

Aprender-com-Big-Data no ensino de ciências

Learning-with-Big-Data in science teaching

Renato P. dos Santos¹, Isadora Luiz Lemes²,

¹ULBRA/PPGECIM, renatopsantos@ulbra.edu.br

²ULBRA/Licenciatura em Física, isa.ulbra@hotmail.com

Resumo

Vivemos num universo digital que se estima conter hoje cerca de 5 zettabytes (aproximadamente 5×10^{21}) de dados. Essa enorme quantidade de dados, processada em computadores extremamente velozes, com técnicas otimizadas, permite encontrar insights em novos e emergentes tipos de dados e conteúdos para responder a perguntas que foram anteriormente consideradas fora de nosso alcance. Essa é a ideia de Big Data. Os objetivos deste projeto, de caráter construcionista, são investigar como se desenvolve o processo de ensino e aprendizagem de Ciências Exatas, quando mediado pelo computador e por softwares aplicativos públicos e gratuitos de Big Data, tais como o *Google Correlate* e o *Google Trends*, desenvolver estratégias de ensino que tirem o melhor proveito dessas ferramentas para o ensino e aprendizagem de Ciências Exatas e, com isso, concluir sobre a viabilidade do uso Big Data como mediador no Ensino de Ciências Exatas, visando uma preparação de nossos estudantes, tanto para os desafios científicos propostos pelo Big Data ao mundo real quanto sobre uma melhor compreensão das noções de fenômeno, observação, medida, leis físicas, teoria e causalidade, dentre outras. O objetivo deste trabalho é investigar a viabilidade dessa proposta através de uma primeira aplicação da utilização de Big Data no Ensino de Ciências. Em termos metodológicos, a aplicação foi desenvolvida dentro da disciplina História e Epistemologia da Física, do curso de Licenciatura em Física da Ulbra, da qual o pesquisador é titular, realizada durante o primeiro semestre letivo de 2014, contando este semestre com 7 alunos de diversos períodos do curso. Os estudantes foram solicitados a pesquisar correlações de termos de busca relativos ao Ensino de Física na *Google Correlate*, à sua livre escolha. Obtidos os termos de busca que melhor se correlacionam com aqueles, os estudantes aprofundaram suas pesquisas sobre essas correlações em várias outras fontes, buscando possíveis explicações científicas (causações) para elas. Como atividade final de avaliação parcial da disciplina, os resultados de suas pesquisas foram apresentados e discutidos em forma de seminários à turma. Ao final do semestre, foi aplicado um questionário focando sobre contribuições das atividades de Big Data realizadas para seu aprendizado e foram feitas entrevistas estruturadas com os participantes sobre sua percepção da atividade em geral. Esta aplicação, por um lado mostrou que a proposta parece ser viável: no geral, os alunos apreciaram as atividades com as ferramentas e demonstraram compreender a distinção entre correlação e causação. Por outro, no entanto, evidenciou a necessidade de seu aperfeiçoamento em vários aspectos que incluem, principalmente, facilitar a apropriação das ferramentas pelos alunos, desenvolver estratégias para contornar dificuldade dos alunos em encontrar termos que tem relação causal e que, por isso, não apresentam o frequente padrão “temas escolares” em seus gráficos e facilitar a conexão da atividade com o processo de construção do conhecimento físico.

Palavras-chave: Aprender-com-Big-Data. Ensino de ciências. Construcionismo. Mineração de dados. Tecnologia educacional.

Abstract

We live in a digital universe that is estimated today to contain about 5 zettabytes (approximately 5×10^{21}) data. This huge amount of data processed on computers extremely fast with optimized techniques, allows one to find insights in new and emerging types of data and content, to answer questions that were previously considered beyond reach. This is the idea of Big Data. The objectives of this project, of constructionist character, are to investigate how the teaching and learning of Exact Sciences develops, when mediated by computer and free public Big Data application software, such as Google Correlate and Google Trends, to develop teaching strategies that take the best advantage of these tools for the teaching and learning of Exact Sciences, and thereby to conclude on the feasibility of using Big Data as a mediator in the teaching of Exact Sciences, aimed at preparing our students for both the scientific challenges proposed for Big Data to the real world and a better understanding of the notions of phenomenon, observation, measurement, physical laws, and causal theory, among others. The aim of this work is to investigate the feasibility of this proposal through the first application of the use of Big Data in Science Teaching. In methodological terms, the application was developed within the 'History and Epistemology of Physics' course, from the pre-service Physics teacher training program at Ulbra - Lutheran University of Brazil, which has this researcher as the lecturer, and was held during the first school semester 2014, counting with 7 students from different periods of the course. The students were

asked to search freely for search terms correlations related to Physics Teaching in Google Correlate. Once the best correlate search terms were obtained, the students deepened their research on these correlations from several other sources, looking for possible scientific explanations (causation) for them. As a final activity for the course partial evaluation, the results of their research were presented and discussed in seminar fashion to the class. At the end of the semester, a questionnaire focusing on the contributions of Big Data activities for their learning and applied structured with participants about their perceptions of overall activity interviews were conducted. In this paper, the results of the first application of our didactic proposal were presented. It is aimed at preparing students to the scientific challenges posed by Big Data to the real world as well on a better understanding of the concepts of phenomenon, observation, measurement, physical laws, theory, and causality, among others. On the one hand, this application showed that the proposal seems feasible: students in general enjoyed the activities with the tools and demonstrated to understand the distinction between correlation and causation. On the other hand, however, it highlighted the need for its improvement in several aspects which include: to facilitate the appropriation of tools by students, to develop strategies to circumvent the students' difficulty of finding terms that have a causal relationship, and to easy the connection of the activity with the process of construction of physical knowledge.

Keywords: Learning-with-Big-Data. Science education. Constructionism. Data mining. Educational technology.

INTRODUÇÃO

O universo digital em que vivemos estima-se conter hoje cerca de 5 zettabytes (aproximadamente 5×10^{21}) de dados, correspondentes a postagens e curtidas nas redes sociais, imagens e vídeos de telefones celulares enviados para o YouTube, filmes digitais de alta definição, movimentações bancárias, imagens de câmaras de segurança, colisões subatômicas registradas pelo LHC do CERN, chamadas telefônicas, mensagens SMS, etc. (GANTZ; REINSEL, 2012). Vivemos na onda do Big Data.

Em nosso entendimento, o parâmetro central que distingue Big Data de processos anteriores de mineração de dados não é o tão enfatizado tamanho das bases de dados, mas o *crowdledge*, que definimos¹ como o conhecimento *emergente* e, portanto, inesperado, a partir de análise Big Data de pegadas digitais espontâneas de indivíduos, deixadas em buscas no Google, posts no Facebook, Twitter, etc. Aqui, utilizamos o conceito de (fracamente) emergente, no sentido de uma propriedade de um sistema complexo que é nova, inesperada e distinta das propriedades das partes desse sistema, embora causada por e dedutível delas (CHALMERS, 2006). Por isso, dentre as muitas definições de Big Data encontradas, preferimos a seguinte, por a julgarmos mais esclarecedora para os propósitos deste trabalho:

Big data é mais do que simplesmente uma questão de tamanho, é uma oportunidade de encontrar insights em novos e **emergentes** tipos de dados e conteúdos, para tornar seu negócio mais ágil e para responder a perguntas que foram anteriormente consideradas fora de seu alcance (IBM, 2011, grifo nosso).

¹ DOS SANTOS, R. P. e DALLA VECCHIA, R. On the Philosophy of Big Data: Reality, Emergence, Crowdledge and Modeling. **em preparação**, 2014.

