

Análise de agrupamento dos cursos de graduação de universidades federais brasileiras

Beli Thomaz Marcatti¹, Andreia Zanella² e Andréa Cristina Konrath³

Universidade Federal de Santa Catarina, Campus Universitário Reitor João David Ferreira, Florianópolis. Brasil. E-mails: ¹belimarcatti@gmail.com; ²andrea.zanella@ufsc.br; ³andrea.ck@ufsc.br

Resumo: A avaliação da educação superior é uma importante ferramenta para o desenvolvimento da qualidade do ensino. Atualmente, o Conceito Preliminar de Curso (CPC) é considerado o principal indicador de qualidade da educação superior brasileira. Uma forma de identificar oportunidades de melhoria da qualidade é a observação das boas práticas entre os pares, ou seja, entre cursos com características similares. Este trabalho tem por objetivo analisar a existência de agrupamentos entre os cursos de graduação em Matemática de instituições federais brasileiras, utilizando as variáveis que compõem o CPC. Para o agrupamento dos cursos, utilizou-se o método não hierárquico k-médias, o qual possibilitou dividir os cursos em três grupos. A análise do perfil dos cursos mostrou que o primeiro *cluster* é formado por cursos com notas altas nas componentes referentes ao aprendizado do aluno e à titulação dos professores; o segundo *cluster* uniu os cursos com notas intermediárias na maioria das componentes; e o terceiro *cluster* agrupou os cursos com notas mais altas nas componentes referentes à percepção dos alunos sobre a estrutura do curso. Os resultados apresentados poderão auxiliar os gestores na identificação de oportunidades de melhoria através da observação das boas práticas entre os cursos com características similares.

Palavras-chave: análise *cluster*, K-médias, avaliação da educação superior, *benchmarking*.

Title: Cluster analysis of undergraduate mathematics courses of Brazilian federal universities.

Abstract: The evaluation of the higher education is an important tool for the development of teaching quality. Currently, the Conceito Preliminar de Curso (CPC) is considered the main quality indicator of the Brazilian higher education. In order to improve the quality, it is possible to observe the best practices examples among the peers, i.e. among the undergraduate courses with a similar profile. This paper aims to analyze the existence of clusters between the undergraduate courses in Mathematics of Brazilian federal institutions, utilizing the variables that comprise the CPC. The non-hierarchical k-means method was used to group the courses, which made it possible to divide them into three clusters. The analysis of the courses' profile revealed that the first cluster consisted of courses with high scores in the components related to the student's learning and to the professors' qualifications. The second cluster consisted of courses with intermediate scores in most components; and, the third cluster comprised

courses with higher scores in components related to the students' perceptions of the course's structure. The results of this study may assist managers in identifying opportunities for improvement through the observation of best practices among courses sharing similar profile.

Keywords: cluster analysis, K-means, evaluation of undergraduate courses, benchmarking.

Introdução

A avaliação das Instituições de Ensino Superior (IES) no Brasil começou a ser realizada em 1983 e, desde então, diferentes programas de avaliação foram implementados. Em 2004, foi instituído o Sistema Nacional de Avaliação da Educação Superior, que é o programa vigente atualmente (Zainko, 2008).

A avaliação de desempenho no ensino superior, quando bem conduzida, produz resultados relevantes para informar a sociedade sobre a qualidade dos cursos e estimula as instituições a melhorar seu desempenho (Schwartzman, 2008). Além de apoiar a elaboração de políticas públicas, as informações advindas das avaliações permitem que as IES e os cursos de graduação aperfeiçoem suas práticas a partir da comparação de seu desempenho com o desempenho de seus pares.

Além disso, um aspecto fundamental dessas avaliações de desempenho é o seu impacto no ensino-aprendizagem. A interconexão entre avaliações em larga escala e o ensino-aprendizagem das ciências é vital no cenário educacional contemporâneo, pois essas avaliações têm potencial de orientar as práticas instrucionais dos educadores, estimulando os professores a oferecer um currículo mais rico e equilibrado (Shimizu, 2011; Suurtamm *et al.*, 2016; Swan e Burkhardt, 2012).

As avaliações de desempenho também fornecem insumos para a elaboração de indicadores educacionais. Um dos indicadores de qualidade da educação superior elaborados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é o Conceito Preliminar de Curso (CPC). Atualmente, o CPC é considerado o principal indicador da qualidade voltado aos cursos superiores brasileiros (Ikuta, 2016). O indicador é calculado a partir de uma média ponderada entre oito componentes, esta média é transformada em um conceito que varia entre 1 e 5. Todos os anos são avaliados cursos diferentes de forma que um mesmo curso seja avaliado a cada três anos (Ministério da Educação Brasil, 2020). O CPC, como indicador de qualidade, é uma ferramenta útil para obter informações a respeito da situação da Educação Superior do país e para analisar e direcionar o desenvolvimento e progresso da qualidade educacional (Ikuta, 2016).

Os cursos de graduação vêm trabalhando para melhorar a qualidade de ensino e, por consequência, obter um conceito alto na avaliação do MEC. Uma forma de identificar oportunidades de melhoria é a observação das boas práticas entre os pares, ou seja, entre cursos com características similares. Como destacado por Piran, Lacerda e Camargo (2018), para que uma análise de *benchmarking* seja eficaz, é importante que as unidades a

serem comparadas sejam relativamente similares. Neste contexto, observa-se a importância de estudar as similaridades entre os cursos de graduação.

O objetivo geral deste trabalho consiste em analisar a existência de agrupamentos entre os cursos de graduação em Matemática, tendo como referência as variáveis que são utilizadas para a construção do Conceito Preliminar de Curso. A identificação e caracterização desses agrupamentos têm o potencial de promover a observação das boas práticas entre cursos com características similares e, conseqüentemente, fornecer uma reflexão mais profunda sobre como os cursos de graduação podem adaptar seus currículos, recursos e estratégias de ensino para melhor atender às necessidades de formação em ciências, particularmente em Matemática, conforme padrões de desempenho e agrupamentos identificados.

Para a condução da pesquisa, primeiramente, foi realizada uma análise descritiva e exploratória das notas dos cursos de Matemática nas componentes do Conceito Preliminar de Curso e aplicado um algoritmo para agrupamento dos cursos de graduação. Para agrupar os cursos de graduação utilizou-se um algoritmo não hierárquico de agrupamento chamado k-médias. Posteriormente, buscou-se caracterizar os agrupamentos obtidos por meio da identificação das similaridades (características comuns) dentro dos clusters e dissimilaridade entre os clusters. Os algoritmos de agrupamento têm contribuído significativamente para extrair conhecimento de dados educacionais, principalmente em relação à identificação de padrões de desempenho e ao estudo das similaridades entre estudantes, cursos ou instituições (Varghese *et al.*, 2011; Huppés *et al.*, 2016).

Este artigo está organizado em cinco seções. A primeira teve como propósito contextualizar o tema, apresentar a motivação e o objetivo do trabalho. Na segunda seção será apresentado o referencial teórico relevante para o desenvolvimento do trabalho. Na terceira seção será apresentada a metodologia a ser seguida. Na quarta seção serão apresentados os resultados do agrupamento aplicado e as análises decorrentes dele. Por fim, a quinta seção traz as principais conclusões do estudo e sugestões para trabalhos futuros.

Referencial teórico

A avaliação da educação superior brasileira e o Conceito Preliminar de Curso

A política da avaliação da Educação Superior em vigência no Brasil é regida pela Lei nº 10.861, de 2004 (Ministério da Educação Brasil, 2004), que instituiu o Sistema Nacional de Avaliação da Educação Superior (Sinaes). O Sinaes tem como objetivo "assegurar o processo nacional de avaliação das instituições de Educação Superior, dos cursos de graduação e do desempenho acadêmico de seus estudantes" (Ministério da Educação Brasil, 2004) e é estruturado por três grandes pilares: a avaliação institucional, composta pela autoavaliação e a avaliação externa; a avaliação de cursos, por meio de visitas *in loco* de avaliadores externos; e a avaliação dos estudantes, pelo Exame Nacional de Desempenho dos Estudantes (Enade).

O Sinaes é um sistema que envolve toda a nação, por isso ele deve envolver todas as instituições de ensino superior, tanto no âmbito público como no privado (Ristoff e Giolo, 2006). Neste sentido, a avaliação deve ter caráter formativo e construtivo, voltada à efetividade científica e social, e não deve ter o foco apenas em medir, classificar e gerar rankings sobre a produtividade e nível de conhecimento das instituições. Mais do que isso, em cada fase avaliativa deve-se reconhecer o meio e as condições reais em que aquela instituição e comunidade envolvida estão inseridas e fornecer subsídios para a melhoria da qualidade (Ministério da Educação Brasil, 2003).

A avaliação do desempenho dos estudantes é realizada por meio do Enade. Este exame analisa o desempenho dos estudantes com relação aos conteúdos programáticos de cada curso, sendo composto por uma prova de formação geral, com peso de 75%, e uma de componente específico, com peso de 25% da nota (Lacerda e Ferri, 2017). A avaliação do desempenho dos estudantes, está diretamente ligada ao Conceito Preliminar de Curso (CPC), pois fornece insumos para o cálculo do CPC.

O CPC, criado no âmbito do Sinaes em 2008, atribui um conceito aos cursos de graduação levando em conta dados advindos principalmente do Enade e do Censo da Educação Superior. De acordo com Griboski e Fernandes (2016), o CPC fornece uma "fotografia" da situação educacional e possibilita traçar um panorama da educação superior brasileira.

