

Ciência de Dados Sociais: uma visão sociológica das Humanidades Digitais

Social Data Science: A Sociological view of Digital Humanities

Ciencia de datos sociales: una visión sociológica de las humanidades digitales

Recebido em 01-09-2021
Modificado em 08-11-2021
Aceito para publicação 02-12-2021

 <https://doi.org/10.47456/simbitica.v8i4.37346>

46

 **Kleberson Toldo Soares**

Licenciado em História e Sociologia pela Universidade Estácio de Sá, Brasil. Licenciando em Computação pela Claretiano e pós-graduando em Data Science & Analytics pela Esalq/USP. Possui MBA em Ciência Política pela Unypública. E-mail: kleberson.toldo.soares@gmail.com

Resumo

O presente trabalho tem como objetivo localizar e contextualizar algumas definições científicas e metodológicas para o emergente campo da Ciência de Dados Sociais. Por ser uma área extremamente nova, a busca abrange orientações para organizar esse novo olhar para as Ciências Sociais, utilizando autores e métodos das áreas constitutivas das Ciências Sociais Computacionais e das Humanidades Digitais, entre elas a Sociologia Digital. Seguindo o processo proposto pela Ciência de Dados, a pesquisa traz reflexões sobre a confecção e utilização de produtos gráficos como efetivadores do processo de extração de informação e condução da tomada de decisão humana ou de aprendizagem de máquina. O desenvolvimento humano ligado à tecnologia é relativamente novo e este trabalho relaciona-se com o atual momento da humanidade.

Palavras-chave: humanidades digitais; sociologia digital; ciência de dados sociais; ciências sociais computacionais.



Introdução

Através da aplicação de técnicas analíticas aos dados coletados nas plataformas disponíveis para pesquisa individual e seguindo o processo proposto pela Ciência de Dados, podemos ter visualmente a noção gráfica de como está a produção científica ligada ao conceito de Ciência de Dados Sociais, além de conhecer algumas técnicas de captura e transformação de informações. O tema principal deste trabalho tem sua dimensão disciplinar estruturada conforme o programa de mestrado em Ciência de Dados Sociais da Universidade de Copenhague¹, que se baseia na relação com as áreas do conhecimento abrangendo principalmente seis disciplinas: Antropologia, Ciência de Dados, Economia, Ciência Política, Psicologia e Sociologia. Essa construção curricular do programa multidisciplinar da instituição também será usada para conduzir a definição de Ciência de Dados Sociais, tendo em vista explicitar sua ligação com o debate acerca dos limites dos campos epistemológicos das Humanidades Digitais e sua estrutura ampla e conectada ao desenvolvimento da era digital. Afinal,

As humanidades digitais (HDs) vêm se projetando no mesmo contexto da sociedade da informação diante da realidade da explosão informacional com foco nas fontes de informação digitais. A área recém-criada, se apresenta como um campo importante de interlocução que se desenvolve rapidamente através da crescente criação de centros acadêmicos para sua reflexão e implementação de projetos desta ordem (Castro, 2020:21).

Para desenvolver a análise de dados e as representações gráficas, foram utilizados o software *Excel*, o software *RStudio*, que também é uma linguagem de programação, e o programa on-line *Infogram*, com o intuito de demonstrar a conectividade das informações criadas na rede.

Cabe às Ciências Humanas o papel de buscar entender as bases das sociedades do passado, presente e futuro e, a partir disso, refletir sobre suas estruturas e sobre o papel de cada indivíduo como sujeito histórico. Às Ciências Sociais, por sua vez, compete, entre outras análises, identificar os funcionamentos da engenhosa sociedade humana e suas diversas ramificações. Já a Sociologia ajusta a lente desse olhar, explorando a reflexão sobre a “ordem” tanto nas hierarquias de cada fragmento que se apresenta em um determinado coletivo, quanto na relação entre todas as sociedades humanas, identificando as individualidades participantes de todo o processo de construção da própria função social.

Em uma sociedade cada vez mais tecnológica, que cria e desenvolve muitas das suas bases nas estruturas computacionais, nada mais óbvio do que voltar o olhar sociológico para esse

¹ Disponível em <https://studies.ku.dk/masters/social-data-science/>. Acesso em 15-03-2021.

fenômeno e analisar o papel do indivíduo nessa etapa de transformação histórica. Esse tipo de investigação é o que denominamos Sociologia Digital, que sugere uma visão ampliada das perspectivas analíticas vinculadas ao desenvolvimento tecnológico da nossa sociedade, trazendo nos campos epistemológicos das Ciências Computacionais e Sociais a provocação e o estranhamento perante a transdisciplinaridade proposta pelas Humanidades Digitais. Essa nova disciplina se faz necessária porque, conforme Abbott (2000:298), a sociologia está lamentavelmente despreparada para lidar com a transformação tecnológica, pois não possui as ferramentas analíticas e a imaginação conceitual necessárias para tal. Este trabalho, além de apresentar um contexto geral sobre a atuação da Sociologia Digital, aponta a reflexão sobre o papel do pesquisador e do professor, especialmente da disciplina de Sociologia, nessa nova direção na qual a humanidade está seguindo.

Metodologia

Buscando as bases da pesquisa científica e do próprio conhecimento, encontram-se nos primeiros filósofos e sociólogos as direções iniciais para organizar uma análise: identificar padrões e categorizar. Nessa origem, situa-se também as diretrizes éticas da doutrina científica e os debates sobre essa questão. Na aplicação dos resultados, identifica-se diferentes entendimentos, como: reestruturar aquilo que porventura possa estar “fora” da sequência apontada como padrão ou até mesmo reestruturar o padrão para que possa comportar aquilo que está sendo indicado como divergente. Pode-se inclusive levantar uma analogia entre o termo “*outsiders*”, utilizado por Erving Goffman para se referir aos indivíduos que estão além dos limites padrões de uma sociedade, e “*outliers*”, termo da Ciência de Dados aplicado na existência de informações extremamente divergentes que são retiradas no momento de limpeza para processamento de dados.

Sobre os *outsiders*, diz Goffman:

Enquanto o estranho está à nossa frente, podem surgir evidências de que ele tem um atributo que o torna diferente de outros que se encontram numa categoria em que pudesse ser incluído, sendo, até, de uma espécie menos desejável [...]. Assim deixamos de considerá-la criatura comum e total, reduzindo-a a uma pessoa estragada e diminuída. Tal característica é estigma, especialmente quando o seu efeito de descrédito é muito grande [...] (Goffman, 1988:06).

