aonde está o processo e também para fazer estudos aprofundados como estes de tempos e movimentos. Apontando-se departamentos com gargalos ou com folgas e até mesmo quando há incorporação de etapas. Em outras diretorias que não foram quebradas desta

mesma forma não foi possível aprofundar nas análises.

6.6.6. Mineração de Processo e Análises – Folha de Pessoal

Na Tabela 27 é apresentada alguns dados gerais do macroprocesso Folha de Pessoal antes e depois da pandemia. Percebe-se que houve uma redução aproximadamente de 65% no número de processos. E também houve redução de 30% na duração média e de 25% número médio de etapas dos processos.

Tabela 27 – Visão Geral do macroprocesso de Folha de Pessoal

Folha de Pessoal	Antes	Depois
Número de processos	103	36
Média de duração (dias)	68,9	48,3
Média do nº de etapas	7,0	5,3

Fonte: Própria (2022)

Para a análise da rede de Folha de Pessoal foram analisadas as métricas apresentadas na Tabela 28. Após a pandemia observa-se que houve menos 11 departamentos envolvidos, 81 menos arcos e 8 menos departamentos com loop. A frequência dos pesos dos arcos também reduziu. O grau médio indica o número de departamentos em média que um departamento se relaciona, ou seja, antes era em torno de 9,42 departamentos e passou a ser apenas 4,92. A densidade e a medida de centralidade intermediação da rede entre os períodos ficaram próximos.

Tabela 28 - Métricas Gerais das redes do macroprocesso de Folha de Pessoal

Métricas gerais das redes	Antes	Depois
Número de departamentos (vértices)	24	13
Número de arcos	113	32
Total dos pesos dos arcos	721	190
Número de departamentos Com loops	9	2
Densidade da Rede	0,20	0,19
Grau médio da rede	9,42	4,92
Intermediação da Rede	0,28	0,32

Fonte: Própria (2022)

As Figura 46 e Figura 47 apresentam as redes sociais derivadas dos *registros* de eventos do macroprocesso Folha de Pessoal e com as métricas baseadas em transferência de trabalho. Verifica-se visualmente, por exemplo, a redução do número de departamento, número de arcos e o grau médio.

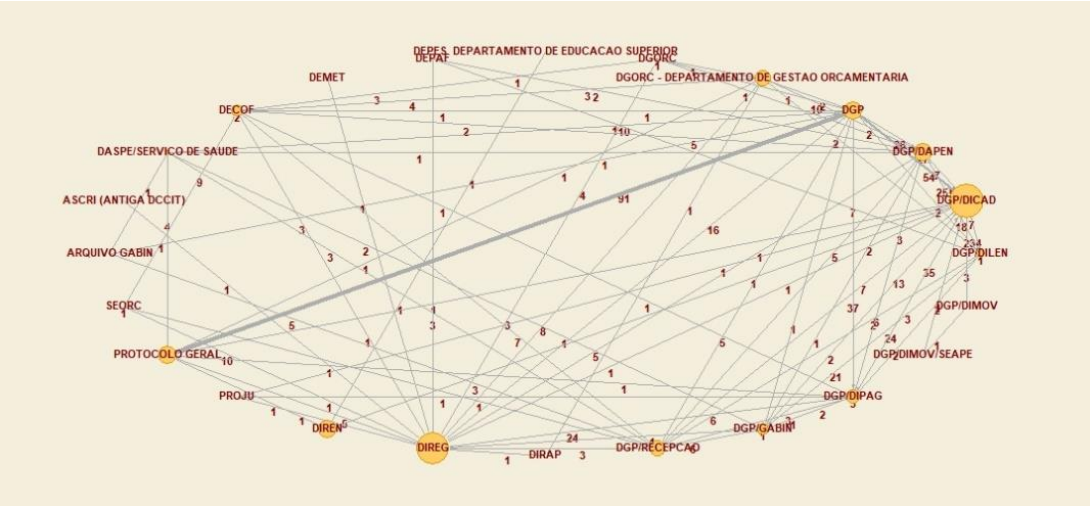


Figura 46 – Rede Social do macroprocesso Folha de Pessoal Antes da Pandemia
Fonte: Própria (2022)

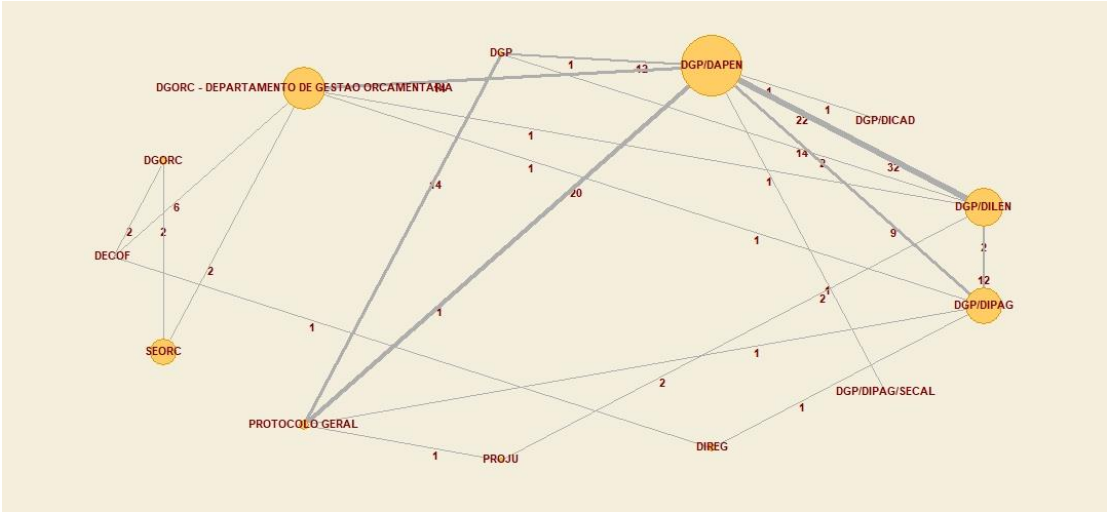


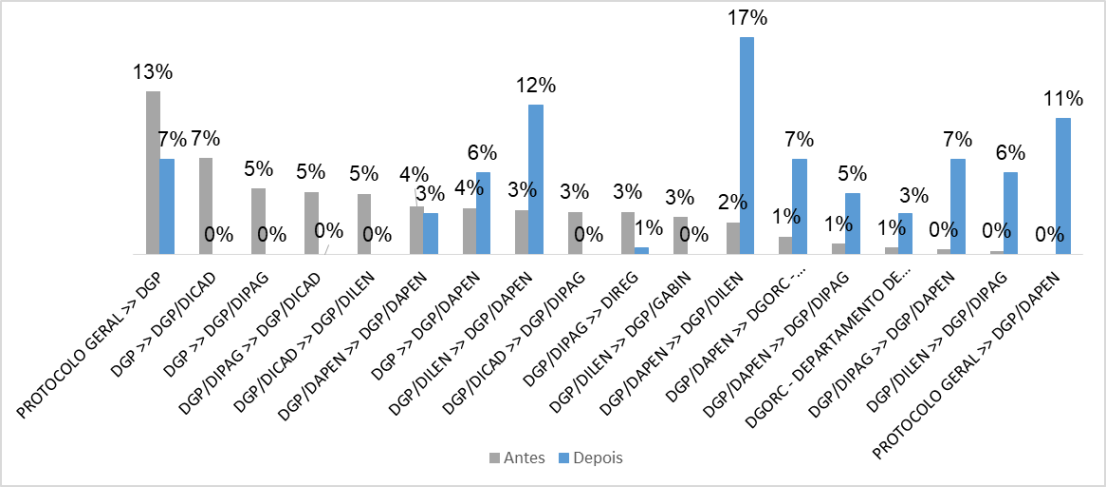
Figura 47 - Rede Social do macroprocesso Folha de Pessoal depois da pandemia
Fonte: Própria (2022)

Através destas redes observou-se que houve alteração no arco de maior peso. Antes da pandemia era “PROTOCOLO GERAL > DGP“ com 91 transferências de trabalho, representando 13% do total. Este arco existe depois da pandemia, porém, ele representa 7% do total de transferências de trabalho. E no período pós pandemia o arco de maior peso é o “DGP/DAPEN > DGP/DILEN” com 32 transferências de trabalho

representando 17% do total. Este arco existia antes da pandemia representa apenas 2% do total.

Para um melhor entendimento desta análise e um detalhamento mais profundo dos arcos, foi feita a frequência relativa dos arcos em cada período, ou seja, calculou-se o peso (a frequência) de cada arco sobre o peso total. E então, foi feita uma extração dos arcos com a frequência maior que 3% antes da pandemia e depois da pandemia. Nesta análise não foi utilizada a lei de Pareto para selecionar os arcos, pois as frequências estão muito distribuídas antes da pandemia. Para se ter 80% dos arcos seriam necessários selecionar 32 arcos, dos quais 21 tem a frequência relativa igual ou menos a 2%. Esta extração está representada no Gráfico 16:

Gráfico 16 - Arcos com frequência relativa igual ou maior que 3% do macroprocesso de Folha de Pessoal antes e depois da pandemia



Fonte: Própria (2022)

Percebeu-se nestas relações os seguintes pontos importantes:

- 7 arcos relevantes antes da pandemia que desapareceram ou reduzindo-se significativamente sua frequência relativa. São eles: DGP >> DGP/DICAD, DGP >> DGP/DIPAG, DGP/DIPAG >> DGP/DICAD, DGP/DICAD >> DGP/DILEN, DGP/DICAD >> DGP/DIPAG, DGP/DIPAG >> DIREG e DGP/DILEN >> DGP/GABIN.
- 2 arcos que tinham a frequência igual a 4% antes da pandemia e aumentaram mais sua frequência relativa depois da pandemia que são: DGP >> DGP/DAPEN e DGP/DILEN >> DGP/DAPEN

- E 7 arcos que não estavam dentro dos arcos mais importantes da lei de Pareto antes da pandemia, tornaram-se relevante no período pós pandemia. Que são: DGP/DAPEN >> DGP/DILEN, DGP/DAPEN >> DGORC - DEPARTAMENTO DE GESTÃO ORÇAMENTÁRIA, DGP/DAPEN >> DGP/DIPAG, DGORC - DEPARTAMENTO DE GESTÃO ORÇAMENTÁRIA >> DECOF, DGP/DIPAG >> DGP/DAPEN, DGP/DILEN >> DGP/DIPAG e PROTOCOLO GERAL >> DGP/DAPEN

De modo geral antes da pandemia eram necessários 32 arcos (representados pelas onze primeiras barras cinzas no Gráfico 16) para cobrir 80% do total dos pesos dos arcos. E após a pandemia são necessários apenas 10 arcos (representados pelas barras azuis no Gráfico 16) para cobrir 84% do total dos pesos dos arcos.

Para analisar no detalhe a relação entre os departamentos e na identificação do departamento mais central das redes calculou-se as métricas de centralidade de cada vértice: grau de entrada, grau de saída e intermediação. A Figura 48 apresenta essas métricas de cada departamento envolvido antes e depois da pandemia e ela foi construída conforme a Figura 36.

DEPARTAMENTO	Grau de Entrada		Grau de Saída		Intermediação	
	Antes	Depois	Antes	Depois	Antes	Depois
DGP/DICAD	11	1	9	1	0,32	0,00
DIREG	11	1	14	1	0,29	0,01
DGP/DAPEN	10	7	11	8	0,09	0,36
DGP/DIPAG	10	5	9	5	0,05	0,13
DGP	8	2	7	2	0,08	0,00
DECOF	8	3	9	0	0,05	0,00
DGP/GABIN	8	0	7	0	0,04	0,00
DGP/DILEN	7	4	7	4	0,02	0,14
DGP/RECEPCAO	6	0	6	0	0,07	0,00
DASPE/SERVICO DE SAUDE	5	0	2	0	0,00	0,00
DGORC - DEPARTAMENTO DE GESTAO ORCAMENTO	5	3	6	3	0,07	0,18
SEORC	3	1	4	1	0,00	0,07
DGP/DIMOV/SEAPE	2	0	0	0	0,00	0,00
PROJU	2	2	1	1	0,00	0,00
DGORC	5	1	4	1	0,00	0,01
PROTOCOLO GERAL	2	1	6	4	0,09	0,01
DEPAF	2	0	1	0	0,00	0,00
DIREN	2	0	2	0	0,08	0,00
ARQUIVO GABIN	1	0	3	0	0,00	0,00
ASCRI (ANTIGA DCCIT)	1	0	1	0	0,00	0,00
DGP/DIMOV	1	0	2	0	0,00	0,00
DIRAP	1	0	1	0	0,00	0,00
DEMET	1	0	0	0	0,00	0,00
DEPES DEPARTAMENTO DE EDUCACAO SUPERIOR	1	0	1	0	0,00	0,00

Figura 48 - Lista de departamentos do macroprocesso de Folha de Pessoal e suas métricas de centralidade antes e depois da pandemia
Fonte: Própria (2022)

Percebe-se que houve mudança no departamento central. O departamento “DIREG” antes da pandemia possuiu os maiores valores nas métricas de grau de entrada e de saída e o segundo maior valor do grau de intermediação. Em segundo lugar, “DGP/DICAD”, com os maiores valores de grau de entrada e intermediação e o terceiro maior valor do grau de saída. E em terceiro lugar é o “DGP/DAPEN”. Após a pandemia “DGP/DAPEN” é o departamento central, pois ele tem os maiores valores de todas as métricas. E segundo e terceiro lugar são “DGP/DIPAG” e “DGP/DILEN”.

