



Programa Interdisciplinar de Pós-Graduação em
Computação Aplicada
Mestrado Acadêmico

Jean Carlos Araújo Costa

**Vis-Scholar – Uma Metodologia de Visualização e Análise de
Dados na Educação**

São Leopoldo, 2016

Jean Carlos Araújo Costa

Vis-Scholar – Uma Metodologia de Visualização e Análise de Dados na Educação

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa Interdisciplinar de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS

Orientador:

Dr. Luiz Paulo Luna de Oliveira

Co-orientadora:

Dra. Marta Becker Villamil

São Leopoldo

2016

C837v

Costa, Jean Carlos Araújo.

Vis-Scholar : uma metodologia de visualização e análise de dados na educação / por Jean Carlos A. Costa. – 2016.

56 f.: il. ; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, RS, 2016.

“Orientação: Dr. Luiz Paulo Luna de Oliveira.”

1. Visualização de dados. 2. Correlação de dados educacionais. 3. Análise de componentes principais. 4. Aplicação web. 5. Índice de desempenho acadêmico. I. Título.

CDU: 004:378

Catálogo na Publicação:

Bibliotecário Alessandro Dietrich - CRB 10/2338

*À minha esposa, Deyse,
e meus filhos, Alberto e Augusto,
pelas alegrias que tem me dado e por
compreenderem que há tempo para todas as coisas.*

RESUMO

Técnicas de visualização de dados podem auxiliar nas mais diversas áreas de atuação humana, em especial na compreensão de dados e informações de diferentes fenômenos que se quer estudar. Quanto mais variáveis estão relacionadas com esse fenômeno, mais desafiador se torna seu tratamento e representação visual. Pensando em educação no Brasil e suas bases de dados abertas, bem como em bases de dados acadêmicas existentes nas instituições, o uso de técnicas matemáticas para correlacionar conjuntos de dados e métodos de visualização para apresentar essas correlações, disponíveis em uma ferramenta de fácil acesso e operação, podem tornar públicas informações sobre a qualidade da educação de determinada região, estado, município e instituição de ensino. Outro benefício pode ser a indicação de fatores que antes eram ignorados, como alvos de investimento e ainda ajudar na elaboração de políticas públicas, nacionais ou regionais, que tornem a educação mais eficiente, abrangente e inclusiva. Iniciativas de organizações não governamentais e algumas vinculadas ao governo brasileiro tem elaborado ferramentas de filtragem de informações e divulgação de dados sobre qualidade e investimento de recursos na educação. O governo brasileiro usa índices de desempenho para avaliar suas Instituições de Ensino Superior. O Conceito Preliminar de Curso é um desses. Este trabalho apresenta uma solução, visando elaborar uma metodologia de visualização de dados através de uma aplicação *web*, com tecnologias *open source*, utilizando o método de análise de componentes principais (ACP) como técnica matemática de correlação de variáveis, e distribuindo resultados sobre um mapa com a utilização da API do *Google Maps*, porém, tendo como foco, a busca do nível de influência de diferentes fatores, inclusive de alguns não ligados diretamente à educação, na performance de instituições de ensino e no rendimento acadêmico de alunos, tendo como estudo de caso, a análise de um índice de desempenho na educação superior.

Palavras-chave: Visualização de dados, Correlação de dados educacionais, Análise de componentes principais, aplicação *Web*, técnicas matemáticas, índice de desempenho acadêmico.

ABSTRACT

Data visualization techniques can help in several areas of human activity, especially in understanding data and information from different phenomena to be studied. The more variables are related to this phenomenon, the more challenging it becomes their treatment and visual representation. Thinking about education in Brazil and its open databases, as well as in existing academic databases in institutions, using mathematical techniques to correlate data sets and visualization methods to present these correlations available in an easy tool access and operation may disclose information on the quality of education in a region, state, county and educational institution. Another benefit could be the indication of factors that were ignored, as investment targets and also help in the development of public policies, national or regional, that make more efficient, comprehensive and inclusive education. Initiatives of non-governmental organizations and some linked to the Brazilian government has prepared information filtering tools and dissemination of data on quality and investment of resources in education. Brazilian government uses performance indicators to assess their undergraduation institutions. Course Preliminar Concept (CPC) is one of those. This paper presents a solution to this profile, aiming to develop a data visualization methodology through a web application with open source technologies, using principal component analysis method (PCA) as mathematical technique of variable correlation, and distributing results on a map using the Google Maps API, however, focusing on the search for the level of influence of different factors, including some not directly related to education, performance of educational institutions and the academic performance of students, taking as a case study, the analysis of a performance index in undergraduation.

Keywords: Data visualization, educational data correlation, principal component analysis, web application, mathematical techniques, academic performance index.

LISTA DE ILUSTRAÇÕES

Figura 1. Gráfico produzido a partir de dados abertos da educação no Brasil.....	18
Figura 2. Dados de ingressantes e concluintes em IES no Brasil em tabela simples.....	18
Figura 3. Detalhe de tela do QEdu	23
Figura 4. Detalhe de um quadro de resultados do Scimago	24
Figura 5. Dados do Inep de matrícula do censo superior de 2012.....	27
Figura 6. Conversão da faixa do CPC para o CPC contínuo.....	30
Figura 7. Pesos em percentuais dos componentes do CPC	30
Figura 8. Visualização Analítica	34
Figura 9. Pesquisa no google maps	35
Figura 10. Amostra de Estados usados na pesquisa, IES e cursos	42
Figura 11. Modelo resumido do protótipo do Vis-Scholar.....	43
Figura 12. Detalhe da tela e gráficos gerados sobre o mapa. Nesta análise o CPC (variável de análise) está sendo expresso como função das variáveis de dependência.....	44
Figura 13. Detalhe do gráfico gerados sobre o mapa do estado de PE.....	44
Figura 14. Detalhe do gráfico gerados sobre o mapa do estado de SP.....	45
Figura 15. Detalhe do gráfico gerados sobre o mapa do estado do RJ.....	45
Figura 16. Detalhe do gráfico gerados sobre o mapa do estado do PA.....	45

LISTA DE SIGLAS

APC	Análise de Componentes Principais
API	Application Programming Interface
AVA	Ambiente Virtual de Aprendizagem
CPC	Conceito Preliminar de Curso
Enade	Exame Nacional de Desempenho dos Estudantes
Enem	Exame Nacional do Ensino Médio
Ideb	Índice de Desenvolvimento da Educação Básica
IES	Instituição de Ensino Superior
IGC	Índice Geral de Curso
Inep	Instituto Nacional de Estudo e Pesquisa Educacional Anísio Teixeira
IRA	Índice de Rendimento Acadêmico
MEC	Ministério da Educação
PC	Personal Computer
PIB	Produto Interno Bruto
SIR	Scimago Institutions Rankings
Sinaes	Sistema Nacional de Avaliação da Educação Superior
TI	Tecnologia da Informação
UNISINOS	Universidade do Vale do Rio dos Sinos

SUMÁRIO

1 INTRODUÇÃO	17
1.1 Relevância	20
1.2 Divulgação de dados abertos na educação	22
1.3 Ranking de IES	23
1.4 Objetivos	24
1.5 Organização do Trabalho	25
2 REVISÃO BIBLIOGRÁFICA.....	27
2.1 Conceito Preliminar de Curso – CPC.....	28
2.2 Pesquisas ligadas ao desempenho de estudantes e IES.....	31
3 CONCEITOS BÁSICOS	33
3.1 Visualização de dados e informações	33
3.2 Visualização de densidade em mapas.....	34
3.3 Análise de Componentes Principais – ACP	35
4 DESENVOLVIMENTO DO MODELO	39
4.1 Hipótese	39
4.2 Metodologia	39
4.3 Resultados	43
5 CONCLUSÃO	49
5.1 Trabalhos futuros	50

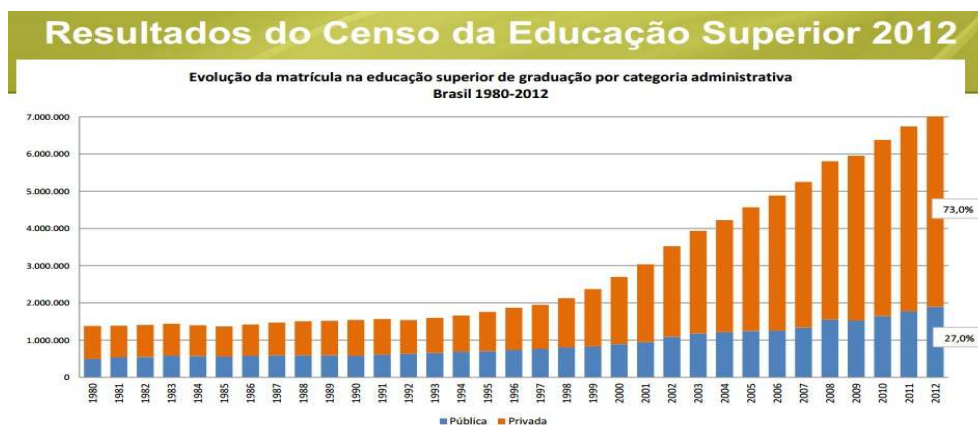
1 INTRODUÇÃO

A educação no Brasil, em especial na esfera pública, é uma das áreas onde a visualização de dados para análise e pesquisa pode apresentar grande relevância social, dado o impacto que políticas públicas e investimentos voltados para educação têm para a população. A avaliação das instituições de ensino é a principal ferramenta da organização e implementação das reformas educacionais (Sobrinho, 2010), e combinada com uma boa técnica de visualização de dados pode ajudar gestores e pesquisadores a enxergarem de forma mais amigável e objetiva informações acerca de sua área de atuação ou de negócio. A educação superior no Brasil é sistematicamente avaliada pelo Ministério da Educação (MEC), que utiliza índices previstos num sistema mais amplo de avaliação conhecido como Sistema Nacional de Avaliação da Educação Superior (Sinaes). Na prática, um desses índices de avaliação de desempenho é o Conceito Preliminar de Curso (CPC). Esse índice de desempenho é utilizado como parâmetro, inclusive, para visitas de avaliadores do MEC e sanções em Instituições de Ensino Superior (IES) com valor resultante insatisfatório, sendo considerado pelo governo como um retrato preliminar de cursos avaliados nas IES (MEC, 2013).

Instituições de Ensino Superior (IES) brasileiras, tem produzido, já há alguns anos, uma quantidade significativa de dados. Parte destes estão disponíveis nas bases de dados públicos brasileiros, censos da educação, estatísticas do Ministério da Educação e agentes relacionados, entretanto muito do que é gerado pelas escolas e universidades ainda fica em bases locais nas próprias instituições. Nesse cenário, o aproveitamento desses dados para análise, investigação e propostas de melhorias de cunho mais locais, nem sempre é efetuado, deixando um amontoado de dados sem aproveitamento, apenas ocupando espaço em discos de servidores. Além disto, existem fatores que não estão diretamente atrelados à educação numa observação superficial, entretanto, num olhar mais aprofundado, podem exercer uma influência significativa tanto para o aluno, quanto para a instituição de ensino. A utilização de dados educacionais, de qualquer base em situação semelhante, combinada com uma ferramenta que provisione visualização desses dados e correlações entre fatores (variáveis) direta ou indiretamente ligados, bem como a necessidade de fomento à políticas públicas e investimento na educação, incrementam a relevância deste tipo de pesquisa. No gráfico da figura 1, podemos visualizar facilmente a ideia de crescimento nos números de matrículas com o acumulado de acadêmicos matriculados, o que levaria mais tempo e atenção se estivéssemos lendo uma tabela com todos esses dados. A figura 1 demonstra como a

visualização dos dados no gráfico fornece facilmente a ideia de crescimento das matrículas ao longo dos anos, se comparado com a leitura de uma simples tabela com os mesmos dados. A figura 2 contém dados apenas de ingressantes e concluintes em IES no Brasil entre 2003 e 2013.