Os *outsiders* são indivíduos estigmatizados pelo padrão de classificação do diferente e, sendo assim, são classificados em uma categoria inferior, que delimita a projeção da qualidade de vida desse indivíduo dentro do esperado pelo classificador. Por mais que classificações – como as de classe social, muito utilizadas para políticas públicas – pareçam úteis ao

desenvolvimento de identificações de grupos envolvidos em determinados contextos sociais, ao analisar mais profundamente percebe-se que se trata de estigmas ou cicatrizes sociais que designam o início e o fim da função social do indivíduo.

Os *outliers* são, por sua vez, dados que se diferenciam drasticamente de todos os outros, são pontos fora da curva. Como um valor que foge da normalidade, esses dados podem causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise. Conforme Hawkins (1980), um *outlier* é uma observação que se diferencia tanto das demais, que levanta suspeitas de que um mecanismo distinto a gerou, isto é, suspeitas de que faz parte de um outro padrão classificatório.

Entende-se, assim, que o conceito de *outliers* parte da construção de um modelo que faz suposições sobre os dados. Os valores extremos são frequentemente indicadores de que o modelo não descreve os dados adequadamente e, portanto, deve-se questionar os resultados do modelo ou a qualidade dos dados o que se assemelha à construção de estigma, proposta por Goffman (1988), a partir da qual se compreende que a sociedade estabelece os meios de categorizar as pessoas e o total de atributos considerados como comuns e naturais para os membros de cada uma dessas categorias.

Levando em consideração que o método de pesquisa quantitativo é influenciado em sua estrutura pela corrente filosófica do positivismo, constata-se que a automação da pesquisa reforça as possibilidades dos dados quantitativos direcionais e que a lógica positivista é necessária em pensamentos computacionais, porém a visão crítica e socioemocional se faz necessária em todas as etapas do processo, levando a uma reflexão sobre as metodologias utilizadas para o levantamento de dados dentro das bases digitais.

Todos os tópicos que os sociólogos agora pesquisam e ensinam estão inevitavelmente conectados às tecnologias digitais, quer eles se concentrem na sociologia da família, da ciência, da saúde e medicina, do conhecimento, da cultura, da economia, do emprego, da educação, do trabalho, do gênero, do risco, do envelhecimento, da raça e etnicidade. Estudar a sociedade digital é se concentrar em muitos aspectos que têm sido preocupações centrais para os sociólogos: individualidade, identidade, embodiment, as relações de poder e as desigualdades sociais, redes sociais, estruturas sociais, instituições sociais e teoria social (Lupton, 2015:8).

A partir disso, e tendo em vista que métodos de análise automatizada de conteúdo são modelos incorretos de linguagem (Grimmer; Stewart, 2013), somos levados ao questionamento sobre a formação dos cientistas de dados. Se no momento da capacitação, da educação básica e profissional, o indivíduo não entrar em contato com o desenvolvimento do senso crítico, conduzirá pesquisas que somente regulamentam a implementação dos resultados através de imagens gráficas que conduzem a interpretação e conseqüentemente a tomada de decisões. Sendo assim, a obtenção de conhecimento não será construtiva, apenas estruturalista e utilitária.

Encontram-se padrões culturais e pessoais dentro das pesquisas com dados estruturados, porém a necessidade de padronização de condutas para seleção de grupos que apresentam homogeneidade entre indivíduos não desenvolve a individualidade de cada participante, categorizando as estruturas sociais como um fluxo lógico e direto, assim com a visão positivista de progresso.

A maioria das áreas de pesquisa utilizam estruturas computacionais para obter êxito em seus estudos. Isso significa que os computadores são usados para modelar, simular e analisar fenômenos. Na biologia computacional, por exemplo, essa estrutura é empregada para verificar padrões de comportamento dos seres vivos. Mas será que podemos manusear da mesma forma para análise de outros tipos de fenômenos?

A Ciência Social Computacional é um exemplo desse tipo de análise. Enquanto subdisciplina acadêmica preocupada com abordagens computacionais para as Ciências Sociais, ela se vale dos computadores para modelar, simular e analisar fenômenos sociais. Para Nascimento (2016), esse movimento vem, gradualmente, alterando a inserção da própria sociologia no cenário público, direcionando os seus resultados para além do público acadêmico.

A ética das pesquisas em sistemas sociais complexos e não simulados circunda a reflexão dos cientistas sociais e, dentro dessa perspectiva, os modelos computacionais e até mesmo as redes sociais trazem um novo olhar sobre esse problema.

A Sociologia Digital não trata apenas de sociólogos pesquisando e teorizando sobre como outras pessoas usam tecnologias digitais ou concentrando-se nos dados digitais produzidos através deste uso. A Sociologia Digital tem implicações muito mais amplas do que simplesmente estudar as tecnologias digitais, levantando questões sobre a prática da sociologia e a pesquisa social. Ela também inclui pesquisas sobre como os próprios sociólogos estão usando as mídias digitais sociais e outros como parte do seu trabalho (Lupton, 2015:15).

Além disso, levanta-se a seguinte questão: será que os ambientes virtuais refletem todas as especificações e individualidades de seus participantes? Percebe-se que não, pois as simulações reforçam os estigmas sociais, porém as humanidades digitais lançam um novo horizonte para a construção, o desenvolvimento e as análises das individualidades.

Dados bibliográficos

Dentro das diretrizes da pesquisa científica, a bibliometria e o estado da arte são norteadores imprescindíveis ao processo, visto que o impacto científico das produções sobre o assunto precisa estar nítido e interligado aos objetivos do trabalho. No entanto, a pesquisa sociológica na era *big data*, ou o olhar sociológico sobre a Ciência de Dados, carece de artigos

científicos, tendo em vista que a bibliografia atual dá ênfase maior às ferramentas da Ciência de Dados, privilegiando uma visão sobre ela que a utiliza apenas como forma geral de mineração e análise de dados em sua área inicial da computação. Nesse contexto, a crítica inicial se estabelece a partir do pressuposto de que o olhar do cientista social não pode ser métrico ou definido por números e estatísticas. O olhar do cientista social deve ser mais amplo, visando o princípio da imaginação sociológica proposta por Charles Wright Mills.

Aprender a pensar sociologicamente – olhando – em outras palavras, de forma mais ampla – significa cultivar a imaginação. Estudar Sociologia não pode ser apenas um processo rotineiro de adquirir conhecimento. Um sociólogo é alguém que é capaz de se libertar das imediatezidades das circunstâncias pessoais e apresentar as coisas num contexto mais amplo. O trabalho sociológico daquilo que o autor norte-americano C. Wright Mills, numa frase famosa chamou de *imaginação sociológica* (Giddens, 2005:24).