De modo geral nota-se que os nós mais centrais fazem parte da DGP (Diretoria de Gestão de Pessoas) a qual é responsável pelos casos dos processos de gestão de pessoas e folha de pagamento. E também é possível perceber “DIREG” e ”DGORC” são departamentos que se destacam nas métricas de centralidade. Já os departamentos com valores menores de grau são de diretorias variadas que desempenha um papel de forma mais isolada e não estão envolvidos em muitas transferências de trabalho.

Para detalhar e aprofundar no entendimento da participação de cada departamento, a Figura 49 apresenta dois cálculos de % de frequência e ela foi construída conforme a Figura 37.

DEPARTAMENTO	Frequência Relativa - Casos		Frequência Relativa - Eventos	
	Antes	Depois	Antes	Depois
DGP/DICAD	54%	3%	83%	3%
DIREG	44%	3%	50%	3%
DGP/DAPEN	27%	92%	77%	181%
DGP/DIPAG	54%	42%	74%	56%
DGP	93%	39%	122%	39%
DECOF	16%	0%	21%	0%
DGP/GABIN	26%	0%	28%	0%
DGP/DILEN	52%	86%	57%	103%
DGP/RECEPCAO	16%	0%	24%	0%
DASPE/SERVICO DE SAUDE	8%	0%	11%	0%
DGORC - DEPARTAMENTO DE GESTAO ORCAMENTO	14%	25%	17%	25%
SEORC	11%	6%	12%	6%
DGP/DIMOV/SEAPE	1%	0%	1%	0%
PROJU	1%	3%	1%	6%
DGORC	6%	6%	8%	6%
PROTOCOLO GERAL	100%	100%	100%	100%
DEPAF	3%	0%	3%	0%
DIREN	1%	0%	2%	0%
ARQUIVO GABIN	5%	0%	5%	0%
ASCRI (ANTIGA DCCIT)	1%	0%	1%	0%
DGP/DIMOV	3%	0%	3%	0%
DIRAP	1%	0%	1%	0%
DEMET	1%	0%	1%	0%
DEPES DEPARTAMENTO DE EDUCACAO SUPERIOR	1%	0%	1%	0%

Figura 49 - Lista de departamentos do macroprocesso Folha de Pessoal e suas frequências relativas antes e depois da pandemia
Fonte: Própria (2022)

Analisou-se que apesar do departamento “Protocolo Geral” não ser destaque nas métricas de centralidades de grau ele é o único que aparece em 100% das instâncias de processos nos períodos antes e depois da pandemia. Os departamentos ”DGP/DAPEN” e “DGP/DILEN” tiveram um aumento no número de casos que passou por eles. E houve pequeno aumento na “DGORC – DEPARTAMENTO DE GESTAO ORCAMENTO”. Os departamentos que deixaram de ser centrais, consequentemente teve menos participação nos casos. E claramente, os departamentos que tornaram mais centrais também aumentou significamente. Importante destacar também que na “Frequência Relativa - Eventos”, no período antes da pandemia o departamento “DGP” e o departamento DGP/DAPEN depois da pandemia, tem valores acima de 100%. Isso

indicar possíveis gargalos e retrabalhos.

Para complementar estas análises, Figura 50 Figura 51 apresentam o fluxo do processo de Folha de Pessoal antes e depois da pandemia respectivamente. Percebe-se que antes da pandemia havia um fluxo muito maior para trâmite das solicitações. Esta Figura 50 é conhecida como fluxo de processo "*Espaguete*", pois são muitos caminhos possíveis não claros. Após a pandemia há um fluxo do tipo lasanha que é legível e são mais claros os caminhos possíveis.

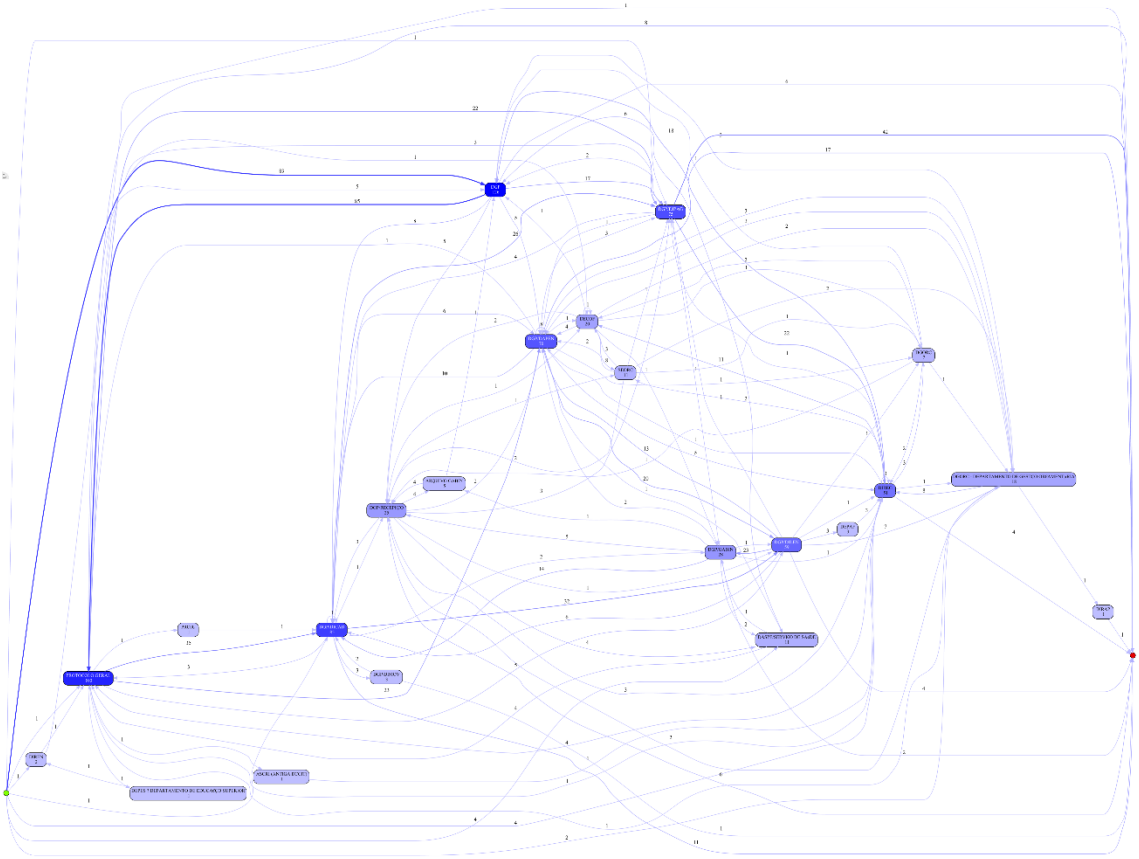


Figura 50 - Fluxo do processo de Folha de Pessoal antes da pandemia
Fonte: Própria (2022)

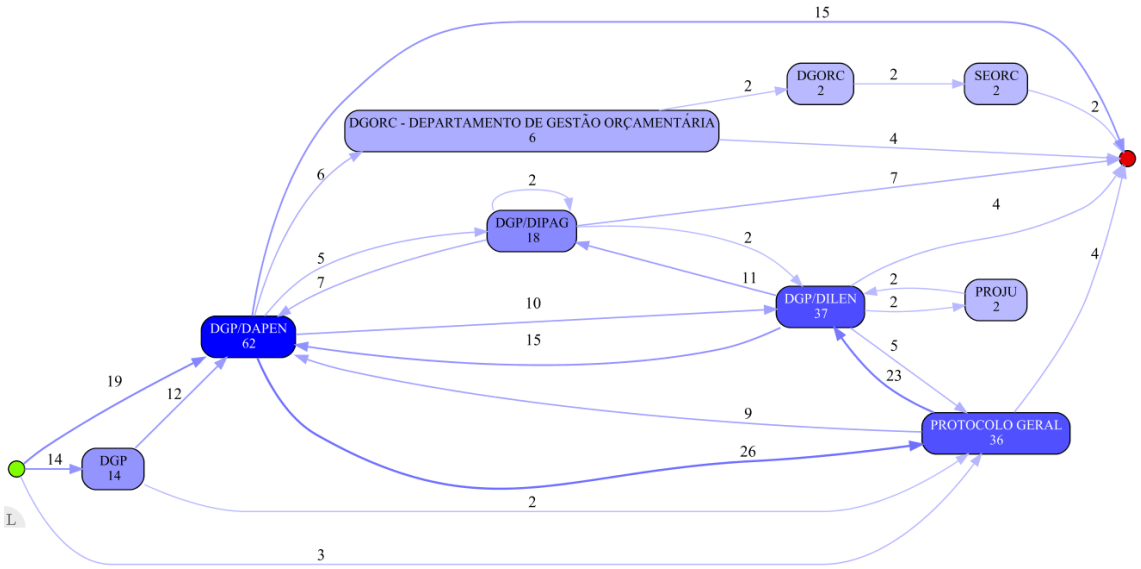


Figura 51 - Fluxo do processo de Folha de Pessoal depois da pandemia
Fonte: Própria (2022)

6.6.7. Avaliação – Folha de Pessoal

Através do fluxo de processo, fica evidente que houve simplificação no processo e que ocorreram trocas entre os departamentos mais centrais e reduziu arcos importantes. Isso resultou na diminuição da duração em dias e número de etapas. É importante dar atenção ao novo departamento central “DGP/DAPEN” E “DGP/DILEN”, indicando possíveis retrabalhos ou um candidato de gargalo no processo. Analisando juntamente com o especialista, estes departamentos são referentes a Divisão de Aposentados e Pensão e Divisão de Legislação e Norma, respectivamente. Logo, há suposição de aumento neste departamento devido a era pandêmica. O mesmo ressaltou que o leve aumento nas transferências de trabalho no DGORC (DEPARTAMENTO DE GESTAO ORCAMENTO) pode ser devido a exigência legal pelo governo de lançar notas de empenho a parte. E outro ponto levantado que também pode contribuir são as assinaturas de papel que foram substituídas por assinaturas de chave pública, uma vez que o Brasil implementou certificados digitais de acordo com a lei comum.

E por fim, para avaliar essa simplificação, a suposição indicada foi de que a organização digitalizou cada etapa desse processo e que também pode ter tido etapas de subnotificação por e-mail. Um ponto importante apontado pelo especialista é que esse arcabouço de processos administrativos é pautado nas leis e são evoluídos a partir delas. O qual houve-se pouca ou nenhuma construção destes com uma visão otimizada

assim como forçou alguma atualização mais brusca com o trabalho em casa na pandemia.

6.6.8. Mineração de Processo e Análises – Reembolso

Na Tabela 29 é apresentada alguns dados gerais do macroprocesso de Reembolso e Pagamentos antes e depois da pandemia. Percebe-se que houve uma redução aproximadamente de 85% no número de processos e de 76% na duração média. O número médio de etapas dos processos ficou próximo entre os períodos, porém houve um pequeno aumento.

Tabela 29 – Visão Geral do macroprocesso de Reembolso e Pagamentos

Reembolso e Pagamentos	Antes	Depois
Número de processos	128	19
Média de duração (dias)	49,4	11,9
Média do nº de etapas	3,4	3,7

Fonte: Própria (2022)

Para a análise da rede de Reembolso e Pagamentos foram analisadas as métricas apresentadas na Tabela 30. Após a pandemia observa-se que houve menos 10 departamentos envolvidos, 23 menos arcos e 4 menos departamentos com loop. A frequência dos pesos dos arcos e o grau de intermediação da rede também reduziu. O a densidade e o grau médio aumentaram mas não foi expressivo, porém houve mais conexões entre os departamentos.