Para o cálculo do CPC, são consideradas as seguintes dimensões: o desempenho dos estudantes, o valor agregado pelo processo formativo do curso, indicadores do corpo docente e as condições ofertadas para o desenvolvimento do processo formativo. Os cursos de graduação que têm o CPC calculado, recebem um conceito que pode variar entre 1 e 5, em que conceito maior ou igual a 3 é considerado satisfatório e menor que 3 é insatisfatório. Essa classificação pelo conceito passou a permitir que cursos com nota satisfatória fossem dispensados da avaliação *in loco*, ou seja, ela só é obrigatória para instituições com CPC menor que três, o que acabou gerando economia para o país (Verhine e Dantas, 2009).

A Tabela 1 detalha as dimensões e as componentes atuais do CPC. O sistema de pesos utilizados no cálculo do indicador também é apresentado.

A primeira dimensão está associada ao desempenho dos estudantes, é dada pela Nota dos Concluintes no Enade (NC) e possui um peso de 20% no CPC.

A segunda dimensão também é composta por uma única componente: a Nota do Indicador de Diferença entre o Observado e o Esperado (NIDD). Ela compara o desempenho dos concluintes do curso no Enade com as características dos estudantes ao ingressar no curso pelo Exame Nacional do Ensino Médio (Enem). Tem como objetivo de verificar o valor agregado pelo processo formativo do curso de graduação. E esta dimensão tem peso de 35% no indicador. O cálculo do NIDD leva em conta o número de estudantes concluintes participantes no Enade com resultados válidos, o desempenho geral dos estudantes participantes do exame, o desempenho dos estudantes no Enem e o número de participantes no Enade com nota do Enem recuperada.

Dimensão	Componentes	Pesos	
Desempenho dos Estudantes	Nota dos Concluintes no Enade (NC)	20%	
Valor Agregado pelo Processo Formativo Oferecido pelo Curso	Nota do Indicador da Diferença entre os Desempenhos Observado e Esperado (NIDD)	35%	
Corpo Docente	Nota de Proporção de Mestres (NM)	7,5%	30%
	Nota de Proporção de Doutores (ND)	15%	
	Nota de Regime de Trabalho (NR)	7,5%	
Percepção Discente sobre as Condições do Processo Formativo	Nota referente à Organização Didático-Pedagógica (NO)	7,5%	15%
	Nota referente à Infraestrutura e Instalações Físicas (NF)	5,0%	
	Nota referente às Oportunidades de Ampliação da Formação Acadêmica e Profissional (NA)	2,5%	

Tabela 1 – Composição do CPC e pesos das suas dimensões e componentes
 Fonte: Ministério da Educação Brasil (2020, p.10)

A terceira dimensão está relacionada ao Corpo Docente, em que 7,5% do peso do CPC é atribuído à Nota de Proporção de Mestres (NM) em relação a todos os professores, 15% à Nota de Proporção de Doutores (ND) e 7,5% à Nota de Regime de Trabalho (NR), totalizando 30% no indicador. As duas primeiras componentes são dadas pela proporção de docentes com titulação igual ou superior a Mestre e Doutor, respectivamente, em relação ao número total de docentes do curso; a nota de regime de trabalho vem da proporção de docentes em regime de trabalho parcial ou integral em relação ao número total de docentes do curso. As notas referentes ao Corpo Docente são obtidas do Censo da Educação Superior.

A quarta dimensão trata da Percepção Discente sobre o Processo Formativo. Observa-se que 7,5% do peso é atribuído à Nota Referente à Organização Didático-Pedagógica (NO), 5% à Nota Referente à Infraestrutura e Instalações Físicas (NF), e 2,5% à Nota referente às Oportunidades de Ampliação da Formação Acadêmica e Profissional (NA), totalizando os últimos 15% de peso na nota do CPC. Essas componentes são estimadas com base no questionário socioeconômico respondido pelos estudantes que realizam o Enade.

Para o cálculo do CPC, as notas das componentes obtidas pelos cursos são padronizadas em z-scores e posteriormente são reescaladas para que variem em uma escala entre 0 e 5. A partir disso, é calculada uma média ponderada com os componentes e pesos descritos na Tabela 1, resultando na Nota Contínua do CPC que posteriormente é transportada em conceito (Ministério da Educação Brasil, 2020).

As avaliações em larga escala e o ensino-aprendizagem das ciências

Recentemente, tem se tornado comum a utilização de avaliações externas por formuladores de políticas em todo o mundo para avaliar o conhecimento dos alunos e compará-lo internacionalmente. De acordo com Suurtamm *et al.* (2016), a interligação entre as avaliações em larga escala

e o ensino-aprendizagem das ciências é fundamental no cenário educacional atual, pois essas avaliações não apenas têm potencial de orientar as práticas de ensino dos educadores, mas também de transmitir mensagens sobre as habilidades ou competências que devem ser desenvolvidas.

Autores como Swan e Burkhardt (2012) e Shimizu (2011) ressaltam a importância de alinhar essas avaliações não apenas com o conteúdo disciplinar, mas também com os processos cognitivos e ações práticas relacionadas ao tema estudado. Além disso, destacam que, se as avaliações refletirem adequadamente o que é fundamental para os alunos aprenderem, elas podem estimular os professores a oferecerem um currículo mais refinado e equilibrado. De acordo com Shimizu (2011) e Paek (2012), essas avaliações não se limitam a medir o conhecimento dos alunos, elas também proporcionam oportunidades de aprendizado, fornecendo aos professores informações valiosas para melhorar sua instrução em sala de aula por meio da conexão explícita entre os padrões de conteúdo de alto nível, o que é medido em avaliações de larga escala, e o que acontece nas salas de aula.

Shimizu (2011) e Fontanive (2013) destacam que um exemplo prático do uso das avaliações externas para o aprimoramento do ensino-aprendizagem em sala de aula é a disponibilização de itens juntamente com documentação sobre os objetivos dos itens e o desempenho dos alunos em cada um deles. O uso desse material em sala de aula apoia o professor na identificação dos tipos de pensamento nos quais os alunos podem estar envolvidos com base na escolha de uma resposta, auxiliando a entender a natureza de concepções equivocadas no pensamento dos alunos com base na resposta incorreta selecionada (Shimizu, 2011).

Além disso, recentemente, a utilização dos dados provenientes das avaliações em larga escala, combinados com a aplicação de técnicas avançadas de análise de dados, tem permitido obter diferentes *insights* sobre como gestores das instituições de ensino superior podem propor políticas que promovam a qualidade e como coordenadores e professores podem ajustar currículos, materiais e estratégias de ensino para atender de forma mais eficaz às demandas de formação em ciências (Zhai, 2021).

Uma forma efetiva de obter esses *insights* é por meio de exercícios orientados de *benchmarking*, ou seja, da observação das boas práticas entre os pares. No contexto educacional, os exercícios de *benchmarking* podem ser realizados por unidades como instituições de ensino, cursos ou grupos de pesquisa que desejam comparar as suas próprias capacidades com as capacidades de outras unidades similares (Nugroho e Jaqin, 2021). A prática de *benchmarking* ajuda a identificar deficiências, estabelecer objetivos de melhoria e inspirar novas ideias, o que tem potencial de contribuir para melhorar o processo de ensino-aprendizagem, desenvolver o currículo dos cursos, melhorar as instalações de infraestrutura física e as operações. Nessa mesma linha, o estudo de Tasopoulou e Tsiotras (2017) indica que a comparação entre pares tem o potencial de promover a excelência nas organizações educacionais. Os autores apresentam metodologias e ferramentas que podem ser aplicadas com sucesso em instituições de ensino superior para alcançar a excelência por meio do uso de práticas de *benchmarking*.

A eficácia de uma análise de *benchmarking* depende da semelhança relativa entre as unidades comparadas (Piran, Lacerda e Camargo, 2018). Algumas técnicas se mostram particularmente promissoras para identificar pares (unidades similares) que podem servir como exemplos de boas práticas. Este é o caso da análise envoltória de dados, que além de avaliar o desempenho, possibilita identificar unidades semelhantes que podem servir como *benchmarks* para as unidades com baixo desempenho. Exemplos do uso dessa técnica para *benchmarking* na área educacional podem ser encontrados em Daultani *et al.* (2021) e Gourishankar e Sai Lokachari (2012). Outra técnica bastante difundida para identificar similaridades ente unidades é a análise de agrupamento (*clustering analysis*). Aplicações em diferentes áreas do conhecimento podem ser encontradas na literatura, inclusive na área educacional, como apresentados nos trabalhos de Urbina-Nájera (2017), Huppel *et al.* (2016) e Scapinelli *et al.* (2022), que usam essa ferramenta com o objetivo de fornecer insumos para a implementação de melhorias desde o nível da sala de aula (ensino-aprendizagem) até níveis de gestão, apoiando tomada de decisão por parte dos gestores.

Zhai *et al.* (2020) ressaltam que essas técnicas podem revelar padrões sutis nos processos cognitivos dos alunos, oferecendo uma visão mais profunda do pensamento e raciocínio dos alunos. De acordo com os autores, ao agregar essas informações ao conhecimento obtido a partir das avaliações em larga escala, podemos avançar significativamente na compreensão da competência dos alunos e, assim, orientar estratégias mais eficazes de ensino e aprendizagem.

A extração de conhecimento de dados educacionais

Atualmente, os dados acadêmicos das instituições de ensino possuem um volume elevado, por isso, para a efetiva análise desses dados, é necessário que haja uma investigação baseada na extração do conhecimento (Hadioui, *et al.*, 2017). As ferramentas de *Educational Data Mining* (EDM), termo que pode ser traduzido para o português como Mineração de Dados em Educação, tem contribuído significativamente para extração de conhecimento de dados educacionais.

