

Ciência de Dados Para Todos (Data Science For All) - 2019.1 - Análise da Produção Científica e Acadêmica da Universidade de São Paulo - Relatório sobre os programas de pós-graduação do Departamento de Ciência da Computação

Andre Garrido Damaceno - 15/0117531

João Marcelo Nunes Chaves - 15/0132085

Lucas Campos Jorge - 15/0154135

30/06/2019

Introdução

A disciplina Data Science For All (Ciência de Dados para Todos), ministrada por Jorge Henrique Cabral Fernandes e Ricardo Barros Sampaio, do departamento de Ciência da Computação da Universidade de Brasília tem como foco integral a aplicação de ciência da dados como ferramenta eficiente de análise.

O tema de estudo escolhido para realização da prática de ciência de dados foi o cenário atual da Pós-Graduação Brasileira, para que seus dados sejam processados e estudados com o objetivo de se retirar análises a respeito da qualidade, relevância, e produtividade dos programas de pós-graduação brasileiros.

A fonte dos dados utilizados são dados disponíveis pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Órgão que já recolhe dados quantitativos e qualitativos sobre a pesquisa nacional, como currículos de professores e pesquisadores, além de avaliações recorrentes sobre os programas de pós-graduação a serem estudados.

A interpretação e manipulação dos dados da CAPES adquiridos seguiram o modelo CRISP-DM, que guiará o projeto e será descrito ao longo do estudo. Sendo o cenário de pesquisa deste projeto a pós-graduação brasileira, foram escolhidos dois programas de pós-graduação como objetos de estudo, de onde serão extraídos seus dados. Serão estes:

- Programa de pós graduação em Ciência da Computação e Matemática Aplicada pela Universidade de São Paulo USP/SC
- Programa de pós graduação em Computação Aplicada pela Universidade de São Paulo USP/RP

O que é ciência?

Ciência pode ser definida como uma investigação de fenômenos desconhecidos por meio de métodos, com o objetivo de buscar a explicação, estrutura e formar uma previsão do comportamento do que é observado.

Os métodos usados para a investigação devem estar em constante atualização, já que são baseados em um conjunto de princípios observados ao longo do tempo, e que não necessariamente estarão corretos para todas as situações ou que são imutáveis. Para uma obtenção fiel da análise proposta, os métodos devem ser cuidadosamente planejados.

As investigações e os métodos são realizados por cientistas, por meio da produção científica, que se baseia em princípios como a projetização da ciência, racionalismo metodológico científico, empirismo científico, reprodutibilidade científica e uma comunidade científica. Como a análise de fenômenos e a consequente

produção científica não possui restrições de acesso, o conhecimento é disseminado por toda a comunidade de modo que todos os interessados sejam capazes de analisar e aprimorar o conhecimento recebido.

O método científico usado por cientistas é um conjunto de regras básicas que devem ser seguidas para o desenvolvimento de um padrão em que possa ser analisado o objeto de estudo de forma controlada. É composto pelos seguintes passos:

- Observação: detecção de um fenômeno.
- Investigação: análise química, física ou matemática do fenômeno.
- Problematização: identificação dos motivos e características do fenômeno.
- Hipótese: formulação de soluções e identificações de como o fenômeno funciona
- Verificação: análise final de todos os itens levantados, reprodutibilidade do fenômeno e obtenção de provas.

Como nem sempre todos os requisitos podem ser seguidos com total precisão, ou justamente devido a alguma falha na coleta ou formulação dos dados, a ciência não é dada como um fato absoluto. Porém, devido a seu rigor e ao avanço dos métodos e revisões, a ciência é considerada muito eficaz e precisa.

Conclui-se assim que a ciência é essencial para o conhecimento humano, por estar em constante evolução, e reflete impactos em todas as áreas da sociedade através das produções científicas que exigem esforço intenso. Também sendo de suma importância na disseminação, verificação e comprovação de fenômenos e conhecimento pelo mundo.

O que é ciência no Brasil?

Ao falar em ciência no Brasil, podemos logo vinculá-la à CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Esse órgão realiza a gestão dos programas de pós graduação *stricto sensu* que engloba mestrados e doutorados. Por gestão entende-se acesso a informação de produções científicas, investimentos de recursos, cooperação científica internacional e outras formas de coordenação relacionadas ao desenvolvimento científico brasileiro.

As produções científicas no Brasil vem, de grande parte, dos programas de pós graduação. Esses programas são formados por cursos de mestrados e doutorados oferecidos por uma instituição de ensino superior. Vale ressaltar que os cursos oferecidos devem fazer parte das áreas de conhecimento atuantes na instituição de ensino superior.

Os mestrados e doutorados podem ser divididos em acadêmicos ou profissionalizantes, ou seja, uma voltada para a área acadêmica e outra voltada para a área profissional. Esses programas visam, principalmente, à formação de professorado e incentivar a pesquisa científica de alto padrão para qualificar pesquisadores e cientistas em alto nível para fazer jus ao desenvolvimento nacional do Brasil.

Para que os programas de pós graduação brasileiros mantenham um nível razoável de qualidade, são necessárias avaliações periódicas realizadas pelo Sistema Nacional de Pós-Graduação Brasileira para os programas já existentes. Já para novos programas, é necessário uma avaliação da proposta inicial do programa.

A CAPES possui um sistema rigoroso de avaliação para os programas nacionais de pós graduação. São realizadas avaliações a cada 4 anos e notas de 1 a 7 são atribuídas aos programas. Notas 1 e 2 são consideradas baixas e portanto não atendem o requisito mínimo para que o programa consiga prosseguir com seu andamento. A nota 3 é considerada o padrão mínimo de qualidade, ou seja, uma nota regular. Já a nota 4 é considerado um bom desempenho. Para programas que possuem apenas mestrado, a nota 5 é a maior nota. Por fim as notas 6 e 7 são consideradas altas e indicam desempenho equivalente a programas de pós graduação de excelência no exterior.

O que é CRISP-DM:

CRISP-DM (Cross-Industry Standard Process for Data Mining) é um modelo de análise de mineração de dados, feita de forma sistemática, sendo amplamente utilizada por ser flexível, podendo ser aplicada em qualquer negócio, e sua execução não ser dependente de ferramentas.

As fases que compõem o CRISP-DM são as seguintes:

1. **Entendimento do negócio:** Identificação das necessidades, objetivos e tipos de soluções do negócio, a fim de transformar a realidade do negócio em um problema de data mining.
2. **Entendimento dos dados:** Reconhecer os tipos de dados disponíveis. Pode ser subdividido em cinco atividades:
 - a) Análise dos dados.
 - b) Realização de coleta dos dados.
 - c) Descrição dos dados.
 - d) Exploração dos dados: Objetivo de aprender e entender melhor a respeito.
 - e) Analisar a qualidade dos dados recolhidos.
3. **Preparação dos dados:** Filtrar, a partir dos dados brutos recolhidos, visando a remoção de partes que não são necessárias, sem qualidade ou fora do contexto da mineração. Para isso, essa fase é subdividida em cinco atividades:
 - a) Seleção dos dados: Identificação dos dados úteis.
 - b) Limpeza dos dados.
 - c) Construção dos dados: Criação de novos dados a partir da análise de outros.
 - d) Integração dos dados: Unir dados de várias fontes em apenas uma.
 - e) Formatação dos dados: Organização e alterações na estrutura de dados para adequação ao método de data mining escolhido.
4. **Modelagem:** É feita a construção e avaliação do modelo. O modelo segue de acordo com as quatro atividades:
 - a) Seleção das técnicas de modelagem: Escolher e ajustar os parâmetros do algoritmo a ser utilizado.
 - b) Realização de testes de modelagem.
 - c) Construção do modelo definitivo.
 - d) Avaliação do modelo e técnicas escolhidas.
5. **Avaliação:** Revisão e análise do modelo usado, organização dos dados e efeitos causados na junção de dados. Possível repetição das fases anteriores para uma maior validação.
6. **Implantação:** Inserção dos produtos desenvolvidos para uso, com um monitoramento regular, sendo feitas adaptações e ajustes quando necessários. Essa fase marca o fim da produção do projeto.

As metodologias propostas pelo CRISP-DM estão apresentados no diagrama abaixo:

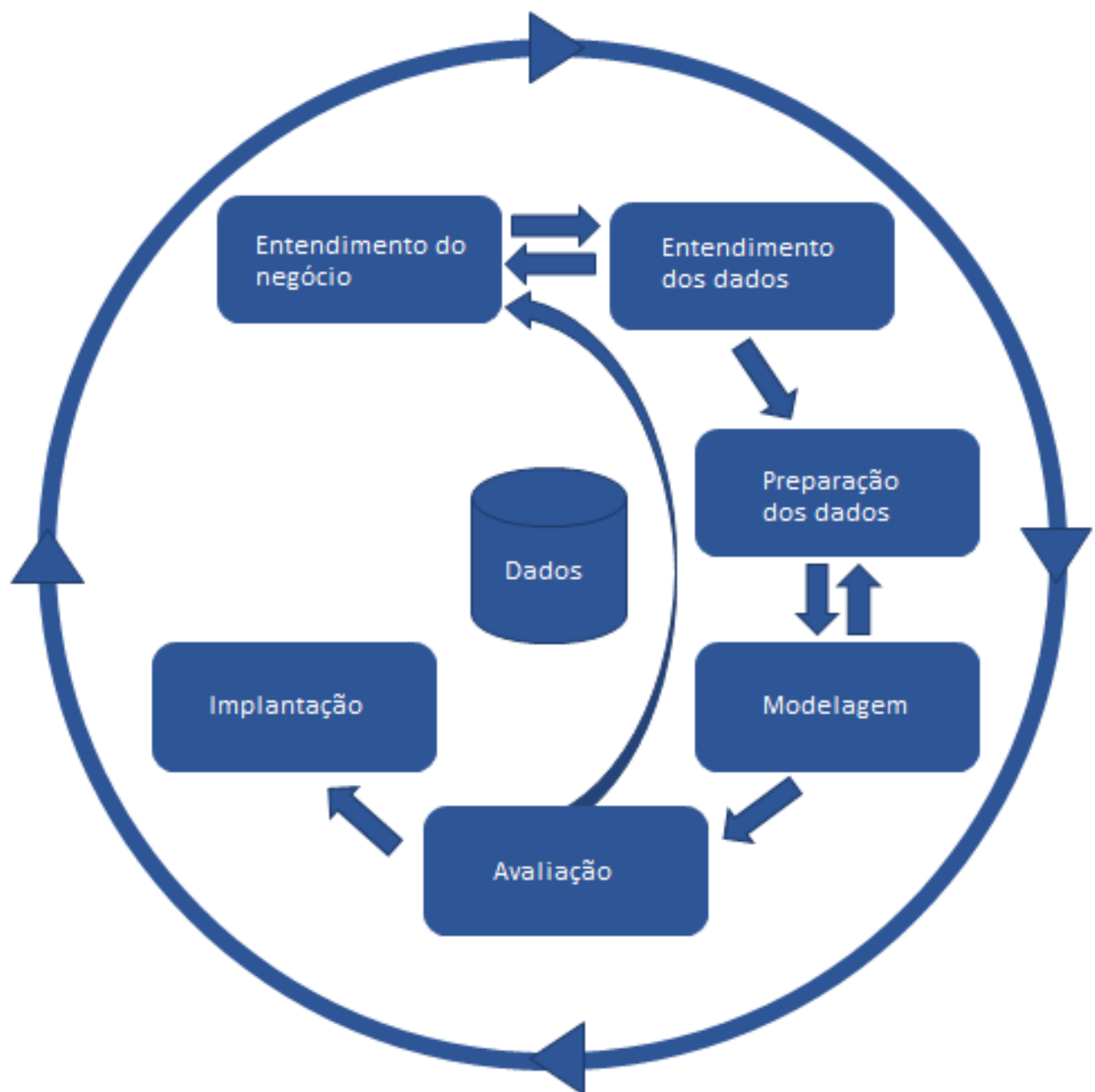


Figura 1: Diagrama CRISP-DM

Contexto dos programas de pós-graduação

Programa de pós graduação em Ciência da Computação e Matemática Aplicada pela Universidade de São Paulo USP/São Carlos