Tabela 30 - Métricas Gerais das redes do macroprocesso de Reembolso e Pagamentos

Métricas gerais das redes	Antes	Depois
Número de departamentos (vértices)	22	12
Número de arcos	5	1
Total dos pesos dos arcos	431	71
Número de departamentos Com loops	58	35
Densidade da Rede	0,12	0,24

Grau médio da rede	5,27	5,83
Intermediação da Rede	0,44	0,27

Fonte: Própria (2022)

As Figura 52 Figura 53 apresentam as redes sociais derivadas dos *registros* de eventos do macroprocesso Reembolso e Pagamentos e com as métricas baseadas em transferência de trabalho. Verifica-se visualmente, por exemplo, a redução do número de departamento e número de arcos.

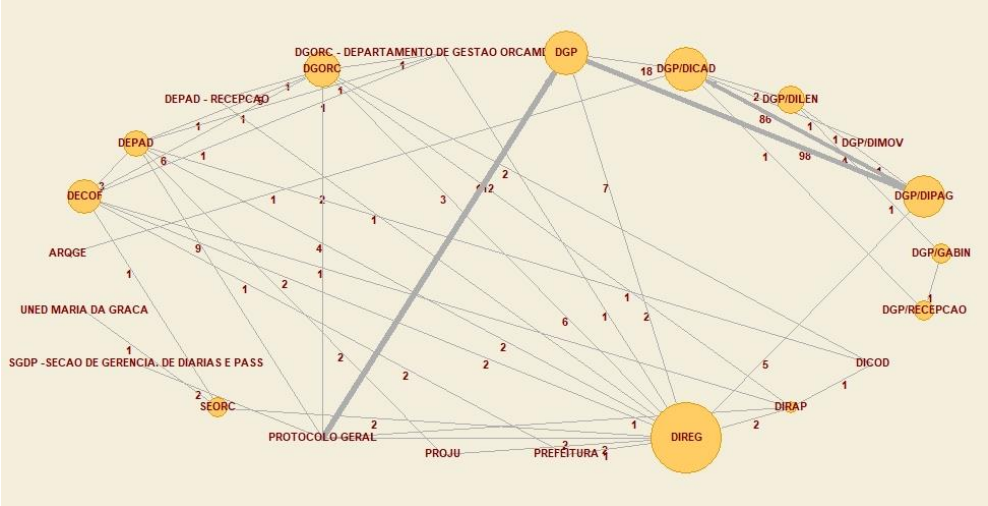


Figura 52 – Rede Social do macroprocesso Reembolso e Pagamentos antes da Pandemia
Fonte: Própria (2022)

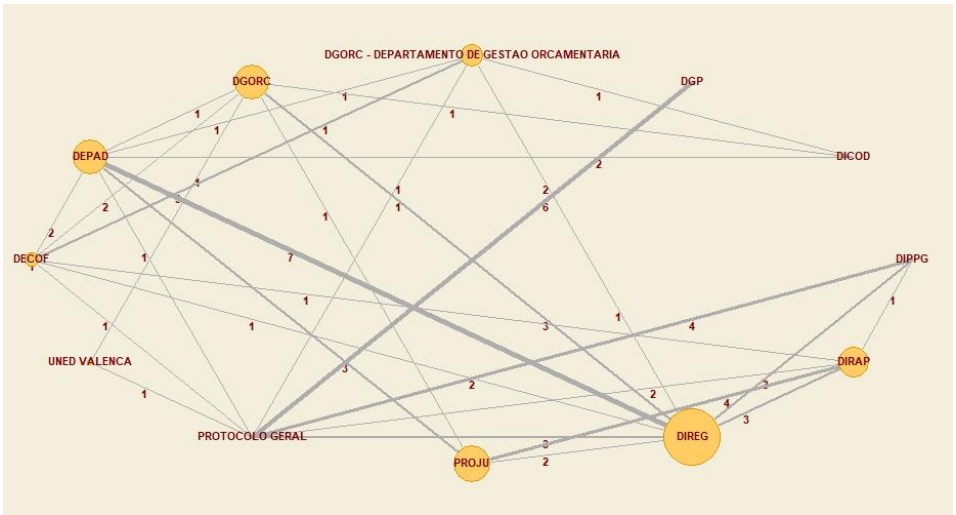


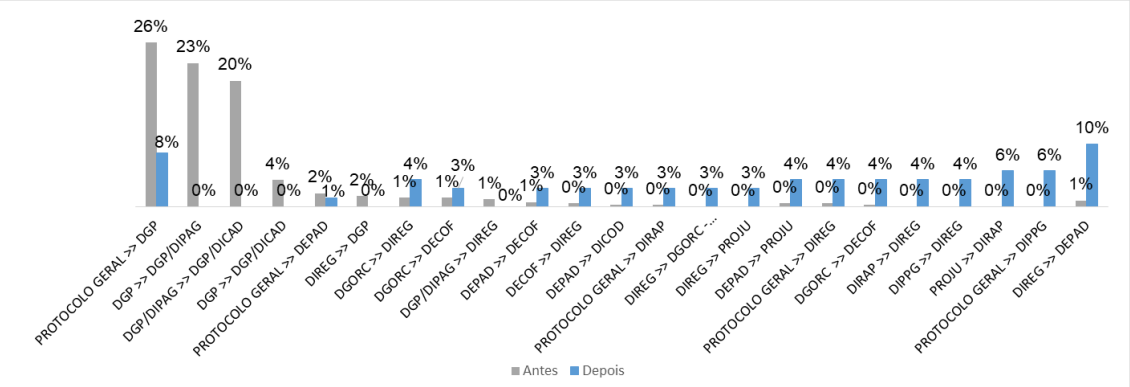
Figura 53 - Rede Social do macroprocesso Reembolso e Pagamentos depois da pandemia

Fonte: Própria (2022)

Através destas redes observou-se que houve alteração no arco de maior peso. Antes da pandemia era “PROTOCOLO GERAL > DGP“ com 112 transferências de trabalho, representando 26% do total. Este arco existe depois da pandemia, porém, ele representa apenas 8% do total de transferências de trabalho. E no período pós pandemia o arco de maior peso é o “DIREG > DEPAD” com 7 transferências de trabalho representando 10% do total. Este arco existia antes da pandemia representando apenas 1% do total.

Para um melhor entendimento desta análise e um detalhamento mais profundo dos arcos, foi feita a frequência relativa dos arcos em cada período, ou seja, calculou-se o peso (a frequência) de cada arco sobre o peso total. E então, foi feita uma extração dos arcos mais relevantes, utilizando a lei de Pareto, antes da pandemia e depois da pandemia. Esta extração está representada no Gráfico 17:

Gráfico 17 - Arcos mais relevantes do macroprocesso de Reembolso e Pagamentos antes e depois da pandemia



Fonte: Própria (2022)

Percebeu-se nestas relações os seguintes pontos importantes:

- 6 arcos selecionados pela lei de Pareto antes da pandemia que desapareceram ou reduziram sua frequência. São eles: DGP >> DGP/DIPAG, DGP/DIPAG >> DGP/DICAD, DGP >> DGP/DICAD, PROTOCOLO GERAL >> DEPAD, DIREG >> DGP e DGP/DIPAG >> DIREG
- 3 arcos que foram selecionados antes da pandemia e aumentaram sua frequência relativa depois da pandemia que são: DGORC >> DIREG, DGORC >> DECOF e DEPAD >> DECOF

- 14 arcos que não foram selecionados da lei de Pareto antes da pandemia, tornaram-se relevante no período pós pandemia. E estes no pós pandemia representam 59% da frequência dos pesos dos arcos. Que são: DEPAD >> DECOF, DECOF >> DIREG, DEPAD >> DICOD, PROTOCOLO GERAL >> DIRAP, DIREG >> DGORC - DEPARTAMENTO DE GESTÃO ORÇAMENTÁRIA, DIREG >> PROJU
- DEPAD >> PROJU, PROTOCOLO GERAL >> DIREG, DGORC >> DECOF, DIRAP >> DIREG, DIPPG >> DIREG, PROJU >> DIRAP, PROTOCOLO GERAL >> DIPPG e DIREG >> DEPAD

De modo geral antes da pandemia eram necessários apenas 3 arcos para cobrir 63% do total dos pesos dos arcos. E após a pandemia são necessários 14 arcos para cobrir 60% do total dos pesos dos arcos. Assim, percebeu-se que a maioria das principais transferências de trabalho foram alteradas e no período pós pandemia houve um maior número de arcos indicando uma descentralização.

Para analisar no detalhe a relação entre os departamentos antes e depois da pandemia e na identificação do departamento mais central das redes calculou-se as métricas de centralidade de cada vértice: grau de entrada, grau de saída e intermediação. A Figura 54 apresenta essas métricas de cada departamento envolvido antes e depois da pandemia e ela foi construída conforme a Figura 36.

DEPARTAMENTO	Grau de Entrada		Grau de Saída		Intermediação	
	Antes	Depois	Antes	Depois	Antes	Depois
DIREG	9	6	9	5	0,50	0,32
DGORC	8	5	5	2	0,12	0,12
DECOF	7	7	4	3	0,11	0,02
DEPAD	4	3	7	5	0,07	0,12
DGP/DIPAG	4	0	4	0	0,18	0,00
DGP/DICAD	4	0	5	0	0,20	0,00
DGORC - DEPART. GESTAO ORCAMENT.	3	5	2	3	0,00	0,05
DGP	2	1	3	0	0,20	0,00
DIRAP	2	2	3	3	0,01	0,10
DICOD	2	1	1	2	0,00	0,00
DGP/DILEN	2	0	2	0	0,07	0,00
PREFEITURA	2	0	1	0	0,00	0,00
ARQGE	1	0	0	0	0,00	0,00
DGP/DIMOV	1	0	0	0	0,00	0,00
PROJU	1	2	1	2	0,00	0,14
UNED MARIA DA GRACA	1	0	0	0	0,00	0,00
DEPAD - RECEPCAO	1	0	1	0	0,00	0,00
SEORC	1	0	2	0	0,04	0,00
DGORC - DEPART. GESTAO ORCAMENT.	1	0	0	0	0,00	0,00
DGP/RECEPCAO	1	0	1	0	0,04	0,00
DGP/GABIN	1	0	1	0	0,04	0,00
PROTOCOLO GERAL	0	0	6	8	0,00	0,00
DIPPG	0	2	0	1	0,00	0,00
UNED VALENCA	0	1	0	1	0,00	0,00

Figura 54 - Lista de departamentos do macroprocesso de Reembolso e Pagamentos e suas métricas de centralidade antes e depois da pandemia
Fonte: Própria (2022)

O departamento “DIREG” antes e pós pandemia possuiu os maiores valores em nas métricas de centralidade analisadas. Logo, ele é o mais departamento central. Antes da pandemia, o segundo departamento mais central é o “DGP/DICAD”, pois teve a segunda maior métrica do grau de intermediação e o quarto maior valor nas métricas de grau de entrada e saída. Em terceiro lugar, é o “DGORC”, pois tem o segundo maior grau de entrada, quarto maior valor do grau de saída e o quinto maior valore do grau de intermediação. Depois da pandemia se destaca em segundo e terceiro lugar mais central os departamentos DEPAD e DECOF, respectivamente.

Para detalhar e aprofundar no entendimento da participação de cada departamento, a Figura 55 apresenta dois cálculos de % de frequência e ela foi construída conforme a Figura 37.

DEPARTAMENTO	Frequência Relativa - Casos		Frequência Relativa - Eventos	
	Antes	Depois	Antes	Depois
DIREG	14%	21%	19%	68%
DGORC	6%	21%	12%	26%
DECOF	4%	16%	5%	21%
DEPAD	8%	26%	13%	47%
DGP/DIPAG	70%	0%	75%	0%
DGP/DICAD	9%	0%	9%	0%
DGORC - DEPART. GESTAO ORCAMENT.	2%	16%	2%	26%
DGP	87%	0%	92%	0%
DIRAP	2%	26%	2%	26%
DICOD	2%	11%	2%	11%
DGP/DILEN	2%	0%	2%	0%
PREFEITURA	1%	0%	1%	0%
ARQGE	1%	0%	1%	0%
DGP/DIMOV	1%	0%	1%	0%
PROJU	2%	21%	2%	26%
UNED MARIA DA GRACA	1%	0%	1%	0%
DEPAD - RECEPCAO	1%	0%	1%	0%
SEORC	2%	0%	2%	0%
DGORC - DEPART. GESTAO ORCAMENT.	1%	0%	1%	0%
DGP/RECEPCAO	1%	0%	1%	0%
DGP/GABIN	1%	0%	1%	0%
PROTOCOLO GERAL	100%	100%	100%	100%
DIPPG	0%	16%	0%	16%
UNED VALENCA	0%	5%	0%	5%

Figura 55 - Lista de departamentos do macroprocesso de Reembolso e Pagamentos e suas frequências relativas antes e depois da pandemia
Fonte: Própria (2022)

Analisou-se que apesar do departamento “Protocolo Geral” não ser destaque nas métricas de centralidades de grau ele é o único que aparece em 100% das instâncias de processos nos períodos antes e depois da pandemia. Na “Frequência Relativa - Casos” DGP/DIPAG e DPG tinham maior frequência e caiu para 0%. Apesar da DIREG ser um nó central antes e depois suas frequências não são muito altas porém, houve um pequeno aumento no pós-pandemia. Também observou-se que houve um aumento da participação do DEORC, DECOF, DEPAD, DIRAP e PROJU.