Figura 1. Gráfico produzido a partir de dados abertos da educação no Brasil



Fonte: <http://blog.portalexamedeordem.com.br/blog/2013/09/censo-da-educacao-superior-2012-direito-tem-o-2o-maior-contingente-de-alunos-no-ensino-superior/>

Figura 2. Dados de ingressantes e concluintes em IES no Brasil em tabela simples

EVOLUÇÃO DO NÚMERO DE INGRESSANTES E CONCLUINTES EM CURSOS DE GRADUAÇÃO		
Ano	Ingressantes	Concluintes
2003	1.554.664	532.228
2004	1.646.414	633.363
2005	1.805.102	730.484
2006	1.965.314	762.633
2007	2.138.241	786.611
2008	2.336.899	870.386
2009	2.065.082	959.197
2010	2.182.229	973.839
2011	2.346.695	1.016.713
2012	2.747.089	1.050.413
2013	2.742.950	991.010

Fonte: Censo da Educação Superior/Inep/MEC

Visualizar correlações entre dados de forma clara e eficiente, em especial para usuários que não atuam no campo da informática diretamente, tem sido um desafio complexo para os profissionais de informática (Marx, 2013), sendo um problema abordado de diferentes formas na literatura acadêmica. Para Lu *et al.* (2013), Visualização e visualização analítica de informações são áreas de estudo muito próximas, ambas objetivam o desenvolvimento de técnicas visuais para revelar informações significativas dos dados em aplicações do mundo real. A visualização, como área de pesquisa tem sua origem na computação gráfica e tornou-se uma área de pesquisa popular ao longo dos anos. Já a visualização analítica é relativamente jovem, com foco em proporcionar um raciocínio analítico através de interfaces visuais interativas

A visualização de informações geralmente se relaciona com a elaboração de representações gráficas que ilustram e destacam determinados aspectos relevantes dos dados analisados. Essa representação pode ser através de diagramas, gráficos, mapas ou outros recursos visuais. Conforme Arabnia (1999), a visualização sintetiza e exhibe imagens derivadas de modelos dinâmicos utilizados para analisar os fenômenos do mundo real. O alvo é empregar o sentido visual humano como um meio para perceber a informação e estimular a criatividade para pensar sobre o seu significado.

Neste trabalho é apresentado um método de exame e visualização de dados, a partir de um grupo complexo de variáveis, a fim de identificar níveis de correlações entre elas, demonstrado através de uma aplicação *Web* aqui chamado de Vis-Scholar. Para tal, foi escolhida a técnica matemática de análise de componentes principais (ACP), para correlacionar grupos de dados distintos, e o uso de técnicas de visualização de informações para facilitar o entendimento dos resultados, estimular reflexão, abstrair detalhes menos relevantes e destacar outros considerados mais interessantes dentro de determinado contexto. As representações visuais dos dados são georreferenciadas em função da região de incidência do evento e o acesso se dá por uma aplicação *Web*, através de um *browser*, o que torna a ferramenta facilmente acessível tanto no aspecto do dispositivo usado quanto da plataforma. Foram selecionados 5 tipos de curso de engenharia (civil, mecânica, elétrica, eletrônica e química) oriundos das IES de 7 Estados brasileiros, com dados obtidos no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), vinculado ao MEC, referentes ao ano de 2011, quando houve avaliação das engenharias no Exame Nacional de Desempenho de Estudantes (Enade) que também traz resultados do Exame Nacional do Ensino Médio (Enem), estando esses dados abertos e disponíveis à população.

1.1 Relevância

A análise de dados educacionais permite a identificação de modelos de estudantes que incluem informações detalhadas como o seu nível de conhecimento em determinado assunto ou disciplina, comportamentos e motivação para aprender. Ela auxilia os educadores na compreensão do processo de aprendizagem e as abordagens que eles podem usar para melhorar os seus métodos de ensino. Os professores podem usar os dados analisados para determinar a forma de organizar e estruturar o currículo, entregar informações sobre o curso e as ferramentas úteis para envolver os seus alunos nos resultados de aprendizagem ideais. Em particular, o exame mais detalhado desses tipos de dados oferece uma oportunidade para que os educadores identifiquem rapidamente os padrões de comportamento, o que pode apoiar os sua metodologia de ensino ao longo de um curso ou para melhorar cursos futuros. Educadores podem determinar os indicadores que mostram a satisfação do aluno e engajamento de material do curso, e também monitorar o progresso de aprendizagem (Card et al. 1999).

As administrações de instituições públicas ou privadas são responsáveis pela atribuição de recursos a estas. Como as instituições estão cada vez mais responsáveis pelo sucesso do aluno, a análise de dados educacionais está cada vez mais comum e tem o intuito de identificar e abordar alunos em risco em tempo e de forma mais eficiente (Romero; Ventura 2010).

Gestores também podem se valer de dados educacionais para traçar ou alcançar metas de desempenho institucionais, identificar fragilidades e utilizar os recursos disponíveis de forma mais precisa e fundamentada em informações concretas, ou seja, avaliar o desempenho da instituição de ensino por eles administrada, contemplando premissas do Sistema Nacional de Avaliação da Educação Superior – Sinaes, instituído pela Lei nº 10.861 (Brasil, 2004a), e regulamentado pela Portaria 2.051 (Brasil, 2004b), de 09 de julho de 2004, que aborda três subsistemas de forma integrada: 1) a avaliação institucional, nas instâncias interna e externa, visando identificar o perfil institucional; 2) a avaliação dos cursos de graduação, objetivando identificar as condições de ensino oferecidas aos estudantes no que diz respeito ao corpo docente, infraestrutura física e organização didático-pedagógica; e 3) a avaliação dos estudantes, realizada mediante aplicação do Exame Nacional de Desempenho dos Estudantes (Enad) (Marchelli, 2007). As dimensões consideradas em cada subsistema apresentam indicadores que receberão conceitos para cada elemento que exprime um padrão de qualidade e um conceito final é atribuído a cada uma das dimensões do Sinaes. O modelo de avaliação do Sinaes em conjunto com outros indicadores avaliados pelo Instituto Nacional de Estudos e

Pesquisas Educacionais Anísio Teixeira (Inep), pode ser usado para simular internamente as avaliações externas as quais a IES é submetida regularmente, antecipando pontos que precisam ser observados pelos seus gestores.

Há casos em que a quantidade de dados educacionais é numerosa e detalhada. Por exemplo, vários Ambientes Virtuais de Aprendizagem (AVA's) disponibilizam informações sobre o acesso de cada aluno no ambiente, quantas vezes ele acessou e quantos minutos ele permaneceu com o objeto de aprendizagem exibido na tela de seu dispositivo (PC, *tablet*, etc). Outro exemplo, são dados de Sistemas Tutores Inteligentes que registram cada vez que um aluno apresente uma solução para um problema; eles podem cobrar no momento da apresentação, ou não, se a solução corresponde à solução esperada, a quantidade de tempo que se passou desde a última apresentação, a ordem em que os componentes da solução foram inseridos no ambiente, etc. A precisão desses dados é tal que mesmo uma relativamente curta sessão com um ambiente de aprendizagem baseado em computador podem produzir uma grande quantidade de dados de processo para análise.

No entanto, há casos onde os dados são mais escassos, menos detalhados e com mais possibilidade de inconsistências. Por exemplo, um estudante de graduação pode ter apenas um cadastro com uma lista de disciplinas cursadas ao longo do tempo, suas notas em cada disciplina cursada, os cursos de graduação que ele participou, no caso de ter passado por mais de um, e quando o estudante selecionado mudou de curso.

Os dados públicos georreferenciados de educação podem apresentar o aproveitamento escolar por região de um município, estado ou região além de determinar relações entre aproveitamento escolar/acadêmico e outras variáveis como por exemplo: distância de locomoção até a escola, escolaridade dos pais, renda per capita da família, deficiências, doenças, inclusive crônicas, horário no emprego para os que trabalham e estudam, entre outros. A mineração de dados educacionais pode aproveitar dados, detalhados ou não, para descobrir informações significativas sobre os diferentes perfis de alunos, como eles aprendem, o efeito de estratégias pedagógicas inseridas dentro de vários ambientes de aprendizagem. Imagine identificar uma relação entre os objetos de aprendizagem que um estudante acessou durante o curso e sua nota final. Ou uma relação entre a sua nota em um curso particular e sua decisão de mudar de curso. Essa informação pode gerar *insights* sobre o *design* de ambientes de aprendizagem, e permitir que estudantes, professores, administradores escolares e os formuladores de políticas públicas para educação tomem decisões embasadas sobre como

interagir, investir e gerenciar recursos, quer sejam verbas públicas, ou sejam capital da iniciativa privada.

O rendimento acadêmico, basicamente, se refere à medição do conhecimento adquirido no âmbito universitário e demonstrado através de alguma forma de aferição. São considerados estudantes com bom rendimento todos os acadêmicos que obtenham qualificações positivas nas avaliações das disciplinas que participam ao longo do curso. Ou seja, esse rendimento é uma medida das capacidades do aluno, que expressa o que este tem aprendido ao longo do processo formativo. Também inclui a capacidade do estudante em responder aos estímulos educativos. Entretanto, vários são os fatores que podem influenciar o rendimento de um aluno (Mascarenhas e Barca, 2012). Desde a dificuldade própria de algumas disciplinas (ou competências) até diversos problemas de ordem pessoal e social. Dessa forma, torna-se útil identificar esses fatores e mensurar o nível de influência que eles exercem sobre o desempenho acadêmico de alunos, bem como na eficiência de instituições de ensino e relevância de políticas públicas aplicadas na educação.

O cerne deste trabalho pode ser resumido num modelo de ferramenta capaz de auxiliar educadores e gestores de instituições de ensino na tomada de decisões e na melhoria do processo pedagógico, através da análise de dados ligados ao desempenho acadêmico de estudantes de graduação e de fatores relacionados com o desempenho institucional das IES.

1.2 Divulgação de dados abertos na educação

Poucas ferramentas tem sido elaboradas para apresentar dados da educação no Brasil, ainda mais que estejam disponíveis na internet, facilmente acessíveis por alunos, pais de alunos, professores, gestores de recursos públicos, pesquisadores ou qualquer um interessado nessas informações. Não é trivial produzir uma ferramenta que possibilite visualizar informações nacionais, por região, por estado, por município e até mesmo dados de uma instituição de ensino específica, utilizando dados das bases públicas de consultas ou uma que demonstre a influência de determinada informação do aluno, seu perfil psicossocial por exemplo, com o rendimento acadêmico desse aluno, ou com indicadores de desempenho institucionais.

No contexto de divulgação visual de dados abertos, destaca-se o QEDu, um portal aberto e gratuito que disponibiliza informações de fontes oficiais do governo sobre a qualidade no ensino básico em cada escola, cidade e estado brasileiro, a fim de que toda sociedade tenha oportunidade de conhecer melhor a educação no Brasil (Meritt 2015a). Esse

projeto, juntamente com o Portal Ideb (Índice de Desenvolvimento da Educação Básica), que é um indicador de qualidade educacional que combina informações de desempenho em exames padronizados (Prova Brasil ou Saeb), obtido pelos estudantes ao final das etapas de ensino (4ª e 8ª séries do ensino fundamental e 3ª série do ensino médio), com informações sobre rendimento escolar (aprovação ou tempo médio numa etapa), desenvolvido pela Meritt e a Fundação Lemann, tem se mostrado uma iniciativa singular em nosso país (Lemann 2015, Meritt 2015b). No entanto, até o presente não foi abordado o ensino superior nessa ferramenta.