Por outro lado, as dificuldades de aprendizado dos alunos em Ciências Exatas em geral, e em Física e Matemática em particular, vêm sendo discutidas na literatura científica desde os anos 70 e, apesar da quantidade de informação acumulada, não há consenso sobre elas nem sobre como solucioná-las. Em parte, essas dificuldades decorrem de os conteúdos serem abordados através de um empirismo ingênuo com afirmações do tipo “a experiência mostra que certa lei é válida” (ROBILOTTA, 1988), ou num racionalismo simplificador que a reduz, de estruturas teóricas autocontidas e autoexplicativas (ROBILOTTA, 1988), a “um conjunto desconexo de fatos expressáveis em fórmulas” (HOFER; PINTRICH, 1997) e procedimentos algorítmicos.

Para tentar minimizar essas dificuldades de aprendizado, pretendemos, com a presente proposta didática (DOS SANTOS, 2014b) que tem embasamento no Construcionismo de Papert (1985), propiciar ao estudante de Ciências Exatas uma reflexão, embasada na prática, sobre a construção do conhecimento científico, tendo o computador e os aplicativos públicos e gratuitos do Big Data como mediadores no ensino de ciências, em vez de *ferramentas de ensino* (PAPERT, 1985, p. 17–18).

Dentre propostas anteriores do uso de aplicativos de Big Data no ensino, podemos citar (BARAM-TSABARI; SEGEV, 2009a, 2009b, 2013; SEGEV; BARAM-TSABARI, 2012) que propõem utilizar os softwares aplicativos² *Google Trends*, *Google Zeitgeist* e *Google Insights for Search* para a pesquisa e a discussão sobre a compreensão da ciência pelo público e sobre a distinção entre ciência e pseudociência em sala de aula. Bülbül (2009) e Yin et al. (2013) propõem determinar e discutir tendências em Física e em Educação através de pesquisas de palavras chave através do *Google Trends*.

Nossa proposta é distinta e baseia-se no aplicativo *Google Correlate*, mais moderna, discutida a seguir.

OS APLICATIVOS GOOGLE TRENDS E GOOGLE CORRELATE

O aplicativo de análise *Google Trends*³, disponibilizado ao público pela empresa Google, permite pesquisar a evolução ao longo do tempo das buscas no Google por termos específicos, comparar a evolução de até cinco termos simultaneamente, ou, simplesmente, descobrir quais são os termos mais pesquisados numa certa região do mundo e num certo

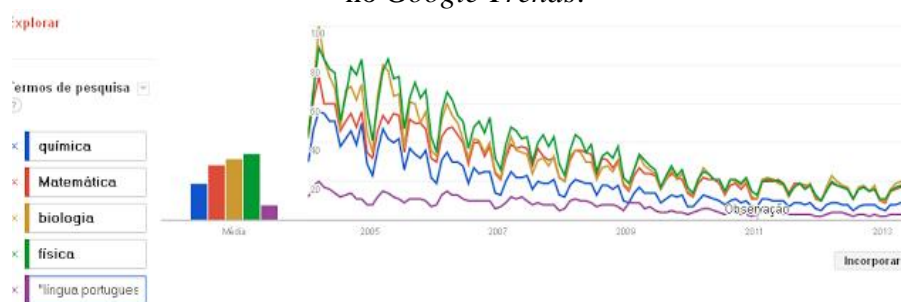
² Software aplicativo (aplicativo ou aplicação) é um programa de computador que, diferentemente de outros tipos de software, tais como sistemas operacionais, jogos e outros, tem o objetivo definido de ajudar o seu usuário a realizar uma tarefa específica, em geral ligada a processamento de dados.

³ <http://www.google.com/trends/>

intervalo de tempo. Milhares de trabalhos científicos já foram realizados utilizando o *Google Trends*, em várias áreas do conhecimento, incluindo Saúde Pública, Medicina, Economia, Educação, Política, dentre outras.

Observamos uma dificuldade com relação a este aplicativo de que frequentemente se observa uma tendência de queda em gráficos de tendência (Figura 1), o que, geralmente, leva à interpretação de um decréscimo no interesse por esses termos pelo público usuário do Google.

Figura 1 - Gráfico comparativo das frequências de buscas de vários termos no *Google Trends*.



Fonte: *Google Trends* (<https://www.google.com.br/trends/explore>).

No entanto, essas tendências de queda nos gráficos já foram interpretadas por Baram-Tsabari e Segev (2009a) como devido ao aumento geral de acesso e uso da Internet, o que leva a uma crescente diversificação nas pesquisas na web via Google e faz com que o volume de pesquisa para termos específicos diminua relativamente ao volume total.

Ao contrário do *Google Trends*, no *Google Correlate*⁴ introduz-se um termo de busca, uma série de dados temporais ou regionais e se obtém uma lista das consultas no Google cujas frequências seguem padrões que melhor se correlacionam com os dados, segundo o coeficiente de determinação⁵ R^2 (MOHEBBI et al., 2011, p. 2). Vários trabalhos baseados no *Google Correlate* podem ser encontrados na literatura científica em vários domínios de conhecimento, incluindo Saúde Pública, Economia, Sociologia e Meteorologia, dentre outros. Em (DOS SANTOS, 2014a), já discutimos alguns recursos do *Google Correlate* e apresentamos vários exemplos de possíveis utilizações deste aplicativo no aprendizado de Física, Química, Biologia e Matemática.

Em seguida, discutiremos resumidamente o suporte teórico para nossa proposta. Para mais detalhes, ver (DOS SANTOS, 2014b).

⁴ <http://www.google.com/trends/correlate>

⁵ O coeficiente de determinação R^2 é uma medida do ajustamento de um modelo estatístico linear aos valores observados e varia entre 0 e 1, sendo que, quanto maior o valor, melhor o modelo se ajusta aos valores.

REFERENCIAL TEÓRICO

Ao contrário da pedagogia tradicional, para Piaget, as crianças não são recipientes vazios a serem preenchidos com o conhecimento, mas construtores ativos de conhecimento – pequenos cientistas que estão constantemente criando e testando suas próprias teorias do mundo (PAPERT, 1999).

Papert, por outro lado, distinguia seu Construcionismo do Construtivismo piagetiano ao acrescentar que essa “construção de estruturas do conhecimento”, que ocorre ‘na cabeça’, muitas vezes acontece de forma especialmente proveitosa quando é apoiada pela construção, de um tipo mais público, ‘no mundo’, de forma que o produto possa ser mostrado, discutido, examinado, sondado e admirado (PAPERT, 1993, p. 142–143).

Papert denunciava que, nos anos 80, em muitas escolas, a frase ‘instrução ajudada por computador’ significava fazer com que o computador ensinasse a criança (PAPERT, 1985, p. 17). Ele, ao contrário, propunha uma aprendizagem da Matemática numa ‘*Matelândia*⁶, num contexto que está para a Matemática como viver na França está para aprender francês (1985, p. 19). Para ele, “a Matelândia alicerçada em computadores estende o tipo de aprendizagem natural, piagetiano, que é responsável pela aquisição da língua materna, ao aprendizado da matemática” (1985, p. 69).

Segundo Papert, seu objetivo foi sempre desenvolver objetos de que as crianças pudessem “se apropriar à sua própria maneira”, objetos em que houvesse uma intersecção de presença cultural, conhecimento implícito, e a possibilidade de identificação pessoal, “objetos-de-pensar-com”, enfim (PAPERT, 1985, p. 26). Aqui, esse autor antecipa a ideia de o aluno *pensar-com*, *aprender-com* e *fazer-com* o computador de Rosa (2008), que, com esta proposta, avançamos para *pensar-com* e *aprender-com-Big-Data* (DOS SANTOS, 2014b).