Com a criticidade desenvolvida pelo cientista social, as ferramentas de pesquisa propostas pela Ciência de Dados são de extrema relevância, visto o atual cenário digital da nossa sociedade.

As humanidades digitais referem-se ao conjunto das Ciências Humanas e Sociais, às Artes e às Letras. As humanidades digitais não negam o passado; apoiam-se, pelo contrário, no conjunto dos paradigmas, no saber fazer e conhecimentos próprios dessas disciplinas, mobilizando simultaneamente os instrumentos e as perspectivas singulares do mundo digital. As humanidades digitais designam uma transdisciplina, portadora dos métodos, dos dispositivos e das perspectivas heurísticas ligadas ao digital no domínio das Ciências Humanas e Sociais (Manifesto das Humanidades Digitais, 2010: s/p).

51

Diante da quantidade de dados que é produzida a todo o instante, o que chamamos de *big data*, é necessário promover recortes dentro dessa imensidão, direcionado a pesquisa para os seus objetivos. Usa-se a representação gráfica dos dados como estrutura para análise geral, em que se utiliza alguma técnica computacional com a finalidade de se obter informação, valor e conhecimento. Todavia, não são apenas estruturas computacionais utilizadas para análise de dados, mas também as habilidades humanas para interpretar essas representações.

Trata-se de um movimento que, ao impactar a área das ciências humanas e sociais, levam os pares a percepção de que as pesquisas agora passam a ser mediadas pelas tecnologias. Amparadas pela digitalização, a tendência irreversível de criação de fontes digitais colocou às ciências humanas o desafio de incorporar novos métodos à sua tradicional metodologia de pesquisa. Esse desafio implica em uma nova forma de trabalhar, tornando essencial sua reflexão pelos envolvidos nela (Castro, 2020:22).

Dentro dessa perspectiva, a condução inicial proposta neste trabalho visa contemplar os dois âmbitos necessários para o seu desenvolvimento: a criticidade do cientista social e as ferramentas do cientista de dados. Ao iniciar uma busca nas bases de dados com o termo “Ciência de Dados Sociais” nenhum artigo científico é obtido. Entretanto, existem postagens em

blogs pessoais sobre a intersecção das áreas de Ciência de Dados e Ciências Sociais vivida pelos autores. Ao ampliarmos a nossa pesquisa através do termo “*Social Data Science*”, é possível desbravar algumas definições e artigos postados em língua inglesa. A principal publicação sobre o assunto é encontrada no site *towardsdatascience.com*, uma plataforma que reúne milhares de pessoas para expandir a compreensão da Ciência de Dados e se refere à implementação do *University of Copenhagen’s MSc of Social Data Science*, já citado na definição de áreas correlacionadas que formam a Ciência de Dados Sociais.² Na definição proposta por Alex Montzau (2019), a Ciência de Dados Sociais é uma nova disciplina que combina as Ciências Sociais e a Ciência da Computação, na qual a análise de *big data* está ligada à teoria e à análise científica social.

Ao encontrar os primeiros artigos publicados, através do Google Acadêmico, com o termo “*Social Data Science*”, percebe-se que a sua produção científica inicia em 2005 com 02 (dois) artigos oriundos de eventos científicos que ocorreram na Lituânia no mesmo ano, chegando a 139 (cento e trinta e nove) em 2020. As informações podem ser observadas no gráfico da figura 1a, que demonstra o crescimento anual dos registros, sendo possível, em uma primeira e rápida análise visual, identificar o ano de 2020 como um *outlier*, visto que o crescimento de artigos foi desproporcional aos outros anos analisados.

Figura 1a - Gráfico de dispersão de trabalhos acadêmicos na área de *Social Data Science*



Elaboração própria.

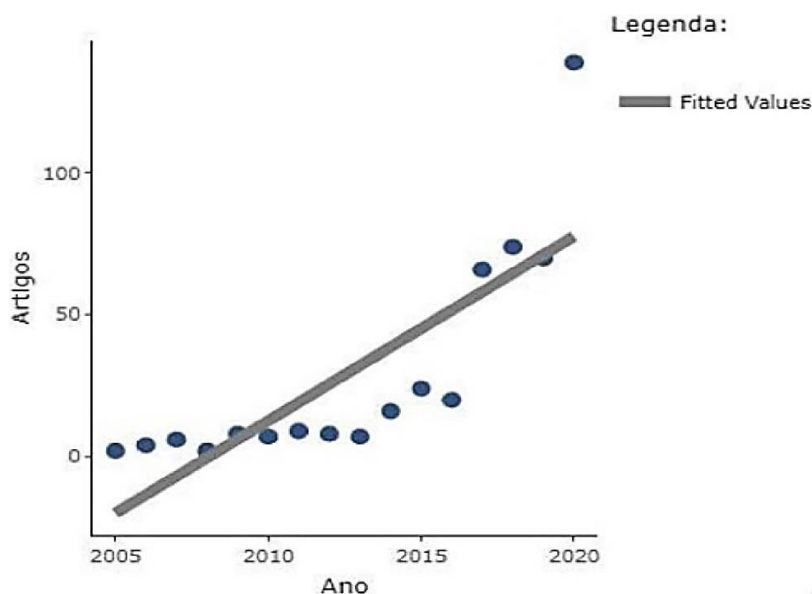
² Disponível em <https://towardsdatascience.com/first-day-of-social-data-science-beedc2174232>. Acesso em 13-03-2021.

A figura 1b demonstra a dispersão em comparação com o modelo linear estimado. O modelo serve de parâmetro para basear a média de crescimento nas publicações dos artigos em relação à variável ano, y e x , respectivamente. Para tentar estabelecer uma equação que represente o fenômeno em estudo, pode-se fazer um gráfico, chamado de diagrama de dispersão, para verificar como se comportam os valores da variável dependente (y) em função da variável independente (x) (Paternelli, 2005). Pode-se utilizar a regressão linear como um instrumento estatístico para, simplesmente, resumir dados e informações (Chein, 2019). A equação linear simples é dada por (Equação 1):

$$y_i = \alpha + \beta \cdot X_i + u_i$$

Onde temos um intercepto (α), um coeficiente de inclinação da reta (β), uma variável explicativa X (Ano) e um termo de erro u .

Figura 1b – Gráfico de dispersão com reta de regressão linear



Elaboração própria.