Figura 3. Detalhe de tela do QEdU



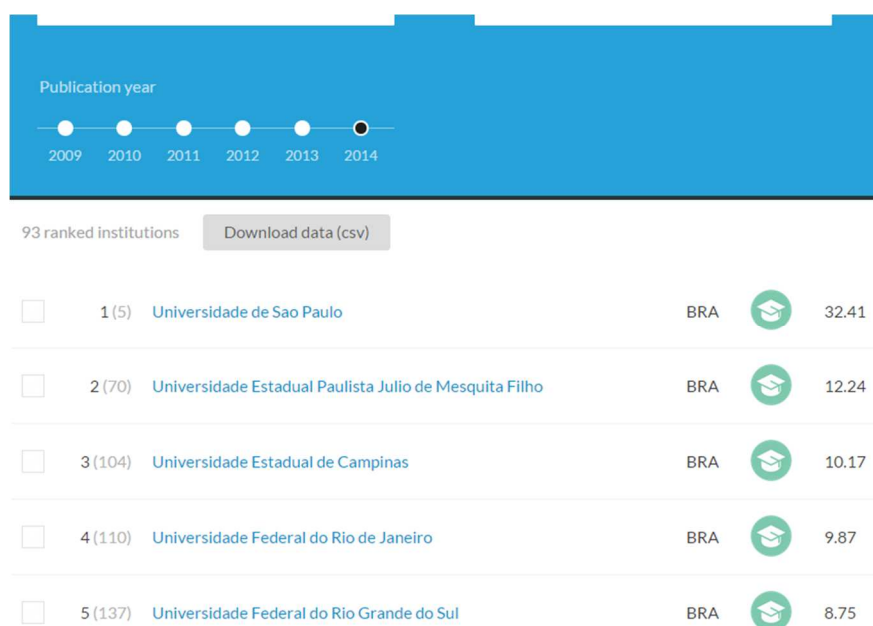
Fonte: <http://www.qedu.org.br>

1.3 Ranking de IES

Apesar de não ser uma prática do governo brasileiro divulgar listas com ranking de desempenho, elas podem ser feitas baseadas em indicadores elaborados pelo próprio MEC, tal como o Índice Geral de Curso (IGC) ou o Conceito Preliminar de Curso (CPC). Existem portais de *ranking* de IES nacionais, como o *Ranking* Universitário da Folha (RUF), e os internacionais, tais como *ranking global Times Higher Education* (THE) (Times, 2015), o *Quacquarelli Symonds* (QS) (Symonds, 2015), o *ranking* de Xangai (ARWU) (ShanghaiRanking, 2015) e o *Scimago Institutions Rankings* (SIR), que é uma ferramenta de avaliação científica usada para classificar universidades e instituições dedicadas à pesquisas

do mundo todo, utilizando a base *Scopus* (Scimagoir, 2015). Este último usa diversos indicadores de desempenho, em diferentes categorias (dimensões), para traçar o *ranking* das instituições e os apresenta num mapa, por países e escala de cores representando, o número de resultados. Junto com o uso de métodos estatísticos para conceituar as instituições, a ferramenta se vale de técnicas de visualização de dados com alto nível de abstração.

Figura 4. Detalhe de um quadro de resultados do Scimago



Fonte: www.scimagoir.com

1.4 Objetivos

O objetivo deste trabalho é desenvolver uma ferramenta computacional usando uma metodologia de análise e visualização de informações decorrentes do estudo da correlação entre variáveis associadas ao desempenho acadêmico e ao desempenho da IES, com outras variáveis. Para atingir este objetivo são necessários os seguintes objetivos específicos:

- Usar ACP para elaborar técnicas para estimar as covariâncias e/ou correlações entre índices de desempenho como o Conceito Preliminar de Curso e variáveis que possam ter alguma influência, obtidas a partir de dados disponibilizados de forma pública pelo governo.
 - Desenvolver tais técnicas de formas implementáveis em uma interface amigável, promovendo a simplificação do seu uso.

- Considerar um estudo de caso, envolvendo variáveis reais obtidas em dados abertos disponibilizados pelo governo.
- Analisar e produzir visualizações das informações de maneira a proporcionar a um número maior de cidadãos, um meio de estudo mais significativo da relação de dependência entre o indicador desempenho institucional selecionado e outras variáveis.
- O método proposto deve permitir a observação de comportamentos evolutivos das IES ao longo do tempo e do espaço geográfico.

1.5 Organização do Trabalho

Este trabalho segue a seguinte organização: no Capítulo 2 tratará do referencial teórico e dos trabalhos relacionados ao tema geral do Vis-Scholar; no Capítulo 3 serão colocados os conceitos básicos utilizados no desenvolvimento da solução proposta; o Capítulo 4 descreve com mais detalhes a ferramenta junto com a metodologia utilizada no seu desenvolvimento, seus objetivos, resultados obtidos com dados reais a partir do censo da educação superior; o Capítulo 5 conclui o trabalho.

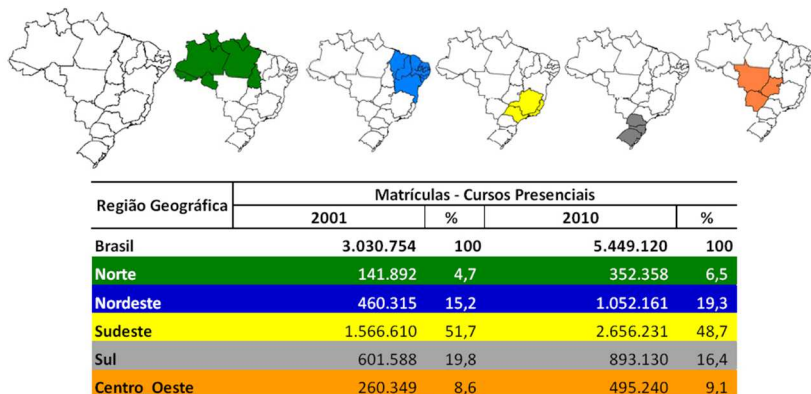
2 REVISÃO BIBLIOGRÁFICA

Sistemas utilizados para visualização analítica de um volumoso e complexo conjunto de dados, integram algoritmos computacionais de descoberta de conhecimento com técnicas de visualização interativa de informações (Schreck; Keim 2013). Esses dados podem ser oriundos de bases públicas ou de bases locais. Programas de monitoramento da educação no Brasil e a política de transparência do setor público tem gerado um grande volume de dados. Entre estes agentes governamentais pode-se citar o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep, que é uma autarquia federal vinculada ao Ministério da Educação, cuja missão é promover estudos, pesquisas e avaliações sobre o sistema educacional brasileiro. O objetivo é subsidiar a formulação e implementação de políticas públicas para a área educacional, a partir de parâmetros de qualidade e equidade, bem como produzir informações claras e confiáveis aos gestores, pesquisadores, educadores e público em geral (Portal do MEC; Inep 2015).

As bases de dados do Inep são atualmente formadas a partir dos dados coletados pelo:

- Censo Escolar (1995-2013);
- Censo da Educação Superior (1995-2012);
- Sistema de Avaliação da Educação Básica - Saeb (1995, 1997, 1999, 2001, 2003, 2005,2011);
- Prova Brasil (2007, 2009, 2011);
- Exame Nacional do Ensino Médio - Enem (1998-2012);
- Exame Nacional de Cursos – ENC/Provão (1997-2003);
- Exame Nacional de Desempenho de Estudantes - Enade (2004-2012).

Figura 5. Dados do Inep de matrícula do censo superior de 2012



Fonte: <http://www.cadsoft.com.br/blog/2011/11/21/censo-da-educacao-superior-2010/>

As pesquisas na área de mineração de dados educacionais se valem de incontáveis técnicas utilizadas em diversos contextos educativos. Em cada caso, a meta é converter dados brutos em informações significativas sobre o processo de aprendizagem, com o propósito de tomar as melhores decisões sobre o projeto e a trajetória de um ambiente de aprendizagem. Assim, a análise de dados educacionais, de forma geral, consiste de quatro fases (Romero; Ventura 2010), (Baker 2010):

- A primeira fase é pesquisar um repositório de dados a partir de um ambiente educacional com o objetivo de encontrar relações consistentes entre as variáveis. Vários algoritmos para identificar essas relações têm sido utilizados, incluindo a sua classificação, regressão, *clustering*, análise fatorial, análise de redes sociais, a mineração de regras de associação, e padrão de mineração sequencial;
- Descobertas essas relações, deve-se validá-las de forma a evitar superajuste;
- Relações validadas são aplicadas para fazer previsões sobre eventos futuros no ambiente de aprendizagem;

As previsões são usadas para apoiar os processos de tomada de decisão.

2.1 Conceito Preliminar de Curso – CPC

No entendimento de (Vasconcelos, Diniz e Andrade, 2012), as IES utilizam de maneiras diferenciadas e com objetivos variados o que eles chamam de índice de rendimento acadêmico (IRA), fazendo deste uma medida de avaliação dos discentes e, conseqüentemente, da própria instituição. No tocante à avaliação de IES, o governo brasileiro também se vale de indicadores de desempenho para avaliar ações realizadas no âmbito da administração acadêmica, entre os quais encontra-se o CPC, acreditando que esses índices devem ser explorados como ferramentas estratégicas para a melhoria da educação superior no Brasil.

O Conceito Preliminar de Curso, é um índice de desempenho para IES elaborado com o objetivo de combinar diferentes medidas de qualidade de cursos de graduação e algumas variáveis de insumo em uma única medida. É constituído originalmente de oito componentes, agrupados em três dimensões visando avaliar a qualidade dos cursos de graduação (Miranda *et al.*, 2013): a) desempenho dos estudantes, b) corpo docente e c) condições oferecidas para o desenvolvimento do processo formativo (Inep, 2015b). Entretanto, um dos componentes, a Nota do Ingressante (NI), foi retirado dos cálculos desde 2011.

O Cálculo do CPC é feito para cada Unidade de Observação, constituída pelo conjunto de cursos que compõe uma área de avaliação específica do Exame Nacional de Desempenho dos Estudantes (Enade), de uma Instituição de Educação Superior, em um determinado município.