No entanto, com o advento das novas tecnologias de interface e a ampliação dos domínios sobre os quais os micromundos foram projetados, a importância das linguagens de programação como a principal semântica para o Construcionismo desapareceu em favor da manipulação dinâmica de representações e das representações gráficas e icônicas de maior qualidade oferecidas pela multimídia e possibilitadas por um conjunto de sistemas de autoria⁷ (HEALY; KYNIGOS, 2009). Para Kynigos (2012), a atual sociedade digital está cheia de

⁶ Conforme nota do tradutor “a palavra Matelândia foi criada pelo autor usando o radical *Mate*, que significa guardar” (PAPERT, 1985, p. 19).

⁷ Uma ferramenta de autoria é um programa de computador usado para a produção de arquivos digitais, geralmente incluindo texto escrito, imagem, som e vídeo.

objetos para serem manipulados e ferramentas para atividades coletivas e comunicação matemática e, assim, o Construcionismo ainda é relevante.

Com este suporte teórico, apresentamos, na sequência, resumidamente a proposta das atividades discutidas neste trabalho. Para mais detalhes, ver (DOS SANTOS, 2014b).

NOSSA PROPOSTA

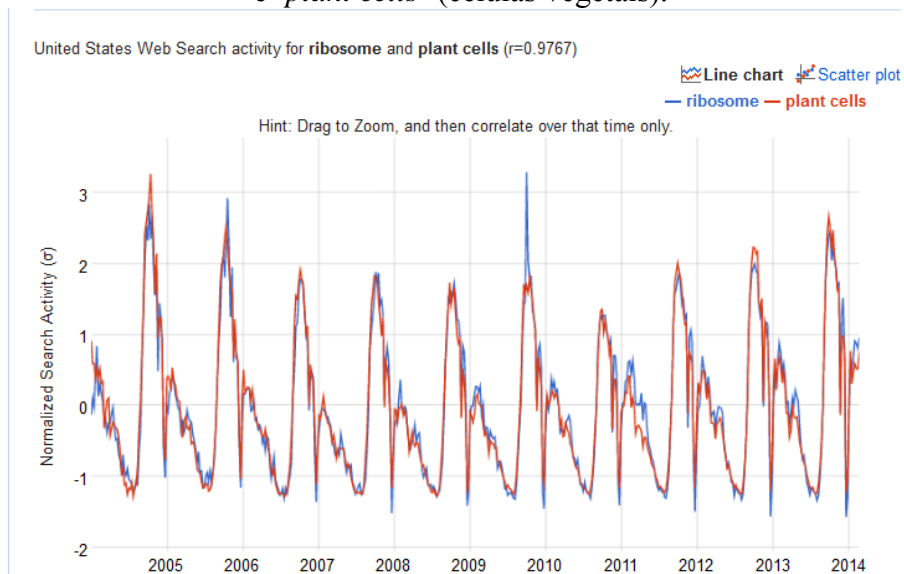
Entendemos que Big Data tem grandes potencialidades frente aos processos de ensino e aprendizagem de Ciências, abrangendo novos modos de conduzir o processo educacional que se mostram em maior consonância com as necessidades atuais. Do que foi discutido acima, mais do que apenas uma nova tecnologia educacional, vemos o *aprender-com-Big-Data* (DOS SANTOS, 2014b) como uma nova maneira de construir o conhecimento científico, de aprender a fazer Ciência, de aprender a pensar como um cientista.

Para demonstrar melhor o aplicativo *Google Correlate* e seu potencial para o Ensino de Ciências, em (DOS SANTOS, 2014b) discutimos um exemplo, no qual introduzimos no *Google Correlate* os dados da variação semanal da atividade solar, medida pela variação do número de manchas solares de 5 jan. 2003 a 31 mar. 2013, e obtivemos uma boa correlação para vários termos de busca no Google, sendo que o que melhor correlacionou ($R^2=0,7523$) foi *'wireless hotspot'*, que significa locais em que a tecnologia *Wi-Fi* está disponível.

Os estatísticos costumam alertar que “correlação não implica em causação” (FIELD, 2003, p. 10). Estabelecer uma correlação entre dois eventos *A* e *B* não implica estabelecer uma relação causal entre *A* e *B*, até porque não sabemos qual delas causou a outra, isto é, qual, *A* ou *B*, é a *variável de resposta* e qual é a *explicativa*. Assim, é possível que haja uma *variável oculta C* que influencie as variáveis *A* e *B* e que ponha de parte a correlação (ARNTZENIUS, 2010).

De fato, muitas correlações encontradas com o *Google Correlate*, exibem um gráfico característico (Figura 2), com um padrão periódico semelhante ao que Mohebbi et al. (2011, p. 4) consideraram que "reflete o ano letivo nos Estados Unidos e sugere que as consultas estão sendo conduzidas por temas escolares", um *terceiro fator* que age como uma causa comum para ambos os termos e não por correlações causais entre estes. Seguindo Mohebbi et al. (2011, p. 4), rotulamos essas correlações não causais, com esse padrão característico, como "temas escolares" (*school topics*). Daí decorre a dificuldade que, como veremos adiante, vários alunos enfrentaram para encontrar um termo de busca cujas correlações tivessem explicação causal (causações) e não o padrão “temas escolares”.

Figura 2 - Comparação entre a frequência de pesquisa dos termos ‘*ribosome*’ (ribossomos) e ‘*plant cells*’ (células vegetais).



Fonte: *Google Correlate* (<http://www.google.com/trends/correlate>).

Inicialmente pode-se não ver relação causal entre as buscas por *wireless hotspots* no Google e as variações do número de manchas solares ou com a atividade solar. No entanto, pela nossa proposta (DOS SANTOS, 2014b), este é um momento frutífero de aprendizado para o estudante de Ciências: observado um novo e inesperado fenômeno emergente (a correlação), buscar uma explicação científica (causação) para ele. Conforme discutimos em (DOS SANTOS, 2014b), sabe-se que a atividade solar tem vários efeitos sobre nossa vida diária, dentre os quais variações nas condições de radiocomunicação, distúrbios e tempestades geomagnéticas, mudanças nas condições climáticas e auroras polares. Desta forma, um mecanismo causal possível relacionando buscas por *hotspots* de *wi-fi* e as variações da atividade solar seria o de que máximos nessa atividade prejudicam o alcance dos *hotspots* e, por isso, usuários acostumados a utilizar determinados *hotspots* se veriam obrigados a procurar novos *hotspots* para se conectarem.

Para confirmar ou não essa hipótese, seria necessário que os estudantes aprofundassem suas pesquisas em várias outras fontes, o que seria extremamente produtivo em termos de aprendizado de Ciências, especialmente em compreensão das noções de fenômeno, observação, medida, leis físicas, teoria, dentre outras.

Assim, interpretamos os ambientes proporcionados pelo *Google Trends* e pelo *Google Correlate* como micromundos: os alunos podem apropriar-se dos recursos desses ambientes e utilizá-los para relacionar as correlações dos termos com suas noções de conceitos físicos e

variáveis, os gráficos correspondentes com sua compreensão de leis científicas e as explicações buscadas para as correlações (causações) com as teorias científicas, fazer alguma coisa nova com eles, brincar com eles e construir com eles; os resultados de suas pesquisas, em forma de gráficos e explicações, podem ser mostrados aos colegas, discutido, examinado, sondado e até admirado como um resultado interessante e relevante. Em suma, um ambiente construcionista onde os estudantes podem *pensar-com* e *aprender-com-Big-Data* (DOS SANTOS, 2014b).