O programa de pós graduação em Ciência da Computação e Matemática Aplicada está organizado em 5 linhas de pesquisa sendo elas:

- Computação Gráfica, Imagens e Visualização
- Engenharia de Software e Sistemas de Informação/Sistemas Web e Multimídia Interativos
- Inteligência Computacional
- Sistemas Distribuídos e Programação Concorrente/Sistemas Embarcados, Evoluídos e Robóticos
- Mecânica dos Fluidos Computacional/Otimização/Modelos Estocásticos

As disciplinas ofertadas para os discentes englobam as principais áreas da Ciência da Computação. Já seu corpo docente é de alto nível, contendo professores formados em ótimas universidades no Brasil e no exterior.

O programa conta como modalidades o mestrado e o doutorado acadêmico, tendo em vista formar pesquisadores e professores na área. O curso é muito bem avaliado e recebeu a nota 7 (máxima) na avaliação trienal da CAPES. Para tanto, o curso foi avaliado como tendo: uma excelente qualificação e inserção internacional dos docentes; dissertações e teses defendidas de ótima qualidade; Ótima produção bibliográfica; e uma gama de atividades de cooperação internacional.

Tabela 1: Dados do programa de pós em Ciência da Computação e Matemática Aplicada (USP/SC)

| | |
|----------------------|-----|
| Discentes | 452 |
| Docentes | 71 |
| Disciplinas | 292 |
| Financiadores | 5 |
| Linhas de Pesquisa | 5 |
| Projetos de Pesquisa | 88 |

Programa de pós graduação em Computação Aplicada pela Universidade de São Paulo USP/RP

O programa de pós graduação em computação aplicada possui como modalidade, apenas, o mestrado acadêmico e contempla duas linhas de pesquisa, sendo elas: Computação aplicada à Biotecnologias e Sistemas Computacionais Complexos.

As disciplinas ofertadas pelo programa abrangem as três áreas da Ciência da Computação, sendo elas:

- Metodologia e Técnicas de Computação
- Teoria da Computação e Análise de Algoritmos e Complexidade da Computação
- Sistemas de Computação

O programa, apesar de recente, conta com um corpo docente experiente onde cada docente possui, pelo menos uma, orientação de mestrado e a maioria possui pelo menos uma orientação de doutorado concluída. Apesar disso, o programa conta com poucos docentes com dedicação exclusiva para o programa, o que é considerado um ponto em que o programa deixa a desejar.

Na última avaliação trienal CAPES, o programa recebeu nota 3. Tal nota foi justificada principalmente pela falta de docentes com dedicação exclusiva, também foi ponderado a falta de disciplinas que fazem com que o discente não garanta aproveitamento em cada uma das áreas do núcleo básico da ciência da computação.

Por ser um programa de pós graduação muito novo, ainda não há publicações discentes e portanto não existe uma maneira de qualificar as produções científicas.

Tabela 2: Dados do programa de pós em Computação Aplicada (USP/RP)

| | |
|----------------------|----|
| Discentes | 23 |
| Docentes | 15 |
| Disciplinas | 21 |
| Financiadores | 5 |
| Linhas de Pesquisa | 2 |
| Projetos de Pesquisa | 15 |

Análise dos dados coletados

Para a coleta de dados dos programas de graduação citados acima, utilizamos a plataforma elattes UnB (<http://unb.elattes.com.br>). Por meio dessa plataforma, obtivemos os dados de 2014 até 2018 das publicações científicas, dos docentes afiliados e sobre as orientações de cada programa de pós-graduação. Com auxílio da linguagem R, utilizamos scripts que nos permitiram interpretar os dados coletados.

Computação Aplicada (USP/SC)

O gráfico abaixo nos apresenta números a respeito dos tipos de orientação concluídos em função dos anos. Podemos observar um significativo aumento nas orientações de mestrado concluídas no ano de 2018. Como o programa de pós-graduação em Computação Aplicada se iniciou no final 2015 - e sabendo que um mestrado tem duração média de 2 anos - podemos concluir que ocorreram 10 orientações concluídas de mestrado nesse período. Vale ressaltar que essas orientações são a soma de todos os programas de pós-graduação em que os docentes fazem parte.

Também, observamos a veracidade do fato de que existem docentes que não tem dedicação exclusiva com o programa de pós-graduação analisado. Isso pode ser observado no alto número de supervisões de pós-doutorado e de doutorado em 2017.

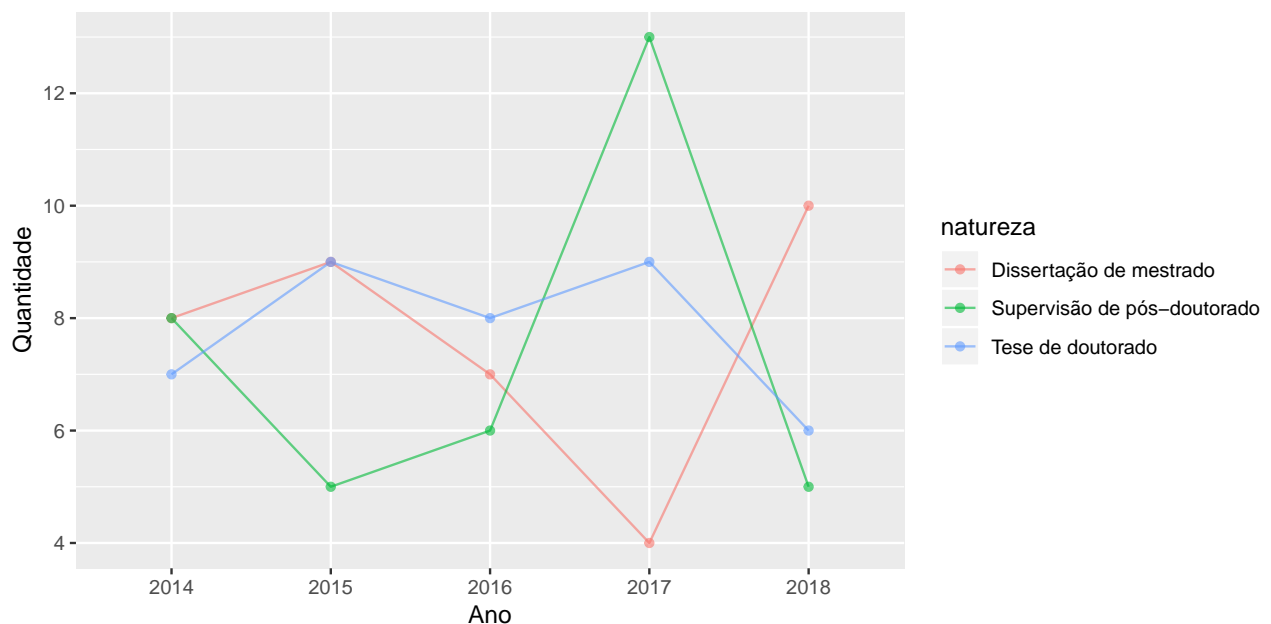


Figura 2: Natureza das supervisões concluídos

O gráfico abaixo nos apresenta números a respeito dos tipos de orientação não concluídos em função dos anos. Podemos observar um significativo aumento nas orientações de mestrado não concluídas no ano de 2017. Isso pode nos indicar que o programa de pós-graduação possui relevância por parte dos docentes que fazem parte desse programa.

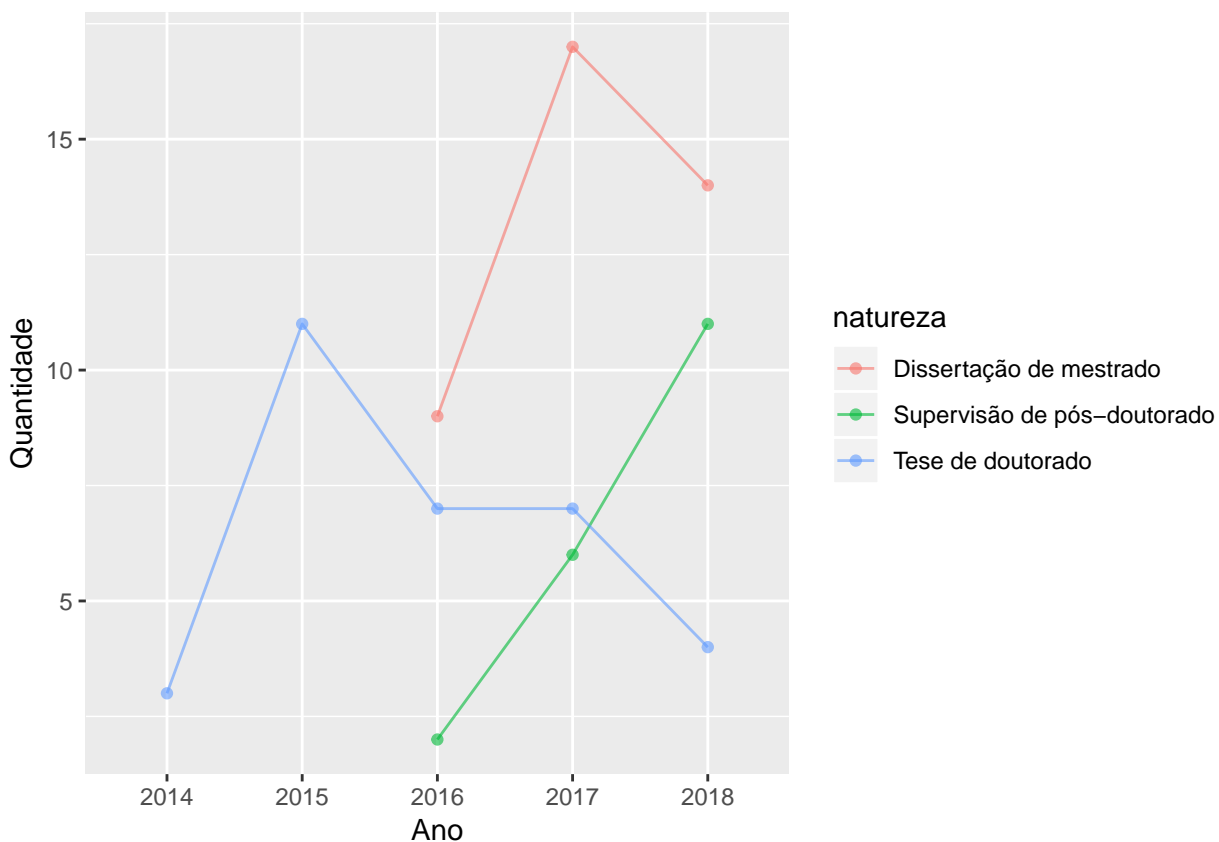


Figura 3: Natureza das supervisões em andamento

Com o gráfico abaixo, observamos que a maioria dos docentes do programa de pós-graduação participou de eventos internacionais e, também, em 126 eventos pelo Brasil. Embora a maioria dos eventos atendidos pelos docentes serem no Brasil, observa-se um bom número de presenças em eventos internacionais principalmente nos Estados Unidos.