A *International Educational Data Mining Society* define a EDM como sendo um campo emergente que se dedica ao estudo e desenvolvimento de métodos para apoiar gestão educacional e entender melhor os alunos e os ambientes em que eles aprendem (IEDMS, 2022). Outros autores acrescentam que a EDM se concentra no aprendizado de padrões latentes em várias situações educacionais e os estudos têm o objetivo final de melhorar o desempenho de aprendizagem do aluno, a gestão dos recursos educacionais e apoiar a tomada de decisão por parte dos gestores (Zhang *et al.*, 2021; Sultana *et al.*, 2017; Baker e Yacef, 2009).

O processo de extração do conhecimento associado à EDM inclui a seleção do conjunto de dados a ser analisado, o pré-processamento e transformação dos dados, a aplicação de técnicas de mineração de dados e a avaliação e interpretação dos resultados (Sultana *et al.*, 2017). As técnicas de mineração podem ser classificadas de acordo com o tipo de aprendizagem: supervisionada e não supervisionada.

A aprendizagem supervisionada cria um modelo através de dados de entrada rotulados e possibilitam fazer previsões sobre novos dados não rotulados, sejam eles numéricos ou categóricos (Nwanganga e Chapple, 2020; Müller e Guido, 2016). As técnicas mais comuns de aprendizagem supervisionada são as árvores de decisão, máquinas de vetores de suporte, regressão linear e regressão logística. Na área educacional, é possível encontrar uma série aplicações desses algoritmos, tal como para prever a evasão escolar (Sivakumar *et al.*, 2016; Silva *et al.*, 2020) ou para prever o desempenho dos estudantes (Hamsa *et al.*, 2016; Salal *et al.*, 2019).

Por outro lado, na aprendizagem não supervisionada, os algoritmos desenvolvem modelos sobre dados não rotulados, são técnicas utilizadas para descobrir padrões ocultos dentro de uma base de dados. Como não há uma saída conhecida, a aprendizagem é feita apenas mostrando os dados de entrada e pedindo para o algoritmo extrair conhecimento dos dados (Nwanganga e Chapple, 2020; Müller e Guido, 2016). Este tipo de aprendizagem é muito utilizado para redução de dimensionalidade (redução do número de variáveis independentes que são utilizados no modelo), para detecção de padrões ocultos nos dados, detecção de anomalias e agrupamento (*clustering*).

Na área educacional, aplicações desses algoritmos podem ser úteis para identificar e caracterizar padrões de desempenho (Varghese *et al.*, 2011; Singh *et al.*, 2016) e identificar similaridades entre estudantes, cursos ou instituições (Urbina-Nájera, 2017; Zheng, 2021; Huppel *et al.*, 2016; Scapinelli *et al.*, 2022).

Especificamente para identificar as similaridades entre cursos usando algoritmos de agrupamento, que é o objeto do presente estudo, foram encontrados na literatura três trabalhos relacionados. O primeiro é o estudo Leite *et al.* (2007), em que os autores utilizaram um método não-hierárquico de agrupamento com o objetivo de agrupar programas de pós-graduação a partir de variáveis medidas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e identificar quais variáveis são mais relevantes na composição do conceito dos programas. Ainda, os autores exploraram as principais diferenças entre os cursos dentro de cada agrupamento.

Utilizando indicadores de desempenho de cursos de graduação de uma universidade federal brasileira, Sales *et al.* (2013) testaram o potencial desses indicadores como classificadores dos cursos. Para isso, conduziram uma análise de agrupamentos utilizando um método hierárquico. A elevada heterogeneidade identificada entre os cursos de cada cluster mostrou que os índices sugeridos são significativos para a classificação dos cursos.

Por fim, Huppel *et al.* (2016) aplicaram um algoritmo hierárquico de agrupamento para identificar a similaridade entre cursos de graduação utilizando as variáveis que compõem o principal índice de alocação de recursos financeiros das IFES brasileiras, o índice aluno-equivalente de graduação. Além disso, os autores exploraram nesse estudo quais variáveis mais impactam no índice.

Metodologia

Descrição dos dados

Para o desenvolvimento deste trabalho, utilizou-se os dados das oito componentes que formam o CPC. Os dados se referem à avaliação conduzida no ano de 2017, ano em que os cursos de Matemática passaram pela avaliação. Os dados são disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Foram selecionados os cursos de Licenciatura em Matemática de instituições públicas federais na modalidade presencial. Outra restrição feita com o objetivo de aumentar a confiabilidade nos resultados da avaliação, foi utilizar dados de cursos em que o número de concluintes inscritos fosse maior ou igual a cinco.

Com base nesses critérios, o conjunto de dados utilizado é composto por 163 cursos de Licenciatura em Matemática, sendo 94 de Universidades e 69 de Institutos Federais. Todos esses cursos pertencem a um total de 78 instituições de Educação Superior.

Análise dos dados

Utilizou-se técnicas de análise exploratória e descritiva de dados para resumir os dados e compreender o comportamento das variáveis estudadas e para explorar as características dos cursos dentro dos agrupamentos. Essas características também foram observadas na intenção de entender melhor quais variáveis foram mais significativas para formar os grupos.

Para estudar as similaridades entre os cursos, aplicou-se a análise de agrupamento, também chamada de *clustering*, que se trata de uma coleção de técnicas em que o principal objetivo é separar objetos em grupos, de acordo com suas características, unindo objetos similares ou próximos e separando os diferentes ou distantes (Linden, 2009). A ideia é agrupar os dados sem a necessidade de conhecimento prévio sobre o assunto e que mantenham características comuns dentro dos grupos e distintas em grupos diferentes.

Os algoritmos de agrupamento podem ser classificados em dois grupos: os hierárquicos e não-hierárquicos. Os procedimentos hierárquicos de agrupamentos possuem uma estrutura do tipo árvore, ou dendrograma, que ilustra as distâncias entre os objetos a serem agrupados. Estes procedimentos não necessitam que o pesquisador defina previamente o número de clusters a serem formados. Diferentemente dos métodos hierárquicos, os não-hierárquicos não envolvem a construção de uma estrutura tipo árvore. Neste processo de clusterização, o resultado do agrupamento depende do número k de clusters que se deseja formar. Primeiro deve-se determinar, em geral de forma aleatória, os pontos de partida, também chamados de sementes de agrupamento. Em seguida, os objetos são associados a uma semente de agrupamento de acordo com a similaridade (Hair *et al.*, 2009).

De acordo com Hair *et al.* (2009), não há uma resposta definitiva para a questão de quais métodos devem ser utilizados - hierárquicos ou não-hierárquicos, a resposta depende do tipo de aplicação em questão. Os métodos hierárquicos têm a vantagem de ser mais simples por sua

estrutura de árvore, trazem resultados rápidos e um grande número de soluções de agrupamento, que vão desde os objetos individualmente até um único agrupamento que abrange todos os objetos. No entanto, eles podem sofrer grandes impactos com observações atípicas e podem se tornar inviáveis para tamanhos amostrais grandes. Por outro lado, os métodos não-hierárquicos são menos suscetíveis às observações anômalas e podem ser usados no agrupamento de um grande número de dados. Em contrapartida, dependem muito da escolha inicial das sementes de agrupamento e não são vantajosos quando se procura uma maior quantidade de soluções, pois geram apenas uma solução de exatamente k agrupamentos a cada análise realizada.

Para o agrupamento dos cursos, proposto no presente estudo, escolheu-se o método não-hierárquico k -médias. Ele foi escolhido por ser menos sensível a possíveis valores discrepantes e anomalias, além de ser mais prático para a interpretação dos resultados do que um esquema de árvore hierárquica, dado o número de objetos a serem agrupados neste estudo.

O k -médias é um dos métodos não-hierárquicos mais conhecidos. Consiste num método iterativo que tem como objetivo encontrar a melhor divisão de n dados em k agrupamentos, de forma que a distância total entre o centro de um grupo e seus respectivos dados, somada para todos os grupos, seja minimizada (Machado, 2011). Partindo de um conjunto de objetos $X = \{p_1, p_2, \dots, p_n\}$, sendo n o número de objetos a agrupar em k agrupamentos, o algoritmo inicia com a determinação dos k centros iniciais e então forma os grupos seguindo uma medida de similaridade que associa cada objeto ao centro mais próximo. A medida de similaridade empregada costuma ser a distância Euclidiana. Após o primeiro agrupamento, os centros são recalculados para que as distâncias entre ele e os objetos do grupo sejam as menores possíveis, em seguida, a divisão é refeita com base nos novos centros e os objetos são rearranjados. Este processo é realizado até que os objetos se estabilizem e não mudem mais de agrupamento, ou até que seja atingido um número pré-definido de iterações (Machado, 2011; Porto, 2017).

Os gráficos apresentados ao longo deste artigo, bem como a análise de agrupamento, foram realizados usando o programa SPSS versão 20.

Resultados

Análise exploratória dos dados

Na Tabela 2 são apresentadas as principais medidas descritivas das oito componentes do CPC para os 163 cursos de Matemática considerados neste estudo. Tal como mencionado na seção 2, as componentes são escalonadas de modo que os valores variem de zero (pior desempenho) a cinco (melhor desempenho).