A fórmula para calcular o CPC de uma IES i , é :

$$CPC = 0,35NIDD + 0,20NC + 0,075NF + 0,075NO + 0,15NPD + 0,075NPM + 0,075NPR, \quad (1)$$

Onde,

NIDD = Nota do indicador de diferença entre o desempenho observado e esperado do aluno, obtido através do Enade (No decorrer de um curso, cada aluno participa do Enade em dois momentos do curso: no início e no fim. Tais exames têm o mesmo nível e abrangem a totalidade dos conhecimentos da grade curricular. Portanto, o objetivo da aplicação desses testes no início e no fim de um curso, é dar uma medida da evolução acadêmica do aluno no curso. Para o cálculo do NIDD de um curso, o mesmo teste Enade é feito por alunos ingressantes, “calouros” naquele ano, e os concluintes naquele ano. O NIDD de um curso num dado ano é a diferença entre a nota média dos alunos concluintes e a nota média dos alunos ingressantes);

NC = Nota dos Concluintes no Enade;

NF = Nota referente à infraestrutura da IES;

NO = Nota referente à organização didático-pedagógica do curso;

NPD = Nota de Professores Doutores;

NPM = Nota de Professores Mestres;

NPR = Nota de Professores com Regime de Dedicção Integral ou Parcial;

Sendo, conforme a figura 6 abaixo, a conversão dos valores contínuos em faixas de 1 a 5 (Inep, 2011):

Figura 6. Conversão da faixa do CPC para o CPC contínuo

CPC (Faixa)	$NCPC_j$ (Valor Contínuo)
1	$0 \leq NC_j < 0,945$
2	$0,945 \leq NC_j < 1,945$
3	$1,945 \leq NC_j < 2,945$
4	$2,945 \leq NC_j < 3,945$
5	$3,945 \leq NC_j \leq 5$

Fonte: Inep, 2011

E na figura 7 abaixo, o peso dos componentes em percentual:

Figura 7. Pesos em percentuais dos componentes do CPC

Componentes	NIDD	NC	NPM	NPD	NPR	NF	NO
Pesos	35%	20%	7,5%	15%	7,5%	7,5%	7,5%

Fonte: Inep, 2015

Existem detalhes técnicos que envolvem o cálculo do CPC, seus componentes, seus pesos e a influência que eles têm quando aplicados em cursos de IES públicas federais, comparando com as IES privadas, cuja finalidade é tentar compensar as diferenças entre esses tipos de instituições, como o perfil dos seus alunos, a concorrência de seus cursos e os recursos disponíveis (Bittencourt *et al.*, 2010). O CPC foi elaborado procurando amenizar essas diferenças e eventualmente sofre mudanças visando essa adequação

O CPC foi o índice selecionado para ser a variável de análise pela importância dada a este na avaliação dos cursos pelo MEC. Ele é utilizado também como base para outros indicadores como o Índice Geral de Curso (IGC), é constituído por componentes resultantes de cálculos estatísticos, aplicados de forma clara em manuais e notas técnicas do Inep, e está disponível nos dados abertos do censo da educação superior.

2.2 Pesquisas ligadas ao desempenho de estudantes e IES

O trabalho de Silva *et al.* (2013) investigou indicadores relacionados à evasão e reprovação no primeiro ano de um curso de engenharia, usando uma ferramenta de processamento e análise de dados baseados em dois modelos matemáticos, um bi linear e outro tri linear, cujos resultados poderiam servir até como indicador de desempenho e qualidade de ensino. O ACP foi utilizado no seu trabalho no intuito de obter possíveis indicadores que pudessem fornecer subsídios para uma análise diferenciada das disciplinas que compõem o primeiro ano do curso de Engenharia de Teleinformática da Universidade Federal do Ceará. A opção por observar dados do primeiro ano do curso foi feita por ser este o período onde a evasão e reprovação dos estudantes são mais críticas (Smaill *et al.*, 2012).

Gramani e Scrich (2012), Estudaram a influência do desempenho em matemática na educação básica para a escolha de carreiras no ensino superior. Analisaram primeiro a relativa eficiência educacional em matemática para cada unidade da federação, se valendo dos dados abertos do Sistema Nacional de Avaliação da Educação Básica e do Exame Nacional do Ensino Médio. Depois, correlacionaram essa eficiência com a procura por cursos no vestibular. O problema motivador desse trabalho foi a escassez de engenheiros no Brasil, a alta taxa de evasão dos alunos de engenharia e o baixo número de professores formados nos cursos de exatas, como matemática, física e química. Os resultados principais indicaram que cursos de engenharia tem maior procura e maior número de ingressos em estados com melhor eficiência educacional em matemática, enquanto os cursos de formação de professores em matemática e física tem maior procura em estados com menor eficiência educacional em matemática.

Dodorico e Coello (2014), propuseram o uso de algoritmos de filtragem colaborativa para prever a nota de um estudante a partir dos dados adquiridos durante a interação de estudantes com sistemas tutores. Essa filtragem consiste na análise do comportamento passado dos usuários, isto é, suas interações registradas, as quais servem para identificar novas associações usuário-item. Além do algoritmo clássico de filtragem colaborativa, User-kNN, foi utilizado o ACP para reduzir a dimensionalidade das matrizes geradas no modelo. O modelo proposto foi comparado com outros modelos e foi verificado ser possível aprimorar a precisão dos resultados em comparação com os modelos clássicos, ainda que o desempenho do modelo proposto não se mostrou melhor.

Prathap e Savithri (2014), realizaram uma análise de desempenho bibliométrico das principais IES na Índia, utilizando os dados divulgados no SIR, e usando ACP para

estabelecer que os indicadores primários são ortogonais e representam uma quantidade e uma dimensão de qualidade / produtividade, respectivamente. As dimensões de quantidade são dependentes do tamanho e as de qualidade / produtividade são independentes do tamanho. Seu trabalho se restringiu a 156 IES classificadas como as melhores da Índia em 2014 no SIR, aproveitado o formato dos dados ali disponíveis, propondo sua avaliação de desempenho para as IES e usando ACP para extrair as dimensões principais.

3 CONCEITOS BÁSICOS

Neste capítulo serão abordados os conceitos elementares a serem utilizados na produção da ferramenta Vis-Scholar, seus contextos e suas definições.

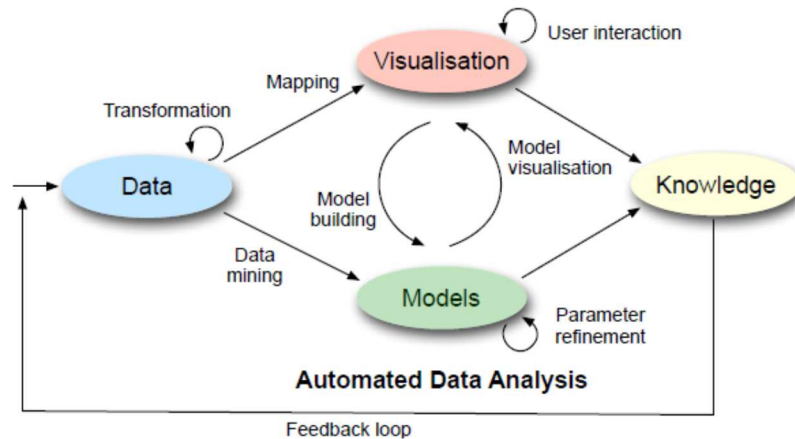
3.1 Visualização de dados e informações

Azevedo, Oliveira e Parreiras, (2014), afirmam que a apresentação visual dos dados de forma clara permite que os usuários possam obter uma melhor compreensão dos mesmos. Já Hidalgo e Almosawi (2014), a entendem como algo bem mais interativo, pois afirmam que as visualizações de dados deixaram a rigidez do design gráfico tradicional, e substituíram tinta por pixels, proporcionando a flexibilidade de trabalhar com projetos que não são totalmente especificados a priori. Na sua encarnação mais moderna, a visualização de dados gerou uma nova forma de *design* gráfico, onde atributos visuais, como linhas, formas e cores se tornam nada mais do que a realidade corporal de objetos gráficos cuja alma é feita de dados. O novo *designer* gráfico já não cria visualizações, escolhendo um conjunto rígido de formas, posições e cores, mas sim, escolhendo as regras necessárias para os dados respirarem na forma de abstrações geométricas.

Estas novas visualizações têm propriedades que estavam ausentes em seus antecessores baseados em tinta. Estas propriedades permitem o surgimento de uma relação mais rica entre a visualização e o leitor, que é agora menos espectador e mais explorador (Hidalgo e Almosawi, 2014).

Segundo Keim Daniel e Mansmann (2010), A visualização analítica de dados é definida como uma ciência do raciocínio analítico baseado por interfaces visuais interativas. A figura 8 demonstra uma visão abstrata de diferentes fases (elipses) e as respectivas transições (setas) do processo de visualização analítica.

Figura 8. Visualização Analítica



Fonte: (Keim Daniel e Mansmann, 2010)

Iliinsky e Steele (2011) explicam porquê a visualização é útil para examinar, compreender e transmitir a informação:

- a) visualização aproveita as capacidades incríveis do sistema visual para mover uma enorme quantidade de informações para o cérebro muito rapidamente;
- b) visualização permite identificar padrões, relacionamentos e seus significados;
- c) visualização ajuda a identificar subproblemas;
- d) visualização é realmente boa para identificação de tendências ou pontos fora de série, descobrindo pontos específicos ou interessantes em um campo maior.

3.2 Visualização de densidade em mapas

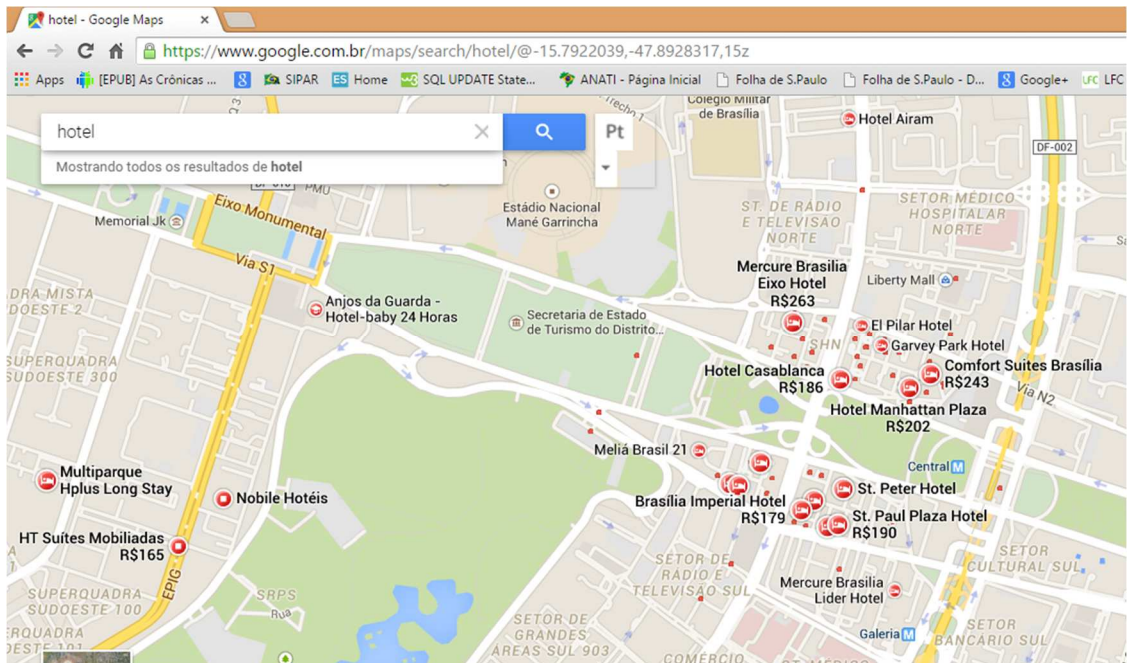
A combinação de informações geográficas com diferentes conjuntos de dados resultou na popularização de sistemas de informações baseados na localização em atlas eletrônicos disponíveis na web. Pode-se, por exemplo, se utilizar de técnicas como mapas de calor para representar a densidade de determinado fenômeno ou entidade numa dada região.

(Azevedo, Oliveira e Parreiras, 2014), afirma que a literatura aponta os Sistemas de Informação Geográficas – SIG como ferramentas importantes para visualização de *linked data*, pois permitem a integração de dados vindo de várias fontes heterogêneas, visando potencializar a descoberta e a divulgação de novos conhecimentos.