Por outro lado, entendemos a correlação discutida acima como um exemplo de uma emergência fraca (cf. CHALMERS, 2006) em Big Data: uma propriedade nova, inesperada e não óbvia (a correlação entre ‘número de manchas solares’ e variações na frequência de busca do termo ‘*wireless hotspot*’) de um sistema complexo (comportamentos de busca dos usuários do Google) que é distinta das propriedades das diferentes partes do sistema (usuários individuais do Google, buscas únicas no Google), embora sejam dedutíveis e causadas por elas, como argumentado acima.

Por outro lado, esta proposta tem viés epistemológico, pois, mais do que meramente em capacitar os estudantes em infraestruturas computacionais ou treiná-los análise preditiva, o objetivo das atividades pedagógicas a serem realizadas durante este projeto é propiciar aos nossos estudantes, futuros profissionais de Ciências, uma preparação, tanto em termos técnicos como em éticos, para os desafios científicos propostos pelo Big Data ao mundo real no qual vão exercer suas profissões, além de uma melhor compreensão, embasada na prática do Big Data, sobre a construção do conhecimento físico, especialmente das noções de fenômeno, observação, medida, leis físicas, teoria, causalidade, dentre outras.

Apresentadas, em termos gerais, a proposta, seu embasamento teórico e as ferramentas a serem utilizadas, passemos, agora, à descrição mais específica da metodologia utilizada nas atividades a que se refere este trabalho.

METODOLOGIA

A proposta foi desenvolvida dentro da disciplina História e Epistemologia da Física, do curso de Licenciatura em Física da Ulbra, da qual o pesquisador é titular, realizada durante o primeiro semestre letivo de 2014, contando este semestre com 7 alunos de diversos períodos do curso. Essas atividades foram apoiadas pela coautora deste trabalho, aluna bolsista de iniciação científica desse curso, encarregada de ajudar os alunos em suas dificuldades e

registrar reações, comentários e dúvidas dos alunos para análise posterior, a qual se apresenta na sequência.

Inicialmente, logo após a aula em que se discutiram os conceitos de fenômeno, observação, medida, leis físicas, teoria, causalidade, dentre outras, foi apresentado o conceito de Big Data, bem como uma breve visão epistemológica sobre seu potencial no ensino de Física, seus principais softwares aplicativos públicos e sua utilização, bem como a proposta de atividades de sua aplicação na disciplina.

Os estudantes foram solicitados a pesquisar correlações de termos de busca relativos ao Ensino de Física no *Google Correlate*, à sua livre escolha, tal como feito no exemplo acima. Obtidos os termos de busca que melhor se correlacionam com aqueles, os estudantes aprofundaram suas pesquisas sobre essas correlações em várias outras fontes, buscando possíveis explicações científicas (causações) para elas. Como atividade final de avaliação parcial da disciplina, eles apresentaram os resultados de suas pesquisas em forma de seminários à turma.

Ao final do semestre, foram feitas entrevistas estruturadas com os participantes sobre sua percepção da atividade em geral. A proposta (DOS SANTOS, 2014b) previa, para uma avaliação do desenvolvimento da compreensão dos conceitos físicos trabalhados, seriam construídos pelos alunos, no início e no fim do semestre, mapas conceituais (NOVAK; GOWIN, 1984) referentes aos vários conceitos epistemológicos tratados nesta disciplina, mapas esses que seriam analisados através de um procedimento, incluindo testes de correlação.

No entanto, essa construção de mapas conceituais não foi possível de ser realizada devido a várias mudanças de datas de aulas por conta da Copa 2014; em vez dela, foi aplicado um questionário, uma parte focando sobre contribuições das atividades de Big Data realizadas para seu aprendizado e outra parte sobre os tópicos focados, tais como o que é Ciência, Método científico, Racionalismo e Empirismo, dentre outros. Na próxima seção, incluiremos uma análise das respostas à primeira parte deste questionário.

DADOS E ANÁLISE

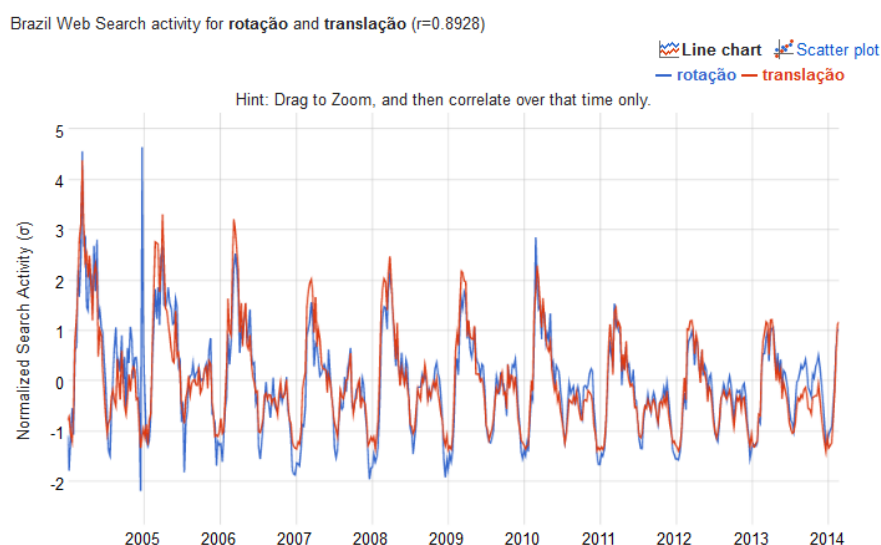
Vamos, aqui, analisar resumidamente as pesquisas dos alunos (identificados por A1, A2, ... até A7) nos aplicativos de Big Data *Google Trends* e *Google Correlate*, juntamente com as respostas à terceira pergunta das entrevistas estruturadas. As versões completas dos trabalhos dos alunos A2, A4 e A6, devido à sua maior extensão, serão objeto de outra

publicação e serão apresentados aqui apenas de forma resumida. Em seguida, analisaremos os dados coletados com o questionário juntamente com as demais perguntas das entrevistas.

Pesquisa do aluno A3

O aluno A3 pesquisou ‘rotação’, obtendo ‘translação’ como o resultado com maior correlação com o termo de busca, com um coeficiente de determinação R^2 de 0,8928 (Figura 3).

Figura 3 - Comparação entre as frequências de pesquisa dos termos ‘rotação’ e ‘translação’ no Google.



Fonte: *Google Correlate* (<http://www.google.com/trends/correlate>) via aluno A3.