Para complementar estas análises, as Figura 56 e Figura 57 apresentam os fluxos do processo de Reembolso antes e depois da pandemia respectivamente. Percebe-se que antes da pandemia, ao contrário dos demais processos, havia um fluxo mais simples para trâmite das solicitações de Reembolso.

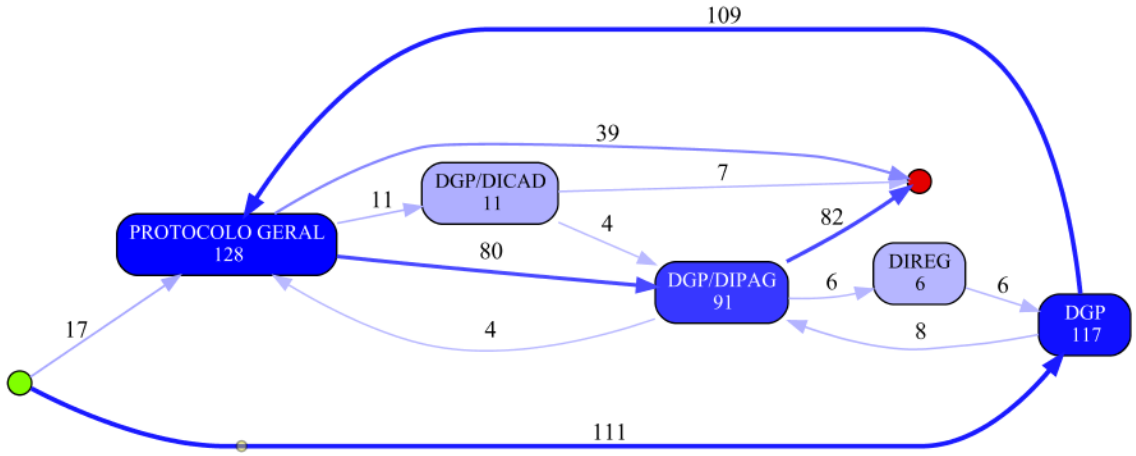


Figura 56 - Fluxo do processo de Reembolso antes da pandemia
Fonte: Própria (2022)

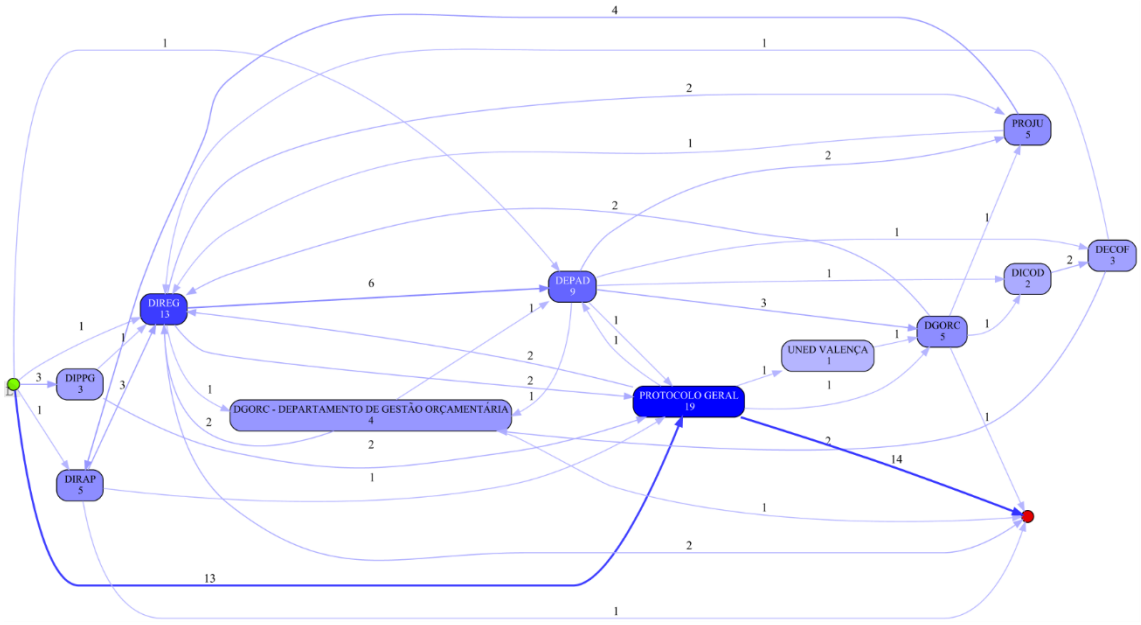


Figura 57 - Fluxo do processo de Reembolso depois da pandemia
Fonte: Própria (2022)

6.6.9. Avaliação – Reembolso

Este processo teve um comportamento diferente dos anteriores, com claras evidência de aumento nas etapas. Nota-se que no número médio das etapas antes e depois da pandemia foi um aumento relativamente pequeno, porém, no desenho do fluxo do processo isso apresentou uma maior complexidade. Na avaliação da centralidade da rede não ficou tão evidente quais os nós mais centrais além do DIREG.

Percebeu-se também que há participações com mais frequência de alguns departamentos referente ao planejamento financeiro, contabilidade e jurídico. Em avaliação juntamente com especialista, estes são fortes indícios sobre uma obrigação leal que fez o processo mudar devendo ser escriturado pelo DORC todo tipo de Reembolso . E as quedas para 0% no DGP/DIPAG pode ser devido a incorporação de pagamento no backoffice. Entretanto, não ficou claro que houve simplificação no processo apesar de ter uma diminuição da duração.

6.6.10. Mineração de Processo e Análises – Compra de Material

Na Tabela 31 é apresentada alguns dados gerais do macroprocesso de Compra de Material antes e depois da pandemia. Percebe-se que houve uma redução aproximadamente de 20% no número de processos. E também houve redução em de 37% na duração média e 76% no número médio de etapas dos processos.

Tabela 31 – Visão Geral do macroprocesso de Compra de Material

Compra de Material	Antes	Depois
Número de processos	84	67
Média de duração (dias)	111,6	70,6
Média do nº de etapas	13,5	3,2

Fonte: Própria (2022)

Para a análise da rede de Compra de Material foram analisadas as métricas apresentadas na Tabela 32. Após a pandemia observa-se que houve menos 23 departamentos envolvidos, 133 menos arcos e 6 menos departamentos com loop. A frequência dos pesos dos arcos também reduziu. O grau médio indica o número de departamentos em média que um departamento se relaciona, ou seja, antes era em torno de 9,27 departamentos e passou a ser apenas 6,76. A densidade e a medida de centralidade intermediação da rede aumentaram, tornando a rede mais conectada e com maior fluxo de informação.

Tabela 32 - Métricas Gerais das redes do macroprocesso de Compra de Material

Métricas gerais das redes	Antes	Depois
Número de departamentos (vértices)	44	21
Número de arcos	204	71
Total dos pesos dos arcos	1136	217
Número de departamentos Com loops	14	8
Densidade da Rede	0,11	0,16
Grau médio da rede	9,27	6,76
Intermediação da Rede	0,46	0,64

Fonte: Própria (2022)

As Figura 58 e Figura 59 apresentam as redes sociais derivadas dos *registros* de eventos do macroprocesso Compra de Material e com as métricas baseadas em transferência de trabalho e o tamanho do vértice é o grau de intermediação. Verifica-se visualmente, por exemplo, a redução do número de departamento, número de arcos e o grau médio dos departamentos.

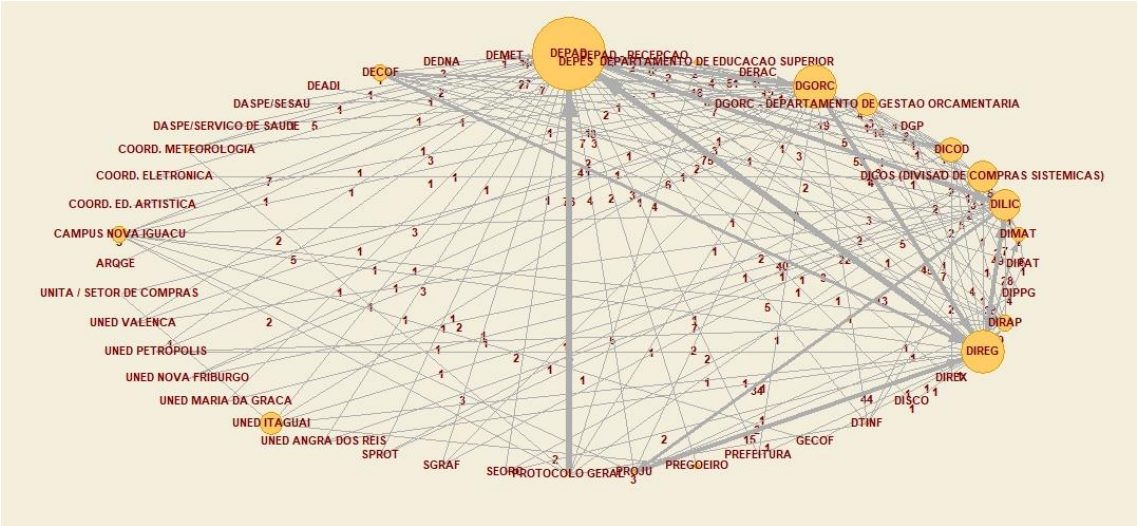


Figura 58 – Rede Social do macroprocesso Compra de Material Antes da Pandemia
Fonte: Própria (2022)

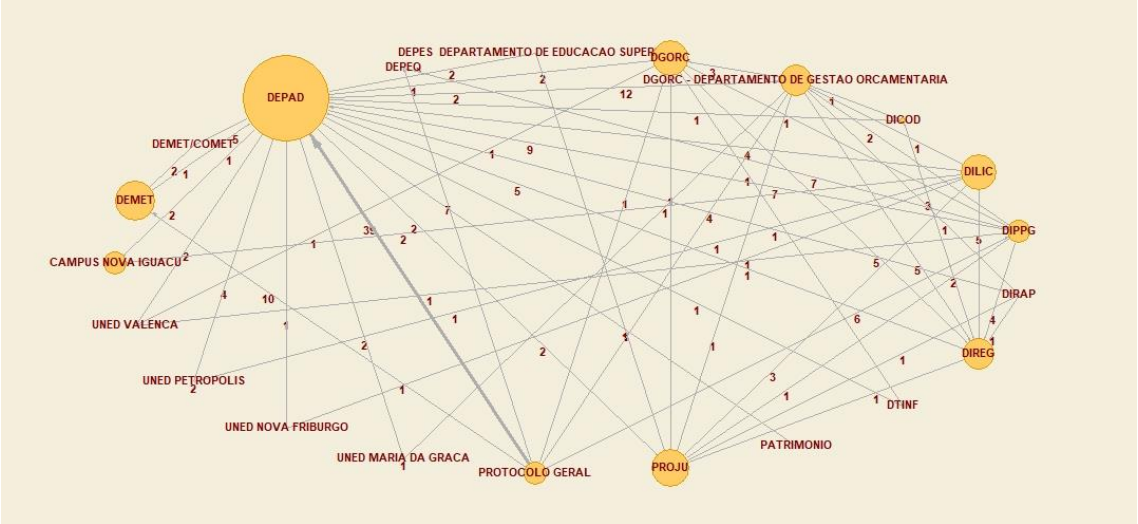
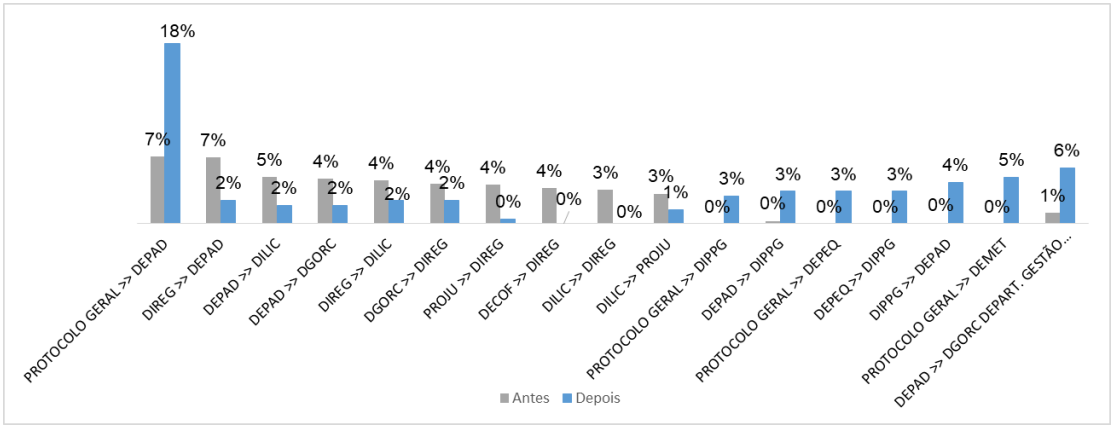


Figura 59 - Rede Social do macroprocesso Compra de Material depois da pandemia
Fonte: Própria (2022)

Através destas redes observou-se não houve alteração no arco de maior peso, o arco com maior peso é “PROTOCOLO GERAL >> DEPAD”. Para um melhor entendimento desta análise e um detalhamento mais profundo dos arcos, foi feita a frequência relativa dos arcos em cada período, ou seja, calculou-se o peso (a frequência) de cada arco sobre o peso total. E então, foi feita uma extração dos arcos com a frequência maior que 3% antes da pandemia e depois da pandemia. Nesta análise não foi utilizada a lei de Pareto para selecionar os arcos, pois as frequências estão muito distribuídas. Para se ter 80% dos arcos antes da pandemia seriam necessários selecionar 52 arcos, dos quais 42 tem a frequência relativa igual ou menos a 2%. Esta extração está representada no Gráfico 18.