Ao se pesquisar por “Hotel”, em Brasília-DF, no google maps, uma série de resultados, já distribuídos no mapa da cidade, incluindo o preço de algumas diárias, surgem facilitando ao usuário a escolha mais vantajosa tanto economicamente, quanto no que diz

respeito à localização do hotel, tudo isso de maneira rápida, prática e com um requisito quase zero de conhecimento da parte do usuário. Esse resultado está ilustrado na figura 9, abaixo.

Figura 9. Pesquisa no google maps



Fonte: <https://www.google.com.br/maps/search/hotel/@-15.7922039,-47.8928317,15z>

3.3 Análise de Componentes Principais – ACP

ACP é uma das técnicas estatísticas mais antigas e conhecidas para a análise multivariada de dados. A ideia central da ACP é reduzir a dimensionalidade do conjunto de dados analisados no qual existe um grande número de variáveis correlacionadas, mantendo o máximo possível da variação dos dados originais. (Jolliffe, 2002)

Tal redução da dimensão dos dados é obtida por meio da transformação em um novo conjunto de variáveis, os componentes principais, que são no mesmo número das variáveis originais, porém, não são correlacionadas e resultam da combinação linear das variáveis originais.

Estas novas variáveis são ordenadas de maneira que as primeiras retêm grande parte da variação presente em todas as variáveis originais. Dessa forma, usando componentes principais, um conjunto significativamente grande de dados complexos pode ser representado por um número significativamente pequeno de variáveis novas. Tal simplificação, aliada a

técnicas de visualização de dados, possibilita avaliar visualmente semelhanças e diferenças entre os dados e avaliar se pode haver um agrupamento de informações (Ringnér, 2008).

Consideremos um conjunto de dados envolvendo V variáveis, ou seja, o conjunto é constituído por V séries temporais. Este conjunto também terá V componentes principais, as quais são combinações lineares das variáveis originais. As combinações lineares são dadas por manipulações algébricas da matriz de covariância das V variáveis. O método de análise por componentes principais encontra o mesmo número de novas variáveis, no entanto, apenas algumas delas normalmente explicam a maior parte da variabilidade do processo. Assim, é possível diminuir o número de variáveis de análise, sem ter perda significativa de informações (Jackson, 1991). Dado um conjunto de dados com V variáveis e N observações, para efeito de análise, Pode-se organizá-los na seguinte matriz:

$$W = \begin{bmatrix} w_1^1 & w_2^1 & \cdots & w_V^1 \\ w_1^2 & w_2^2 & \cdots & w_V^2 \\ \vdots & \vdots & \ddots & \vdots \\ w_1^N & w_2^N & \cdots & w_V^N \end{bmatrix}, \quad (3.1)$$

onde, o subscrito indica a variável e o sobrescrito indica a observação. Sendo assim, obtemos nas colunas da matriz as variáveis de análise e nas linhas as observações. Em geral, consideramos as variáveis de análise como *vetores-coluna N -dimensionais*, isto é:

$$\mathbf{w}_1 = [w_1^1 \ w_1^2 \ \dots \ w_1^N]' \quad (3.2)$$

$$\mathbf{w}_2 = [w_2^1 \ w_2^2 \ \dots \ w_2^N]' \quad (3.3)$$

⋮

$$\mathbf{w}_V = [w_V^1 \ w_V^2 \ \dots \ w_V^N]' \quad (3.4)$$

Onde os apóstrofes indicam transposições de matrizes / vetores.

O método de determinação das componentes principais, levando em conta este processo com V variáveis, se dá encontrando a matriz de covariância de W , dada por:

$$S = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1V} \\ s_{12} & s_2^2 & \cdots & s_{2V} \\ \vdots & \vdots & \ddots & \vdots \\ s_{V1} & s_{V2} & \cdots & s_V^2 \end{bmatrix}, \quad (3.5)$$

Onde, s_k^2 é a variância da variável k e s_{ij} é a covariância entre as variáveis i e j . Observe que s_k^2 e s_{ij} são produtos internos entre as variáveis de análise. Portanto, assumindo

que essas variáveis de análise estejam centralizadas, ou seja, que cada observação de cada variável foi diminuída de sua média, podemos definir s_k^2 e s_{ij} como:

$$s_k^2 = \frac{(\mathbf{w}_k, \mathbf{w}_k)}{N-1} \quad s_{ij} = \frac{(\mathbf{w}_i, \mathbf{w}_j)}{N-1} \quad (3.6)$$

Onde $(.,.)$ representa o produto interno usual. Note que S é uma matriz simétrica, o que implica numa matriz diagonalizável, visto que a ACP está baseada na álgebra linear. Então, existe uma matriz U , ortonormal, tal que:

$$U'SU = L, \quad (3.7)$$

Onde U' indica a transposta da matriz U , e L é a matriz diagonal, onde os elementos da diagonal principal são os autovalores $l_1 > l_2 > \dots > l_V$ de S . Já as colunas $\mathbf{u}_1, \mathbf{u}_2 \dots \mathbf{u}_V$ de U são os autovetores normalizados de S , respectivamente associados àqueles autovalores, naquela ordem. Esses autovetores são mutualmente ortogonais entre si. Sendo assim, as matrizes U' e U são matrizes de mudança de base entre as variáveis antigas e as variáveis novas, em relação às quais os dados mostram maior independência. Muitas outras informações sobre ACP podem ser encontradas em (Jackson, 1991).

Sob o ponto de vista geométrico, o último componente principal, definido pelo último autovetor \mathbf{u}_V , nos dá a equação do hiperplano de dimensão $V-1$ paralelo às demais componentes principais, definidas pelos autovetores $\mathbf{u}_1, \mathbf{u}_2 \dots \mathbf{u}_{V-1}$, segundo a equação abaixo:

$$\sum_{v=1}^V u_{vV} w_v = \mathbf{u}_V \cdot \bar{\mathbf{w}}, \quad (3.8)$$

onde $\bar{\mathbf{w}}$ é o vetor médio das observações (amostras).

4 DESENVOLVIMENTO DO MODELO

O modelo adotado se apresenta através de uma ferramenta, o Vis-Scholar, que seja útil para dar base às decisões de educadores e gestores de instituições de ensino na melhoria do processo pedagógico, antecipar resultados de avaliações externas aplicadas periodicamente às IES, e apresentar informações essenciais aos agentes governamentais que investem recursos e estabelecem políticas públicas em prol da educação de forma estratégica, bem como à sociedade civil que está cada dia mais atenta à gestão dos bens públicos, através da análise matemática de fatores ligados ao desempenho acadêmico de estudantes de graduação e de fatores relacionados com o desempenho institucional da IES, utilizando para isso índices de desempenho estabelecidos pelo governo brasileiro e dados abertos da educação superior.

4.1 Hipótese

Uma metodologia que se vale de uma solução de visualização de dados na educação, construída sobre um modelo matemático robusto de análise de correlação de variáveis, pode auxiliar na percepção de indicações de correlação, com maior ou menor intensidade, entre fatores direta e indiretamente ligados a índices usados pelo governo brasileiro para aferir o desempenho das IES.

4.2 Metodologia

Para alcançar os objetivos geral e específicos, foram definidas as variáveis a serem utilizadas como foco desta pesquisa, trazendo um estudo de caso com dados reais para apresentação do método. As variáveis levadas em consideração para a elaboração da ferramenta foram obtidas a partir de dados públicos e abertos da educação superior no Brasil. Estes dados foram selecionados principalmente por estarem disponíveis em tempo para a pesquisa. São os seguintes:

- Dependentes
 - Conceito Preliminar de Curso – CPC
- Independentes
 - Concorrência por curso
 - População municipal por número de alunos matriculados no curso
 - Despesas gastas na IES por curso

- Carga horária do curso
- Nota no Enem dos ingressantes no curso

A metodologia de análise consiste na aplicação da equação (3.8), reescrita abaixo para o caso em que os dados estão centralizados ($\bar{\mathbf{w}} = \mathbf{0}$),

$$\sum_{v=1}^V u_{vV} w_v = 0 \quad (3.8a)$$

Entre as variáveis $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_V$ encontra-se uma das variáveis dependentes, enquanto as demais são independentes, ambas escolhidas pelo usuário.

Para fins de clareza, consideremos o caso em que $V = 3$, e digamos que \mathbf{w}_3 seja a dependente.

Neste caso, podemos escrever (3.8a), explicitando-se w_3 , como:

$$\mathbf{w}_3 = -\frac{u_{13}}{u_{33}} \mathbf{w}_1 - \frac{u_{23}}{u_{33}} \mathbf{w}_2 \quad (3.8b),$$

isto é, temos a variável dependente escrita como função linear das outras (independentes). Com isto, podemos definir os percentuais de explicabilidade de \mathbf{w}_1 e \mathbf{w}_2 para com \mathbf{w}_3 como

$$p_{13} = \frac{\alpha_{13}}{\alpha_{13} + \alpha_{23} + \epsilon_3} \quad \text{e} \quad p_{23} = \frac{\alpha_{23}}{\alpha_{13} + \alpha_{23} + \epsilon_3}$$

onde

$$\alpha_{13} = -\frac{u_{13}}{u_{33}}, \quad \alpha_{23} = -\frac{u_{23}}{u_{33}} \quad \text{e} \quad \epsilon_3 = \frac{l_3}{l_1 + l_2 + l_3},$$

em que l_1, l_2 e l_3 são os autovalores da matriz de covariâncias S , e portanto ϵ_3 mede a proporção em que o modelo (3.8b) não explica os dados. Com isso, temos três percentuais que relacionam o modelo aos dados: dois de explicabilidade (p_{13} e p_{23}) e um de não explicabilidade (ϵ_3). Tais percentuais usualmente são mostrados numa representação gráfica do tipo “pizza”.

O argumento acima pode ser generalizado para um número arbitrário V de variáveis, sendo uma escolhida como dependente, e as demais, independentes. Assumindo-se que a última variável \mathbf{w}_V seja aquela a ser expressa como combinação linear das outras, teremos uma expressão análoga à (3.8b),

$$\mathbf{w}_V = -\frac{u_{1V}}{u_{VV}} \mathbf{w}_1 - \frac{u_{2V}}{u_{VV}} \mathbf{w}_2 - \dots - \frac{u_{(V-1)V}}{u_{VV}} \mathbf{w}_{(V-1)} \quad (3.8c),$$

com os percentuais de explicabilidade,

$$p_{vV} = \frac{\alpha_{vV}}{\alpha_{1V} + \alpha_{2V} + \dots + \alpha_{(V-1)V} + \epsilon_V}, \quad v = 1, 2, \dots, (V - 1),$$

onde,

$$\alpha_{vV} = -\frac{u_{vV}}{u_{VV}} \text{ e } \epsilon_V = \frac{l_3}{l_1 + l_2 + \dots + l_V}$$

Com isso, temos V percentuais que relacionam o modelo aos dados: $(V - 1)$ de explicabilidade $(p_{1V}, p_{2V}, \dots, p_{(V-1)V})$ e um de não explicabilidade (ϵ_V) .

A validação do procedimento acima, se deu com dados simulados. Com efeito, com o uso de uma ferramenta como o Matlab, simulando dados a partir de uma perturbação ruidosa de uma relação do mesmo tipo da formula (3.8b), aqui chamado de modelo de prova. Os dados obtidos da simulação foram submetidos à técnica acima constando em que grau se encontram os coeficientes do modelo de prova, a partir da técnica. Foram testados então os limites da técnica, variando-se os níveis de ruídos e as proporções dos coeficientes.