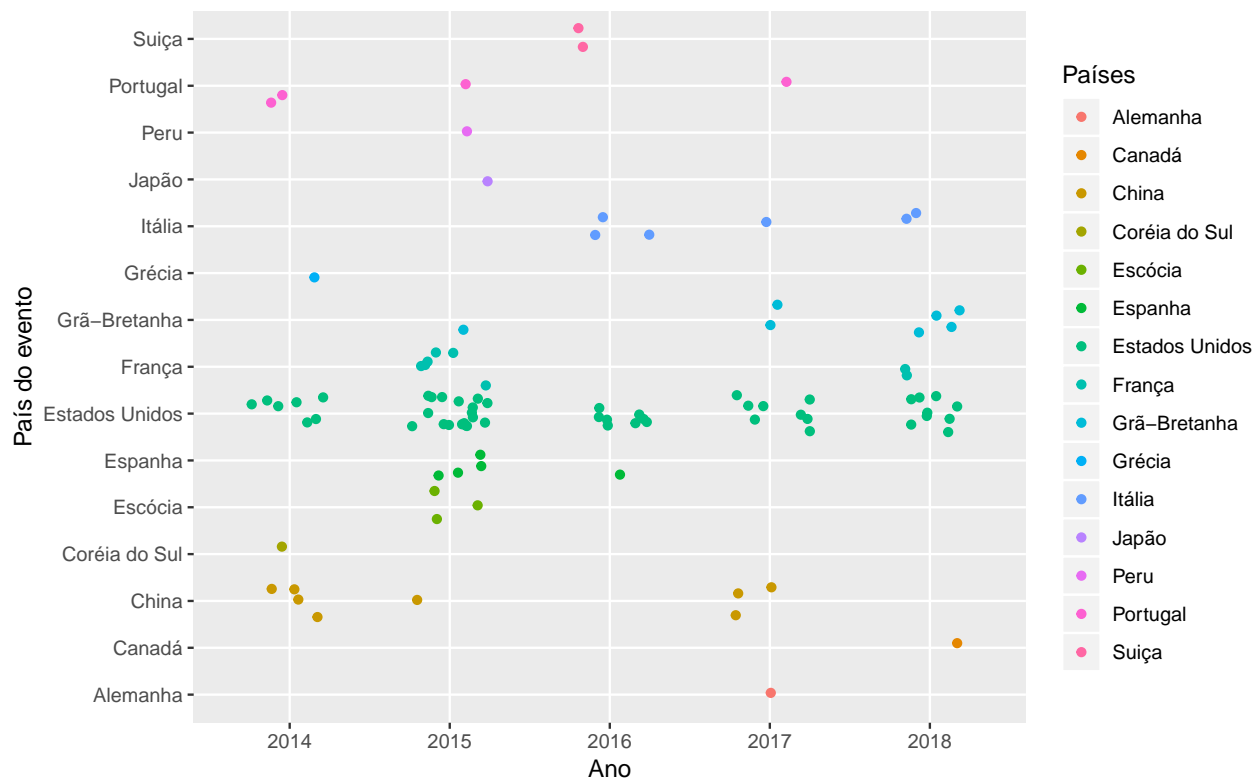


Figura 4: Eventos internacionais comparecidos por docentes

Com as análises abaixo, percebemos uma leve queda, com o passar dos anos, nos números de publicações em eventos e em periódicos. Essa queda foi relativamente expressiva quando comparado ao ano de 2014 - que foi o período de início das análises.

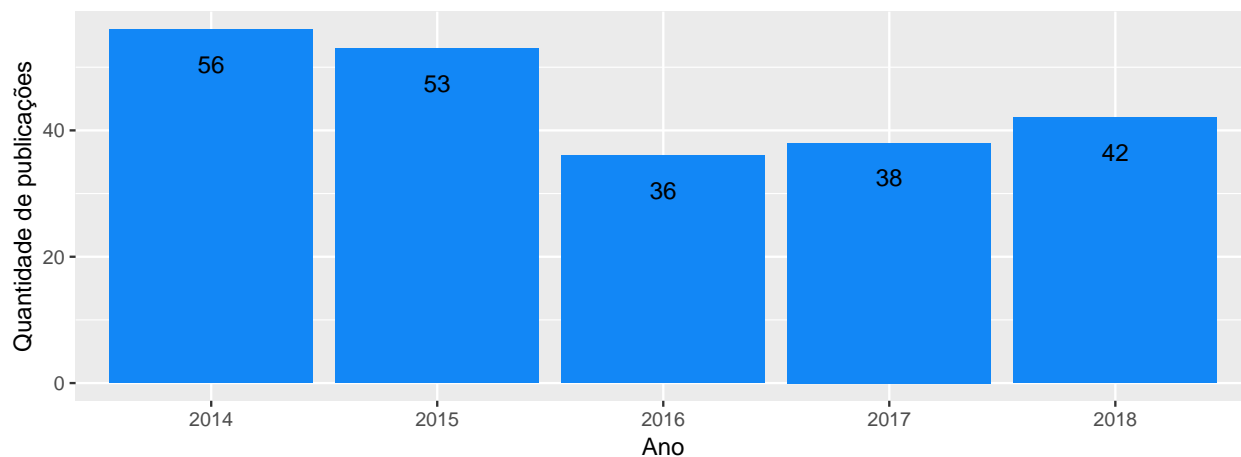


Figura 5: Quantidade de publicações em eventos por ano

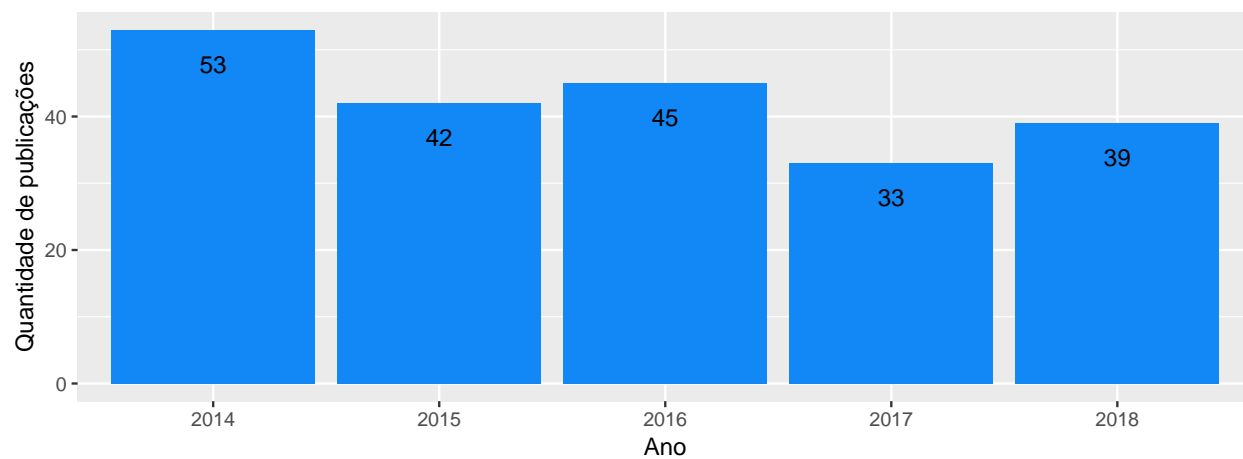


Figura 6: Quantidade de publicações em periódicos por ano

Abaixo, podemos notar que grande parte dos docentes desse programa possuem como grande área a de Ciências Exatas e da Terra seguido por Engenharias. Esse fato nos mostra que as grandes áreas dos pesquisadores envolvidos no programa estão condizentes com a proposta do programa de pós-graduação.