Observando os valores do desvio-padrão, que é uma medida descritiva que capta o grau de afastamento dos dados em relação à média, é possível notar que as maiores discrepâncias de valores estão nas componentes relacionadas à Proporção de Doutores (ND), seguida pela Proporção de Mestres (NM). Já a nota referente ao Regime de Trabalho (NR) é a que possui a menor variação. Sobre essa componente, também é possível

observar que os valores da média e da mediana são bastante elevados, então compreende-se que grande parte dos cursos apresentaram nota do regime de trabalho próxima ou igual ao valor máximo. As demais componentes apresentam desvios-padrões com valores intermediários.

	Mínimo	Máximo	Mediana	Média	D. Padrão
NC	0,39	5,00	2,40	2,50	0,86
NIDD	0,00	4,61	2,48	2,56	0,66
NO	0,77	4,43	2,89	2,84	0,81
NF	1,04	4,68	2,85	2,89	0,70
NA	1,21	4,52	3,06	3,03	0,67
NM	0,00	5,00	4,40	4,11	0,90
ND	0,00	5,00	2,57	2,46	1,45
NR	3,73	5,00	5,00	4,96	0,17

Tabela 2 – Medidas descritivas para as variáveis.

Além da Tabela 2 com as medidas descritivas, é possível gerar histogramas que refletem o comportamento de cada uma das variáveis. A Figura 1 apresenta os histogramas das oito componentes do CPC para os 163 cursos considerados.

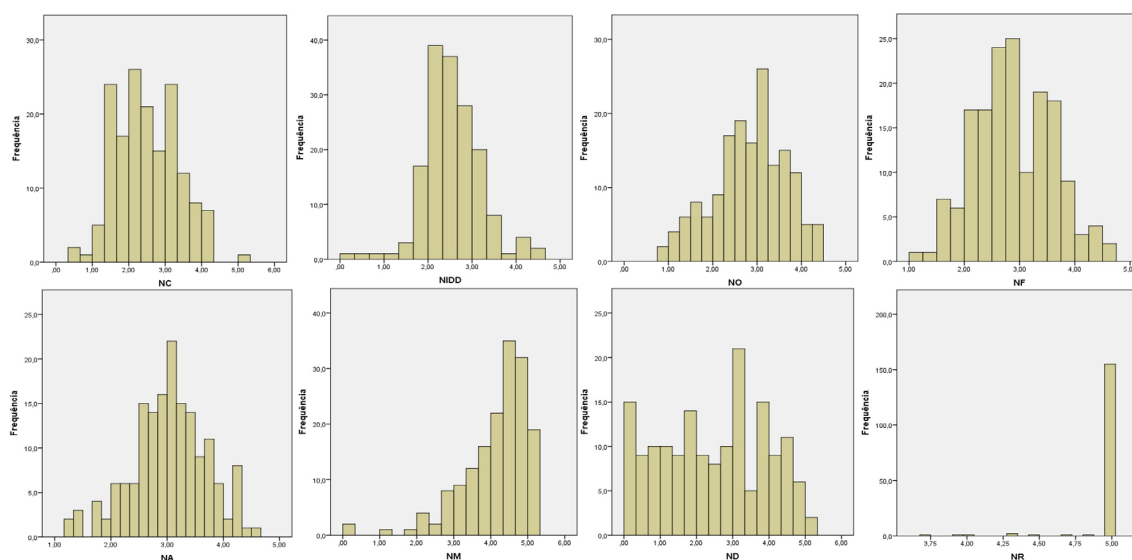


Figura 1 – Histogramas das variáveis NC, NIDD, NO, NF, NA, NM, ND e NR, respectivamente.

Analisando os histogramas das variáveis, pode-se observar que as variáveis das notas dos Concluintes no Enade (NC), da Diferença de Desempenhos Observado e Esperado (NIDD), da Organização Didático Pedagógica (NO), da Infraestrutura (NF) e das Oportunidades de Ampliação (NA) tendem a ter grande parte dos valores centralizados e não têm assimetrias muito elevadas. As informações se complementam ao relacionar com as medidas descritivas apresentadas pela Tabela 2.

A nota referente à Proporção de Mestres (NM) possui uma grande assimetria à esquerda, mostrando que a maioria dos cursos possui um grande número de docentes com titulação mínima de mestre, o que faz sentido já que estão sendo consideradas apenas as instituições públicas

federais, em que, na maioria das vezes, a titulação é um requisito para o cargo. Ainda, nota-se que a componente Nota de Regime de Trabalho (NR) teve apenas oito cursos com nota diferente de 5. Como foram considerados apenas os cursos de universidades ou institutos federais, era esperado que esta variável apresentasse pouca ou nenhuma variabilidade. Por isso, essa variável será excluída da etapa de análise de agrupamento, pois não agrega informações para a análise e pode ser um fator de confusão.

Análise de agrupamento e caracterização dos grupos

Na base de dados considerada, todas as variáveis (componentes do CPC) são disponibilizadas pelo Inep já transformadas em uma escala de 0 a 5, portanto, já estão na mesma unidade de medida. No entanto, as componentes têm variabilidades muito distintas, como pode-se observar pela medida do desvio-padrão mostrada na Tabela 2. Ao conduzir a análise de agrupamento, em geral, variáveis com maior dispersão têm maior impacto sobre o valor de similaridade final (Hair *et al.*, 2009). Portanto, antes de conduzir a análise de agrupamento, as componentes foram padronizadas de maneira que cada componente tenha média igual a zero e desvio-padrão igual a 1. Ou seja, para cada componente, os valores de desempenho de cada curso foram subtraídos da média da componente e, posteriormente, essa diferença foi dividida pelo desvio padrão da componente.

A partir das sete componentes padronizadas, realizou-se o agrupamento dos cursos pelo método não-hierárquico k -médias. Para decidir o número de clusters a serem formados, foram gerados alguns cenários utilizando métodos de agrupamentos hierárquicos, que possibilitam ver os agrupamentos em forma de árvore e então definir o melhor número de clusters a serem formados. Os resultados destes testes mostraram que faria mais sentido dividir os cursos em três agrupamentos. Assim, definido $k=3$ clusters finais, $n = 163$ cursos a serem agrupados, aplicou-se o método k -médias.

Como resultado, obteve-se que o Cluster 1 é formado por 49 cursos, o Cluster 2 por 66 cursos e o Cluster 3 por 48 cursos. A distribuição detalhada dos cursos nos três clusters é apresentada no Anexo 1.

Um gráfico interessante de ser analisado é o que mostra os valores do centro final de cada uma das sete componentes avaliadas em cada um dos clusters. A Figura 2 apresenta este gráfico, em que cada cor representa uma componente diferente. Deve-se lembrar que os valores das componentes estão padronizadas, então a notação "Zscore(NC)" corresponde ao valor da Nota dos Concluintes no Enade (NC) na padronização descrita anteriormente, da mesma forma acontece para as outras componentes. Devido à padronização, a média é zero e, portanto, os valores próximos de zero estão perto da média, os valores positivos estão acima da média e os negativos abaixo.

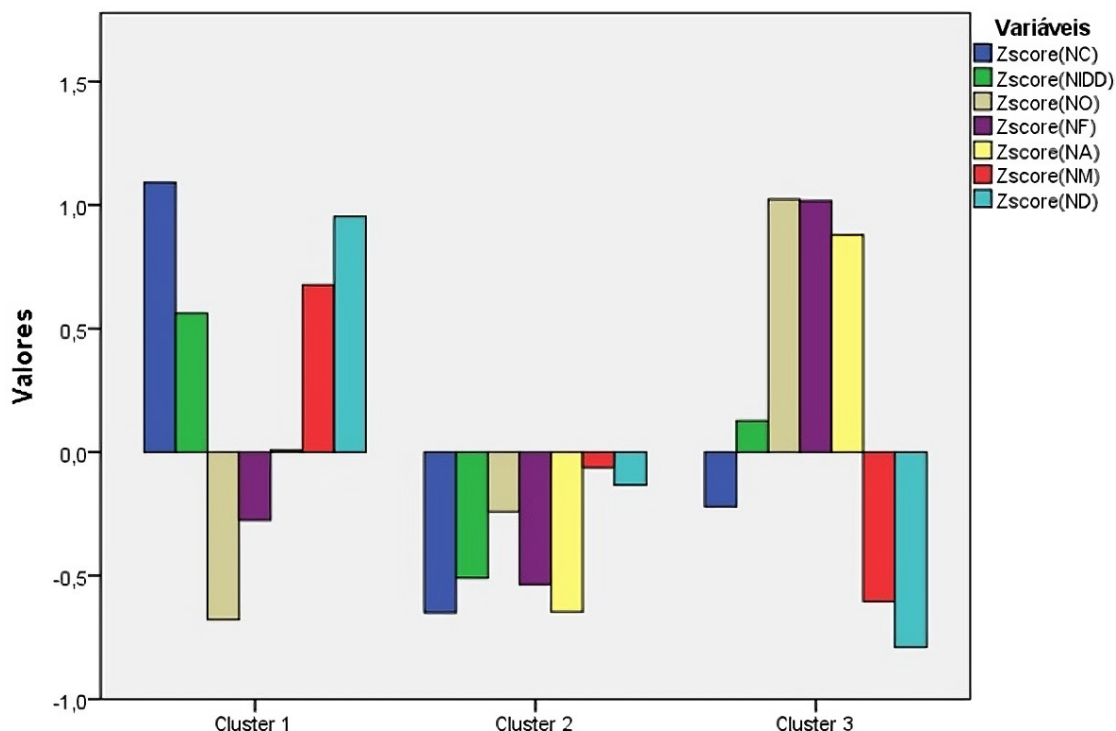


Figura 2 - Componentes dos centros dos clusters finais.