Uma grande vantagem do uso da técnica acima é que, dado que já se tenha a expressão (3.8a) já calculada dos dados disponíveis, pode-se explicitar qualquer das variáveis em função das outras com um custo computacional baixíssimo. Assim, o câmbio de *status* de variáveis entre independentes e dependentes pode ser feito de forma muito eficiente.

Para estudo de caso, o CPC foi o índice escolhido por ser constituído de variáveis significativas quando pensamos no desempenho do estudante (inclui o conceito geral do Enade) e da IES, mas também pelo seu protagonismo em retratar a educação superior para o governo brasileiro. Das variáveis independentes utilizadas, apenas a carga horária do curso e nota no Enem do ingressante no curso foram encontradas em seu estado original diretamente nos dados abertos, sem a necessidade de algum cálculo ou tratamento. A concorrência por curso foi obtida usando o número de candidatos daquele ano no curso dividido pelo número de vagas ofertadas por curso em cada instituição. A população municipal por número de alunos matriculados no curso foi elaborada com auxílio de dados do Instituto Brasileiro de Geografia e Estatística (IBGE), também referentes ao ano de 2011. A despesa gasta no curso foi calculada a partir do total gasto no ano pela IES, dividido pelo seu total de alunos ativos e multiplicado pelo número de alunos matriculados no curso.

O número de cursos utilizados para a pesquisa totalizou 259 de 120 IES diferentes. As IES pesquisadas foram dos Estados do Pará, Pernambuco, Espírito Santo, Rio de Janeiro, São Paulo, Goiás e Rio Grande do Sul, conforme ilustrado na figura 10, selecionados como amostra, procurando equilibrar Estados possuem mais IES com os que possuem menos. Os

curso se restringiram às civil, mecânica, elétrica, eletrônica e química, visando delimitar o escopo da pesquisa, sendo selecionados a partir da área descrita no Enade.

Figura 10. Amostra de Estados usados na pesquisa, IES e cursos

Estado	Número de IES	Número de Cursos
PA	5	10
PE	5	7
GO	7	11
ES	6	13
RJ	19	47
SP	59	121
RS	19	50
Totais:	120	259

Fonte: Próprio autor

Uma vez inseridos esses dados, o Vis-Scholar utiliza o modelo matemático de correlacionamento de variáveis e apresenta os dados em gráficos de fácil compreensão para os usuários comuns enxergarem correlações complexas.

Com o objetivo de facilitar o acesso ao maior número de usuários, independente de dispositivos e plataformas, o Vis-Scholar foi desenvolvido voltado para o ambiente *web*, para ser acessado através da internet, requerendo apenas que o usuário possua um navegador em seu equipamento e esteja conectado com a internet. A ferramenta utiliza tecnologias que tornam sua visualização responsiva, usando as novidades implementadas no HTML 5. A linguagem de programação da regra de negócio é PHP, que possui uma grande quantidade de pacotes com as mais diversas funcionalidades, incluindo matemáticas e estatísticas, além de ser robusta e amplamente usada por uma grande comunidade de desenvolvedores ao redor do mundo. Para armazenamento de dados, optou-se pelo MySQL, Sistema Gerenciador de Banco de Dados (SGBD) na sua edição *open source*, que também é um dos mais populares no mundo (DB-Engines, 2015).

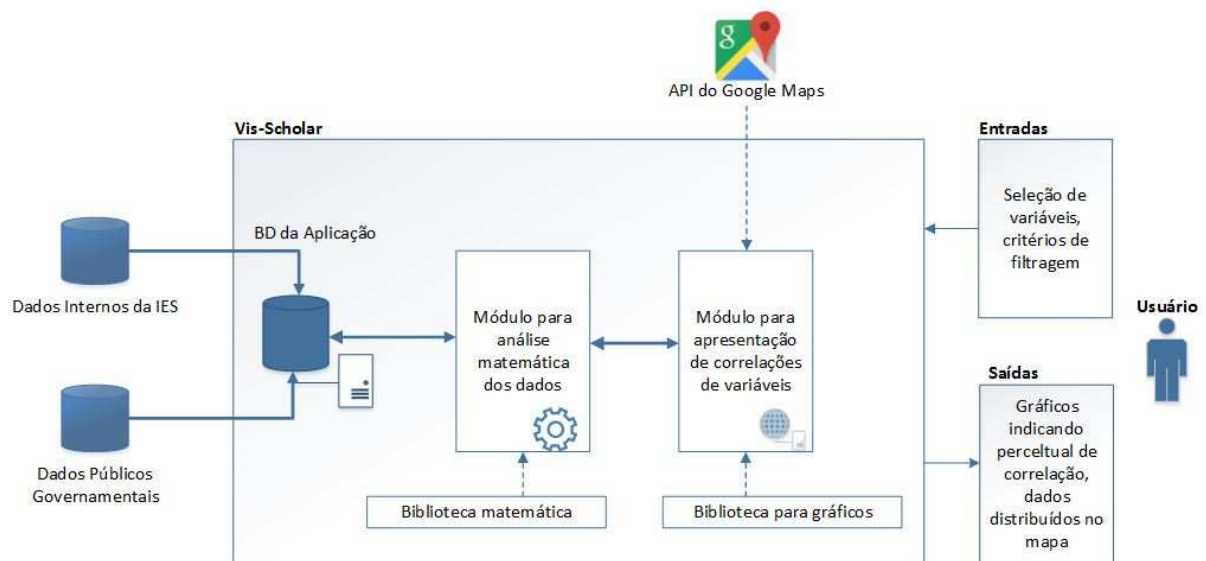
O Vis-Scholar se vale da *Google Maps Application Programming Interface* (API) para apresentar geograficamente dados relacionados com a educação, instituições de ensino e informações de alunos. *Google Maps* é um serviço de mapas online, que utiliza tecnologias interativas, possibilitando a personalização dos recursos que se deseja usar, adaptando seu conteúdo às páginas dos usuários de maneira mais simples (Google, 2015).

A análise e correlação dos dados do Vis-Scholar é feito através do método matemático clássico de Análise de Componentes Principais. Através da ACP, a ferramenta provê ao

usuário a possibilidade de correlacionar dados em tempo real, escolhendo as variáveis a serem processadas pelo sistema.

A figura 11 demonstra o modelo resumido do Protótipo do Vis-Scholar, indicando a origem dos dados, que serão inicialmente inseridos no banco de dados do protótipo manualmente, consultadas a partir de bases públicas e ocasionalmente nas bases internas da IES, tratadas e formatadas antes de serem agregadas à base do protótipo. Também é indicado nesta figura a presença de bibliotecas específicas para implementação dos algoritmos para realização dos cálculos matemáticos, outras para desenho de gráficos em ambiente web e o auxílio da API do *Google Maps*. Naturalmente, outras tecnologias relevantes estão envolvidas na construção da ferramenta, sendo abstraídas desta figura por questões de simplificação do modelo.

Figura 11. Modelo resumido do protótipo do Vis-Scholar



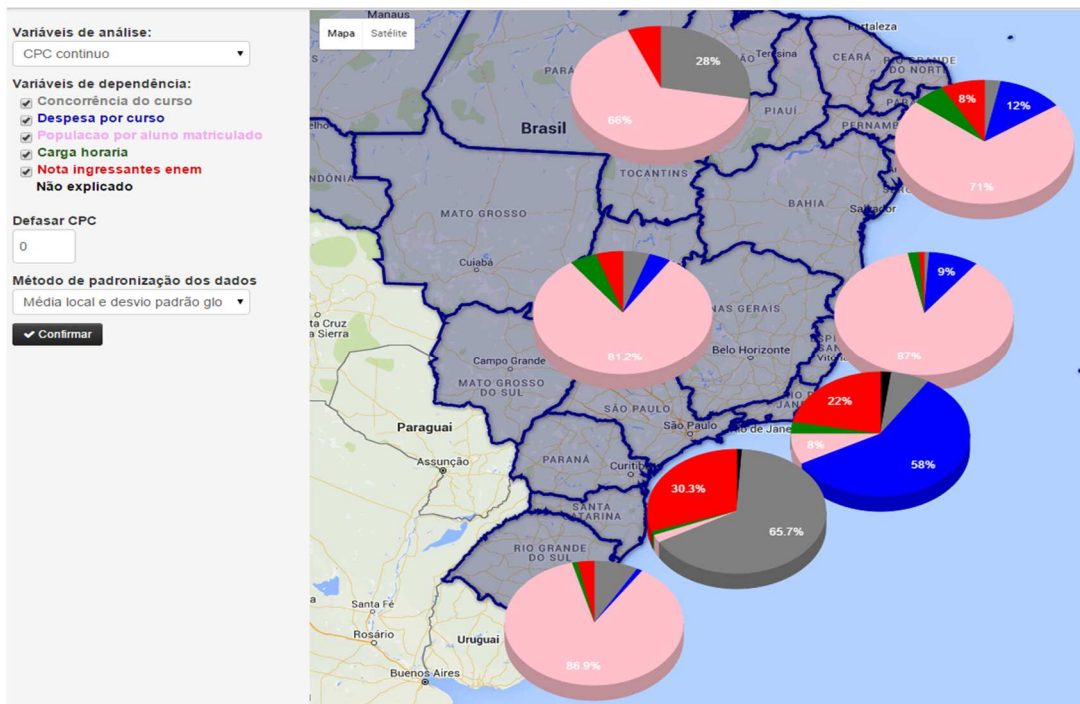
Fonte: Próprio autor

4.3 Resultados

Os resultados gerados pelo sistema são apresentados no formato de gráficos de pizza, representando o percentual de correlação das variáveis independentes com a variável de análise, distintas através de cores diferentes, incluindo um percentual não explicado. Os resultados também são identificados como diretamente correlacionados ou anti-correlacionados através dos sinais [+] e [-] respectivamente. Estes gráficos são posicionados sobre um mapa (georreferenciados), indicando o Estado a que ele se refere. Há também duas

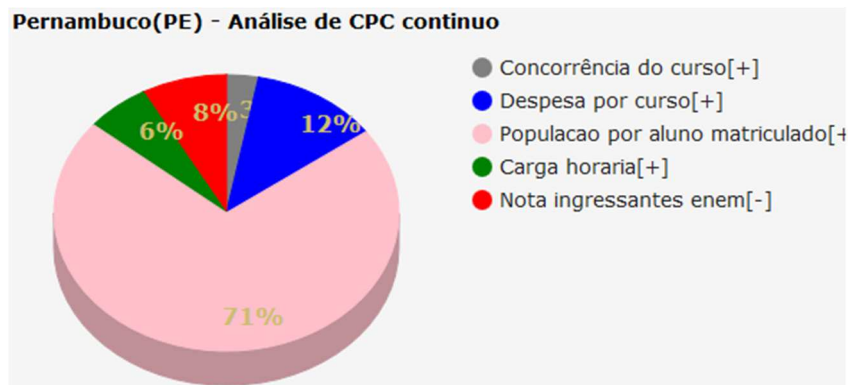
opções de padronização dos dados: usando o desvio padrão de cada Estado respectivamente, ou seja, um desvio padrão local, e outro usando o desvio padrão global, ou seja, levando em conta todo o universo da amostra independentemente do Estado. Os detalhes de cada gráfico resultante podem ser vistos ao se clicar no gráfico sobre o mapa. As figuras de 13 a 16 demonstram esses detalhes.

Figura 12. Detalhe da tela e gráficos gerados sobre o mapa. Nesta análise o CPC (variável de análise) está sendo expresso como função das variáveis de dependência.