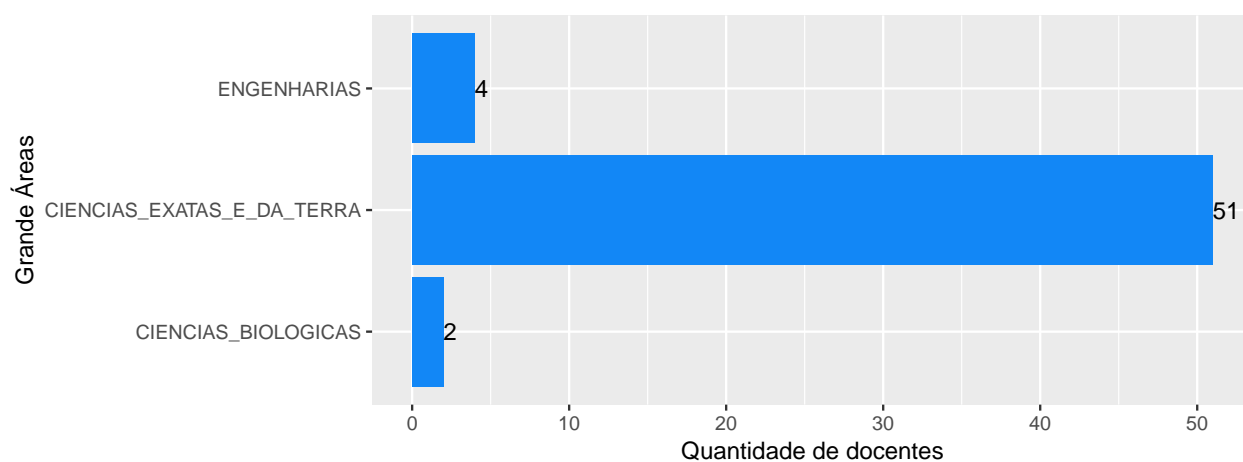


Figura 7: Quantidade de docentes por Grande Área

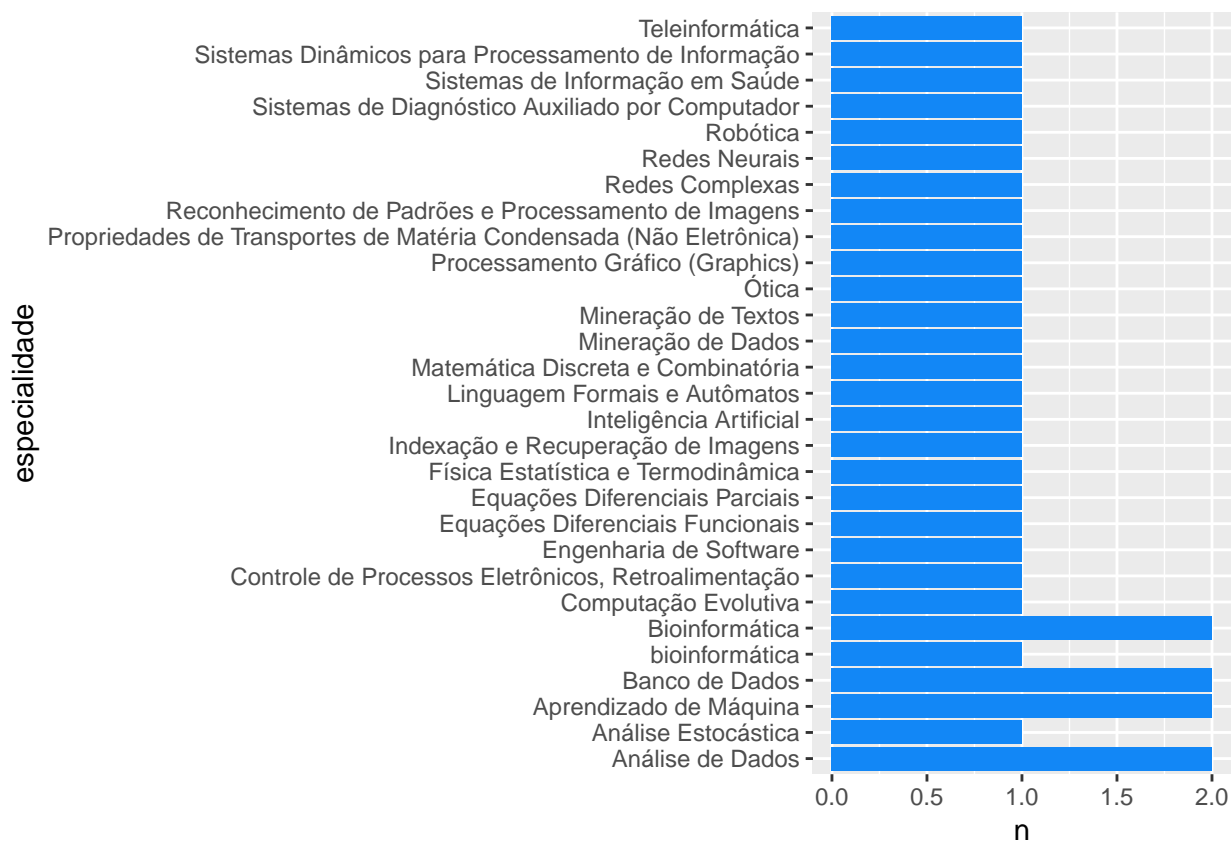


Figura 9: Especialidade dos docentes

Também podemos notar que é predominante a existência de pesquisadores nas áreas ligadas a exatas e a tecnologia.

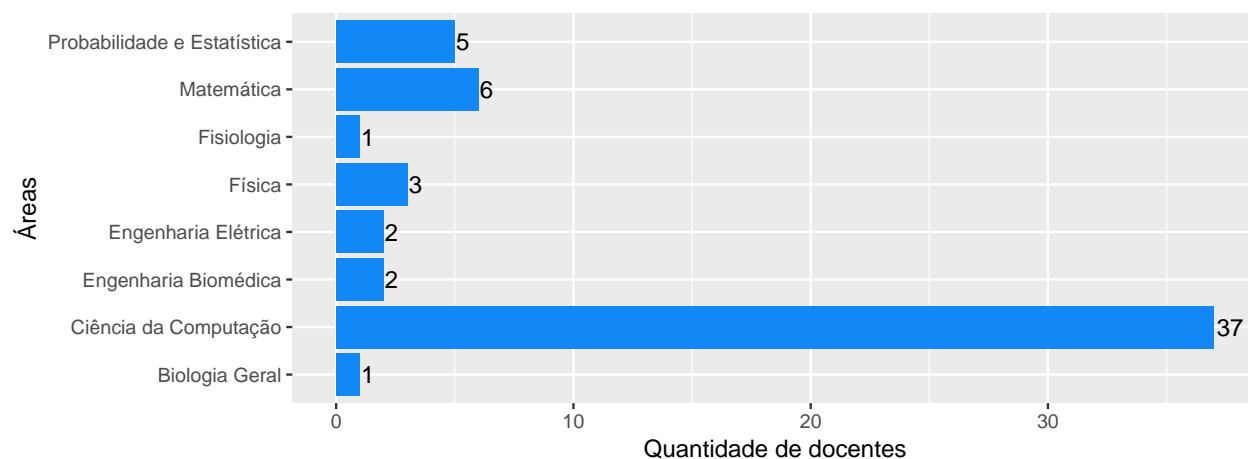


Figura 8: Quantidade de docentes por Área

Por fim, com as análises abaixo, podemos notar a entrada de algumas empresas privadas financiadoras de importância no ano de 2018, como a Intel e a Algar Telecom, em orientações de pós-graduação não concluídas.

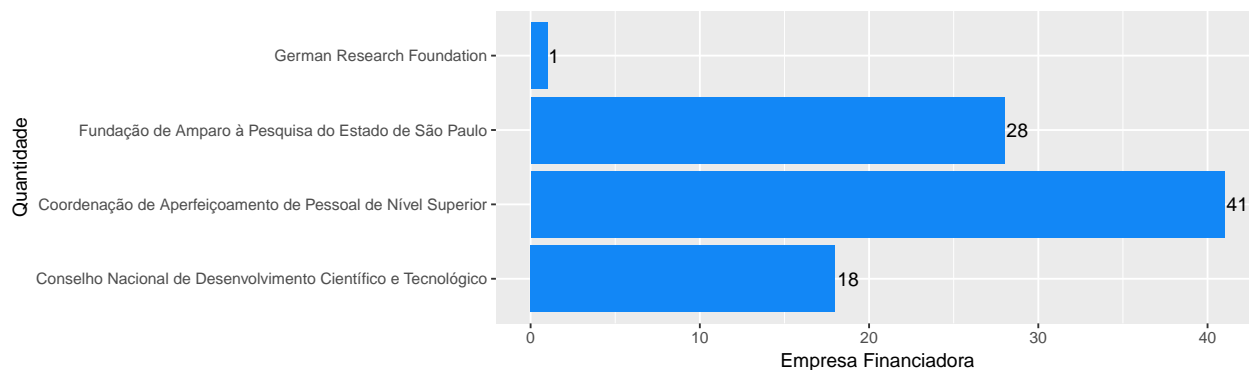


Figura 10: Financiamento de Orientações concluídas de pós-graduação

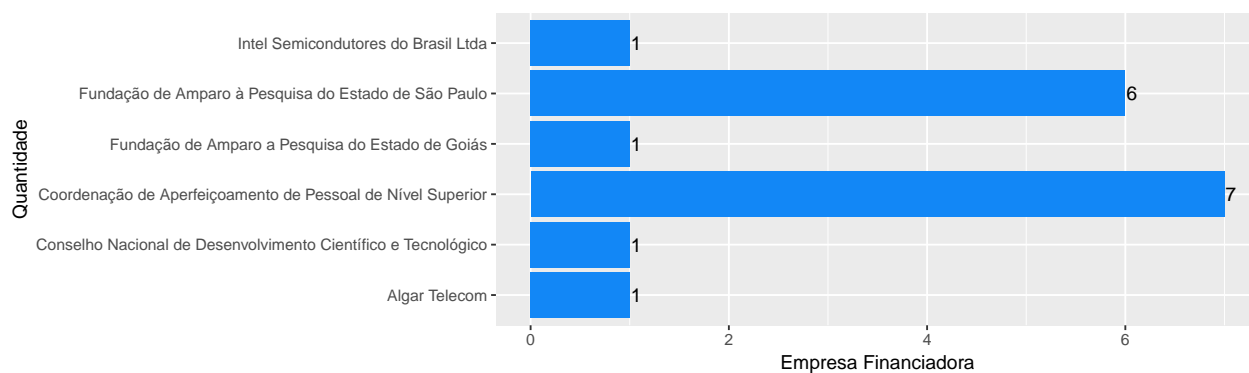


Figura 11: Financiamento de Orientações em andamento de pós-graduação do ano de 2018

Ciência da Computação e Matemática Aplicada (USP/SC)

O gráfico abaixo nos apresenta números a respeito dos tipos de orientação concluídas em função dos anos. Podemos observar uma significativa queda nas orientações de mestrados concluídas no ano de 2017. Em contrapartida, temos um aumento significativo nas orientações concluídas de doutorados. Curiosamente, no ano de 2018, os números de orientações concluídas de doutorado e mestrado se aproximam bastante. Vale ressaltar que essas orientações são a soma de todos os programas de pós-graduação em que os docentes fazem parte.

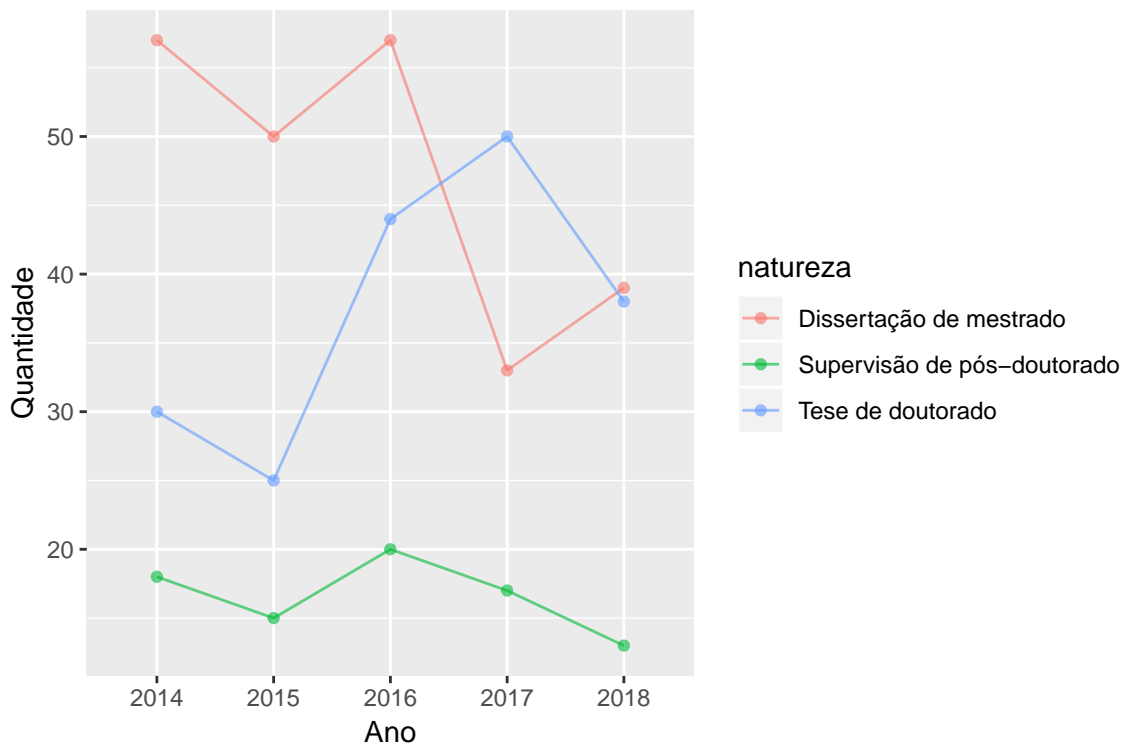


Figura 12: Natureza das supervisões concluídos

O gráfico abaixo nos apresenta números a respeito dos tipos de orientação não concluídos em função dos anos. Podemos observar um significativo aumento, a cada ano, nas orientações de mestrado não concluídas, chegando a seu maior valor do período em 2018. Com sua maior variação ocorrendo em 2017. Em contrapartida, no período de 2017 a 2018, observa-se uma grande queda nas teses de doutorado.