A projeção média de crescimento foi de aproximadamente 29%. O modelo não segue o padrão de forma homogênea, tendo uma queda no comparativo entre os anos de 2007 e 2008 e um crescimento de 22% no comparativo dos anos 2019 e 2020. Os diagramas ou gráficos de dispersão permitem uma análise inicial de possibilidades. É usado para verificar se existe relação de causa e efeito entre duas variáveis de natureza quantitativa. Isto não tem o intuito de provar que uma observação afeta diretamente a outra, mas de determinar a intensidade da relação entre

os valores. Assim, ainda que os diagramas sejam uma ferramenta de diagnóstico útil para determinar essa associação, a análise computacional nunca pode “provar” uma relação de causa e efeito. Após uma série de avaliações e testes provenientes das metodologias científicas, somente o pesquisador pode concluir que a causalidade realmente existe³.

O crescimento não uniforme da variável “artigos” em relação à variável “ano” pode trazer reflexões sobre o fomento à pesquisa da área que está sendo observada, bem como sobre a relação com a política pública voltada para a área científica em geral em determinado país ou, ainda, sobre o comparativo com outras áreas.

Isso acontece, devido ao fato do fenômeno que está em estudo, não ser um fenômeno matemático, e sim um fenômeno que está sujeito à influências que acontecem ao acaso. Assim, o objetivo da regressão é obter um modelo matemático que melhor o ajuste aos valores observados de Y em função da variação dos níveis da variável X (Paternelli, 2005:03).

O *output* do modelo de regressão linear pode servir de base estrutural para o desenvolvimento de novas variáveis e observações, pois as interferências na variável X (Ano) são essenciais para determinação dos resultados.

O estudo não buscou abranger outras possibilidades, mantendo-se na análise quantitativa da relação entre os artigos publicados, na área de “Social Data Science” no decorrer do tempo. A construção feita na linguagem R, para gerar o gráfico da figura 1b, está registrada conforme script constante na figura 1c.

Figura 1c – Fragmento do Script para Gráfico de Dispersão

```

1
2
3 *****GRÁFICO DE DISPERSÃO*****
4
5 #utiliza-se um modelo linear estimado para verificar a relação entre as variáveis
6
7 ggplotly(
8   ggplot(Data_Base_Dispersao, aes(x =Ano, y = Quantidade_de_Artigos)) +
9     geom_point( color = "339568CFF", size = 2.5)+
10    geom_smooth(aes(color = "Fitted Values"),
11               method= "lm", se = F, size 2) +
12    xlab("Ano") +
13    ylab("Artigos") +
14    scale_color_manual("Legenda",
15                      values = "grey50") +
16    theme_classic ()
17 )
18

```

Elaboração própria.

Na linguagem R podemos utilizar a função *summary* para obter noções importantes sobre a base de dados. Isso nos ajuda a visualizar algumas questões, antes de definir o modelo que será

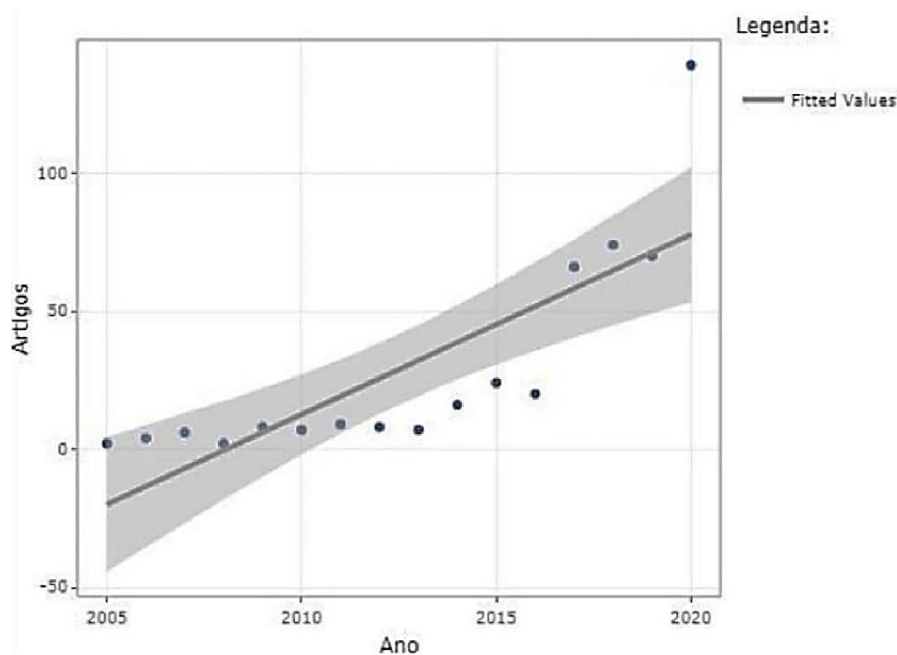
³ Disponível em <https://www.itl.nist.gov/div898/handbook/eda/section3/scatterp.htm>. Acesso em 19-11-2021.

utilizado. As definições, como média e mediana, trazem a possibilidade de uma primeira análise, mesmo que superficial. Existem, no entanto, outras formas de obter as mesmas informações a partir da linguagem R, o que a torna uma excelente ferramenta estatística.

O R torna-se, portanto, uma importante ferramenta na análise e na manipulação de dados, com testes paramétricos e não paramétricos, modelagem linear e não linear, análise de séries temporais, análise de sobrevivência, simulação e estatística espacial, entre outros, além de apresentar facilidade na elaboração de diversos tipos de gráficos, no qual o usuário tem pleno controle sobre o gráfico criado (Paternelli e Mello, 2007:05).

Após estimar o modelo e salvar as variáveis de valores ajustados e valores de erro, consideradas *yhat* e *residuals*, respectivamente, uma nova base de dados é gerada. Em resumo, por esse método, com a soma de quadrados das distâncias entre os pontos do diagrama e os respectivos pontos na curva da equação, são minimizadas as discrepâncias, obtendo-se, desta forma, uma relação funcional entre X e Y para o modelo escolhido, com o mínimo de erro possível (Paternelli, 2005:03). A figura 1d representa um novo gráfico do modelo ajustado para um grau de confiança de 95%.