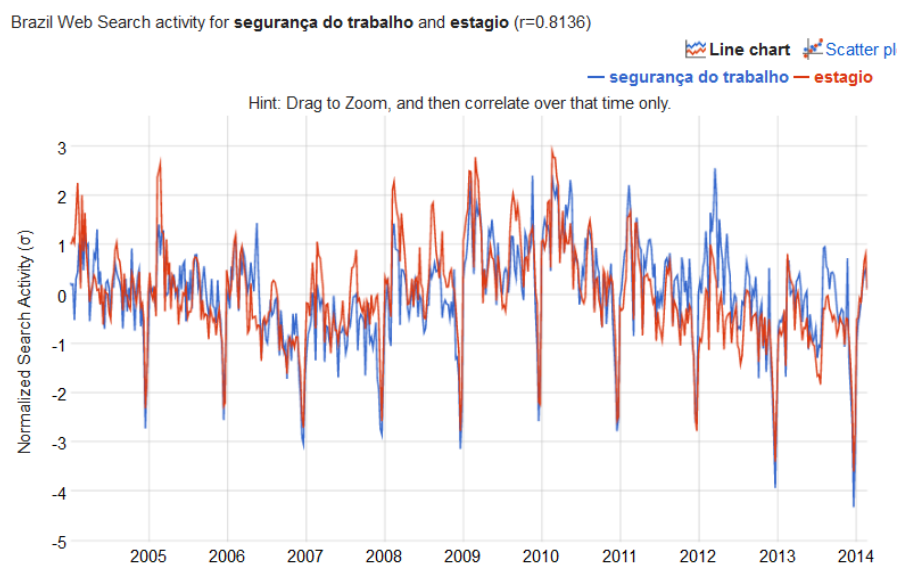
Observa-se, todavia, no gráfico da Figura 3, o padrão "temas escolares", discutido anteriormente, isto é, essas consultas estão sendo conduzidas por temas escolares e não por correlações causais entre esses termos. O aluno A3 declarou não ter encontrado um termo de busca que não apresentasse esse padrão e também não aventou qualquer hipótese para essa correlação que não fosse a mera conexão por temas escolares. Apesar disso, respondeu que “*com certeza*” os aplicativos auxiliam na construção do conhecimento, acrescentando que “*agora sei que existe esta ferramenta⁸ e para que ela serve. Além do mais conseguimos medir, ter ideia de como se comporta um determinado estudo e sua relação em lugares diferentes*”. Isso sugere que, a partir da apropriação desses meios, conseguirá utilizá-los em outras situações de medida e análise de variáveis.

⁸ Naturalmente, o aluno, não familiarizado com a teoria que embasa esta pesquisa, refere-se ao *Google Trends* e ao *Google Correlate* como ‘ferramentas’, algo que *auxilia* a aprendizagem, em vez de um mediador, um recurso que participa da aprendizagem, algo que ele *pensa-com*.

Pesquisa do aluno A5

O aluno A5 pesquisou na *Google Correlate* vários termos que apareceram no conteúdo das aulas, tais como ‘Positivismo’, ‘Racionalismo’ e ‘sistema solar’, dentre muitos outros, todos apresentando o padrão "temas escolares", discutido acima. Finalmente, o aluno A5 decidiu-se pelo termo ‘segurança do trabalho’, obtendo ‘estágio’ como o termo com maior correlação com o termo de busca, com um coeficiente de determinação R^2 de 0,8136 (Figura 4).

Figura 4 - Comparação entre as frequências de pesquisa dos termos ‘segurança do trabalho’ e ‘estágio’ no Google.



Fonte: *Google Correlate* (<http://www.google.com/trends/correlate>) via aluno A5.

Para tentar entender essa correlação, o aluno A5 introduziu no *Google Trends* os termos correlacionados ‘segurança do trabalho’ e ‘estágio’, obtendo o gráfico da Figura 5.

Comparando o gráfico do *Google Correlate* (Figura 4) com o da *Google Trends* (Figura 5), o aluno A5 concluiu que

apesar de ter uma correlação, trata-se de assuntos muito diferentes, sendo apresentado muitos tópicos de notícias não relacionadas sobre segurança no trabalho e poucas notícias no termo estágio.

Analisando a pesquisa do aluno A5, ele, também, apesar de pesquisar muitos termos de busca, só encontrou correlações apresentando o padrão “temas escolares” em seus gráficos, incluindo aquela entre os termos ‘segurança do trabalho’ e ‘estágio’, apresentada acima.

Embora tenha comparado os gráficos do *Google Correlate* com o *Google Trends*, nada em seus comentários indicou ter aventado alguma hipótese para essa correlação e de ter feito essa comparação de gráficos com o objetivo de testar alguma hipótese; aparentemente, esperava que o gráfico do *Google Trends* lhe desse alguma inspiração para formular uma hipótese. Da mesma forma, também não pesquisou fontes externas, conformando-se com os resultados até então encontrados, alegando não ter conseguido encontrar termos que não apresentassem o padrão “temas escolares” em seus gráficos, mesmo fugindo do contexto ‘Ensino de Física’ designado para a tarefa. Consideramos que, nesta primeira aplicação, teríamos que tolerar essa fuga ao contexto até que tivéssemos meios de garantir aos alunos de encontrar termos que não apresentassem o padrão “temas escolares”.

Figura 5 - Gráfico comparativo das frequências de buscas dos termos 'segurança do trabalho' e 'estagio' no Google.



Fonte: *Google Trends* (<https://www.google.com.br/trends/explore>) via aluno A5.

Pesquisa do grupo dos alunos A4 e A6

O grupo formado pelos alunos A4 e A6 obteve um conjunto de dados⁹ referentes à variação quinzenal do campo magnético terrestre, coletados de 1 jan. 2010 a 15 dez. 2011 na Califórnia, e o introduziram no *Google Correlate*, obtendo ‘lyrics’ (letras de música, em inglês) como o termo com maior correlação, com um coeficiente de determinação R^2 de 0,9876, já que, pela fonte dos dados, a correlação foi realizada sobre o universo de utilizadores do Google nos EUA.

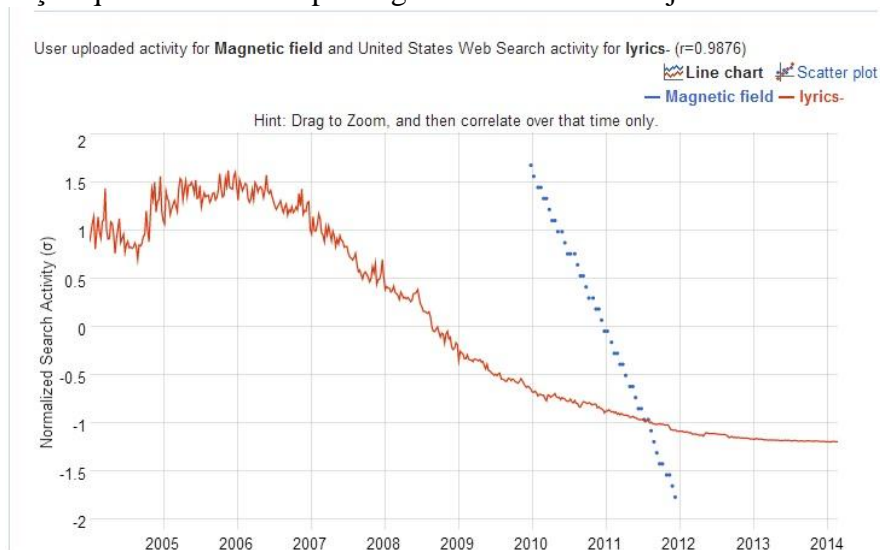
O grupo analisou o gráfico obtido (Figura 6) e considerou que

⁹ <http://www.noaa.gov/>

neste caso, poucos dados foram utilizados, por este motivo o gráfico não apresentou coerência, logo não podemos estabelecer uma correlação causal.

De fato, o conjunto de dados restringia-se a um período muito curto, de apenas dois anos, para uma grandeza de variação tão lenta quanto o campo magnético terrestre.

Figura 6 - Comparação entre a frequência de pesquisa do termo ‘lyrics’ no Google e a variação quinzenal do campo magnético terrestre de 1 jan. 2010 a 15 dez. 2011.