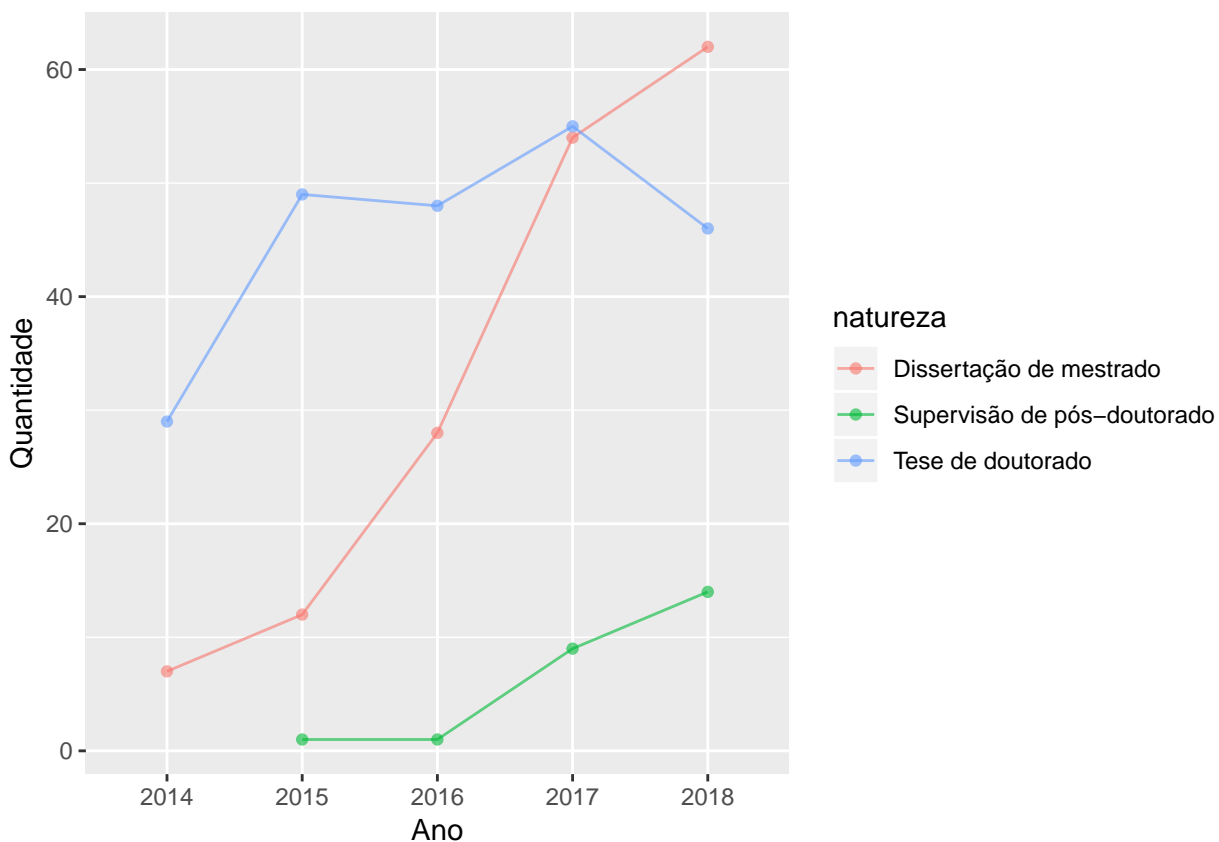


Figura 13: Natureza das supervisões em andamento

Com o gráfico abaixo, observamos que a maioria dos docentes do programa de pós-graduação participou de eventos internacionais. Embora a maioria dos eventos atendidos pelos docentes serem no Brasil, observa-se um grande número de presenças em eventos internacionais principalmente nos Estados Unidos.

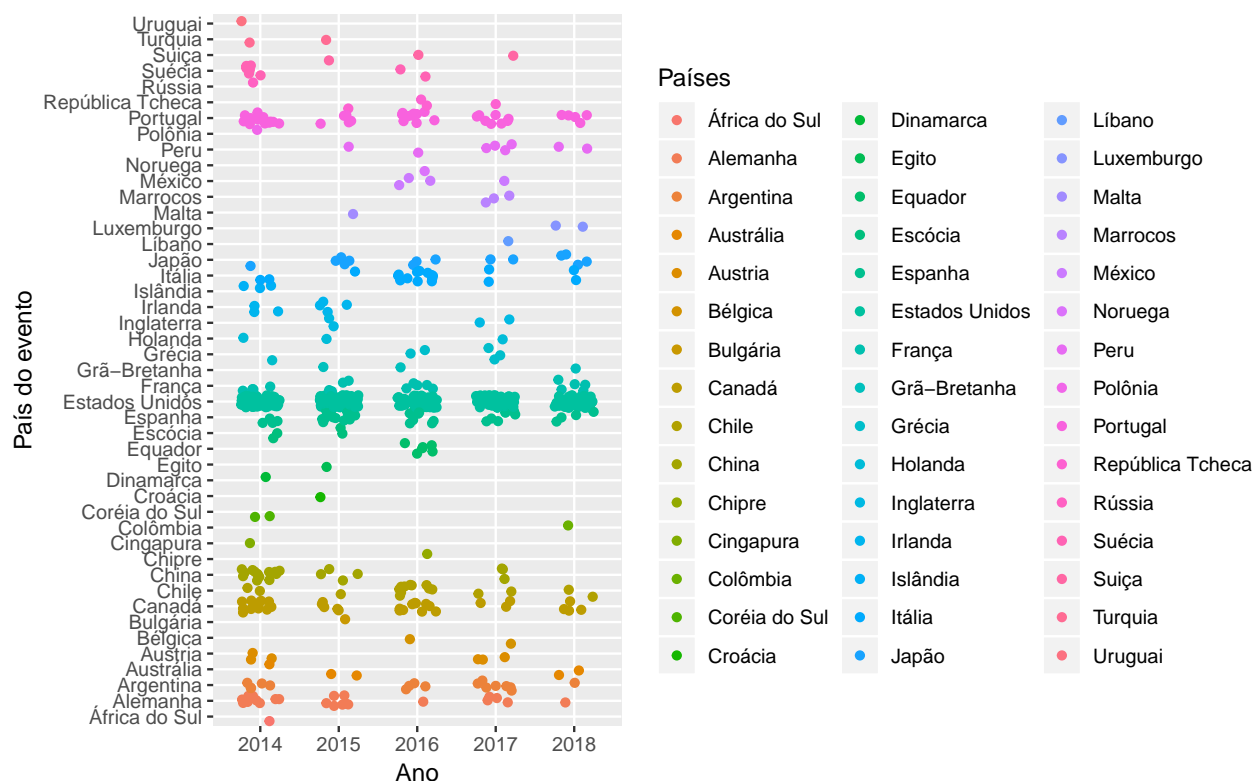


Figura 14: Eventos internacionais comparecidos por docentes

Com as análises abaixo, percebemos uma leve queda, com o passar dos anos, nos números de publicações em eventos. Essa queda foi relativamente expressiva quando comparado ao ano de 2014 - que foi o período de início das análises -. Já o número de publicações em eventos por ano apenas oscilou decrescentemente com o decorrer dos anos.

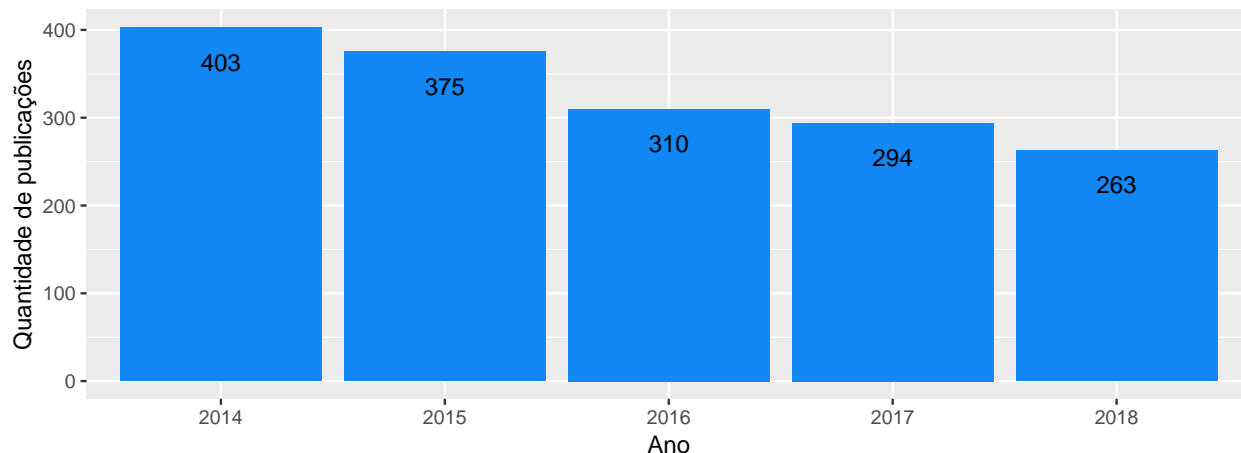


Figura 15: Quantidade de publicações em eventos por ano

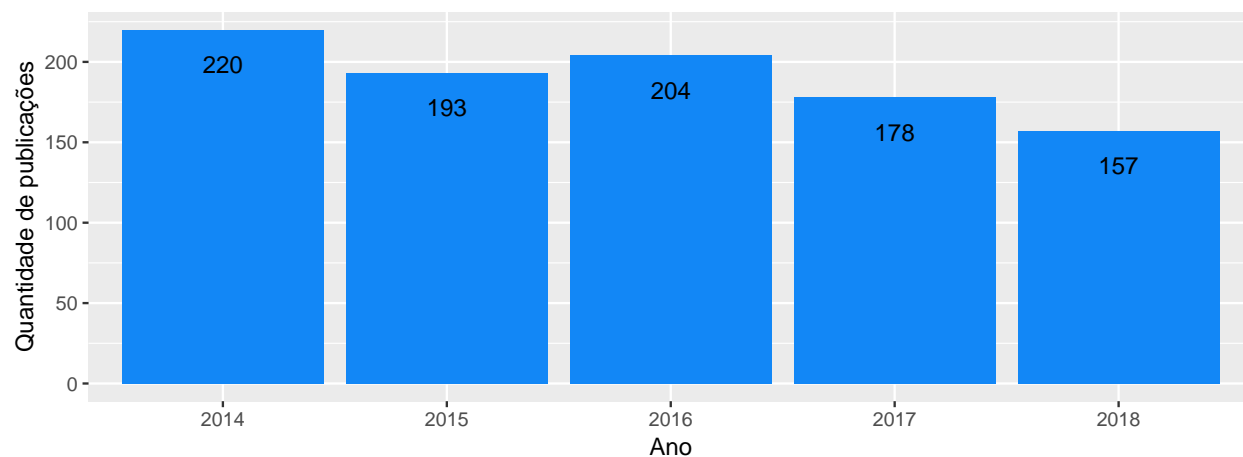


Figura 16: Quantidade de publicações em periódicos por ano

Abaixo, podemos notar que grande parte dos docentes desse programa possuem como grande área a de Ciências Exatas e da Terra seguido por Engenharias. Esse fato nos mostra que as grandes áreas dos pesquisadores envolvidos no programa estão condizentes com a proposta do programa de pós-graduação.