Algumas observações interessantes podem ser feitas ao analisar esse gráfico. Como os valores dos centros dos clusters correspondem às médias dos valores dos objetos de cada agrupamento, eles dão uma noção muito boa das características dos cursos de cada cluster.

O Cluster 1 é o que possui os cursos com maiores notas NC e NIDD. Ambas as componentes estão relacionadas aos alunos dos cursos de Matemática, seja pelo conhecimento dos alunos concluintes dos cursos, mensurado pelo exame Enade, seja pelo valor agregado pelo processo formativo oferecido pelo curso, mensurado pelo NIDD. Este cluster também é o que possui as maiores proporções de professores com titulação de Mestre e Doutor (NM e ND), observa-se que para essas duas componentes, é o único cluster que possui as médias das componentes maiores que as médias gerais. Em contrapartida, as médias referentes às componentes que têm a ver com a percepção do aluno sobre a organização didático-pedagógica, as oportunidades de ampliação da formação e a infraestrutura e instalações físicas (NO, NA e NF) estão abaixo da média geral, ou no máximo muito próximas à média.

É possível notar que o Cluster 2 é formado por cursos que possuem todas as notas observadas abaixo da média. As proporções de Mestres (NM) e de Doutores (ND) estão próximas à média geral, mas ainda se encontram abaixo dela. As notas dos Concluintes no Enade (NC) e a Diferença dos Desempenhos Observado e Esperado (NIDD) estão muito abaixo da média geral, assim como as notas advindas das opiniões dos alunos sobre a Oportunidade de Ampliação da Formação Acadêmica (NA) e Infraestrutura e Instalações Físicas (NF).

O Cluster 3 por sua vez, quase que ao contrário do Cluster 1, é formado por cursos que apresentam médias altas nas componentes que se referem à percepção do estudante sobre o curso (NO, NA e NF); médias abaixo ou bem próximas das médias gerais nas componentes NC e NIDD que se referem ao conhecimento adquirido pelos alunos concluintes; e as piores médias nas componentes proporções de Mestres e Doutores (NM e ND) em relação aos outros dois agrupamentos.

A Figura 3 mostra os diagramas em caixa de cada uma das sete componentes consideradas em relação a cada um dos três clusters. Quando colocados lado a lado, os diagramas facilitam bastante a comparação dos valores típicos das componentes por cluster, complementando assim a informação comparativa das médias, apresentada na Figura 2.

Na variável Nota dos Concluintes no Enade (NC), pode-se confirmar que o cluster 1 agrupou cursos que tinham notas significativamente melhores que os demais cursos, inclusive pertence ao cluster 1 o curso da UFABC de Santo André que aparece como um *outlier* (valor atípico na distribuição) por ser o único com a nota máxima 5; é possível observar que a mediana é aproximadamente 3,4. Por outro lado, os cursos dos clusters 2 e 3 tiveram desempenhos razoáveis ou baixos no Enade em comparação com os cursos do cluster 1.

Na variável Nota do Indicador de Diferença entre os Desempenhos Observado e Esperado (NIDD), o comportamento é similar à componente anterior, com o cluster 1 sendo o que obtém os cursos com as notas mais altas, mas nesse caso a mediana está em torno de 3. Os clusters 2 e 3 também possuem cursos com desempenhos razoáveis ou baixos, em especial o cluster 2 que contém os cursos da UNB de Brasília, UFT de Palmas e UFBA de Salvador com as piores notas. Ainda, vale observar que a amplitude mostrada no diagrama do cluster 2 é consideravelmente menor que a amplitude observada nos demais clusters, sugerindo que, excluindo os valores atípicos, os cursos do cluster 2 apresentaram valores de NIDD mais homogêneos do que os cursos dos demais agrupamentos.

Quando se trata da componente Nota de Organização Didático-Pedagógica (NO) o cenário muda. De modo geral, o cluster 1 contém os cursos que receberam as piores avaliações dos alunos, o cluster 2 teve avaliações intermediárias e o cluster 3 concentrou os que tiveram as melhores notas nesse quesito. Nesta componente, a nota mínima observada no cluster 3 é superior ao valor do quartil 3 do cluster 1, portanto, ao menos 75% dos cursos do cluster 1 tiveram nota inferior à nota mínima observada nos cursos do cluster 3.

O comportamento dos clusters em relação à Nota Referente à Infraestrutura e Instalações Físicas (NF) é muito similar à componente NO, no entanto, no cluster 2 aparece um *outlier* com a pior nota entre todos os cursos, sendo da UFPB de João Pessoa.

Em relação à Nota Referente à Oportunidade de Ampliação da Formação Acadêmica (NA), os cursos com maiores notas também se concentraram no cluster 3, porém, diferentemente das componentes anteriores, o cluster 1 teve avaliações razoáveis e o cluster 2 foi o que concentrou os cursos com as piores avaliações.

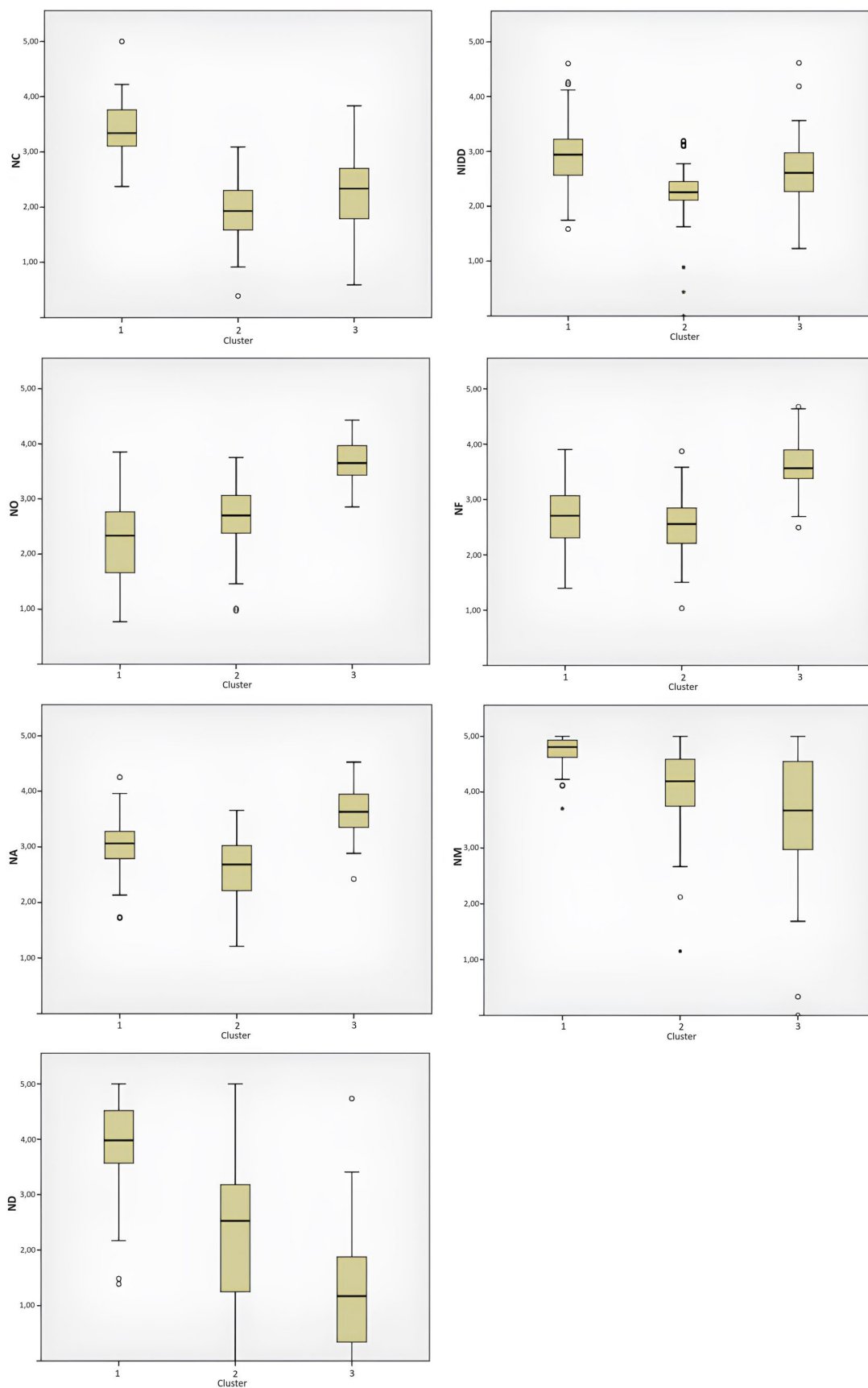


Figura 3 – Diagramas em caixa das componentes NC, NIDD, NO, NF, NA, NM e ND, respectivamente.

Observando os diagramas da componente Nota de Proporção de Mestres (NM), nota-se que o cluster 1 agrupou quase que apenas cursos com nota maior que 4, com exceção do curso da UFTM de Rondonópolis que tem nota nesta componente em torno de 3,6. Os clusters 2 e 3 também contém cursos com nota máxima, porém estão inclusos cursos superiores com notas intermediárias e baixas, inclusive o cluster 3 contém o curso do IFPI de Uruçuí que teve nota igual a zero. Esta é a componente que apresentou as maiores discrepâncias na comparação entre os clusters. Pode-se observar que o valor do quartil 1 do cluster 1 (aproximadamente 4,5) é superior aos valores do quartil 3 dos demais clusters, indicando que 75% dos cursos do primeiro cluster tiveram notas superiores a 75% dos cursos dos cluster 2 e 3.