Gráfico 18 - Arcos mais relevantes do macroprocesso de Compra de Material antes e depois da pandemia



Fonte: Própria (2022)

Percebeu-se nestas relações os seguintes pontos importantes:

- 4 arcos antes da pandemia que desapareceram ou reduzindo-se sua frequência relativa para 1% ou menos. São eles: DECOF >> DIREG, DILIC >> DIREG, PROJU >> DIREG e DILIC >> PROJU
- O arco PROTOCOLO GERAL >> DEPAD que já era importante antes da pandemia e aumentou ainda mais sua frequência relativa depois da pandemia.
- E 7 arcos que não estavam dentro dos arcos mais importantes da lei de Pareto antes da pandemia, tornaram-se relevante no período pós pandemia. Que são: PROTOCOLO GERAL >> DIPPG, DEPAD >> DIPPG, PROTOCOLO GERAL >> DEPEQ, DEPEQ >> DIPPG, DIPPG >> DEPAD, DEPAD >> DGORC DEPART. GESTÃO ORÇAMENTÁRIA e PROTOCOLO GERAL >> DEMET.

De modo geral antes da pandemia eram necessários 52 arcos para cobrir 80% do total dos pesos dos arcos. E após a pandemia são necessários apenas 32 arcos para cobrir 80% do total dos pesos dos arcos. Com as mudanças apresentadas acima é um indicativo de mudança no processo de compras. Para analisar no detalhe a relação entre os departamentos antes e depois da pandemia e na identificação do departamento mais central das redes calculou-se as métricas de centralidade de cada vértice: grau de entrada, grau de saída e intermediação. A Figura 60 apresenta essas métricas de cada departamento envolvido antes e depois da pandemia e ela foi construída conforme a Figura 36.

DEPARTAMENTO	Grau de Entrada		Grau de Saída		Intermediação	
	Antes	Depois	Antes	Depois	Antes	Depois
DEPAD	28	10	31	18	0,48	0,69
DIREG	19	8	15	4	0,18	0,09
DGORC	17	8	19	4	0,17	0,11
DILIC	12	5	16	3	0,09	0,12
DICOD	11	2	12	2	0,05	0,00
DICOS (DIVISAO DE COMPRAS SISTEMICAS)	10	0	9	0	0,00	0,00
DECOF	9	0	4	0	0,02	0,00
DGORC - DEPART. DE GESTAO ORCAMENT.	8	6	5	7	0,00	0,09
PROJU	8	5	4	4	0,00	0,13
DIMAT	7	0	7	0	0,02	0,00
CAMPUS NOVA IGUAÇU	6	3	7	2	0,03	0,05
DEPES DEPART DE EDUCACAO SUPERIOR	5	1	4	1	0,00	0,00
DTINF	5	1	2	1	0,00	0,00
DIRAP	4	3	12	1	0,03	0,00
DISCO	3	0	3	0	0,00	0,00
DIPAT	3	0	2	0	0,00	0,00
DEMET	3	3	4	4	0,00	0,14
UNED VALENCA	3	3	3	1	0,00	0,00
UNED PETROPOLIS	3	3	2	2	0,00	0,00
UNED ITAGUAÍ	3	0	3	0	0,04	0,00
PROTOCOLO GERAL	3	1	5	6	0,00	0,05
DERAC	2	0	2	0	0,00	0,00
PREFEITURA	2	0	3	0	0,00	0,00
DIREX	2	0	2	0	0,00	0,00
DEPAD - RECEPCAO	2	0	2	0	0,00	0,00
DIPPG	2	3	3	6	0,00	0,05
UNED NOVA FRIBURGO	2	1	1	1	0,00	0,00
SPROT	2	0	2	0	0,00	0,00
COORD. METEOROLOGIA	2	0	3	0	0,00	0,00
SGRAF	2	0	1	0	0,00	0,00
UNED ANGRA DOS REIS	2	0	2	0	0,00	0,00
UNED MARIA DA GRACA	2	2	2	2	0,00	0,00
SEORC	1	0	2	0	0,00	0,00
DGP	1	0	2	0	0,00	0,00
UNITA / SETOR DE COMPRAS	1	0	1	0	0,00	0,00
DASPE/SERVICO DE SAUDE	1	0	1	0	0,00	0,00
DEDNA	1	0	1	0	0,00	0,00
COORD. ED. ARTISTICA	1	0	1	0	0,00	0,00
ARQGE	1	0	0	0	0,00	0,00
GECOF	1	0	1	0	0,00	0,00
DEADI	1	0	0	0	0,00	0,00
PREGOEIRO	1	0	1	0	0,00	0,00
DASPE/SESAU	1	0	1	0	0,00	0,00
COORD. ELETRONICA	1	0	1	0	0,00	0,00
DEPEQ	0	1	0	1	0,00	0,00
DEMET/COMET	0	1	0	1	0,00	0,00
PATRIMONIO	0	1	0	0	0,00	0,00

Figura 60 - Lista de departamentos do macroprocesso de Compra de Material e suas métricas de centralidade antes e depois da pandemia
Fonte: Própria (2022)

Percebe-se que não houve mudança no departamento central. O departamento “DEPAD” é o mais central antes e depois da pandemia pois ele apresenta os maiores valores de todas métricas de centralidade. O segundo e o terceiro mais centrais são “DIREG” e o “DGORC”, respectivamente. De modo geral apesar das relações entre o departamento terem mudado os departamentos mais centrais se mantiveram os mesmos.

Para detalhar e aprofundar no entendimento da participação de cada departamento, a Figura 61 apresenta dois cálculos de % de frequência e ela foi construída conforme a Figura 37. Analisou-se que apesar do departamento “Protocolo Geral” não ser destaque nas métricas de centralidades de grau ele é o único que aparece em 100% das instâncias de processos nos períodos antes e depois da pandemia. Na “Frequência Relativa - Casos” DGP/DIPAG reduziu-se consideravelmente número de casos que passam pelos departamentos centrais “DEPAD”, de 98% para 61%, “DIREG”, de 83% para 19%, e “DGORC”, de 71% para 10%. Importante destacar também que na “Frequência Relativa - Eventos”, no período antes da pandemia, os departamentos “DEPAD”, “DIREG”, “DGORC” e “DILIC” tem um alto percentual. Isso mostra um número de eventos muito maior que o número de instâncias de processos, podendo indicar possíveis gargalos e retrabalhos. Já para os processos depois da pandemia este comportamento não ocorreu, mesmo nos nós mais centrais.

DEPARTAMENTO	Frequência Relativa - Casos		Frequência Relativa - Eventos	
	Antes	Depois	Antes	Depois
DEPAD	98%	61%	318%	72%
DIREG	83%	19%	262%	21%
DGORC	71%	10%	119%	12%
DILIC	49%	9%	139%	9%
DICOD	25%	4%	49%	4%
DICOS (DIVISAO DE COMPRAS SISTEMICAS)	14%	0%	21%	0%
DECOF	43%	0%	52%	0%
DGORC - DEPART. DE GESTAO ORCAMENT.	20%	19%	21%	24%
PROJU	45%	7%	76%	7%
DIMAT	18%	0%	19%	0%
CAMPUS NOVA IGUAÇU	13%	3%	23%	3%
DEPES DEPART DE EDUCACAO SUPERIOR	5%	1%	13%	1%
DTINF	6%	1%	8%	1%
DIRAP	25%	1%	33%	1%
DISCO	2%	0%	4%	0%
DIPAT	5%	0%	5%	0%
DEMET	10%	10%	19%	18%
UNED VALENCA	4%	1%	4%	1%
UNED PETROPOLIS	5%	4%	6%	4%
UNED ITAGUAÍ	5%	0%	5%	0%
PROTOCOLO GERAL	100%	100%	100%	100%
DERAC	1%	0%	4%	0%
PREFEITURA	6%	0%	7%	0%
DIREX	2%	0%	5%	0%
DEPAD - RECEPCAO	2%	0%	2%	0%
DIPPG	5%	24%	6%	27%
UNED NOVA FRIBURGO	1%	1%	1%	1%
SPROT	1%	0%	4%	0%
COORD. METEOROLOGIA	2%	0%	6%	0%
SGRAF	1%	0%	1%	0%
UNED ANGRA DOS REIS	2%	0%	2%	0%
UNED MARIA DA GRACA	1%	3%	2%	3%
SEORC	2%	0%	2%	0%
DGP	2%	0%	2%	0%
UNITA / SETOR DE COMPRAS	1%	0%	1%	0%
DASPE/SERVICO DE SAUDE	1%	0%	1%	0%
DEDNA	1%	0%	1%	0%
COORD. ED. ARTISTICA	1%	0%	1%	0%
ARQGE	1%	0%	1%	0%
GECOF	1%	0%	1%	0%
DEADI	1%	0%	1%	0%
PREGOEIRO	2%	0%	2%	0%
DASPE/SESAU	1%	0%	1%	0%
COORD. ELETRONICA	1%	0%	1%	0%
DEPEQ	0%	10%	0%	10%
DEMET/COMET	0%	1%	0%	1%
PATRIMONIO	0%	1%	0%	1%

Figura 61 - Lista de departamentos do macroprocesso de Compra de Material e suas frequências relativas antes e depois da pandemia
Fonte: Própria (2022)

Para complementar estas análises, as Figuras 62 e 63 apresentam os fluxos do processo de Compra de Material antes e depois da pandemia respectivamente. Ambos estão maiores que os anteriores que podem ser conferidos através do Grau médio da rede. Este fluxo antes da pandemia, assim como o Processo Folha de Pessoal antes da Pandemia possuem o Grau médio em torno de 9. Estes são caracterizados por fluxos "*Espaguete*".