Fonte: *Google Correlate* (<http://www.google.com/trends/correlate>) via alunos A4 e A6.

Quanto à pesquisa dos alunos A4 e A6, estes foram os únicos que tentaram a opção de trabalhar com correlações entre dados temporais, em vez de entre termos. Infelizmente, o gráfico correspondente, no entanto, resultou pouco útil, pois, como os alunos A4 e A6 compreenderam posteriormente, o conjunto de dados era muito pequeno. Apesar disso, o aluno A4 respondeu à pergunta da entrevista sobre as ferramentas auxiliarem na construção do conhecimento de forma bastante afirmativa, dizendo que “*com elas podemos, além de perceber fenômenos interessantes, desenvolver a capacidade de análise, evoluindo nossa percepção consideravelmente*”, o que sugere que aprendeu algo com a atividade, mesmo que seu resultado não fosse o desejado.

Pesquisa do aluno A2

Após várias pesquisas de termos de busca que realizou no *Google Correlate*, todos apresentando o padrão “temas escolares” visto acima, o aluno A2 também se viu obrigado a

fugir do contexto 'Ensino de Física' designado para a tarefa, selecionando o termo 'cooperativa' para seu trabalho, por relação com interesses pessoais, como declarou.

Figura 7 - Gráfico da frequência de buscas do termo 'cooperativa' no Brasil no Google.



Fonte: *Google Trends* (<https://www.google.com.br/trends/explore>) via aluno A2.

O aluno A2 pesquisou, então, a evolução da frequência de buscas do termo 'cooperativa' no *Google Trends*, obtendo o gráfico da Figura 7.

Numa primeira análise desse gráfico, o aluno A2 observou que a frequência de busca no Google do termo 'cooperativa' teve “uma curva praticamente decrescente” no período, levando-o à hipótese causal de isso se dever a um decréscimo do número de cooperativas no Brasil.

Desconhecendo o fato visto acima de que essa tendência de queda nos gráficos, na verdade, decorre de uma característica do algoritmo dessas ferramentas, para testar sua hipótese, o aluno A2 buscou dados¹⁰ sobre o crescimento do cooperativismo no Brasil na forma do gráfico da Figura 8.

Por comparação do gráfico da Figura 7 com o da Figura 8, o aluno A2 concluiu que sua hipótese não se confirmou, raciocinando que

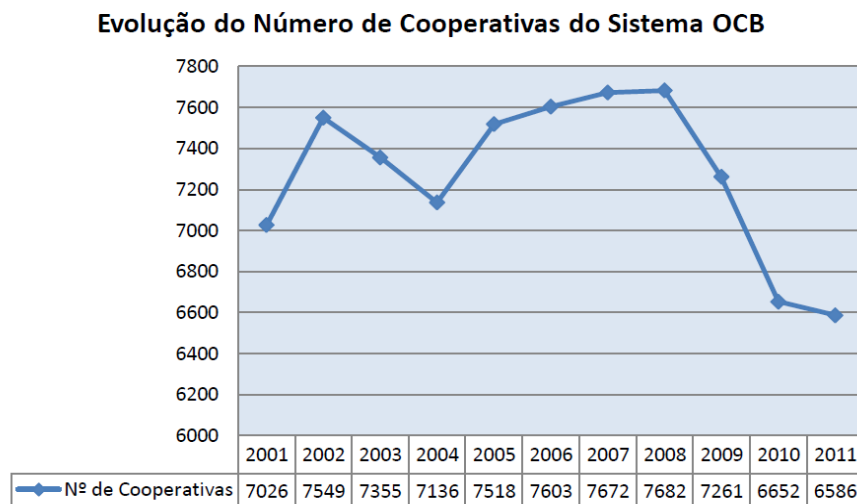
“não podemos definir uma possível causa da queda do número de pesquisas do termo ‘Cooperativa’ no Google, pois no gráfico da OCB temos um aumento do surgimento de novas cooperativas no Brasil a partir de meados de 2004 até meados de 2008”.

¹⁰ PARADELO, B.; FARDINI, G; VILELA, S. M.; ALVES, M. C. Relatório da Gerência de Monitoramento - "Panorama do Cooperativismo Brasileiro - Ano 2011". Brasília: SESCOOP - Serviço Nacional de Aprendizagem do Cooperativismo. mar. 2012.

e concluiu que

“os gráficos se assemelham apenas em alguns períodos em que não podemos afirmar se existe uma correlação causal.”

Figura 8 - Evolução do número de cooperativas do sistema OCB¹¹ de 2001 a 2011.



Fonte: PARADELO, B.; FARDINI, G; VILELA, S. M.; ALVES, M. C. (2012) via aluno A2.

Para avançar com sua análise, o aluno A2 selecionou no *Google Correlate*, dentre os termos com maior correlação com o termo de pesquisa ‘cooperativa’, o termo correlacionado ‘assessoria de imprensa’ que apresenta um coeficiente de determinação R^2 de 0,9853, obtendo o gráfico da Figura 9.

Analisando esse gráfico, o aluno A2 disse que

“vemos que as curvas dos dois termos pesquisados se assemelham muito. Apesar destes gráficos serem muito parecidos, não temos como afirmar que esta correlação tem uma explicação física.”

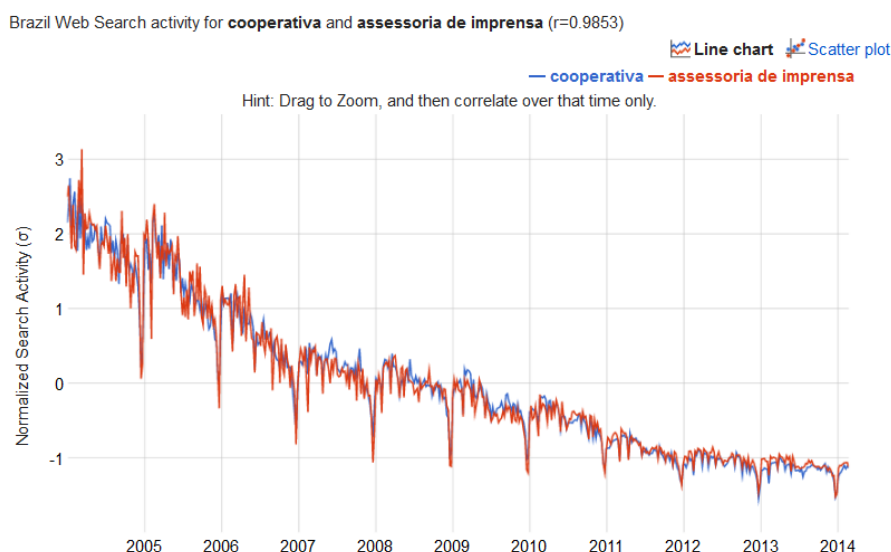
quase citando literalmente o conhecido alerta dos estatísticos, já mencionado acima.

Consideramos a pesquisa do aluno A2 ter sido a que mais se aproximou da proposta: ele pesquisou vários termos de busca até encontrar um cujo gráfico não apresentasse o padrão “temas escolares”, extraiu uma conclusão a partir dos gráficos, formulou uma hipótese causal para ela, pesquisou em outras fontes para confirmar ou não a hipótese e concluiu pela não

¹¹ Organização das Cooperativas Brasileiras.

validade dela a partir da análise dos dados secundários obtidos, extraindo disso a confirmação do preceito de que nem sempre existe uma explicação causal para uma correlação observada num gráfico. Perguntado se recursos tais como *Google Correlate* e *Google Trends* auxiliam na construção do conhecimento, este aluno respondeu que “*podem auxiliar, mas junto com outras ferramentas*¹²”, o que indica que não lhe terão auxiliado, pelo menos nesta ocasião e dessa forma.