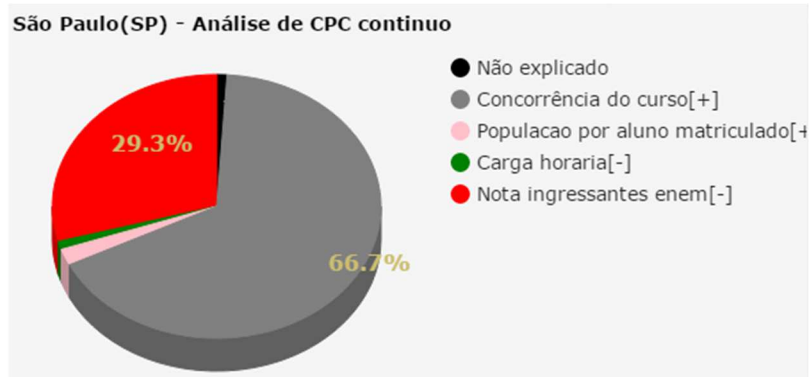
Fonte: Próprio autor

Figura 13. Detalhe do gráfico gerados sobre o mapa do estado de PE



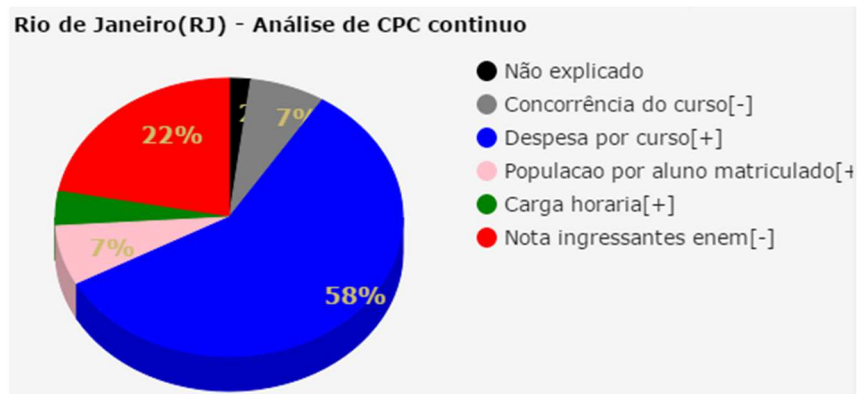
Fonte: Próprio autor

Figura 14. Detalhe do gráfico gerados sobre o mapa do estado de SP



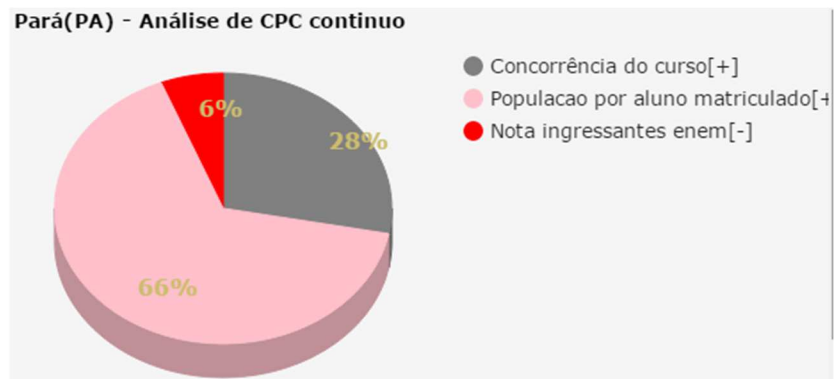
Fonte: Próprio autor

Figura 15. Detalhe do gráfico gerados sobre o mapa do estado do RJ



Fonte: Próprio autor

Figura 16. Detalhe do gráfico gerados sobre o mapa do estado do PA



Fonte: Próprio autor

No tocante ao cerne do trabalho, os resultados podem ser descritos, em termos de fenômenos relevantes observados, como significados práticos das correlações com o índice de desempenho de IES por cursos, o CPC. Tais considerações não são exaustivas, podendo haver muito mais implicações nos resultados do que aqueles aqui descritos:

a) Variável dependente CPC, Variável independente: população municipal por aluno matriculado

Análise: Média local e desvio global

Fenômeno: correlação (+) em todos os casos.

Possível explicação:

As diferenças entre a correlação das variáveis com o CPC, quando analisadas no âmbito global, parecem ser um reflexo do baixo número de estudantes de graduação (vagas ofertadas) e a população geral. Quanto mais População por vaga correlacionada com o índice de desempenho, mais concorrência, conseqüentemente, êxito dos mais preparados para a prova de ingresso. Se os melhores estudantes foram selecionados na prova de ingresso, isso se reflete em melhores resultados nos indicadores de desempenho acadêmico.

b) Variável dependente CPC, Variável independente: Despesa por curso.

Análise: Média local e desvio global

Fenômeno: baixa correlação (+) em todos os casos e nível alto de correlação com o CPC apenas no RJ (58%).

Possível explicação:

Considerando que recursos gastos numa IES resultem no bom desempenho institucional e na qualidade do aluno formado (Szentirmai e Radács, 2013), pressupõe-se serem esses recursos bem administrados, visando atender corretamente o negócio da instituição. Isso implica que, recursos mal geridos, mesmo que em grande volume, não garantem o efeito desejado no sucesso de qualquer empreendimento, inclusive na educação. Dessa forma, quando a variável despesas por curso se destaca, pode significar que o recurso está sendo bem administrado ao ponto de ser um fator de explicação bastante relevante do desempenho de uma IES, o que já é esperado. Mas, quando essa variável é pouco relevante para o desempenho de uma IES, então pode haver mais possibilidades, entre elas, uma gestão frágil dos recursos (gastos desordenados, não planejados, etc.),

c) Variável dependente CPC, Variável independente: concorrência do curso.

Análise: Média local e desvio global

Fenômeno: baixa correlação em GO (4%), PE (3%), RJ (7%), ES (1%) e RS (8%); alto nível de correlação em SP (65%).

Possível explicação:

O baixo nível de correlação pode indicar, que a concorrência da prova de ingresso não tem gerado estudantes com desempenho acadêmico na mesma proporção das vagas ofertadas. Ou seja, há uma lacuna entre o desempenho do estudante de ensino médio que foram classificados na prova de ingresso e do estudante de graduação. Se for esse o caso, a consequência é a pouca influência da concorrência na explicação do CPC no modelo.

d) Variável dependente CPC, Variável independente: carga horária do curso.

Análise: Média local e desvio global

Fenômeno: baixa correlação com a variável dependente.

Possível explicação:

O nível de correlação da carga horária do curso não foi significativa, talvez pela pouca variação do total de horas dos cursos de engenharia praticados pelas IES, que tem seu valor mínimo regulado pelo MEC.

e) Variável dependente CPC, Variável independente: nota do ingressante no Enem.

Análise: Média local e desvio global

Fenômeno: baixa correlação em todos os estados, menos em SP, com nível de correlação em SP (30%).

Possível explicação:

O Enem avalia estudantes de ensino médio e desde 2009 também passou a ser usado como critério para ingresso no ensino superior. Seu resultado pode ser aproveitado totalmente ou como apenas uma parte do processo de seleção dos candidatos à graduação nas IES públicas e ainda na participação em programas de financiamento de despesas de mensalidades em IES particulares. De qualquer modo, mais uma vez, parece haver uma distância entre o desempenho do aluno de ensino médio e o do estudante de graduação. A eficiência do Enem, pelo menos na edição de 2011, em aferir a qualidade do aluno objetivamente pode ser questionada. Uma das mudanças trazidas pelo Enem foi a mobilidade. Mobilidade significa que estudantes de um estado do Brasil podem realizar o Enem visando entrar numa IES de

outro estado, inclusive com a possibilidade de escolher mais de uma delas em mais de um estado. No entanto, segundo Silveira, Barbosa, & Silva, (2015), dados de 2012 mostram que a mobilidade entre estados aconteceu com apenas 13.2% dos candidatos e consistiu principalmente de candidatos oriundos de estados mais ricos (tendo o PIB como referência), com um destaque para São Paulo. Então temos o seguinte cenário: Um grande número de alunos de ensino médio se inscrevem no Enem selecionando algumas IES diferentes em alguns estados. Os candidatos com melhores desempenhos no Enem selecionam algumas IES, inclusive de outros estados também, mas só podem fazer um curso em um lugar. Dão preferência à determinada IES e vão deixando as outras vagas para os colocados abaixo deles. Temos, portanto boas notas no Enem para várias IES, mas os candidatos que as produziram não entram para o corpo discente destas, subindo a expectativa de bons indicadores de desempenho (média do ingressante), mas não as médias esperadas dos acadêmicos matriculados nas avaliações dos cursos de graduação, como o Enade e conseqüentemente nos indicadores, como o CPC. Ainda existe o baixo número de candidatos que conseguiram ingressar no curso, fazendo o Enem e que se identificaram informando seu CPF. Tudo isso combinado pode ter mascarado o resultado dessa variável, trazendo baixos níveis de correlação.

5 CONCLUSÃO

Neste trabalho foi apresentado um método para análise da correlação entre variáveis. Tal método pode ser usado em diferentes cenários, com diferentes temáticas independentemente do conjunto de dados. A temática da educação superior se mostrou oportuna para a utilização do método e os dados abertos da educação superior foram úteis para validação do modelo, que se mostrou capaz de identificar correlações entre as variáveis analisadas. As possibilidades de explicação dos fenômenos observados não são exaustivas e abrem um leque de questões que podem ser investigadas.

A concorrência para ingresso num curso superior, em tese, promove a entrada de estudantes mais preparados para a prova de ingresso, ou seja, os melhores alunos de ensino médio. Normalmente, a qualidade dos estudantes que entram numa IES não é o único fator responsável pelo sucesso da IES e nem sempre as escolas de ensino médio preparam seus alunos para o ambiente acadêmico, que demanda autodisciplina, automotivação, iniciativa e atenção às oportunidades. A avaliação também deve levar em conta questões estruturais, pedagógicas, sociais e docentes.

Fator mais significativo na de definição do CPC é sobre o desempenho do estudante (resultado do Enade). O CPC é o parâmetro pelo qual se verifica a necessidade da visita de um avaliador do MEC na IES, e ainda se estabelece sanções, como a redução do número de vagas ou até mesmo fechamento de cursos (Inep, 2011). Quando a nota do CPC é superior a 3 (faixa do CPC), não há avaliação *in loco*, sob a premissa de que o CPC está alto. Notas inferiores a 3 são consideradas insatisfatórias e tornam a IES candidata a uma visita de avaliadores do MEC. Então resta a pergunta: Seria o CPC realmente um bom critério para avaliar as IES, nestes termos?

A imprensa recentemente divulgou os resultados do último ciclo de avaliação do MEC, realizado em 2014 e disponibilizado apenas ao final de 2015, com manchetes do tipo: “os melhores cursos de engenharia civil, segundo o MEC” (Exame, 2015) e “MEC avalia 756 cursos como ‘insatisfatórios’...” (G1, 2015), estas informações foram geradas com base no CPC dos cursos e são tidas como um reflexo legítimo da realidade das IES.