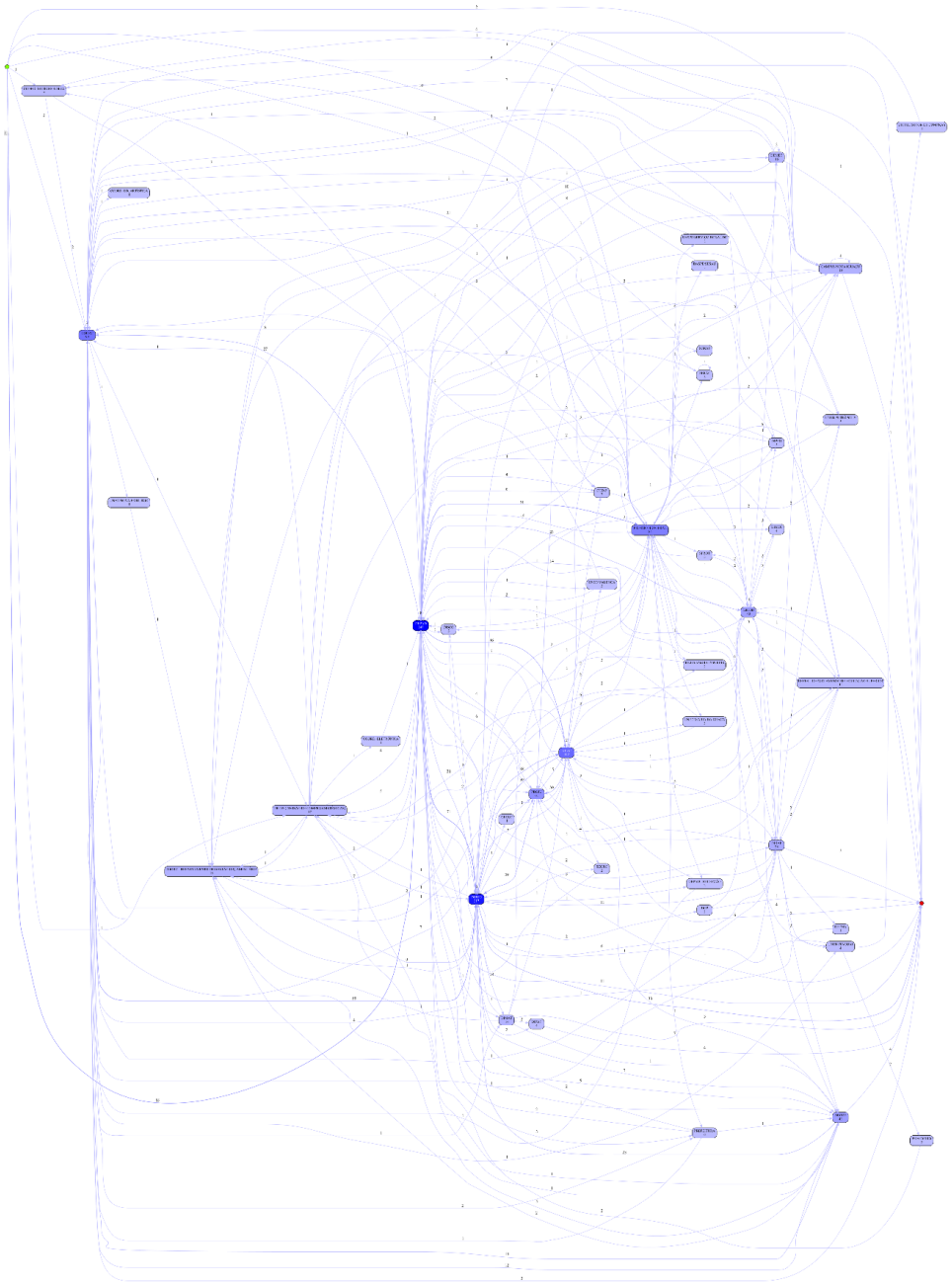


Figura 62 - Fluxo do processo de Compra de Material antes da pandemia
Fonte: Própria (2022)

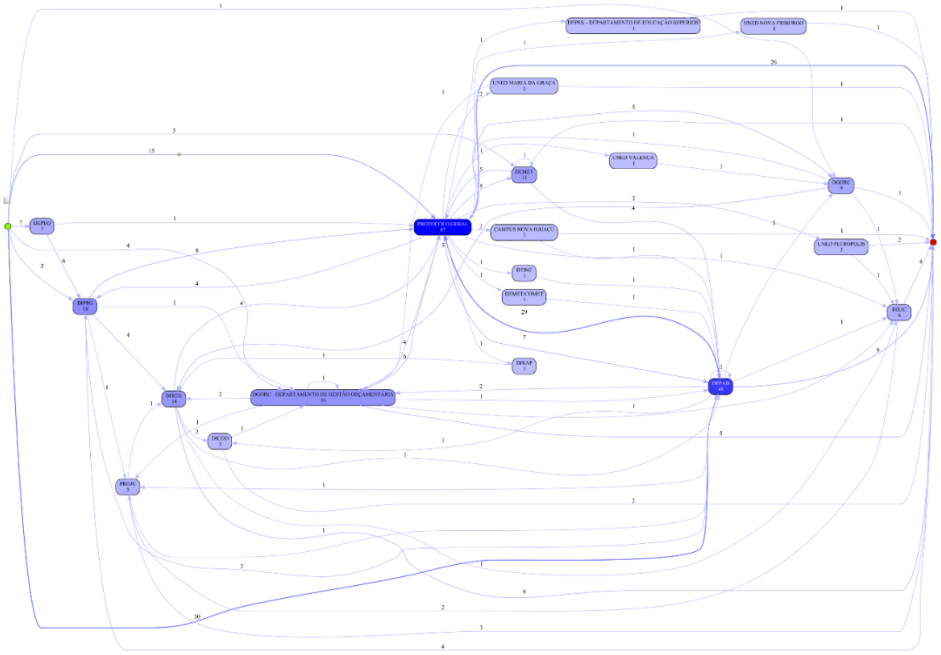


Figura 63 - Fluxo do processo de Compra de Material depois da pandemia
Fonte: Própria (2022)

6.6.11. Avaliação – Compra de Material

Observou-se uma mudança para a simplificação neste processo sendo a maior redução do número de etapas entre todos analisados, ou seja menor número de departamento envolvidos apesar de ter tido a menor redução do número de processo. Os departamentos centrais se mantiveram mas houve uma grande redução de suas participações nos casos e o grau de intermediação também foi um destaque a qual aumentou, apresentando assim, uma maior troca de informação entre os departamentos. É preciso entender se houve algum tipo de restrição de compra no período depois da pandemia

A suposição levantada pelo especialista, é novamente a indicação de subnotificações por e-mail, logo não registrados todos os lançamentos no sistema e, também incorporação de etapas por outros setores. Notou-se também que neste processo há muitas unidades externas ao CEFET Maracanã, como por exemplo, Campus Valença e Petrópolis. Entendeu-se que este comportamento é devido a instituição objeto deste estudo é a sede de compras para as demais unidades. Assim há transferências de trabalho com outros departamentos externos a sede.

6.6.12. Discussão sobre o Estudo de Caso

Na análise para o entendimento dos processos gerais em todo os períodos de tempo houve a necessidade de um longo tratamento da nomenclatura dos nomes dos processos. Esta limpeza foi fundamental para a simplificações das análises posteriores onde houve uma redução de aproximadamente de 73% das variações nominais. E em seguida, houve uma agregação em macroprocesso onde diferentes tipos de instância de processos pertencentes a uma classe de processos equivalente foram agrupados. Com esses macroprocessos agrupados foi possível verificar a distribuição de frequências dos processos administrativo, a duração média e a distribuição do número de etapas de cada um. Pode-se observar que 80,6% das instâncias tratam de 24 macroprocessos. Destacou-se como os cinco primeiros macroprocessos como: Gestão de Carreira, Convênio, Compra de Material, Diploma e Folha de Pessoal. Foi possível observar que desde 2016 havia sido crescente as solicitações nos processos administrativos até 2019. Em 2020, ano da pandemia, houve uma queda significativa em torno de 50% de redução comparando-se com o ano anterior. E que no ano de 2021 ainda continuou bem abaixo esse número. Na série histórica percebeu-se que no mês de abril de 2020 o número de novos processos caiu drasticamente ficando bem próximo de zero. Observou-se que cerca de 50% dos processos tinham a duração de até 28 dias e que aproximadamente 83% dos processos tinham duração entre 168 e 196 dias. Ou seja, a grande maioria dos processos são tramitados em torno do período de 6 meses. Este corte foi utilizado como o recorte da base para a análise seguinte para os períodos antes e depois da pandemia. Essa análise geral dos processos administrativos apontou que houve um forte impacto da pandemia reduzindo as solicitações do protocolo geral.

Na segunda iteração, a análise abrange 3.144 e 1.177 processos nos períodos pré-pandêmico e pandêmico e, em vista geral, os processos de negócios administrativos de protocolo geral do CEFET mudou drasticamente. Reduziu-se suas durações, frequências e também número de etapas nos períodos comparados antes e depois da companhia. O que foram indícios concretos de que os trabalhadores implementaram uma nova estrutura de processos durante a pandemia. Vinte macroprocessos foram identificados como tendo a maior ocorrência antes da pandemia representando 80,92% do total (3.144 macroprocessos). O macroprocesso “Diploma” representa 17,56% do total de processos seguido de “Gestão de Carreira do Corpo Docente” e a “Gestão de Carreira da TAE”. E dezoito macroprocessos tiveram a maior ocorrência. Estes representam 80,54% do total (1.177 macroprocessos) onde os três primeiros foram:

“Gestão de Carreira TAE”, “Convênio” e “Diploma”. Entretanto, identificou-se que o macroprocesso de “Colação de Grau” foi o único que teve aumento na frequência dos macroprocessos principais em 25% pós pandemia. E também houve aumento na duração média do macroprocesso de “Aposentaria”, mesmo com a sua redução da sua frequência de instâncias de processos em 90%. O que levantou-se a suposição de que neste cenário pandêmico outros macroprocessos poderiam ter tido aumentado suas durações também, não são tão óbvios os resultados de reduções significativas.

Na terceira e última iteração, a mineração organizacional e a descoberta do fluxo dos seguintes macroprocessos conhecidos foram realizadas: “Gestão de Carreira Docente”, “Gestão de Carreira TAE”, “Folha de Pessoal”, “Reembolso de Pagamentos”, e “Compra de Material”. Ressalta-se que todos os aprofundamentos foram feitos juntamente com um especialista do Protocolo Geral e que já atuou diretamente com as solicitações dos processos administrativos do CEFET/RJ. Através destas novas análises de rede social ficou evidente que houve mudança indicando a simplificação nos processos não apenas em termos de frequência, duração e etapas mas também percebeu-se que departamentos centrais antes da pandemia foram impactados e alterados em novos departamentos centrais. Pode-se perceber a redução dos departamentos tramitados através de um menor número de vértices e uma menor transferência de trabalhos entre os departamentos. Também foi possível identificar departamentos que eram gargalos, pois, passavam por ele o mesmo processo mais de uma vez. E através dos fluxos de processos, como uma análise complementar utilizando a técnica de MP de descoberta de processos, pode-se visualizar essa simplificação em quatros dos cinco processos, exceto no processo de “Reembolso”.

Nos macroprocessos de Gestão de Carreira TAE, Gestão de Carreira Docente e Folha de Pessoal, os quais tramitam principalmente pela DGP (Diretoria de Gestão de Pessoas) e seus departamentos observou-se essa redução no trâmite entre departamentos desburocratizando as transferências de trabalho e possíveis revisões de gargalos. Através dos fluxos de processos notou-se uma forma de trabalho mais direta, com menos loops entre os departamentos e assim, redução de retrabalho nos processos. Já no macroprocesso de “Compra de Material” observou-se departamentos de unidades externas ao CEFET/RJ havendo transferência de trabalho com outros departamentos da sede, assim como, uma simplificação no processo apesar de ter tido a menor redução do número de processo e a maior redução do número de etapas entre todos analisados. Já no macroprocesso de “Reembolso”, o comportamento foi diferente onde através do fluxo de processos não foi observada uma simplificação. Foi mantido

do número de etapas antes e depois da pandemia apesar de uma significante redução das solicitações (em torno de 85%) e redução de 75% na duração.

Durante a pandemia, a organização forçou a execução dos processos registrados em papel em uma implementação virtual. Juntamente com o especialista, levantou-se as suposições que podem ter sido criadas listas de verificação no site da instituição onde o funcionário tem com antecedência o que é preciso para seguir com maior fluidez à sua solicitação. E que a evolução do processo por digitalização através do e-mail e com subnotificações onde os registros feitos no sistema de Protocolo Geral são marcados apenas nas etapas principais e as demais etapas ficam registradas apenas na caixa do e-mail. Mas ainda assim, algumas tarefas foram movidas e relatadas as suas atividades no sistema. Ressalta-se que há um padrão rigoroso e que não foi alterado, onde 100% dos processos são registrados no setor “Protocolo Geral” antes e depois da pandemia. Ou seja, todos inícios de processo ainda são lançados neste departamento, exceto o macroprocesso Gestão de Carreira Docente que há um padrão de iniciar no CPPD e depois segue para o Protocolo Geral. Também contribuiu para essa simplificação a aceitação de assinaturas digitais liberadas pelo governo federal. Observou-se que alguns departamentos se destacaram após a pandemia, pois estão relacionados ao financeiro (orçamento, contabilidade) o qual apontam um maior controle de gastos devido a exigência legal pelo governo de lançar notas de empenho a parte. E outra suposição é que parte de processos podem ter sido foi incorporado para outros departamentos. Deseja-se que esta pesquisa contribua para novas estratégias da diretoria desta instituição e na evolução de seus processos administrativos. Assim como na tratativa das lacunas identificadas e ainda seja fonte para outros pesquisadores e profissionais de negócio.

Entendeu-se que é uma boa prática para identificar com mais precisão aonde está o processo quando os departamentos são mais detalhados a nível dos departamentos/divisão do que apenas a nível de diretoria. Houve uma melhor compreensão na Diretoria DGP que teve mais quebras do que nas demais que não haviam esse detalhamento. Uma limitação deste trabalho é não ter essas informações para analise para profunda nos demais apartamentos. Identificou-se também como sugestões de melhoria a padronização dos nomes dos processos no Sistema de Protocolo Geral e ter uma validação no campo de data para evitar inconsistências.