Figura 1d – Gráfico com intervalo de confiança de 95%



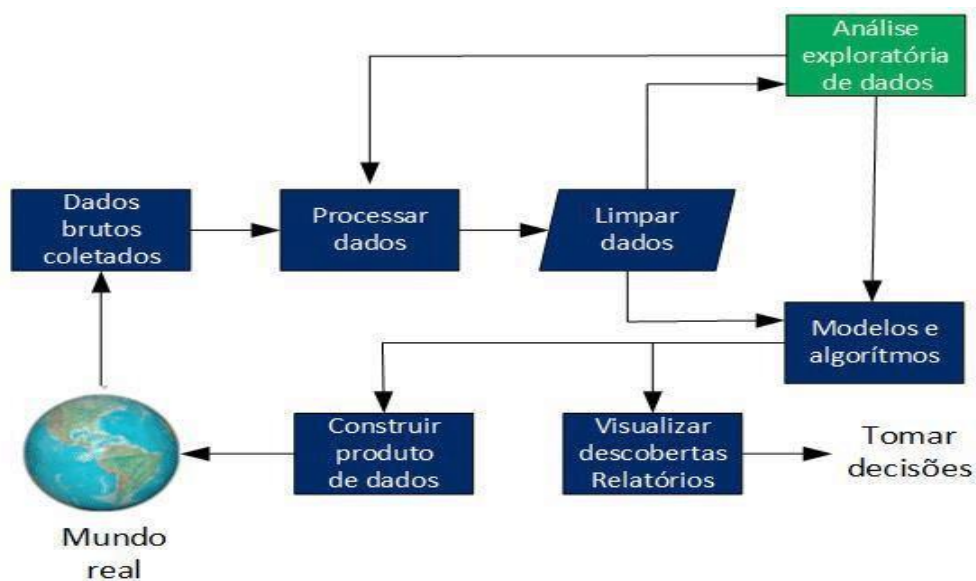
Com as informações do gráfico acima, é possível verificar que mesmo com um ajuste no modelo e com um intervalo de confiança de 95%, os anos de 2012, 2013, 2014, 2015, 2016 e 2020 ficaram fora da estimativa. A demonstração computacional proposta retoma, assim, a reflexão de possíveis situações que influenciam o comportamento de dispersão dos dados. As

situações sendo exploradas devem ser categorizadas como variáveis explicativas. No modelo anterior, a observação principal foi considerada a quantidade de artigos e a explicativa o ano da publicação. O modelo deve ser ajustado ao número de relações que serão observadas e à qualificação dos valores como qualitativos ou quantitativos. A análise de regressão que envolve apenas uma variável explicativa é chamada de regressão simples, utilizada na demonstração, enquanto a análise envolvendo duas ou mais é denominada regressão múltipla (HAIR Jr. *et al.*, 2005). A regressão linear múltipla é dada por (Equação 2):

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

Para um estudo mais aprofundado, outras constatações devem ser inseridas no modelo. Como mencionado anteriormente, existem várias situações que podem explicar o fenômeno, como fomento à pesquisa, eventos científicos e a própria intensificação do digital na sociedade do séc. XXI. Toda a metodologia de pesquisa utilizada neste trabalho segue o proposto na fig.2.

Figura 2 - Metodologia do Processo de Ciência de Dados



Fonte: João Pedro Albino (2017).

O processo de Ciência de Dados parte da análise do mundo real, levantando os questionamentos e definindo os objetivos. Após essa primeira avaliação, os dados brutos são coletados, ou minerados, das diferentes estruturas disponíveis. Para isso, no estudo apresentado foi utilizado o Google Acadêmico e para apurações em toda rede o buscador Google. A análise exploratória dos dados comporta, inicialmente, as fases de processamento e limpeza dos dados,

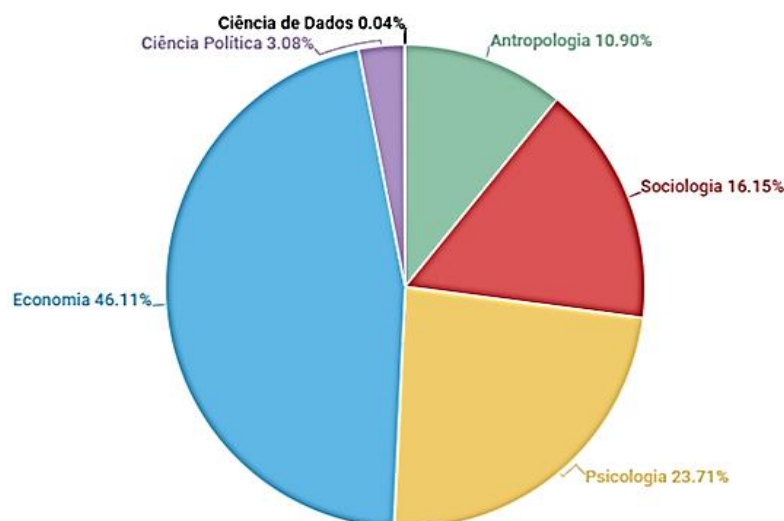
para então escolher o modelo e os algoritmos que farão a obtenção de valores ou conhecimento. Esse processo pode ou não utilizar a inteligência artificial e, a depender da escolha por resultados direcionados ou de uma análise exploratória, define-se a utilização de métodos supervisionados e não supervisionados. Existem inúmeras abordagens e métodos de aprendizado de máquina. Eles podem ser classificados como métodos de aprendizado supervisionado ou não supervisionado. O que difere as duas classificações é a presença ou não de um rótulo ou classificador nos dados a serem analisados (De Oliveira, 2020).

Após o desenvolvimento de produtos gráficos, vislumbra-se as descobertas, que são úteis para o processo de tomada de decisão, e construir produtos de dados, que podem ser softwares, aplicativos, entre outros, que impactem o mundo real e apontem uma solução para o questionamento inicial, retornando ao ciclo do processo, agora com a estrutura alterada pela solução.

Produção e análise gráfica

A partir dos resultados obtidos após o processo de Ciência de Dados proposto na metodologia apresentada na figura 2, desenvolvemos os produtos gráficos representados nas figuras 3a e 3b. Na figura 3a, que apresenta a frequência relativa dos resultados acadêmicos, a visualização gráfica foi construída on-line no programa Infogram.

Figura 3a - Gráfico de frequência relativa dos resultados acadêmicos

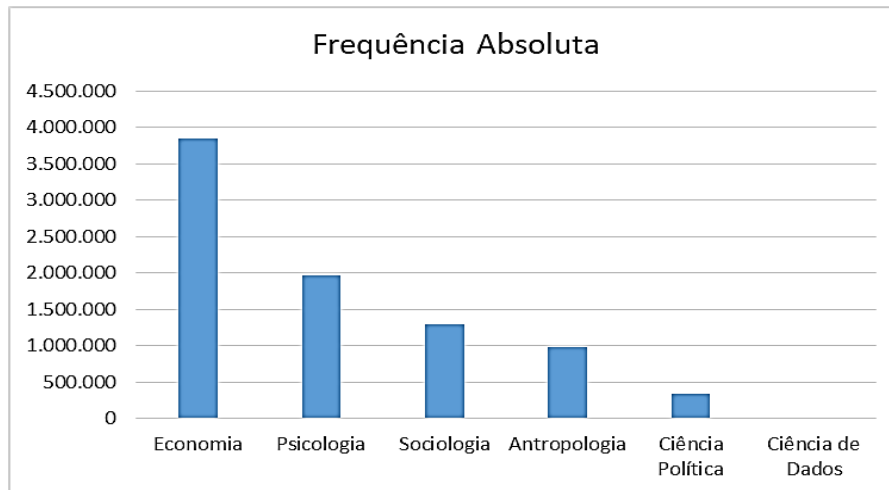


Elaboração própria.