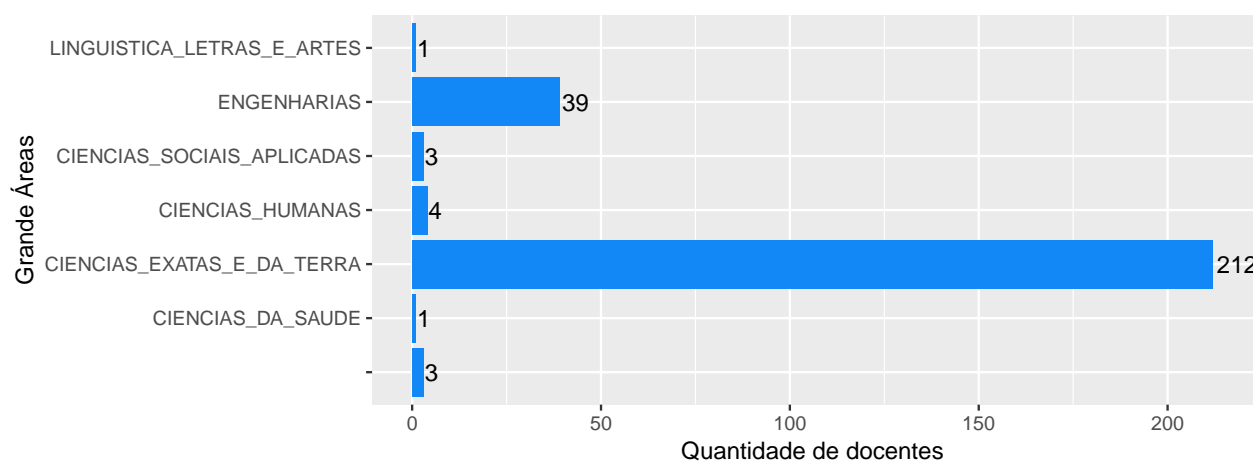


Figura 17: Quantidade de docentes por Grande Área

Também podemos notar que é predominante a existência de pesquisadores nas áreas ligadas a exatas e a tecnologia. Nota-se, também, uma grande pluralidade de áreas, incluindo áreas das ciências humanas e da saúde.

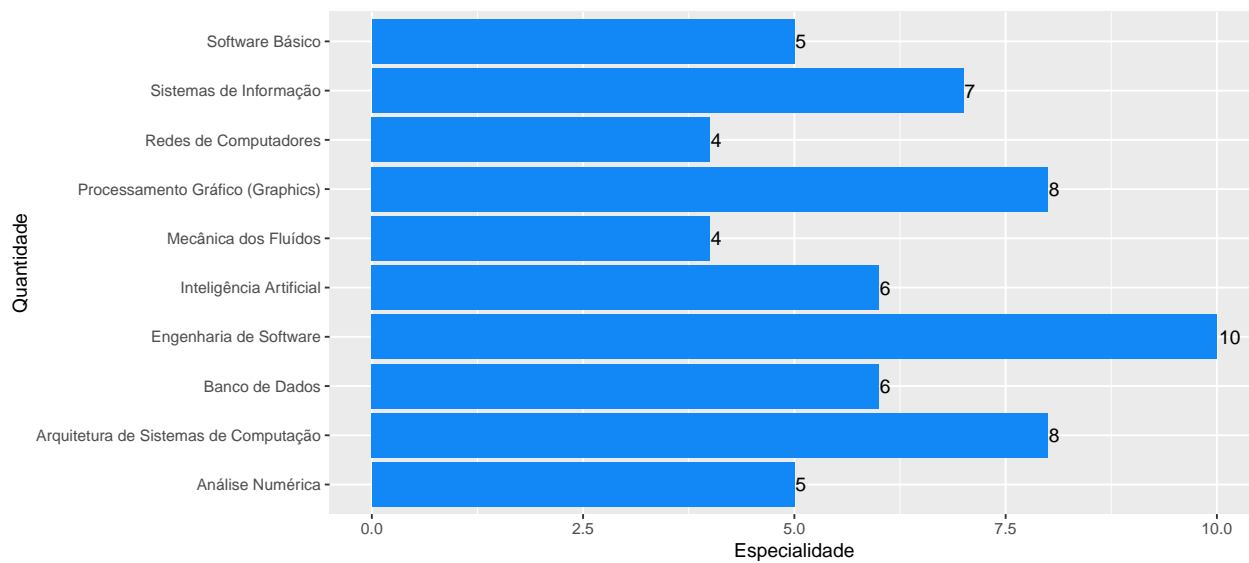


Figura 19: Quantidade de docentes por especialidade

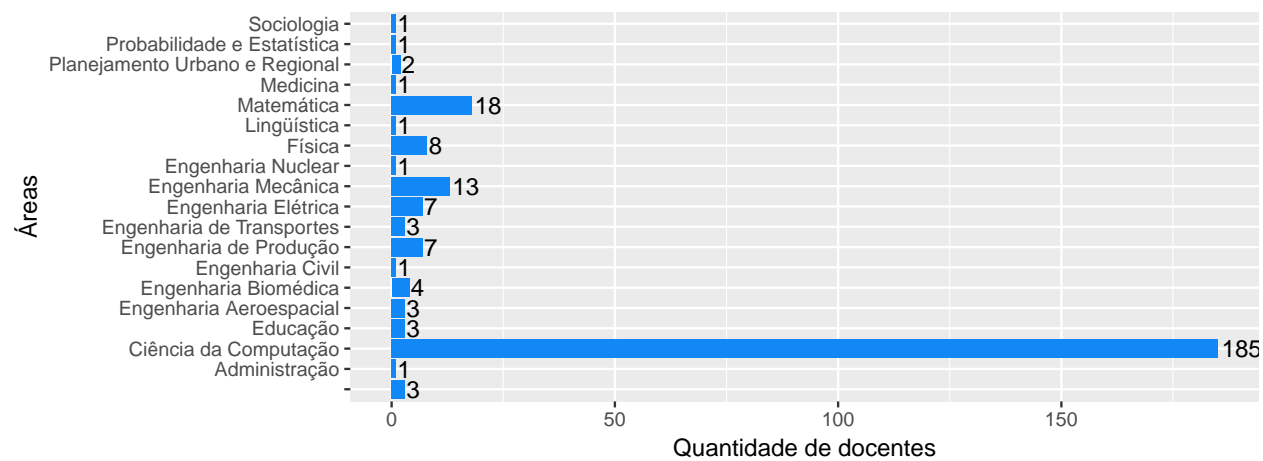


Figura 18: Quantidade de docentes por Área

Por fim, com as análises abaixo, podemos notar a entrada de algumas empresas privadas financiadoras de importância no ano de 2018, como a Intel e a Algar Telecom, em orientações de pós-graduação não concluídas. É evidente a importância dos docentes desse programa de pós-graduação, pois existem investimentos de grandes empresas internacionais como a Google, Samsung e até mesmo a União Europeia. Ainda assim, é notório a grande participação no investimento público do Governo Federal e do Governo do Estado de São Paulo.

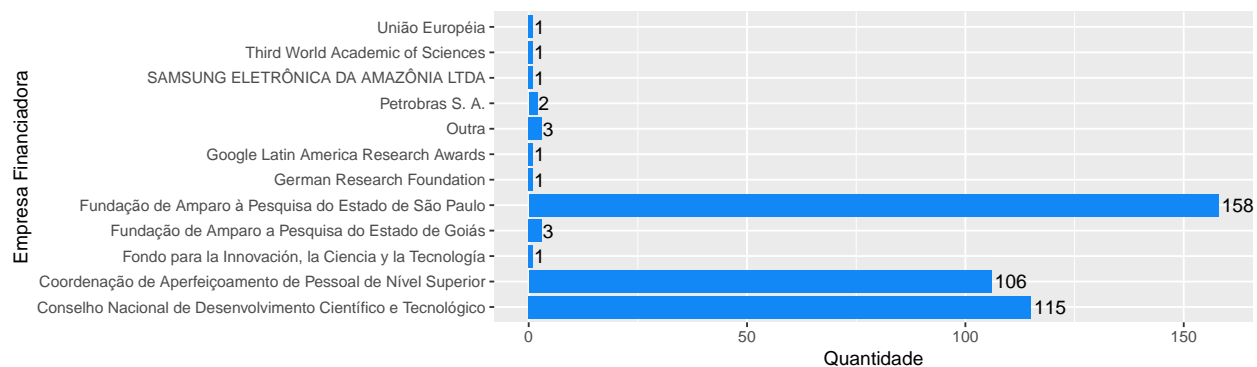


Figura 20: Financiamento de Orientações concluídas de pós-graduação do ano de 2018

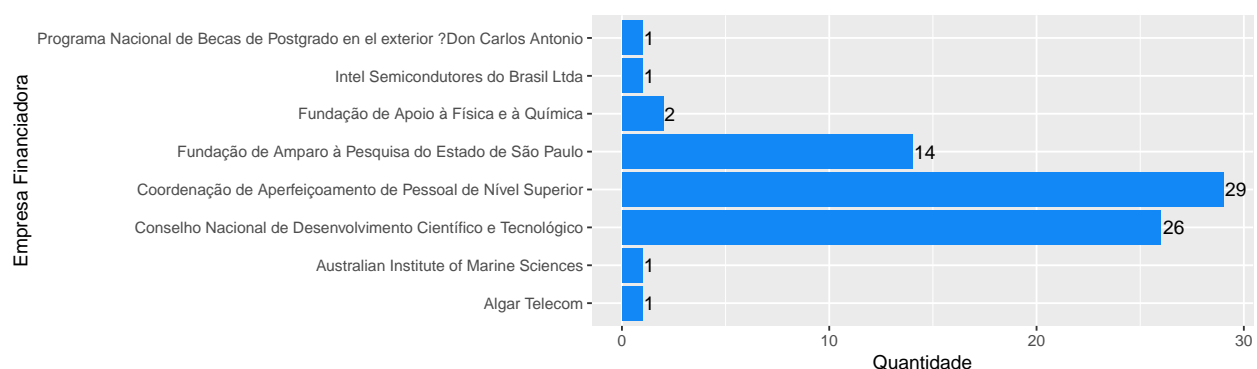


Figura 21: Financiamento de Orientações em andamento de pós-graduação do ano de 2018

Mineração de texto

A coleta de dados realizada para a produção das nuvens de palavras foi feita utilizando os títulos de publicações em periódicos de cada programa de pós graduação estudado. Para refinar essa análise, utilizamos alguns filtros para remover conectivos, pontuações e palavras indesejadas.