Conclusões

Esta pesquisa objetivou analisar o impacto antes e depois da pandemia nos processos administrativos tramitados no protocolo geral do CEFET/RJ, utilizando técnicas da mineração de processos sob a perspectiva organizacional. A pandemia da COVID-19 afetou muitos processos e modelos de negócios nos últimos dois anos onde as organizações tiveram que se adaptar ao novo formato de trabalho em casa impostos pelas novas condições pandêmicas. A instituição objeto deste estudo possui procedimentos administrativos regulados por um arcabouço jurídico rigoroso que visa à prestação de contas ao governo. Estes procedimentos são executados em formato tradicional de papel, contudo possui o Sistema Eletrônico de Protocolo Geral o qual registra dados para identificar as movimentações do processo em qualquer momento.

Nestes dados há os chamados registros de eventos que são as entradas básicas para a mineração de processos e através de suas técnicas e algoritmos é possível fazer diversas análises como descoberta de processos, comparações com a conformidade do processo real versus o modelo, previsão de comportamento dos processos e até estudos sobre a forma como os recursos (pessoas, departamentos entre outros) estão se relacionando. Para esta última prática, são apresentados fortes conceitos em análise de redes sociais para realizar a mineração organizacional e, assim, obter entendimento dessas relações entre os recursos das empresas. Para atingir seu objetivo este estudo se propôs a responder às duas perguntas de pesquisa: “Qual é o panorama da produção científica sobre Mineração de Processos, em termos gerais e da perspectiva organizacional, segundo parâmetros da bibliometria?” e se “Houve alteração nos processos administrativos do CEFET/RJ no período anterior e posterior a pandemia?”.

Para responder a primeira pergunta utilizou-se do estudo bibliométrico para avaliar o estado da arte sobre mineração de processos e mineração organizacional. Encontrou-se os principais autores, temas, comunidades científicas, países foco. E reforçou que no Brasil ainda é pouco explorado este tema principalmente com o foco na perspectiva organizacional. Assim como, este último ainda é um tema pouco explorado pelos pesquisadores diantes das outras perspectivas.

Para responder a segunda pergunta de pesquisa utilizou-se da metodologia de estudo de caso envolvendo as seis fases do framework PM^2 para Projeto de Mineração de Processo. Ressalta-se foram utilizados os dados do o Sistema Eletrônico de Protocolo Geral no qual ficam registradas as tramitações das solicitações do público tanto interno quanto externa ao CEFET/RJ. Esses dados gerais foram coletados entre

2013 a junho de 2021 o qual contêm dois grandes arquivos onde em um deles há 30.871 instâncias de processos e no outro há 193.227 registros de eventos. Estes foram os dados básico para detalhar as três interações do PM^2 que foram: análise para o entendimento dos processos gerais em todo os períodos de tempo, depois análise dos processos comparando o recorte no período antes e depois da pandemia e por último a análise aprofundado em processos específicos utilizando-se da mineração organizacional através de redes sociais dos departamentos e descoberta de seus fluxos antes e depois da pandemia. Entendeu-se que houve simplificações significativas nos processos não apenas pois reduziu a demanda mas pelas mudanças no trâmite entre um departamento e outro e das estruturas organizacionais alteradas analisadas pelas redes sociais. Não sendo resultados óbvios pois tiveram processos os quais aumentaram o tempo de processo, número de etapas e até o número de processos.

As implicações práticas dos resultados obtidos podem ser explorados em três aspectos principais: Adaptação a um novo formato de trabalho: com os resultados observados onde houve mudanças significativas nos processos de negócios dado o novo modelo de trabalho remoto sugere-se que a instituição de ensino continue a implementar mudanças em seus processos administrativos para acomodar o trabalho remoto, mesmo após a pandemia; Simplificação de processos: este trabalho apresenta que houve uma simplificação na maioria dos processos administrativos, resultando em maior eficiência. A instituição de ensino pode considerar simplificar ainda mais seus processos para reduzir custos e aumentar a eficiência; Mudanças na estrutura organizacional: observou-se que vários departamentos centrais foram modificados durante a pandemia, logo a instituição pode considerar a adoção de uma nova estrutura organizacional que permita maior flexibilidade e adaptação a possíveis crises futuras. Essas sugestões podem ser úteis não apenas para a instituição de ensino estudada na dissertação, mas também para outras organizações que enfrentam desafios semelhantes devido à pandemia.

Sugestões de trabalhos e discussões futuras a médio e longo prazo voltados ao estudo de caso e mineração de processos podem indicar se o “novo normal” imposto pela pandemia de COVID-19 dentro desta instituição se tornou a regra ou foram apenas mecanismos utilizados para manter o funcionamento da organização durante o período em que foram aplicadas medidas restritivas mundialmente. E outra possibilidade de desdobramentos para melhorar aprofundamento de mineração organizacional é avaliar a nível de recurso e criar uma estrutura organizacional informal.

Ademais, sugere-se de trabalhos futuros voltados ao estudo bibliométrico é investigar ainda mais como esse processo de revisão de literatura acontece através da ARS criando um processo que possa ser seguido por qualquer autor com o mínimo de conhecimento nessas técnicas e assim facilitar a sua visão geral rapidamente sobre o assunto. E também conhecer o grupo de artigo dos temas que lhe interessa e em contrapartida disseminar mais o conhecimento sobre análise de redes sociais que poderá permear qualquer área do conhecimento. Além destes, explorar mais sobre a mineração de texto trabalhando com análise de redes sociais de forma a melhorar os resultados dos *clusteres* encontrados para os temas, pois partindo de ARS o pesquisador já inicia com um conhecimento prévio de comunidade, coautoria e relação entre artigos e atores. Assim, enriquecer todo o estudo bibliométrico e a coleção de publicações estado da arte para os pesquisadores tomar conhecimento.

Referências

AALST, Wil M P Van Der; ROSA, Marcello La. Business Process Management. **Business & Information Systems Engineering**, [*s. l.*], v. 58, p. 1–6, 2016. Disponível em: https://link.springer.com/article/10.1007/s12599-015-0409-x#citeas

ACCENTURE. **Outmaneuver uncertainty: Navigating the human and business impact of Covid-19**. [*S. l.*], 2020. Disponível em: https://www.accenture.com/fi-en/about/company/coronavirus-business-economic-impact. Acesso em: 2 nov. 2021.

ALBUQUERQUE, D.F; MONTEIRO, A.F.C. Análise das interconexões entre conselho de administração de empresas brasileiras usando redes sociais. *In*: , 2019, Santos, São Paulo,. **XXXIX Encontro Nacional de Engenharia de Produção**. Santos, São Paulo,: [*s. n.*], 2019.

APPICE, Annalisa *et al.* Discovering and tracking organizational structures in event logs. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 9607, p. 46–60, 2016. Disponível em: https://doi.org/10.1007/978-3-319-39315-5_4

BAILEY, Katie; BRESLIN, Dermot. The COVID-19 Pandemic: What can we learn from past research in organizations and management? **International Journal of Management Reviews**, [*s. l.*], v. 23, n. 1, p. 3–6, 2021. Disponível em: https://doi.org/10.1111/ijmr.12237

BARBASTEFANO, Rafael Garcia *et al.* Impactos dos nomes nas propriedades de redes sociais: um estudo em rede de coautoria sobre sustentabilidade. **Perspectivas em Ciência da Informação**, [*s. l.*], v. 18, n. 3, p. 78–95, 2013. Disponível em: https://doi.org/10.1590/s1413-99362013000300006

BARBASTEFANO, Rafael Garcia *et al.* Influência da ambiguidade de nomes na centralidade de redes de coautoria. **Transinformacao**, [*s. l.*], v. 27, n. 3, p. 189–198, 2015. Disponível em: https://doi.org/10.1590/0103-37862015000300001

BARTSCH, Silke *et al.* Leadership matters in crisis-induced digital transformation: how to lead service employees effectively during the COVID-19 pandemic. **Journal of Service Management**, [*s. l.*], v. 32, n. 1, p. 71–85, 2021. Disponível em: https://doi.org/10.1108/JOSM-05-2020-0160

BATAGELJ, Vladimir; ; ; MRVAR, Andrej. **Programs for Analysis and Visualization of Very Large Networks**. Ljubljana: [*s. n.*], 2008.

BBC. **Coronavirus: How the pandemic has changed the world economy**. [*S. l.*], 2021. Disponível em: https://www.bbc.com/news/business-51706225. Acesso em: 1 nov. 2021.

BELLI, Simone *et al.* Coronavirus mapping in scientific publications: When science advances rapidly and collectively, is access to this knowledge open to society? **Scientometrics**, [*s. l.*], v. 124, n. 3, p. 2661–2685, 2020. Disponível em: https://doi.org/10.1007/s11192-020-03590-7

BLONDEL, Vincent D; GUILLAUME, Jean-loup; LEFEBVRE, Etienne. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics Theory and Experiment**, [*s. l.*], p. 1–12, 2008. Disponível em: https://doi.org/http://dx.doi.org/10.1088/1742-5468/2008/10/P10008

BORDIN, Andréa Sabedra; GONÇALVES, Alexandre Leopoldo; TODESCO, José Leomar. Análise da colaboração científica departamental através de redes de coautoria. **Perspectivas em Ciência da Informação**, [*s. l.*], v. 19, n. 2, p. 37–52, 2014. Disponível em: https://doi.org/10.1590/1981-5344/1796

BOSE, R P Jagadeesh Chandra; AALST, Wil M P Van Der. Abstractions in Process Mining : A Taxonomy of. *In*: , 2009. **International Conference on Business Process Management**. [*S. l.*]: Springer, Berlin, Heidelberg, 2009. p. 159–175. Disponível em: https://doi.org/10.1007/978-3-642-03848-8_12

CERQUEIRA, Carlos; COSTA, Juliana; CARVALHO, Diego. Aplicação de Análise de Redes Sociais em uma Cadeia de Suprimentos de uma Epresa do Setor Elétrico Brasileiro. **Sistemas & Gestão**, [*s. l.*], v. 9, n. 4, p. 418–429, 2014. Disponível em: https://doi.org/10.7177/sg.2014.v9.n4.a1

CHANG, Hansol *et al.* Impact of COVID-19 Pandemic on the Overall Diagnostic and Therapeutic Process for Patients of Emergency Department and Those with Acute Cerebrovascular Disease. **Journal of Clinical Medicine**, [*s. l.*], v. 9, n. 12, p. 3842, 2020. Disponível em: https://doi.org/10.3390/jcm9123842

CHIAVENATO, Idalberto. **Administração geral e pública**. 2. eded. Rio de Janeiro: Elsevier, 2008.

CHIAVENATTO, Idalbert. **Administração nos Novos Tempos**. 2. eded. Rio de Janeiro: Elsevier, 2004.

DA COSTA, Henrique Pais; RODRIGUES, Thiago Sales. Mineração de processos aplicada à gestão orçamentária da Marinha do Brasil. **Revista Brasileira de Planejamento e Orçamento**, [*s. l.*], v. 10, p. 97–119, 2020.

DAVENPORT, Thomas H; SPANYI, Andrew. **What Process Mining Is, and Why Companies Should Do It?**. [*S. l.*], 2019. Disponível em: https://hbr.org/2019/04/what-process-mining-is-and-why-companies-should-do-it.

DELOITTE. **The Acceleration of Digitization as a Result of COVID-19**. [*S. l.*], 2020. Disponível em: https://www2.deloitte.com/global/en/blog/responsible-business-blog/2020/acceleration-of-digitization-as-result-of-covid-19.html. Acesso em: 2 nov. 2021.

FERREIRA, Diogo R.; ALVES, Cláudia. Discovering user communities in large event logs. **Lecture Notes in Business Information Processing**, [*s. l.*], v. 99 LNBIP, n. PART 1, p. 123–134, 2012. Disponível em: https://doi.org/10.1007/978-3-642-28108-2_11

FERREIRA, Ismael De Vasconcelos *et al.* Operacionalização E Logística Dos Processos Administrativos Na Gestão Pública Para Enfrentamento À Pandemia Da Covid-19. **SANARE - Revista de Políticas Públicas**, [*s. l.*], v. 20, p. 28–38, 2021. Disponível em: https://doi.org/10.36925/sanare.v20i0.1505

FREEMAN, Linton C. A Set of Measures of Centrality Based on Betweenness. **Sociometry**, [*s. l.*], v. 40, n. 1, p. 35–41, 1977. Disponível em: https://doi.org/10.2307/3033543

FREEMAN, Linton C. Centrality in Social Networks Conceptual Clarification. [*s. l.*], v. 1, n. 1968, p. 215–239, 1979. Disponível em: https://doi.org/http://dx.doi.org/10.1016/0378-8733(78)90021-7

FREEMAN, Linton C. Visualizing Social Networks. **Journal of Social Structure**, [*s. l.*], v. 1, 2000.