Na figura 3b, sobre a frequência absoluta e cujos dados foram obtidos no Google Acadêmico, os resultados trazem as quantidades de trabalhos acadêmicos por cada área: “Sociologia”, “Economia”, “Antropologia”, “Psicologia”, “Ciência Política”, “Ciência de

Dados”. As áreas pesquisadas são consideradas constituintes da Ciência de Dados Sociais, conforme o já citado *University of Copenhagen’s MSc of Social Data Science*.

Figura 3b - Gráfico de frequência absoluta dos trabalhos acadêmicos das áreas constitutivas dentro das Ciência de Dados Sociais



Elaboração própria.

Já na tabela da figura 3c estão os valores referenciais para construção do gráfico de frequência absoluta apresentado na figura 3b, onde as investigações foram feitas no repositório do Google Acadêmico, dentro dos limites de um pesquisador independente. A estratégia de mineração foi “artesanal”, visto que a automatização da coleta de informações, ainda está em desenvolvimento, principalmente no uso e proteção dos dados. No processo é possível vislumbrar algumas questões que circundam o debate sobre a ética em relação ao objeto de pesquisa.

58

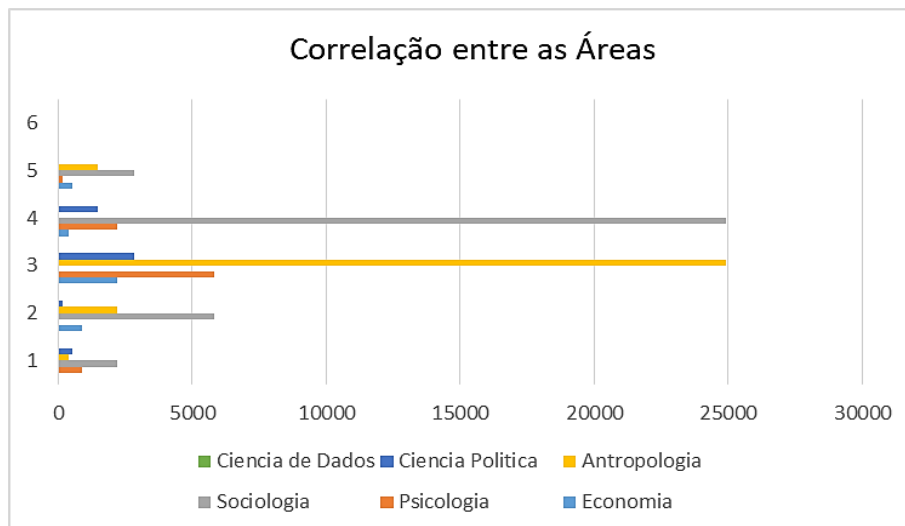
Figura 3c - Tabela de frequência absoluta dos trabalhos acadêmicos das áreas constitutivas dentro das Ciência de Dados Sociais

TABELA DE FREQUÊNCIA ABSOLUTA	
ÁREAS PESQUISADAS	QUANTIDADE DE TRABALHOS ACADÊMICOS
Economia	3.850.000
Psicologia	1.970.000
Sociologia	1.290.000
Antropologia	982.000
Ciência Política	331.000
Ciência de Dados	2.400
TOTAL DE RESULTADOS	8.425.400

Elaboração própria.

A figura 4a (Correlação entre as áreas) teve seus dados obtidos no Google Acadêmico. Os resultados são as correlações entre todas as áreas constitutivas e a visualização gráfica foi construída no programa Excel.

Figura 4a - Correlação entre todas as áreas constitutivas da proposta de Ciência de Dados Sociais



Elaboração própria.

59

A figura 4b (Fragmento Script para Dendograma) traz o código utilizado para gerar os dendogramas da figura 4c (Dendogramas de Correlação), que apresenta uma exemplificação de métodos para construção de dendogramas. Os dendogramas nos auxiliam a verificar as possibilidades de agrupamentos. Com essa estratégia, a análise do dendograma é realizada a cada nível da hierarquia para tentar identificar os conceitos que descrevem os exemplos nos *clusters* (Metz, 2006). Ambas as figuras foram geradas na linguagem R.