- Computação aplicada (USP/RP)

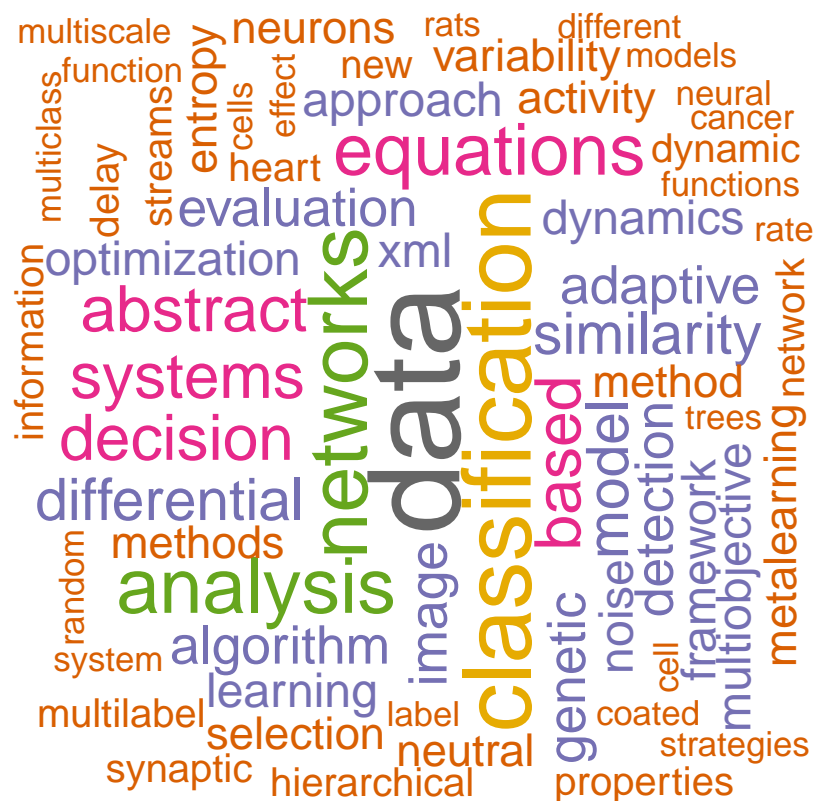


Figura 22: Nuvem de palavras - Computação Aplicada (USP/RP)

- Ciência da computação e matemática aplicada (USP/SC)

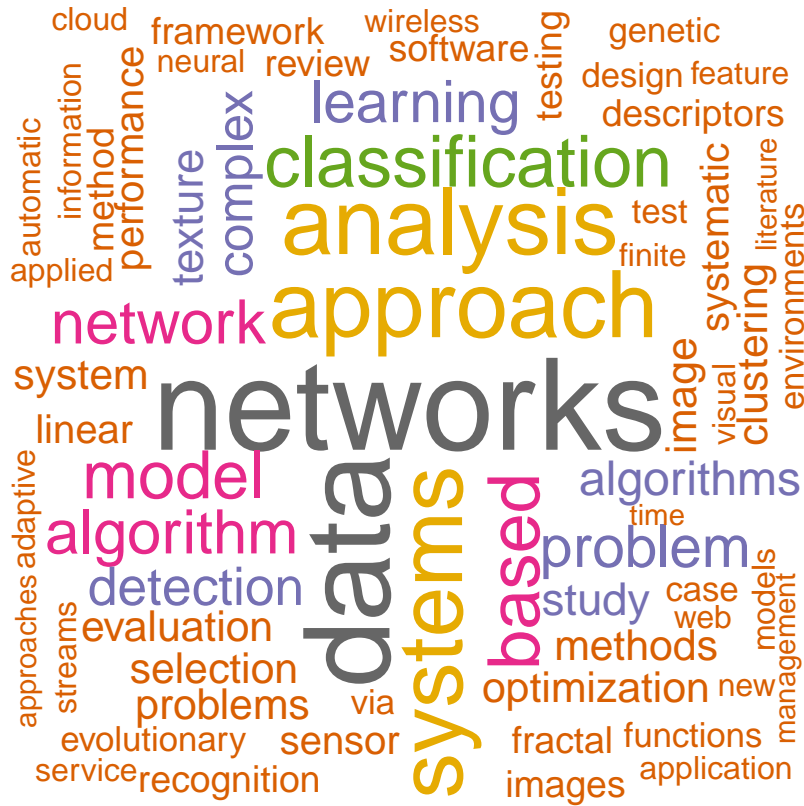


Figura 23: Nuvem de palavras de Ciência da Computação e Matemática Aplicada (USP/SC)

Análise de Redes

A análise de redes é a monitorização de padrões entre nós de uma rede, com o fim de estabelecer uma estrutura que pode ser usada para rastrear possíveis relações de quantidade de comunicação e de agrupamentos.

Para cada programas de pós graduação, foram feitos três grafos para essa análise.

- O primeiro mostra o número de colaborações de cada autor pelo tamanho do nó relacionado a seu índice da tabela de autores.
- O segundo mostra o número de publicações de cada autor pelo tamanho do nó relacionado a seu índice da tabela de autores
- O terceiro mostra as comunidades entre os autores com agrupamentos no grafo por cores

Seguem os grafos obtidos:

- Computação aplicada (USP/RP)

Tabela 3: Tabela de referência da análise de redes - Computação aplicada (USP/RP)

| Index | IdLattes | Docente |
|-------|------------------|--|
| 1 | 0145887238250184 | Edson dos Santos Moreira |
| 2 | 0241862487783754 | Roseli Aparecida Francelin Romero |
| 3 | 0359360551958223 | Claudio Fabiano Motta Toledo |
| 4 | 0756373063566786 | Maristela Oliveira dos Santos |
| 5 | 1201079310363734 | Alexandre Cláudio Botazzo Delbem |
| 6 | 1278004515460973 | Maria Cristina Ferreira de Oliveira |
| 7 | 1361624597372651 | Fabício Simeoni de Sousa |
| 8 | 1531713258988427 | Zhao Liang |
| 9 | 1743726998195607 | Roberto Federico Ausas |
| 10 | 1916886621452638 | Eduardo Marques |
| 11 | 2046970162506930 | Marina Andretta |
| 12 | 2153014839354888 | Francisco Aparecido Rodrigues |
| 13 | 2844974351441051 | Márcio Eduardo Delamaro |
| 14 | 2854771102810220 | Rudinei Goularte |
| 15 | 3030047284254233 | Seiji Isotani |
| 16 | 3185627819021012 | Vanderlei Bonato |
| 17 | 3197500103141618 | Franklina Maria Bragion de Toledo |
| 18 | 3521622923263530 | Gustavo Enrique de Almeida Prado Alves Batista |
| 19 | 3549648389944031 | Maria das Graças Volpe Nunes |
| 20 | 3559042497669898 | Kalinka Regina Lucas Jaquie Castelo Branco |
| 21 | 3794241680729178 | Luis Gustavo Nonato |
| 22 | 4254001833729264 | Paulo Sérgio Lopes de Souza |
| 23 | 4328003866597876 | Fernando Vieira Paulovich |
| 24 | 4462292447937016 | Murilo Francisco Tome |
| 25 | 4694142315497364 | Sarita Mazzini Bruschi |
| 26 | 4793072701914550 | Sandra Maria Aluísio |
| 27 | 4796921913434370 | Odemir Martinez Bruno |
| 28 | 4962820320879891 | Maria da Graca Campos Pimentel |
| 29 | 5041497500746910 | Moacir Antonelli Ponti |
| 30 | 5106492431042243 | Eduardo Fontoura Costa |
| 31 | 5118629875846648 | Caetano Traina Junior |
| 32 | 5136155977351408 | Agma Juci Machado Traina |
| 33 | 5138799833042314 | Rosane Minghim |

| Index | IdLattes | Docente |
|-------|------------------|---|
| 34 | 5337976093923686 | Luciano da Fontoura Costa |
| 35 | 5368167403557640 | Gustavo Carlos Buscaglia |
| 36 | 5433967267727516 | Júlio Cezar Estrella |
| 37 | 5434724108176150 | Elias Salomão Helou Neto |
| 38 | 5877415967509579 | Rosana Teresinha Vaccare Braga |
| 39 | 6253898580239376 | Elaine Parros Machado de Sousa |
| 40 | 6806241288652019 | Dilvan de Abreu Moreira |
| 41 | 6840478133476887 | Rodrigo Fernandes de Mello |
| 42 | 7029133375366014 | Marcio Fuzeto Gameiro |
| 43 | 7078737101024368 | Thiago Alexandre Salgueiro Pardo |
| 44 | 7082447037374118 | Renata Pontin de Mattos Fortes |
| 45 | 7137178343756327 | Simone do Rocio Senger de Souza |
| 46 | 7320090199496500 | José Alberto Cuminato |
| 47 | 7396818382676736 | Fernando Santos Osório |
| 48 | 7489482613903725 | Francisco José Monaco |
| 49 | 7491976072797902 | Leandro Franco de Souza |
| 50 | 7494142007764616 | Elisa Yumi Nakagawa |
| 51 | 7514074129301673 | João Luís Garcia Rosa |
| 52 | 7879216332273376 | Tiago Pereira da Silva |
| 53 | 7913302545613108 | Ellen Francine Barbosa |
| 54 | 8098209307634371 | Jó Ueyama |
| 55 | 8526960535874806 | Solange Oliveira Rezende |
| 56 | 8598262641668520 | Marcelo Garcia Manzato |
| 57 | 8713285339065842 | Cristina Dutra de Aguiar Ciferri |
| 58 | 8763967270082717 | Diego Raphael Amancio |
| 59 | 8807333466702951 | Jose Carlos Maldonado |
| 60 | 8839646082565060 | Ricardo José Gabrielli Barreto Campello |
| 61 | 8948846267711214 | Denis Fernando Wolf |
| 62 | 9129748440846907 | João do Espirito Santo Batista Neto |
| 63 | 9140990601961322 | Alneu de Andrade Lopes |
| 64 | 9274817658247444 | Afonso Paiva Neto |
| 65 | 9347760158370905 | José Fernando Rodrigues Júnior |
| 66 | 9426650405097319 | João Porto de Albuquerque |
| 67 | 9546302784474665 | Alysson Machado Costa |
| 68 | 9674541381385819 | André Carlos Ponce de Leon Ferreira de Carvalho |
| 69 | 9836776931160228 | Adenilso da Silva Simão |
| 70 | 9869412005300784 | Robson Leonardo Ferreira Cordeiro |
| 71 | 9890260144844047 | Antonio Castelo Filho |