Quando se estabelece variáveis e utiliza-se ACP para calcular a correlação dessas variáveis com um índice de desempenho, uma referência do quanto estas variáveis explicam o índice é obtida. No caso analisado, o CPC é explicado principalmente pela relação da “população municipal por aluno matriculado no curso”, pela “concorrência do curso” e pelo somatório das “despesas gastas pela IES por curso”. Entre essas variáveis, aquela que é

diretamente gerenciável a curto e médio prazo, especialmente entre as instituições públicas, é a despesa por curso. População por aluno matriculado está com o maior nível de correlação e diretamente correlacionado em todos os casos, o que significa que quanto maior a população por alunos matriculados (vagas ofertadas ao longo do tempo), melhores os resultados do indicador. As sanções impostas pelo MEC às IES que não atingem resultados satisfatórios em seus indicadores, seja suspensão ou cancelamento de cursos, resultam numa redução de vagas ofertadas nos respectivos cursos. Isto faz a variável *população municipal por alunos matriculados* aumentar a médio e longo prazo, uma vez que não se permite criar novas turmas e as antigas vão sendo concluídas. A diminuição nas vagas ofertadas, portanto, provavelmente irá mascarar a análise do resultado do CPC de uma IES a médio e longo prazo, o que aponta uma falha no modo como o governo avalia e sanciona as IES. O modelo mostrou que quanto mais se investe em educação, melhores serão os resultados dos indicadores. Conseqüentemente, resultados insatisfatórios nos índices de desempenho indicam falta de recursos ou má gestão dos recursos. A partir desta perspectiva, pode-se encontrar instituições com melhores resultados nos indicadores utilizados pelo governo, e até posições mais elevadas em *rankings* da imprensa, do que outras que gastam mais recursos. No entanto, esta não é a regra geral.

5.1 Trabalhos futuros

O Vis-Scholar pode ser utilizado com mínimas adaptações para um número muito superior de variáveis e formatos de apresentação dos dados. Em relação a isto, adaptações do Vis-Scholar para usos diferentes se constituem em opções de trabalhos futuros. Além disso, métodos estatísticos alternativos podem ser usados e comparados com o presente modelo, que utiliza ACP. Também se pretende de utilizar o modelo atual com quantidade bem maior de variáveis e apresentar os resultados através de outros tipos de gráficos que mais se adequem ao número de dados nos resultados. Outro tópico de trabalhos futuros seria quando aos métodos de visualização utilizados. Enquanto que no presente trabalho usou-se o método *standard* de fatias de pizza (*Pie Chart*), métodos mais sofisticados podem ser pensados para melhoria das visualizações. Ligado a isso, o método de modelagem abordado, quando aplicado a conjuntos de dados de alta densidade espacial, faz com que o método das pizzas se torne inviável. Portanto, métodos mais apropriados devem ser concebidos. Mais ainda, além das estruturas de covariância entre as variáveis consideradas, surgirá a estrutura de covariâncias espaciais, mostrando a variação espacial da estrutura de correlação das variáveis

envolvidas. Entre os métodos pensados estão alguns baseados em representações por campos vetoriais. Entretanto, infelizmente, não se pode explorar tais métodos de visualização devido à pouca densidade espacial dos dados considerados, e esse tópico ficou para um trabalho futuro.

REFERÊNCIAS

ARABNIA, H. Information Visualization. 1999.

AZEVEDO, P. C. N.; OLIVEIRA, J. E. E. DE; PARREIRAS, F. S. **Visualização de dados abertos vinculados em sistemas de informações geográficas: Uma revisão sistemática da literatura.** Florianópolis, SC: LOD Brasil, 2014

BAKER, R. S. J. D. Data mining for education. **International Encyclopedia of Education**, v. 7, p. 112–118, 2010.

BITTENCOURT, H. R. *et al.* Mudanças nos pesos do CPC e seu impacto nos resultados de avaliação em universidades federais e privadas. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 15, p. 147–166, 2010.

BRASIL. **Lei nº 10.861, de 14 de abril de 2004** Brasília, 2004a. Disponível em: <http://download.inep.gov.br/download//superior/2004/Legislacao/LEI_n10861_14_4_04_SINAES.doc>

_____. **Portaria nº 2.051, de 09 de julho de 2004**, 2004b. Disponível em: <http://download.inep.gov.br/download//superior/2004/Legislacao/Portaria_2051_2004_SINAES_regulamentacao.doc>

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. Readings in Information Visualization: Using Vision to Think. *In: Information Display*. [s.l: s.n.]. v. 1stp. 686.

DB-ENGINES. **DB-Engines**. Disponível em: <<http://db-engines.com/en/ranking/relational+dbms>>. Acesso em: 20 fev. 2015.

DODORICO, B. H. S.; COELLO, J. M. A. **Previsão do desempenho de estudantes usando algoritmos de filtragem colaborativa baseados em fatoração de matrizes.** Anais do XIX Encontro de Iniciação Científica. **Anais...** Campinas, SP: Pontifícia Universidade Católica, 2014

EDUCAÇÃO, M. DA. **Portal MEC**. Disponível em: <http://portal.mec.gov.br/index.php?option=com_content&id=208&Itemid=267>. Acesso em: 20 fev. 2015.

EXAME. **Os melhores cursos de engenharia civil, segundo o MEC**. Disponível em: <<http://exame.abril.com.br/carreira/noticias/os-melhores-cursos-de-engenharia-civil-segundo-o-mec>>. Acesso em: 18 dez. 2015.

G1, P. **MEC avalia 756 cursos como “insatisfatórios”**. Disponível em: <<http://g1.globo.com/educacao/noticia/2015/12/mec-avalia-756-cursos-como-insatisfatorios-e-preve-enade-digital.html>>. Acesso em: 18 dez. 2015.

GOOGLE. **Documentação API Google Maps**. Disponível em: <<https://developers.google.com/maps/documentation/javascript/?hl=pt-br>>. Acesso em: 20 fev. 2015.

GRAMANI, M. C. N.; SCRICH, C. R. Influência Do Desempenho Na Escolha Da the Influence of Educational Performance on the Choice of. **Caderno de Pesquisa - Fundação Carlos Chagas**, v. 42, p. 868–883, dez. 2012.

HIDALGO, C. A.; ALMOSSAWI, A. The Data-Visualization Revolution. **Scientific American**, p. 1–4, 2014.

ILIINSKY, N.; STEELE, J. **Designing Data Visualizations**. Sebastopol: O’Reilly Media, Inc., 2011.

INEP. **Manual dos Indicadores de Qualidade 2011**, 2011.

_____. **Portal Inep**. Disponível em: <<http://portal.inep.gov.br/aceso-a-informacao/suporte-a-pesquisa>>. Acesso em: 20 fev. 2015a.

_____. **Nota Técnica 72/2014 revisada em 2015 - Cálculo do CPC** Brasília, DFBrasil, , 2015b. Disponível em: <http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2013/nota_tecnica_n_72_2014_calculo_cpc_2013.pdf>

JACKSON, J. E. **A User’s Guide to Principal Components**. 2. ed. New York: Wiley Intescience Publications, 1991. v. 43

JOLLIFFE, I. T. Principal Component Analysis. **Journal of the American Statistical Association**, v. 98, p. 487, 2002.

KEIM DANIEL, K. J.; MANSMANN, G. REY E. AND F. Mastering the Information Age Solving Problems with Visual Analytics. *In: Mastering the Information Age Solving Problems with Visual Analytics*. [s.l: s.n.]. p. 57 – 86.

LEMANN. **Fundação Lemann**. Disponível em: <<http://www.fundacaolemann.org.br>>. Acesso em: 20 fev. 2015.

LU, A. *et al.* **Guest editorial: Special issue on visualization and visual analytics** *Tsinghua Science and Technology*, 2013.

MARCHELLI, P. S. O Sistema de Avaliação Externa dos Padrões de Qualidade da Educação Superior no Brasil: considerações sobre os indicadores. **Estudos em Avaliação Educacional**, n. 37, v. 18, 2007.

MARX, V. Data visualization: ambiguity as a fellow traveler. **Nature Methods**, v. 10, n. 7, p. 613–615, 2013.

MASCARENHAS, S.; BARCA, A. Descobrimo Estudantes Talentosos e Superdotados no Ensino Superior Brasileiro Analisando Efeitos dos Hábitos de Estudo Sobre o Rendimento Acadêmico. **Revista AMazônica, LAPESAM/GMPEPPE/UFAM/CNPq/EDUA**, v. X, n. ISSN 1983-3415, p. 280–301, 2012.

MEC, M. DA E. **Instrução Normativa 03/2013 - Alteração de número de vagas de cursos superiores de graduação** Brasília, DFBrasil, , 2013.

MERITT. **QEDu**. Disponível em: <<http://www.qedu.org.br/sobre>>. Acesso em: 20 fev. 2015a.

_____. **Meritt**. Disponível em: <<http://meritt.com.br>>. Acesso em: 20 fev. 2015b.

MIRANDA, G. J. *et al.* Relações entre desempenho acadêmico e acesso aos programas de mestrado em ciências contábeis. **Revista Ambiente Contábil**, v. 6, p. 141–162, 2013.

PRATHAP, G.; SAVITHRI, S. Principal component analysis of end-to-end performance evaluation of leading Indian Higher Educational Institutions. **CSIR National Institute for Interdisciplinary Science and Technology**, p. 14, 2014.

RINGNÉR, M. What is principal component analysis? **Nature biotechnology**, v. 26, p. 303–304, 2008.

ROMERO, C.; VENTURA, S. Educational Data Mining: A Review of the State of the Art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 40, 2010.

SCHRECK, T.; KEIM, D. Visual analysis of social media data. **Computer**, v. 46, p. 68–75, 2013.

SCIMAGOIR. **Scimago Institutions Rankings**. Disponível em: <www.scimagoir.com>. Acesso em: 1 jan. 2015.

SHANGHAIRANKING, C. **Academic Ranking of World Universities**. Disponível em: <www.shanghairanking.com>. Acesso em: 18 nov. 2015.

SILVA, T. E. V. *et al.* **Ferramentas De Processamento e Análise De Informação Aplicadas Ao Ensino De Engenharia** XXXI Simpósio Brasileiro de Telecomunicações - SBRT 2013. **Anais...** Fortaleza, CE: 2013 Disponível em: <<http://www.sbrt.org.br/artigos?id=256>>

SILVEIRA, F. L.; BARBOSA, M. C. B.; SILVA, R. Exame Nacional do Ensino Médio (ENEM): Uma análise crítica. **Revista Brasileira de Ensino de Física**, v. 37, n. 1806-9126, 2015.

SMALL, C. R. *et al.* An investigation into the understanding and skills of first-year electrical engineering students. **IEEE Transactions on Education**, v. 55, n. 1, p. 29–35, 2012.

SOBRINHO, J. D. Avaliação e transformações da educação superior brasileira (1995-2009): do provão ao Sinaes. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 15, n. 1, p. 195–224, 2010.

SYMONDS, Q. **QS Top Universities**. Disponível em: <<http://www.topuniversities.com/university-rankings>>. Acesso em: 18 nov. 2015.

SZENTIRMAI, L.; RADÁCS, L. World university rankings qualify teaching and primarily research. **ICETA 2013 - 11th IEEE International Conference on Emerging eLearning Technologies and Applications, Proceedings**, n. vii, p. 369–374, 2013.

TIMES. **Times Higher Education World University Ranking**. Disponível em: <<https://www.timeshighereducation.com/>>. Acesso em: 18 nov. 2015.

VASCONCELOS, A. I. T.; DINIZ, G.; ANDRADE, T. **Determinantes socioeconômicos do índice de rendimento acadêmico dos discentes de instituições de ensino superior em um município cearense** Anais do V encontro de Pesquisa e Extensão da Faculdade Antônio Feijão. **Anais...**Sobral-CE: 2012