Figura 4b – Fragmento Script para Dendograma

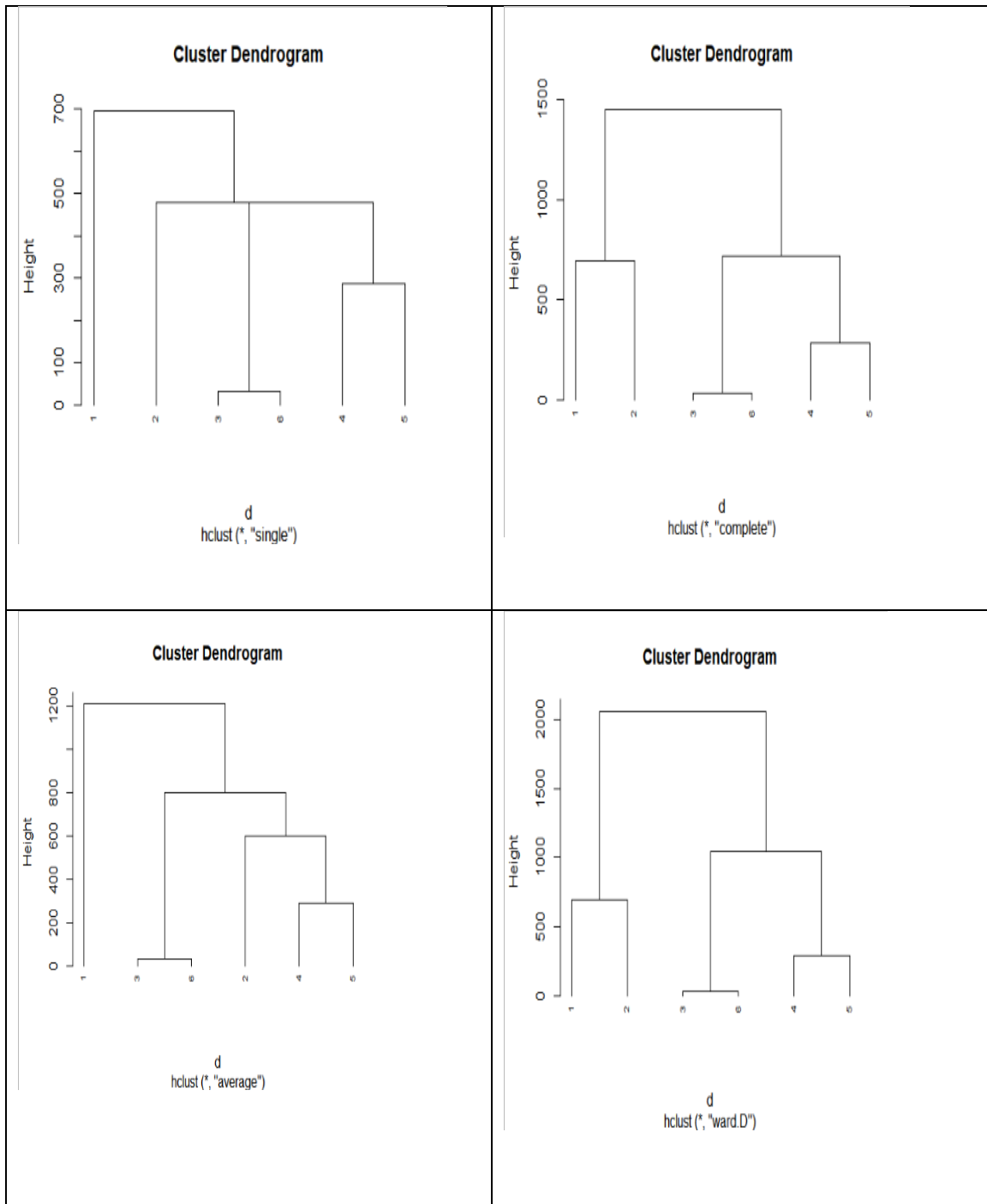
```
#CALCULANDO MATRIZ DE DISTANCIAS
d <- dist(areas_totalidade, method = "euclidean")
d

#DEFININDO O CLUSTER A PARTIR DO METODO ESCOLHIDO
#metodos disponiveis "average", "single", "complete" e "ward.D"
hc1 <- hclust(d, method = "single" )
hc2 <- hclust(d, method = "complete" )
hc3 <- hclust(d, method = "average" )
hc4 <- hclust(d, method = "ward.D" )

#DESENHANDO O DENDOGRAMA
plot(hc1, cex = 0.6, hang = -1)
plot(hc2, cex = 0.6, hang = -1)
plot(hc3, cex = 0.6, hang = -1)
plot(hc4, cex = 0.6, hang = -1)
```

Elaboração própria.

Figura 4c - Dendogramas de correlação



Elaboração própria.

Análise das representações gráficas

Na figura 3a, é possível perceber a disposição dos trabalhos acadêmicos dentro das áreas que formam a Ciência de Dados Sociais, conforme proposta apresentada. Dentro da análise gráfica, fica explícito que as produções das áreas de Ciências Políticas e Ciência de Dados,

3,08% e 0,04% respectivamente, são as que possuem menor produtividade. Essa informação nos sugere um ponto de atenção devido à grande importância desses dois campos de pesquisa, principalmente relacionados à era de dados pessoais que constituem uma parte importante da identidade digital.

As identidades construídas a partir da virtualização no ciberespaço desconstruem a percepção de tempo e espaço, e este indivíduo, sujeito à interface do computador, torna-se onipresente e onisciente, em virtude de poder estar em qualquer lugar em qualquer tempo num clique, e no mesmo instante se comunicar com diversas pessoas em qualquer lugar do planeta não levando em conta o tempo cronológico e a distância física – que também se torna relativa em virtude da interface (Azevedo, 2014:10).

Além disso, o mesmo autor, no estudo sobre a identidade digital, reforça que essas mudanças desconstruíram o entendimento de espaço e lugar, transformando o que era concreto e físico nas relações sociais em uma representação simbólica e virtual (Azevedo, 2014).

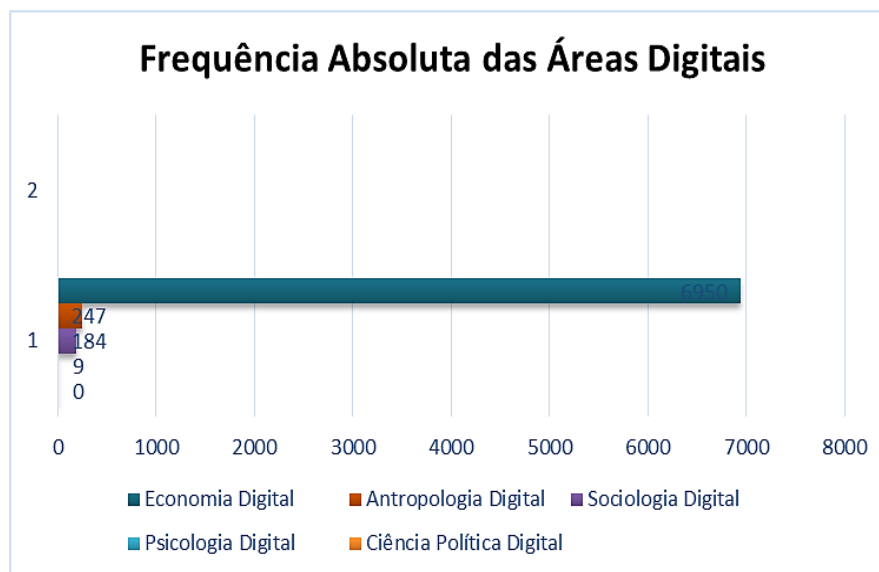
Na figura 3b e 3c essa percepção é ainda mais evidente, tendo a área da Economia cerca de 3.850.000 resultados, enquanto a Ciência de Dados apenas 2.400.

Na figura 4a, as correlações entre as áreas propostas são analisadas, evidenciando a percepção de que as áreas de Sociologia e Antropologia são as que mais possuem relação, seguidas pelas áreas de Sociologia e Psicologia. A mesma percepção é possível na demonstração dos modelos de dendogramas presentes na figura 4c.

Cabe aqui uma ênfase na questão da área de Sociologia e Ciência política, que durante as buscas relacionadas à Ciência de Dados retornou o endereço eletrônico de *startups* que trabalham com levantamento de dados para projetos sociais. As empresas foram categorizadas e organizadas em uma tabela, porém não serão divulgadas, pois o intuito do trabalho vai em direção contrária ao que essa ação ocasionaria. A reflexão gira em torno da crítica à falta de formação específica para as Ciências de Dados Sociais e à existência de empresas que atuam nesse setor, mas que não devolvem à sociedade a orientação científica da estrutura de pesquisa.