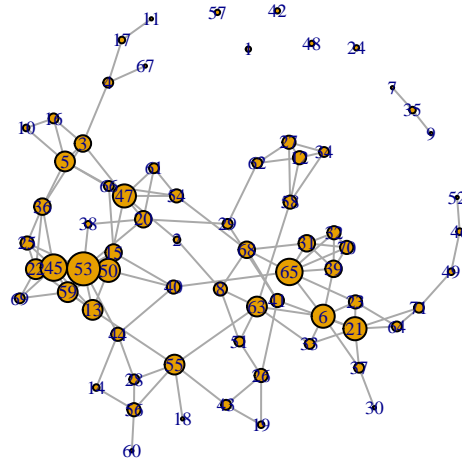


Figura 24: Grafo de colaboração por quantidade de relações - Computação Aplicada (USP/RP)

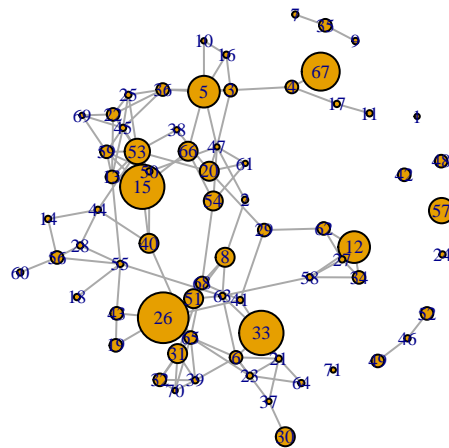


Figura 25: Grafo de colaboração por quantidade de publicações - Computação Aplicada (USP/RP)

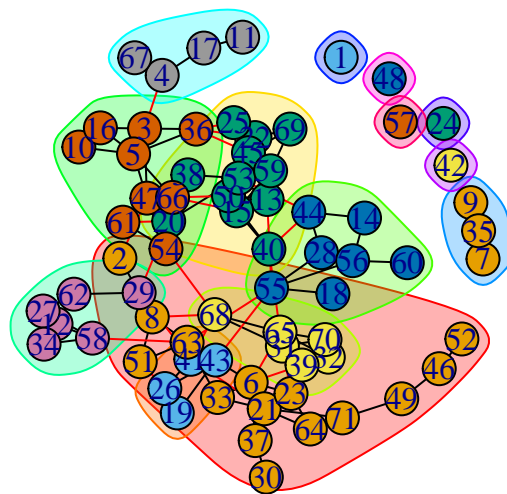


Figura 26: Comunidades geradas pelas colaborações - Computação Aplicada (USP/RP)

- Ciência da computação e matemática aplicada (USP/SC)

Tabela 4: Tabela de referência da análise de redes - Ciência da computação e matemática aplicada (USP/SC)

| Index | IdLattes | Docente |
|-------|------------------|---|
| 1 | 0276583403489241 | Luiz Otavio Murta Junior |
| 2 | 0525430082581472 | Evandro Eduardo Seron Ruiz |
| 3 | 0652038346929546 | Cléver Ricardo Guareis de Farias |
| 4 | 1273134370963830 | Renato Tinós |
| 5 | 1531713258988427 | Zhao Liang |
| 6 | 2407277993285186 | Alessandra Alaniz Macedo |
| 7 | 3262036759491294 | Eduardo Alex Hernández Morales |
| 8 | 3278315914566734 | Ricardo Zorzetto Nicoliello Vêncio |
| 9 | 4050202318600269 | José Augusto Baranauskas |
| 10 | 4609096963207702 | Antonio Carlos Roque da Silva Filho |
| 11 | 5136155977351408 | Agma Juci Machado Traina |
| 12 | 7367695769675241 | Vanessa Rolnik Artioli |
| 13 | 8428230811718070 | Joaquim Cezar Felipe |
| 14 | 8703740676075525 | Alexandre Souto Martinez |
| 15 | 9674541381385819 | André Carlos Ponce de Leon Ferreira de Carvalho |

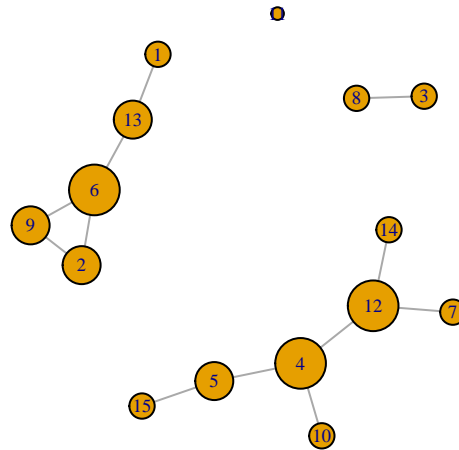


Figura 27: Grafo de colaboração por quantidade de relações - Ciência da Computação e Matemática Aplicada (USP/SC)

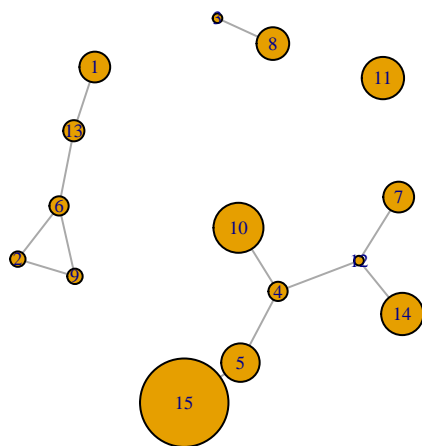


Figura 28: Grafo de colaboração por quantidade de publicações - Ciência da Computação e Matemática Aplicada (USP/SC)

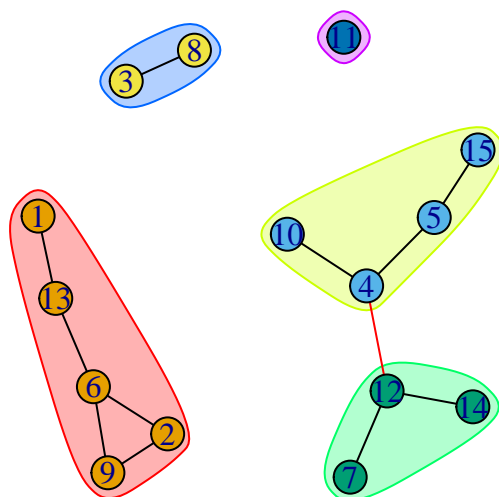


Figura 29: Comunidades geradas pelas colaborações - Ciência da Computação e Matemática Aplicada (USP/SC)

Análise dos resultados

Analisando as orientações concluídas do programa de pós-graduação podemos perceber uma ampla diferença, tanto na quantidade de docentes quanto na produção científica em si. O programa Ciência da Computação e Matemática Aplicada, além de mais antigo, possui um número significativo de docentes, consequentemente, maior participação internacional, observando um foco nos Estados Unidos - juntamente no programa de pós-graduação de Computação Aplicada -, e também em diversos países da Ásia e Europa. Além disso, observamos um maior número de grandes empresas financiadoras, em comparação ao programa de pós-graduação de Computação Aplicada.

Pelas nuvens de palavras geradas é possível observar que existem certas similaridades entre os temas tratados em cada programa de pós graduação. Nota-se que as palavras *network* e *data* são amplamente utilizadas nos periódicos produzidos pelos docentes. Esse fato é esperado, visto que existe certa similaridade entre os principais campos de estudo abordados entre os programas estudados nesse relatório.

Com as redes geradas, observamos certos comportamentos curiosos. Existem docentes que possuem um alto numero de publicações, porém esses mesmos docentes não possuem muitas conexões entre eles. Exemplificando, o docente André Carlos Ponce de Leon Ferreira de Carvalho do programa de pós graduação Ciência da computação e matemática aplicada, possui poucas conexões, porém quando observamos o grafo de quantidade de publicações, verifica-se um grande número de publicações para esse mesmo docente.

Todas essas observações comprovam as disparidades entre os programas, justificando uma nota 7 no programa Ciência da Computação e Matemática Aplicada e 3 no programa Computação Aplicada.

Conclusão

Os programas de pós graduação de Computação aplicada (USP/RP) e de Ciência da computação e Matemática aplicada (USP/SC) possuem bastantes informações para análise, e as suas semelhanças e diferenças são observadas ao filtrar os dados e ao gerar os gráficos. Como a pesquisa a respeito dos resultados desses programas é a manipulação e o estudo de big data, a ciência de dados foi tido como recurso ideal para filtrar as informações essenciais da massa de dados e fornecer o meio de predições e decisões necessárias.

Para obter um bom resultado nas análises, foram utilizadas as metodologias do CRISP-DM, técnicas de mineração e manipulação de dados em R e análise crítica das melhores formas de gerar os gráficos de acordo com os tipos de dados.

Por fim, as análises obtidas foram satisfatórias para o conjunto de dados, e os resultados são coerentes ao comparar as pontuações de avaliação quadrienal, que colocam o programa Ciência da computação e Matemática aplicada (USP/SC) como um programa nota 7 que possui um alto número de publicações, e também consegue justificar o programa Computação aplicada (USP/RP) com nota 3 por ser um programa recente, e ainda não possuir tantas publicações.

Referências

- [1] Jorge Henrique Cabral Fernandes, Ricardo Barros Sampaio. Unb Aprender, 2019. Sobre a Ciência e sua Avaliação. Disponível em https://aprender.ead.unb.br/pluginfile.php/474549/mod_resource/content/2/SobreCiencia.pdf. Acesso em: 13/04/2019.
- [2] IBM. IBM, 2012. CRISP-DM Help Overview Disponível em https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm. Acesso em: 13/04/2019.
- [3] Wikipedia. Wikipedia, 2019. Cross Industry Standard Process for Data Mining Disponível em: https://pt.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining. Acesso em: 13/04/2019.
- [4] ICMC-USP. ICMS USP, 2019 Disponível em: <https://icmc.usp.br/pos-graduacao/ppgccmc>. Acesso em: 13/04/2019.
- [5] Avaliação Quadrienal. Ficha de Avaliação do Programa CIÊNCIAS DA COMPUTAÇÃO E MATEMÁTICA COMPUTACIONAL (33002045004P1) UNIVERSIDADE DE SÃO PAULO (USP) Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/consultaFichaAvaliacao.xhtml>. Acesso em: 13/04/2019.
- [6] Avaliação Quadrienal. Ficha de Avaliação do Programa COMPUTAÇÃO APLICADA (33002029052P5) UNIVERSIDADE DE SÃO PAULO (USP) Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/consultaFichaAvaliacao.xhtml>. Acesso em: 13/04/2019.
- [7] Brasil Escola. Método científico Disponível em: <https://brasilecola.uol.com.br/quimica/metodo-cientifico.htm>. Acesso em: 22/05/2019.