Figura 9 - Comparação entre as frequências de pesquisa dos termos ‘cooperativa’ e ‘assessoria de imprensa’ no Google.



Fonte: *Google Correlate* (<http://www.google.com/trends/correlate>) via aluno A2.

Infelizmente, devido a um defeito posterior no *pendrive*, não foi possível recuperar a atividade de pesquisa realizada pelo grupo de alunos A1 e A7. De qualquer forma, ambos apenas encontraram termos cujos gráficos apresentavam o padrão “temas escolares”. O aluno A1 desculpou-se, afirmando que, quanto aos recursos *Google Correlate* e *Google Trends* auxiliarem na construção do conhecimento, eles, na opinião dele, “*não acrescentaram em nada*”, não achando neles “*nada de especial*”.

Análise geral da atividade

Conforme dissemos na seção sobre nossa proposta, esperávamos que os alunos alcançassem uma melhor compreensão, embasada na prática do Big Data, sobre a construção

¹² Vide nota de rodapé ⁸.

do conhecimento físico, especialmente das noções de fenômeno, observação, medida, leis físicas, teoria, causalidade, dentre outras.

Embora vários alunos (A2, A5, A4, A6) tenham, através de suas pesquisas, chegado à importante conclusão de que correlação não implica em causação, eles centraram-se nas correlações em si; em nenhum momento de suas falas emerge alguma sugestão de que tenham atingido o objetivo principal da proposta, o qual era desenvolverem suas compreensões sobre construção do conhecimento físico ou dos conceitos de fenômeno, observação, medida, leis físicas, teoria, causalidade, ou qualquer outro.

Apesar disso, perguntados sobre como avaliavam sua construção de aprendizagem nesta disciplina, um total de 4 dos 7 alunos (57%) responderam que elas tinham sido satisfatórias, tendo cumprido todos os objetivos que tinham para si mesmos com relação ao conhecimento científico, enquanto 3 alunos (43%) consideraram que elas tinham sido regulares, tendo deixado a desejar com relação aos objetivos.

Por outro lado, perguntados sobre com que regularidade estudaram para esta disciplina, apenas 2 alunos (29%) declararam que estudaram semanalmente, revisando o material em casa, participando em aula e buscando informações extraclasse, 1 aluno (13%) declarou que estudou semanalmente, revisando o material em casa e participando em aula, sem, no entanto, buscar informações extraclasse, 2 alunos (29%) declararam ter estudado apenas por participação nas aulas, enquanto 2 (29%) confessaram não ter estudado para a disciplina.

Quanto a como as aulas mostraram-se a eles, nada menos do que 5 alunos (71%) consideraram as aulas importantíssimas, tendo modificado suas visões sobre a Ciência, enquanto 2 alunos (29%) consideraram as aulas relevantes, porque o conteúdo foi interessante.

Perguntados sobre se acharam que as atividades relacionadas à Big Data lhe ajudaram de alguma forma, a maioria dos alunos considerou que as atividades de Big Data foram de alguma utilidade. Um aluno (A2) respondeu que *“ajudaram no sentido da descoberta que podemos trabalhar com esta imensidão de dados que dispomos em qualquer sistema informatizado e tirar algum proveito”*. O aluno A3 respondeu que *“ajudaram no sentido de conhecer a ferramenta, posso utilizar ela para diversos fatores de estudo”*. O aluno A4 respondeu que *“Sim, claro”*. No entanto, o aluno A1 considerou que não ajudou *“em nada”*, não achou *“nada de mais”*.

Quanto a quais foram as maiores dificuldades encontradas ao lidar com os aplicativos *Google Correlate* e *Google Trends*, vários alunos relataram a dificuldade em encontrar

correlações com explicação causal (causações), conforme discutimos antes. O aluno A2 comentou que *“achar uma correlação que tenha causa e se justifique é muito difícil”*, ao que o aluno A4 concordou que *“a maior dificuldade dessas ferramentas¹³ é não encontrar sempre correlações causais”*, acrescentando que *“alguns padrões persistem em sobressair-se”*, referindo-se, possivelmente, a padrões temporais decorrentes de periodicidades de calendário escolar, por exemplo. O aluno A3 confessou que deixou a atividade *“para a última hora achando que era fácil”*, até porque entendeu mal o funcionamento do aplicativo *Google Correlate* e que ele *“é que iria colocar a relação e não a ferramenta me mostrar”*. Por outro lado, o aluno A1 declarou que não considerou o *Google Correlate* *“uma ferramenta útil”*; aparentemente, não atingiu a compreensão do significado das correlações entre conceitos, dizendo que não lhe *“ficou bem claro essas correlações”* *“e nem entendi porque tem correlações”* e, por isso, afirmou que não achou *“nada de mais”*, *“nada a ver”*.

De fato, vários alunos recorreram à colega bolsista, mencionada na metodologia, por e-mail ou outros meios de contato, com dúvidas tais como *“estou tentando fazer o trabalho, mas não estou achando correlação com que eu pesquiso; é difícil assim mesmo ou eu estou fazendo errado?”*, *“tentei fazer com positivismo e ele [o Google Correlate] relacionou com capitalismo”*, *“Quando você entra no Correlate e aí coloca o título [termo de busca] lá, já aparecem os gráficos pra ti? Ou tem que ir a outra área?”* e *“encontrei relação entre viagem e pacotes de viagem, será que vale isto?”*. Um aluno referiu não compreender o porquê do gráfico, como ele se forma; outros relataram dificuldades com o layout da página da *Correlate* e para inserir termos de busca. Vários alunos em dificuldade pediram exemplos de correlações causais à bolsista.

Finalmente, perguntados se eles acreditavam que incluir aplicativos de Big Data no ensino de ciências possa ser positivo para as aulas, o aluno A2 respondeu que sim, desde que *“com ferramentas¹⁴ bem estruturadas”*. O aluno A3 concordou, lamentando que *“se não ouvirmos falar desta ferramenta na faculdade, será muito mais difícil conhecê-la fora da instituição”* e acrescentou que *“esta ferramenta pode ajudar qualquer um. Por exemplo, [num] estudo para abrir um negócio”*. O aluno A7 comentou até que *“esta ferramenta me fez pesquisar pela primeira vez, nunca tinha pegado revistas científicas para olhar”*. O aluno A4 também concordou, afirmando que *“a inclusão desta ferramenta será o novo diferencial no ensino, algo que emerge de forma rápida e que deve ser adaptado às salas de aula”*. Já o

¹³ Vide nota de rodapé ⁸.

¹⁴ Vide nota de rodapé ⁸.

aluno A1 discordou, dizendo que “*no Big Data sinceramente não achei nada de interessante*” e afirmou que “*não utilizaria nas aulas*”.

Avaliando os resultados desta aplicação, consideramos que não demos suficiente atenção a atividades preparatórias de familiarização dos alunos aos conceitos de Big Data, correlação e causalidade e aos aplicativos *Google Correlate* e *Google Trends* a serem utilizados. Acreditamos que vários alunos, especialmente o A1, pelas suas respostas, não conseguiram se *apropriar* desses recursos, o que, aparentemente, contribuiu para as dificuldades encontradas por eles e para um resultado inferior ao esperado, por alguns deles. A ideia corrente de que os nascidos em ou depois de 1980 são ‘nativos digitais’, com habilidades sofisticadas no uso de tecnologias digitais, não tem base empírica (MARGARYAN; LITTLEJOHN; VOJT, 2011). Por isso, numa próxima realização destas atividades, pretendemos, após a aula em que se discutiram os conceitos de leis físicas, teoria e causalidade e foi apresentado o conceito de Big Data, incluir uma aula prática com esses aplicativos de Big Data no laboratório de informática, por forma a sanar dúvidas sobre sua utilização.