Sobre a Nota de Proporção de Doutores (ND), o cluster 2 chama atenção, pois ele engloba toda a extensão de notas, o que significa que neste agrupamento é possível encontrar cursos em toda a faixa de valores de 0 a 5, especificamente neste quesito os cursos são heterogêneos. Enquanto isso, no cluster 3 observa-se que os 75% inferiores tiveram nota inferior a 2 e no cluster 1 os 50% superiores tiveram nota superior a 4. A partir dessas constatações, observa-se que uma característica muito forte do cluster 1 é ter um grande número de professores com formação de mestre ou doutor, já o cluster 3 é quase que o oposto com relação à formação de doutor.

Em resumo, a partir dessa análise exploratória foi possível observar que os cursos do cluster 1 possuem notas altas nas componentes relacionadas ao desempenho do aluno (NC e NIDD), enquanto que as notas que dependem da percepção dos alunos sobre o curso e a instituição (NO, NF e NA) costumam ser as mais baixas comparadas aos outros cursos. Quase que de uma forma antagônica, os cursos do cluster 3 possuem as notas mais baixas quando se trata do desempenho dos alunos e da proporção de mestres e doutores, e receberam as melhores avaliações por parte dos alunos em relação à estrutura geral do curso. No meio disso, o cluster 2 reuniu os cursos com notas intermediárias em praticamente todas as componentes avaliadas, sem apresentar relação direta com as dimensões observadas.

Tirando o foco das componentes, uma comparação que vale a pena fazer é sobre os tipos de organização acadêmica – Universidade ou Instituto Federal – dos cursos de cada cluster. A Figura 4 mostra o número de instituições de cada tipo de organização acadêmica dentro dos grupos.

Nota-se que o cluster 1 que é majoritariamente formado por cursos de universidades e inclui poucos institutos federais. Já o cluster 3 que é formado predominantemente por institutos federais e o cluster 2 é um pouco mais heterogêneo neste aspecto.

Com todas essas observações, não poderia deixar de ser feita a comparação dos clusters com o Conceito Preliminar de Curso que cada curso recebeu. A Figura 5 mostra o gráfico da relação dos conceitos dos cursos de cada cluster.

Observa-se na Figura 5 que o cluster 1 é o que possui os cursos com os maiores CPCs, nele estão todos os cursos com conceito 5 e a maioria dos cursos com conceito 4. Este resultado poderia ser esperado tendo em vistas

as análises anteriores, pois este agrupamento apresenta notas boas em muitas variáveis, principalmente nas variáveis que possuem um peso alto no CPC. O cluster 3, por sua vez, possui um pouco mais de equilíbrio com cursos conceito 3 e 4, e alguns conceito 2. O cluster 2 é o que possui o maior número de cursos que tiveram conceito 3 e possui poucos cursos com conceito 2 e 4.

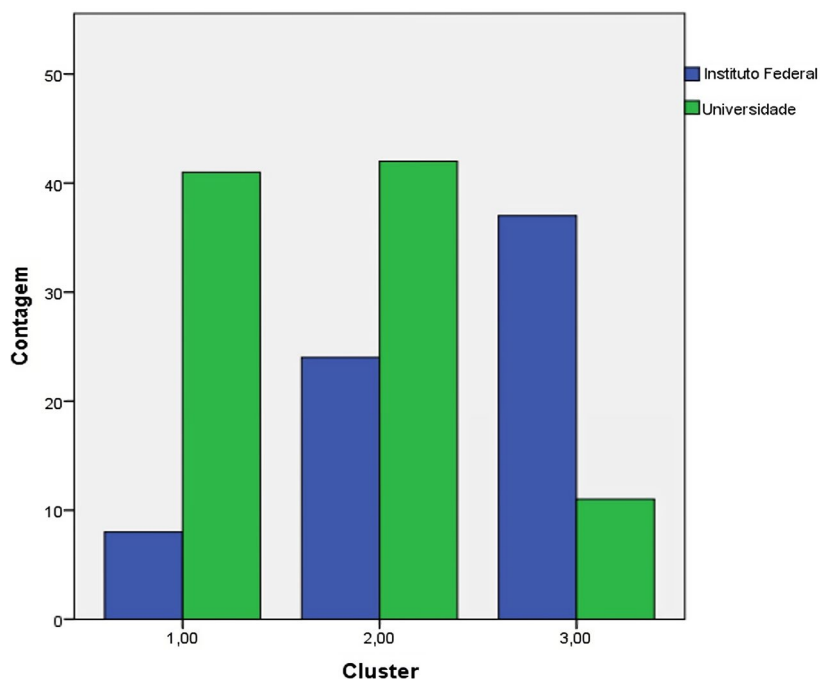


Figura 4 – Gráfico dos clusters de acordo com o tipo de organização acadêmica.

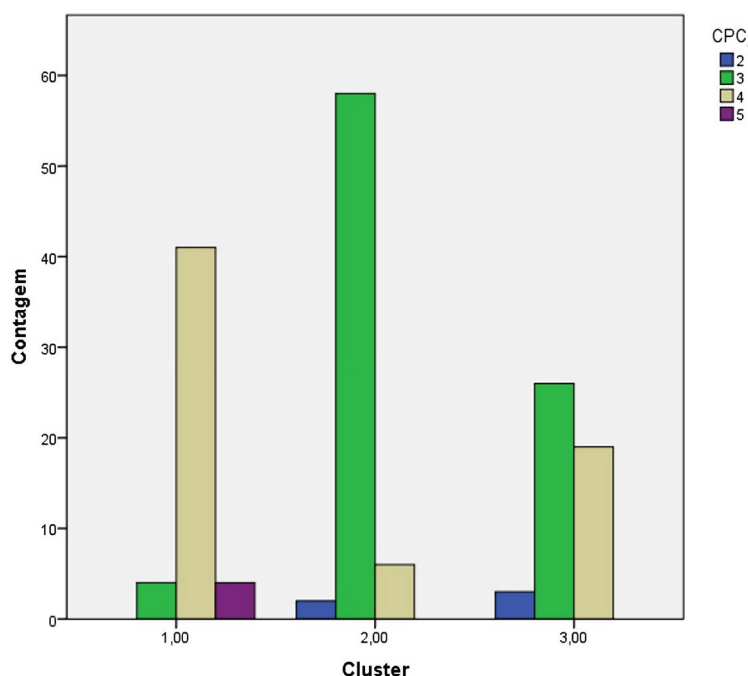


Figura 5 – CPC dos cursos por cluster.

É interessante observar que nenhum curso de Matemática considerado teve conceito 0 ou 1, apenas cinco tiveram conceito 2, a maioria – oitenta e

oito – tiveram conceito 3, sessenta e seis tiveram conceito 4 e quatro tiveram conceito 5.

Apesar dessas comparações, é importante lembrar que a fórmula do cálculo do CPC inclui uma variável que foi descartada neste trabalho por apresentar variabilidade muito baixa. Além disso, cada componente possui um peso diferente no cálculo do CPC, enquanto que a clusterização deste trabalho utilizou inclusive ferramentas de padronização na tentativa de que as variáveis tivessem o mesmo peso.

As técnicas de agrupamento aplicadas possibilitaram agregar dentro de um mesmo grupo cursos que apresentam características semelhantes. A identificação de cursos semelhantes pode apoiar gestores na identificação de oportunidades de melhoria por meio da observação das boas práticas entre os cursos que tem características similares e que tiveram boa avaliação no CPC.

Conclusões

A literatura recente registra a importância da avaliação de desempenho como um importante recurso à gestão universitária e também ao suporte à decisão no contexto de formulação nas políticas públicas (Hoffmann *et al.*, 2014). Portanto, gerar informações relevantes ao monitoramento de desempenho dos cursos de graduação em IES pode promover a melhoria da qualidade destas instituições brasileiras.

O presente trabalho abordou a avaliação dos cursos de Licenciatura em Matemática de acordo com as componentes do Conceito Preliminar de Curso. A partir dessas componentes, foi aplicado o método de clusterização *k*-médias com a finalidade de formar agrupamentos de cursos com elevada homogeneidade interna (dentro dos grupos) e heterogeneidade externa (entre os grupos). Como resultado obteve-se 3 grupos de cursos, a partir disso, foi feita uma análise exploratória a fim de caracterizar cada cluster de acordo com o perfil dos cursos.

O cluster 1 ficou com 49 cursos, o cluster 2 com 66 cursos e o cluster 3 com 48 cursos. A análise dos clusters mostrou que o cluster 1 agrupou os cursos com maiores notas nas componentes das dimensões Desempenho dos Estudantes, Valor Agregado pelo Processo Formativo Oferecido pelo Curso e Corpo Docente. O cluster 3 agrupou os cursos com maiores notas na dimensão Percepção Discente sobre as Condições do Processo Formativo. O cluster 2, por sua vez, agrupou os cursos com notas intermediárias na maioria das componentes. Observando o Conceito Preliminar de Curso de cada curso, nota-se que, de modo geral, o cluster 1 agrupou os cursos com conceito mais elevado.

A identificação dos agrupamentos, ou seja, dos cursos que possuem características similares é crucial para promover a prática de *benchmarking*. Ao analisar cursos com características similares, as instituições de ensino superior podem identificar pontos fortes e áreas de melhoria em seus próprios programas acadêmicos. Essa abordagem colaborativa não apenas incentiva a inovação e a excelência, mas também facilita a adaptação de currículos, métodos de ensino e recursos para melhor atender às necessidades dos estudantes. Como resultado, o *benchmarking* pode

desempenhar um papel significativo na melhoria contínua da qualidade dos cursos de graduação, contribuindo para uma experiência educacional mais eficaz e enriquecedora para os alunos.