GANDULFO, Pablo I. *et al.* Redução de custos nas organizações públicas com método de mineração de processos: Um estudo de caso no controle de férias. **SBSI 2016 - 12th Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era, Proceedings**, [*s. l.*], p. 589–596, 2016. Disponível em: https://doi.org/10.5753/sbsi.2016.6010

GOVERNO. **Governo Eletrônico**. [*S. l.*], 2021. Disponível em: www.governoeletronico.gov.br. Acesso em: 20 maio 2021.

GÜNTHER, Christian W.; VERBEEK, Eric. **XES Standard DefinitionEindhoven University of Technology**. [*S. l.: s. n.*], 2014. Disponível em: http://bpmcenter.org/%0Awp-content/uploads/reports/2014/BPM-14-09.pdf.

HAKVOORT, Ron; SLUITER, Alexander. Process Mining : Conformance analysis from a financial audit perspective. **International Journal of Business Process Integration and Management**, [*s. l.*], v. X, p. 1–26, 2006.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts andTechniques**. 3rd. ed. [*S. l.*]: Morgan Kaufmann, 2012. Disponível em: https://doi.org/10.1016/b978-0-12-381479-1.00001-0

JANS, Mieke; ALLES, Michael; VASARHELYI, Miklos. The case for process mining in auditing: Sources of value added and areas of application. **International Journal of Accounting Information Systems**, [*s. l.*], v. 14, n. 1, p. 1–20, 2013. Disponível em: https://doi.org/10.1016/j.accinf.2012.06.015

KAMAL, Muhammad Mustafa. The triple-edged sword of COVID-19: understanding the use of digital technologies and the impact of productive, disruptive, and destructive nature of the pandemic. **Information Systems Management**, [*s. l.*], v. 37, n. 4, p. 310–317, 2020. Disponível em: https://doi.org/10.1080/10580530.2020.1820634

KERREMANS, Marc. **Market Guide for Process Mining**. [*S. l.: s. n.*], 2019.

LEEMANS, Sander J J. **Inductive visual Miner manual**. [*S. l.: s. n.*], 2017. Disponível em: http://promtools.org.

LEEMANS, Sander J.J.; POPPE, Erik; WYNN, Moe T. Directly follows-based process mining: A tool. **CEUR Workshop Proceedings**, [*s. l.*], v. 2374, p. 9–12, 2019a.

LEEMANS, Sander J.J.; POPPE, Erik; WYNN, Moe T. Directly follows-based process mining: Exploration & a case study. **Proceedings - 2019 International Conference on Process Mining, ICPM 2019**, [*s. l.*], p. 25–32, 2019b. Disponível em: https://doi.org/10.1109/ICPM.2019.00015

LEONI, Massimiliano De; AALST, Wil M P Van Der; DEES, Marcus. A General Framework for Correlating Business Process Characteristics. *In*: , 2014. (vol 8659 Lecture Notes in Computer Science, Org.)**International Conference on Business Process Management**. [*S. l.*]: Springer, Cham, 2014. p. 250–251. Disponível em: https://doi.org/https://doi.org/10.1007/978-3-319-10172-9_16

MARCONDES, Carlos Henrique; JARDIM, José Maria. Políticas de Informação Governamental: a construção de Governo Eletrônico na Administração Federal do Brasil. **Revista de Ciência da Informação**, [*s. l.*], v. 4, n. 2, p. 1–14, 2003.

MARCONI, Maria; LAKATOS, Eva. **Fundamentos de metodologia científica**. [*S. l.: s. n.*], 2003. ISSN 9788522457588. Disponível em: https://doi.org/10.1590/S1517-97022003000100005

MARCZYK, Geoffrey; DEMATTEO, David; FESTINGER, David. **Essentials of Research Design and Methodology**. 1st. ed. Canada.: John Wiley & Sons, 2005.

MCKINSEY. **The great acceleration**. [*S. l.*], 2020. Disponível em: https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-great-acceleration. Acesso em: 5 nov. 2021.

MERRIS, Russell. **Graph Theory**. 1st editioed. [*S. l.*]: Wiley-Interscience, 2000.

MORENO, J. L. **Who Shall Survive? A New Approach to the Problem of Human Interrelations**. Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934.

MOURA, Luzia Menegotto Frick de; BRAUNER, Daniela Francisco; JANISSEK-MUNIZ, Raquel. Blockchain e a Perspectiva Tecnológica para a Administração Pública: Uma Revisão Sistemática. **Revista de Administração Contemporânea**, [*s. l.*], v. 24, n. 3, p. 259–274, 2020. Disponível em: https://doi.org/10.1590/1982-7849rac2020190171

NEWMAN, Mark. **Networks: a introduction**. New York: Oxford University Press, 2010. Disponível em: https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

NEWMAN, M E J. Modularity and community structure in networks. [*s. l.*], v. 103, n. 23, p. 8577–8582, 2006.

NEWMAN, M. E.J.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, [*s. l.*], v. 69, n. 2 2, p. 1–16, 2004. Disponível em: https://doi.org/10.1103/PhysRevE.69.026113

NOOY, Wouter; MRVAR, Andrej; BATAGELJ, Vladimir. **Exploratory Social Network Analysis with Pajek**. 1st ed.ed. New York: Cambridge University Press, 2005. Disponível em: https://doi.org/10.1017/CBO9780511806452

OMS. **Conselhos sobre doença coronavírus (COVID-19) para o público**. [*S. l.*], 2021. Disponível em: https://www.who.int/pt/emergencies/diseases/novel-coronavirus-2019/advice-for-public. Acesso em: 30 out. 2021.

OPENREFINE. **OpenRefine Documentation**. [*S. l.*], 2021. Disponível em: https://docs.openrefine.org/next/manual/cellediting. Acesso em: 15 out. 2021.

PIKA, Anastasiia *et al.* Mining Resource Profiles from Event Logs. **ACM Transactions on Management Information Systems**, [*s. l.*], v. 8, n. 1, p. 1–30, 2017. Disponível em: https://doi.org/10.1145/3041218

PINHO, José Antonio Gomes de. Investigando portais de governo eletrônico de estados no Brasil: muita tecnologia, pouca democracia. **Revista de Administração Pública**, [*s. l.*], v. 42, n. 3, p. 471–493, 2008. Disponível em: https://doi.org/10.1590/S0034-76122008000300003

PINZARU, F.; ZBUCHEA, A.; ANGHEL, L.. The Impact of The COVID-19 Pandemic on Business. A Preliminary Overview. *In*: , 2020, Bucharest, ROMANIA. **8th International Academic Conference on Strategica - Preparing for Tomorrow, Today**. Bucharest, ROMANIA: TRITONIC PUBL HOUSE, 2020. p. 721–730. Disponível em: https://doi.org/2392-702X

ROJAS, Eric *et al.* Process mining in healthcare: A literature review. **Journal of Biomedical Informatics**, [*s. l.*], v. 61, n. April, p. 224–236, 2016. Disponível em: https://doi.org/10.1016/j.jbi.2016.04.007

SCOTT, John. **Social Network Analysis**. 2 ed.ed. London: [*s. n.*], 2000.

SONG, Minseok; VAN DER AALST, Wil M P. Towards comprehensive support for organizational mining. **Decision Support Systems**, [*s. l.*], v. 46, n. 1, p. 300–317,

2008. Disponível em: https://doi.org/10.1016/j.dss.2008.07.002

SOUZA, Cristina Gomes; BARBASTEFANO, Rafael Garcia. Knowledge diffusion and collaboration networks on life cycle assessment. **International Journal of Life Cycle Assessment**, [*s. l.*], v. 16, n. 6, p. 561–568, 2011. Disponível em: https://doi.org/10.1007/s11367-011-0290-x

SOUZA, Cristina Gomes; BARBASTEFANO, Rafael Garcia; DE LIMA, Leonardo Silva. Redes de colaboração científica na área de química no Brasil: Um estudo baseado nas coautorias dos artigos da revista Química Nova. **Quimica Nova**, [*s. l.*], v. 35, n. 4, p. 671–676, 2012. Disponível em: https://doi.org/10.1590/S0100-40422012000400003

THIEDE, Malte; FUERSTENAU, Daniel; BEZERRA BARQUET, Ana Paula. How is process mining technology used by organizations? A systematic literature review of empirical studies. **Business Process Management Journal**, [*s. l.*], v. 24, n. 4, p. 900–922, 2018. Disponível em: https://doi.org/10.1108/BPMJ-06-2017-0148

TIWARI, A.; TURNER, C. J.; MAJEED, B. A review of business process mining: State-of-the-art and future trends. **Business Process Management Journal**, [*s. l.*], v. 14, n. 1, p. 5–22, 2008. Disponível em: https://doi.org/10.1108/14637150810849373

VAN DER AALST, Wil *et al.* Process Mining Manifesto. *In*: BUSINESS PROCESS MANAGEMENT WORKSHOPS. [*S. l.*]: Springer Berlin Heidelberg, 2012. p. 169–194. Disponível em: https://doi.org/10.1007/978-3-642-28108-2_19

VAN DER AALST, W. M.P. *et al.* Business process mining: An industrial application. **Information Systems**, [*s. l.*], v. 32, n. 5, p. 713–732, 2007. Disponível em: https://doi.org/10.1016/j.is.2006.05.003

VAN DER AALST, Wil; WEIJTERS, Ton; MARUSTER, Laura. Workflow mining: Discovering process models from event logs. **IEEE Transactions on Knowledge and Data Engineering**, [*s. l.*], v. 16, n. 9, p. 1128–1142, 2004. Disponível em: https://doi.org/10.1109/TKDE.2004.47

VAN DER AALST, W.M.P *et al.* The Process Mining Toolkit: ProM. **Process Mining Group, Eindhoven University of Technology**, [*s. l.*], v. 489, 2013. Disponível em: http://www.promtools.org/

VAN DER AALST, Wil M. P. **Process Mining: Discovery, Conformance and Enhancement of Business Processes**. 1st editioed. [*S. l.*]: Incorporated, 2011. ISSN 2093-3681. Disponível em: https://doi.org/10.4258/hir.2014.20.2.157

VAN DER AALST, Wil M. P.; REIJERS, Hajo A; SONG, Minseok. Discovering Social Networks from Event Logs. **Computer Supported Cooperative Work (CSCW)**, [*s. l.*], v. 14, n. 6, p. 549–593, 2005. Disponível em: https://doi.org/10.1007/s10606-005-9005-9

VAN DER AALST, Wil M.P. Process discovery from event data: Relating models and logs through abstractions. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [*s. l.*], v. 8, n. 3, 2018. Disponível em: https://doi.org/10.1002/widm.1244

VAN DER AALST, Wil M.P. **Process mining Data Science in Action**. 2nd editioed. Netherlands: Springer, 2016. ISSN 01706012.v. 35 Disponível em: https://doi.org/10.1007/s00287-012-0641-4

VAN DER AALST, Wil M.P. TomTom for business process management (TomTom4BPM). **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*],

v. 5565 LNCS, p. 2–5, 2009. Disponível em: https://doi.org/10.1007/978-3-642-02144-2_2

VAN DER AALST, Wil M.P.; SONG, Minseok. Mining Social Networks: Uncovering Interaction Patterns in Business Processes. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 3080, n. June, p. 244–260, 2004. Disponível em: https://doi.org/10.1007/978-3-540-25970-1_16

VAN DONGEN, B. F. *et al.* The ProM framework: A new era in process mining tool support. **Lecture Notes in Computer Science**, [*s. l.*], v. 3536, n. June, p. 444–454, 2005. Disponível em: https://doi.org/10.1007/11494744_25

VAN ECK, Maikel L *et al.* PM2: A Process Mining Project Methodology. *In*: , 2015. **Advanced Information Systems Engineering**. [*S. l.*]: Springer International Publishing, 2015. p. 297–313. Disponível em: https://doi.org/10.1007/978-3-319-19069-3_19

WASSERMAN, Stanley; FAUST, Katherine. **Social Network Analysis**. 1st editioed. [*S. l.*]: Cambridge University Press, 1994. Disponível em: https://doi.org/10.1017/CBO9780511815478

YANG, Jing; OUYANG, Chun; VAN DER AALST, WM P. OrgMining 2.0: A Novel Framework for Organizational Model Mining from Event Logs. [*s. l.*], 2020.

YIN, Robert K. **Case Study Research_ Design and Methods**. Third Edited. California: [*s. n.*], 2002.

ZHAO, W., Zhao X. Process Mining from the Organizational Perspective. **Advances in Intelligent Systems and Computing**, [*s. l.*], v. 277, p. 701–708, 2014. Disponível em: https://doi.org/10.1007/978-3-642-54924-3