Além disso, pretendemos, em próximos trabalhos, pesquisar mais os aplicativos *Google Correlate* e *Google Trends* em si, especialmente quanto às suas características de frequente aparecimento de padrão “temas escolares” nos gráficos de correlação e queda aparente nos gráficos de tendência, que se tornam fatores que também dificultam a apropriação desses recursos e prejudicam os resultados dos alunos.

CONCLUSÃO

Acreditamos que esta é a primeira proposta de utilização de Big Data no Ensino de Ciências no Brasil, com um viés que não é meramente de uma capacitação em infraestruturas computacionais ou de treinamento em análise preditiva, mas uma preparação de nossos estudantes, futuros profissionais de Ciências, tanto em termos técnicos como em éticos, para os desafios científicos propostos pelo Big Data ao mundo real, no qual vão exercer suas profissões, além de uma melhor compreensão, embasada na prática do Big Data, sobre a construção do conhecimento físico, especialmente numa melhor compreensão das noções de fenômeno, observação, medida, leis físicas, teoria, causalidade, dentre outras.

Esta aplicação, por um lado mostrou que a proposta parece ser viável: no geral, os alunos apreciaram as atividades mediadas pelos aplicativos *Google Correlate* e *Google Trends* e demonstraram compreender a distinção entre correlação e causalidade. Por outro, no entanto, evidenciou a necessidade de seu aperfeiçoamento em vários aspectos que incluem,

principalmente, facilitar a apropriação desses recursos pelos alunos, desenvolver estratégias para contornar dificuldade dos alunos em encontrar termos que tenham relação causal e que, por isso, não apresentem o frequente padrão “temas escolares” em seus gráficos e facilitar a conexão da atividade com o processo de construção do conhecimento físico, especialmente das noções de fenômeno, observação, medida, leis físicas, teoria, causalidade, dentre outras.

REFERÊNCIAS

- ARNTZENIUS, F. Reichenbach's Common Cause Principle. In: E. N. ZALTA (Ed.). **The Stanford Encyclopedia of Philosophy**. Stanford, CA: Stanford University, 2010.
- BARAM-TSABARI, A. e SEGEV, E. Just Google it! Exploring New Web-based Tools for Identifying Public Interest in Science and Pseudoscience. In: Y. ESHET-ALKALAI et al. (Eds.). **THE CHAIS CONFERENCE ON INSTRUCTIONAL TECHNOLOGIES RESEARCH 2009: LEARNING IN THE TECHNOLOGICAL ERA, FEB. 18, 2009, RAANANA, ISRAEL. Proceedings...** Raanana: The Open University of Israel, 2009a.
- BARAM-TSABARI, A. e SEGEV, E. Exploring new web-based tools to identify public interest in science. **Public Understanding of Science**, v. 20, n. 1, p. 130-143, 2009b.
- BARAM-TSABARI, A. e SEGEV, E. The half-life of a “teachable moment”: The case of Nobel laureates. **Public Understanding of Science**, p. 83-89, 2013.
- BÜLBÜL, M. Ş. Google Centered Search Method in Pursuit of Trends and Definitions in Physics and Education. **fizikli.com**, 2009. Disponível em: <www.fizikli.com/piwi/fizikli6.pdf>. Acesso em: 7 fev. 2014.
- CHALMERS, D. J. Strong and weak emergence. In: CLAYTON, P.; DAVIES, P. (Eds.). **The reemergence of emergence**. New York: Oxford University Press, 2006. p. 244-256.
- DOS SANTOS, R. P. Aprender Física, Biologia, Química e Matemática com Big Data. **Educação Matemática em Revista - RS**, v. 2, n. 15, 2014a.
- DOS SANTOS, R. P. Big Data as a Mediator in Science Teaching: A Proposal. **Innovation Educator: Courses, Cases & Teaching eJournal**, v. 2, n. 25, p. 1, 2014b.
- FIELD, H. Causation in a Physical World. In: M. J. LOUX; D. W. ZIMMERMAN (Eds.). **Oxford Handbook of Metaphysics**. Oxford: Oxford University Press, 2003.
- GANTZ, J. e REINSEL, D. **The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East**. Framingham, MA: IDC - International Data Corporation, 2012.
- HEALY, L. e KYNIGOS, C. Charting the microworld territory over time: design and construction in mathematics education. **ZDM Mathematics Education**, v. 42, n. 1, p. 63-76, 2009.

HOFER, B. K. e PINTRICH, P. R. The Development of Epistemological Theories: Beliefs About Knowledge and Knowing and Their Relation to Learning. **Review of Educational Research**, v. 67, n. 1, p. 88-140, 1997.

IBM. **What is big data?** Armonk, NY: IBM, 2011. Disponível em: <<http://www-01.ibm.com/software/data/bigdata/>>. Acesso em: 10 maio. 2013.

KYNIGOS, C. Constructionism: Theory of Learning or Theory of Design? In: 12TH ICME - INTERNATIONAL CONGRESS ON MATHEMATICAL EDUCATION, JULY 8TH-15TH, 2012, SEOUL, KOREA. **Proceedings...** Cheongju: Korea National Univ. of Education, 2012. Disponível em: <http://www.icme12.org/upload/submission/2020_F.pdf>. Acesso em: 25 jul. 2012.

MARGARYAN, A. M.; LITTLEJOHN, A. e VOJT, G. Are digital natives a myth or reality? University students' use of digital technologies. **Computers & Education**, v. 56, n. 2, p. 429-440, 2011.

MOHEBBI, M. H. et al. **Google Correlate Whitepaper**. Menlo Park, CA: Google Inc., 2011. Disponível em: <<http://www.google.com/trends/correlate/whitepaper.pdf>>. Acesso em: 28 abr. 2013.

NOVAK, J. D. e GOWIN, D. B. **Learning how to Learn**. Cambridge, MA: Cambridge University Press, 1984.

PAPERT, S. A. **Logo: Computadores e Educação**. São Paulo: Brasiliense, 1985.

PAPERT, S. A. **The Children's Machine: Bringing the Computer Revolution to Our Schools**. New York: Basic Books, 1993.

PAPERT, S. A. Papert on Piaget. **TIME**, special issue: "The Century's Greatest Minds", p. 105, 1999.

ROBILOTTA, M. R. O cinza, o branco e o preto da relevância da história da ciência no ensino da física. **Caderno Brasileiro de Ensino de Física**, v. 5, n. 5, p. 7-22, 1988.

ROSA, M. A **Construção de Identidades online por meio do Role Playing Game: relações com o ensino e aprendizagem de matemática em um curso à distância**. Tese (Doutorado em Educação Matemática) - Universidade Estadual Paulista, Rio Claro, SP, 2008.

SEGEV, E. e BARAM-TSABARI, A. Seeking science information online: Data mining Google to better understand the roles of the media and the education system. **Public Understanding of Science**, v. 21, n. 7, p. 813-829, 2012.

YIN, C. et al. Learning by Searching: A Learning Environment that Provides Searching and Analysis Facilities for Supporting Trend Analysis Activities. **Journal of Educational Technology & Society**, v. 16, n. 3, p. 286-300, 2013.