Como sugestão para trabalhos futuros, recomenda-se inserir na análise mais variáveis explicativas que reflitam as características dos cursos, por exemplo, número de docentes, número de alunos matriculados e diplomados, índices de evasão e reprovação e turno em que o curso é oferecido, de maneira a aprimorar a identificação de cursos com características similares. Assim, quanto mais similares forem os cursos dentro de cada cluster, mais efetiva poderá ser a observação entre pares em busca de exemplos de boas práticas.

Referências

Baker, R. S. J. D., e Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3–17. <https://doi.org/10.5281/zenodo.3554657>.

Daultani, Y., Dwivedi, A., e Pratap, S. (2021). Benchmarking higher education institutes using data envelopment analysis: capturing perceptions of prospective engineering students. *OPSEARCH*, 58, 773–789. <https://doi.org/10.1007/s12597-020-00501-5>.

Fontanive, N. S. (2013). A divulgação dos resultados das avaliações dos sistemas escolares: limitações e perspectivas. *Ensaio: Avaliação e Políticas Públicas Educacionais*, Rio de Janeiro, 21(78), 83-100. <https://doi.org/10.1590/S0104-40362013005000005>.

Gourishankar, V., e Sai Lokachari, P. (2012). Benchmarking educational development efficiencies of the Indian states: A DEA approach. *International Journal of Educational Management*. 26(1), 99-130. <https://doi.org/10.1108/09513541211194400>.

Griboski, C. M., e Fernandes, I. R. (2016). Avaliação da Educação Superior: Como avançar sem desqualificar. *Observatório Universitário*, 1, 1-31. Recuperado de: http://www.observatoriouniversitario.org.br/documentos_de_trabalho/avaliacao-da-educacao-superior-como-avancar-sem-desqualificar.pdf

Hadioui, A., El Faddouli, N., Touimi, Y. B., e Bennani, S. (2017). Machine Learning Based on Big Data Extraction of Massive Educational Knowledge. *International Journal on Emerging Technologies in Learning*, 12(11), 151-167. <https://doi.org/10.3991/ijet.v12i11.7460>.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., e Atham, R. L. (2009). *Análise Multivariada de Dados*. 6. ed. Porto Alegre: Bookman.

Hamsa, H., Indiradevi, S., e Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. *Procedia Technology*, 25, 326-332. <https://doi.org/10.1016/j.protcy.2016.08.114>.

Hoffmann, C., Zanini, R. R., Corrêa, Â. C., Siluk, J. C., Schuch, Jr. V. F., e Ávila, L. V. (2014). O desempenho das universidades brasileiras na perspectiva do Índice Geral de Cursos (IGC). *Educação e Pesquisa*, 40(3),

651-665. <http://dx.doi.org/10.1590/s1517-97022014041491>.

Huppés, J. C., Souza, A. M., e Ansuaj, A. P. (2016). Classificação da Similaridade dos Cursos de Graduação da UFSM em Relação ao Índice do Aluno Equivalente – Uma Ferramenta de Gestão. *Revista Espacios*, 37(13), 22-30. Recuperado de: <https://www.revistaespacios.com/a16v37n13/16371322.html>.

IEDMS (2022). *International Educational Data Mining Society*. Recuperado de: <https://educationaldatamining.org/>.

Ikuta, C. Y. S. (2016). Sobre o Conceito Preliminar de Curso: concepção, aplicação e mudanças metodológicas. *Estudos em Avaliação Educacional*, 27(66), 938-969. <https://doi.org/10.18222/eae.v27i66.4039>.

Lacerda, L. V., e Ferri, C. (2017). Conceito Preliminar de Curso: conceito único para uma realidade educacional múltipla. *Estudos em Avaliação Educacional*, 28 (69), 748-772. <https://doi.org/10.18222/eae.v0ix.3970>.

Leite, M. F. B., Viana, A. B. N., Pereira, G. G., e Mantovani, D. M. N. (2007). Análise dos Programas de Pós-graduação em Administração Reconhecidos pela Capes: uma Abordagem Multivariada por Análise de Cluster. *FACEF Pesquisa*, 10(1), 20-34. Recuperado de: <http://periodicos.unifacef.com.br/facefpesquisa/article/view/87>.

Machado, R. L. Desenvolvimento de um Algoritmo Imunológico para Agrupamento de Dados. 2011. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade de Caxias do Sul, Caxias do Sul, 2011. Recuperado de: <https://repositorio.ucs.br/handle/11338/1486>.

Ministério da Educação Brasil (2004). *Lei 10.861, de 14 de abril de 2004*. Institui o Sistema Nacional de Avaliação da Educação Superior – Sinaes e dá outras providências. Diário Oficial da União. Brasília, DF, 15 abr. 2004. Recuperado de: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.861.htm.

Ministério da Educação Brasil (2003). *Sistema Nacional de Avaliação da Educação Superior (Sinaes): bases para uma nova proposta de avaliação da educação superior*. Brasília, DF: MEC, 2003. Recuperado de: <http://portal.mec.gov.br/arquivos/pdf/sinaes.pdf>.

Ministério da Educação Brasil (2020). *Nota técnica n. 58, de 04 de dezembro de 2020*. Brasília, DF: Inep, 2020. Recuperado de: https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2019/NOTA_TECNICA_N_58-2020_CGCQES-DAES_Metodologia_de_calculo_do_CPC_2019.pdf.

Ministério da Educação Brasil (2004). *Portaria n. 2.051, de 9 de julho de 2004*. Regulamenta os procedimentos de avaliação do Sistema Nacional de Avaliação da Educação Superior (Sinaes), instituído na Lei no 10.861, de 14 de abril de 2004. Recuperado de: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=533-portaria-2051&category_slug=documentos-pdf&Itemid=30192.

Müller, A. C., e Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. Primeira edição. Sebastopol, CA: O'Reilly

Media, Inc.

Nugroho, B. H., e Jaqin, C. (2021). Implementation of Benchmarking Method for Higher Education Institution: A Literature Review. *Indonesian Journal of Industrial Engineering & Management*, 2(2), 81-93. <https://doi.org/10.22441/ijiem.v2i2.10797>.

Nwanganga, F., e Chapple, M. (2020). *Practical machine learning in R*. 1st ed. Indianapolis: John Wiley and Sons.

Paek, P. L. (2012). Using learning trajectories in large-scale mathematics assessments. *Proceedings of the 12th International Congress on Mathematical Education*, 6711-6720. Seoul, Korea. Recuperado de: <https://unglueit-files.s3.amazonaws.com/ebf/dcd3d69b6b6f4715a095c1863d7e0391.pdf>.

Piran, F. S., Lacerda, D. P., e Camargo, L. F. R. (2018). *Análise e Gestão da Eficiência*. 1 ed. Rio de Janeiro: Elsevier.

Porto, J. P. S. Geometria do Modelo Estatístico das Distribuições Normais Multivariadas. 2017. Tese (Doutorado em Matemática Aplicada) - Universidade Estadual de Campinas, Campinas, 2017. Recuperado de: <https://www.ime.unicamp.br/pos-graduacao/geometria-modelo-estatistico-distribuicoes-normais-multivariadas>.

Ristoff, D., e Giolo, J. (2006). O Sinaes como Sistema. *Revista Brasileira de Pós-Graduação*, 3(6), 193-213. <https://doi.org/10.21713/2358-2332.2006.v3.106>.

Salal, Y., Abdullaev, S., e Kumar, M. (2019). Educational data mining: Student performance prediction in academic. *International journal of Engineering and Advanced Tech*, 8(4C) 54-59. <https://doi.org/10.20368/1971-8829/1135578>.

Sales Jr, J. S., Penholato, J. P., Erler, I. da S., e Carneiro, T. C. J. (2013). Proposição de Indicadores para o Corpo Docente e Análise de Agrupamentos Aplicada aos Cursos de Graduação da UFES. *Revista Gestão Universitária na América Latina*, 6(2), 106-125. <https://doi.org/10.5007/1983-4535.2013v6n2p106>.

Scapinelli, H., Flach, L., e Mattos, L. K. (2022). Indicadores de desempenho na gestão universitária: um estudo com análise de componentes principais e análise de clusters. *Revista da Educação Superior do Senac-RS*, 15(1), 1-12. Recuperado de: https://www.researchgate.net/publication/367691484_Indicadores_de_desempenho_na_gestao_universitaria.

Shimizu, Y. (2011). Building bridges between large-scale external assessment and mathematics classrooms: A Japanese perspective. Em B. Kaur & K. Y. Wong (Eds.), *Assessment in the mathematics classroom: 2011 Association of Mathematics Educators Yearbook* (pp. 217-235), Singapore: World Scientific Publishing. https://doi.org/10.1142/9789814360999_0010.

Schwartzman, S. (2008). O Conceito Preliminar e as boas práticas de avaliação do ensino superior. *Estudos - Revista da Associação Brasileira de Mantenedoras de Ensino Superior*, 26(38), 9-32. Recuperado de: <https://www.schwartzman.org.br/simon/prelim.pdf>.

Silva, A. F., Almeida, A. T. C., e Ramalho, H. M. B. (2020). Predição do risco de reprovação no ensino superior usando algoritmos de machine learning. *Teoria e Prática em Administração*, 10(2), 58-80. <https://doi.org/10.21714/2238-104X2020v10i2-51124>.

Singh, A., Sabitha A. S., e Bansal, A. (2016). Student performance analysis using clustering algorithm. 6th International Conference - Cloud System and Big Data Engineering (Confluence). *Anais [...]*. Noida, India, p. 294-299. <https://doi.org/10.1109/CONFLUENCE.2016.7508131>.

Sivakumar, S., Venkataraman, S., e Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, 9(4), 1-5. <https://doi.org/10.17485/ijst/2016/v9i4/87032>.

Sultana, S., Khan, S., e Abbas, M. A. (2017). Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. *International Journal of Electrical Engineering Education*, 54(2), 105-118. <https://doi.org/10.1177/0020720916688484>.

Suurtamm, C., Thompson, D. R., Kim, R. Y, Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., e Vos, P. (2016). *Assessment in Mathematics Education: Large-Scale Assessment and Classroom Assessment*. Berlin: Springer International Publishing.

Swan, M., e Burkhardt, H. (2012). A designer speaks: Designing assessment of performance in mathematics. *Educational Designer: Journal of the International Society for Design and Development in Education*, 2(5), 1-41. Recuperado de: <https://www.educationaldesigner.org/ed/volume2/issue5/article19/>.

Urbina-Nájera, A.B., Calleja, J., e Medina, M.A. (2017). Associating students and teachers for tutoring in higher education using clustering and data mining. *Computer Applications in Engineering Education*, 25, 823-832. <https://doi.org/10.1002/cae.21839>.

Varghese, B., Unnikrishnan, A., e Jacob, K. P. (2011). Clustering Student Data to Characterize Performance Patterns. *International Journal of Advanced Computer Science and Applications. Special Issue on Artificial Intelligence*, p. 138-140. <https://doi.org/10.14569/SpecialIssue.2011.010322>.

Verhine, R. E., e Dantas, L. M. (2009). A avaliação do desempenho de alunos de educação superior: uma análise a partir da experiência do ENADE. Em Dazzani, M. V., e Lordêlo, J. A. C. *Avaliação Educacional: destando e reatando nós* (pp. 173-199). Salvador: SciELO Books.

Zhai, X. (2021). Practices and theories: How can machine learning assist in innovative assessment practices in science education. *Journal of Science Education and Technology*, 30(2), 1-11. <https://doi.org/10.1007/s10956-021-09901-8>.

Zhai, X., Haudek, K. C., Shi, L., Nehm, R., e Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching*, 57(9), 1430-

1459. <https://doi.org/10.1002/tea.21658>.

Zainko, M. A. S. (2010). Avaliação da Educação Superior no Brasil: Processo de Construção Histórica. *Revista da Avaliação da Educação Superior*, 13(3), 827-831. <https://doi.org/10.1590/S1414-40772008000300012>.

Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H, e Shang, X. (2021). Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis. *Frontiers in Psychology*, 12, 1-19. <https://doi.org/10.3389/fpsyg.2021.698490>.

Zheng, W. (2022). Cluster Analysis Algorithm in the Analysis of College Students' Mental Health Education. *Advances in Computing and Engineering for Bionics and Medical Applications*, 1-9. <https://doi.org/10.1155/2022/6394707>.

Anexo 1 - Distribuição dos cursos nos três clusters.

Cluster 1:	
FURG Rio Grande (1035)	UFPR Curitiba (49478)
IF SUL MINAS Inconfidentes (1117203)	UFRGS Porto Alegre (13741)
IFBA Salvador (1122743)	UFRJ Rio de Janeiro (14328)
IFES Vitória (108730)	UFRN Natal (312333)
IFMG Formiga (116820)	UFRPE Recife (14505)
IFRJ Nilópolis (102430)	UFRRJ Nova Iguaçu (96158)
IFSP Bragança Paulista (1128375)	UFRRJ Seropédica (12918)
IFSP Guarulhos (122122)	UFSC Florianópolis (14233)
IFSP São Paulo (113490)	UFSCAR São Carlos (624)
UFABC Santo André (1102402)	UFSCAR São Carlos (637)
UFBA Salvador (13276)	UFSCAR Sorocaba (115086)
UFES Vitória (12806)	UFSJ São João del Rei (48946)
UFF Niterói (12702)	UFSM Santa Maria (13873)
UFG Goiânia (1131697)	UFSM Santa Maria (41069)
UFG Goiânia (1304557)	UFTM Uberaba (114556)
UFJF Juiz de Fora (1166038)	UFU Uberlândia (1428)
UFLA Lavras (101556)	UFV Florestal (122358)
UFMG Belo Horizonte (12969)	UFV Viçosa (696)
UFMS Campo Grande (15833)	UNB Brasília (44372)
UFMS Três Lagoas (15865)	UNIFAL-MG Alfenas (96955)
UFMT Rondonópolis (34)	UNIFEI Itajubá (120389)
UFPE Recife (13619)	UNIRIO Rio de Janeiro (1101770)
UFPEL Capão do Leão (15000)	UTFPR Curitiba (1128169)
UFPEL Pelotas (122746)	UTFPR Pato Branco (14542)
UFPR Curitiba (12553)	
Cluster 2:	
IF Catarinense Camboriú (1103025)	UFG Catalão (14121)
IF Catarinense Concórdia (1103031)	UFGD Dourados (15868)
IF Catarinense Rio do Sul (1102795)	UFMA São Luís (11439)
IFAC Cruzeiro do Sul (5000487)	UFMS Aquidauana (18382)
IFAL Maceió (1103556)	UFMT Cuiabá (26)
IFAM Manaus (5000493)	UFMT Pontal do Araguaia (18354)
IFBA Barreiras (116730)	UFOPA Santarém (12039)
IFBA Valença (123519)	UFPA Abaetetuba (114853)
IFCE Canindé (1160979)	UFPA Belém (11999)
IFCE Crateús (5000138)	UFPA Bragança (12033)
IFCE Fortaleza (63383)	UFPA Breves (12044)
IFG Goiânia (1103461)	UFPA Cametá (12074)
IFMA São Luís (21443)	UFPA Castanhal (12035)
IFNMG Januária (99503)	UFPB João Pessoa (13402)
IFNMG Salinas (1102641)	UFPB Rio Tinto (99045)
IFPA Belém (48289)	UFPE Caruaru (118100)
IFPI Angical do Piauí (1103910)	UFPI Parnaíba (102576)
IFPI Teresina (63062)	UFPI Picos (102588)
IFRJ Volta Redonda (121013)	UFPI Teresina (300520)
IFRN Mossoró (123787)	UFRB Amargosa (100412)
IFRN Natal (1123270)	UFRN Caicó (12354)
IFRS Ibirubá (1156565)	UFS Itabaiana (95055)
IFS Aracaju (96913)	UFS São Cristóvão (297)
IFTO Paraíso do Tocantins (1106721)	UFT Araguaína (1105218)
UFAL Arapiraca (102152)	UFT Arraias (17154)
UFAL Maceió (107520)	UFT Palmas (1121464)
UFAM Manaus (379)	UFVJM Teófilo Otoni (100916)

UFBA Salvador (117052)	UNB Brasília (161)
UFC Fortaleza (38246)	UNIFAP Macapá (17196)
UFCG Cajazeiras (1152880)	UNIFESSPA Marabá (12037)
UFCG Campina Grande (13448)	UNIPAMPA Bagé (104282)
UFCG Cuité (99806)	UNIR Ji-Paraná (16010)
UFF Santo Antônio de Pádua (12713)	UNIR Porto Velho (16009)
Cluster 3:	
IF Catarinense Sombrio (1103032)	IFRJ Paracambi (1153801)
IFBA Camaçari (1161404)	IFRN Santa Cruz (1184450)
IFCE Cedro (85320)	IFRO Cacoal (1292902)
IFCE Juazeiro do Norte (84374)	IFRO Vilhena (1182765)
IFES Cachoeiro Itapemirim (1103657)	IFRR Boa Vista (1152312)
IF Farroupilha Alegrete (1128936)	IFRS Bento Gonçalves (116144)
IF Farroupilha J. Castilhos (1103185)	IFRS Caxias do Sul (1126144)
IF Farroupilha Santa Rosa (1128938)	IFSEMG Rio Pomba (113589)
IF Farroupilha São Borja (1168573)	IFSP Araraquara (1129073)
IF Fluminense C. Goytacazes (52442)	IFSP Birigui (1128360)
IFMA Buriticupu (1264336)	IFSP Campos do Jordão (1181042)
IFMA Caxias (1327336)	IFSP Caraguatatuba (1128378)
IFMA Codó (1103914)	IFTO Palmas (123494)
IFMA Zé Doca (1160371)	UFAC Rio Branco (11540)
IFMG São João Evangelista (1142320)	UFES São Mateus (116882)
IFMT Campo Novo do Parecis (114388)	UFG Jataí (18954)
IFMT Juína (1160424)	UFMS Corumbá (15862)
IFPB Cajazeiras (1128096)	UFMS Paranaíba (52139)
IFPB Campina Grande (1128030)	UFMS Ponta Porã (121792)
IFPE Pesqueira (100690)	UFMT Sinop (100775)
IFPI Corrente (1103911)	UFU Ituiutaba (102944)
IFPI Floriano (83929)	UNIPAMPA Itaqui (5000912)
IFPI Piripiri (1103813)	UTFPR Cornélio Procópio (1127676)
IFPI Uruçuí (1103912)	UTFPR Toledo (1152637)

Nota: Para cada curso é apresentado a sigla do nome da universidade, cidade em que está localizado e código *e-mec* do curso.