Tendo em vista que as análises propostas estão estruturadas para o direcionamento de resultados específicos, e com o intuito de identificar produções científicas na área de Ciência de Dados Sociais ou os conceitos que organizam as estruturas iniciais, o processo para obtenção de dados é retomado, agora relacionando as áreas propostas com o conceito “digital”, devido à ênfase na utilização do conceito dentro dos entendimentos da atual sociedade. Os resultados são as pesquisas de “Economia Digital”, “Psicologia Digital”, “Sociologia Digital”, “Antropologia Digital”, “Ciência Política Digital”, conforme figura 5.

Figura 5 - Gráfico da frequência absoluta das áreas constitutivas da proposta com o termo “digital”



Elaboração própria.

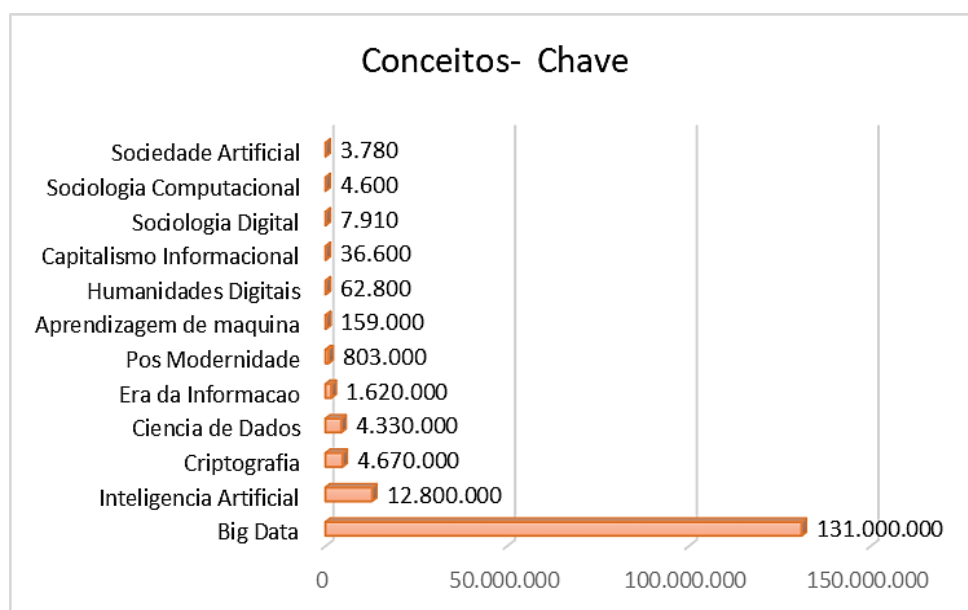
Novamente fica perceptível a disparidade entre as produções científicas nas áreas digitais, tendo a Economia Digital um número de 6.950 publicações, seguida da área de Antropologia Digital, com 247 publicações. A Sociologia Digital aparece com 184 publicações e a Psicologia Digital com 09 publicações. A área da Ciência Política Digital não retorna resultados. A reflexão indica novamente o fato de que as ferramentas da Ciência de Dados estão sendo utilizadas, inclusive em vários projetos da Ciência Política, como acompanhamento de agendas, controle de gastos públicos, acesso à informação, porém sem uma estrutura metodológica que oriente as diretrizes éticas dessas pesquisas, muitas vezes organizadas e patrocinadas por empresas e instituições ligadas ao objeto de pesquisa.

Para entender a relação da nossa atual sociedade com os meios digitais, interligando a conexão entre as Humanidades Digitais, Ciência de Dados Sociais e a estrutura proposta neste trabalho, foi orientada e supervisionada uma investigação simples no Google com o intuito de relacionar os conceitos, ou palavras-chave, que ativem os subsunçores ou “gatilhos” da cultura digital. A definição de subsunçores faz parte da Teoria da Aprendizagem significativa, proposta por David Ausubel, que defende que o aprendizado ocorra de maneira que o indivíduo participe da concepção da nova informação, relacionando-a com conceitos pré-existentes em sua estrutura cognitiva – chamados de subsunçores –, aos quais ela irá se ancorar (Ausubel, Novak & Hanesian, 1978).

Em termos simples, subsunção é o nome que se dá a um conhecimento específico, existente na estrutura de conhecimentos do indivíduo, que permite dar significado a um novo conhecimento que lhe é apresentado ou por ele descoberto. Tanto por recepção como por descobrimento, a atribuição de significados a novos conhecimentos depende da existência de conhecimentos prévios especificamente relevantes e da interação com eles (Moreira, 1999:02).

Na figura 6 está o gráfico de frequência absoluta dos conceitos relacionados com a estrutura neural pré-existente. Os dados foram obtidos no buscador Google, em toda a rede. Os resultados formam a proposta de conceitos-chave para a construção do conhecimento.

Figura 6 - Gráfico de frequência absoluta dos conceitos-chave



Elaboração própria.

Desenvolvimento

Em trabalhos científicos e acadêmicos, é comum o destaque do estado da arte como mecanismo para se estabelecer um ponto de partida para o estudo proposto, pois permite que seus leitores compreendam o conhecimento e as ferramentas disponíveis ao autor do estudo no momento em que o trabalho foi realizado, assim como os acréscimos sendo ofertados ao estado de conhecimento atual. No entanto, devido à já mencionada escassez de artigos sobre o assunto, o levantamento de dados bibliográficos para direcionar a construção da análise do estado da arte da Ciência de Dados Sociais no Brasil atual é muito difícil. Fica explícito, porém, que a área da Economia, inclusive em aliança com a Ciência de Dados, está em desenvolvimento, o que traz a urgência da definição da área e suas metodologias para a construção de uma filosofia ética e moral em relação a todo o processo. A reflexão se dá para além das questões do âmbito da política de dados, como em qualquer entendimento social que se estende para a base da

sociedade, para aquela parcela da população que está desamparada da proteção judicial, marginalizada desde a fundação do Estado moderno.

O entendimento ético e moral da atuação de um cientista de dados deve ser construído nos moldes da visão sociológica, alimentado pela digitalização da própria disciplina, buscando se enraizar na metodologia e na estrutura disciplinar da formação profissional. Não é correto limitar o entendimento às “*soft skills*”, ou habilidades interpessoais, utilizadas pelo setor de recursos humanos. Esse entendimento superficial conduz o indivíduo que busca formação profissional na área de Ciência de Dados para as mãos do charlatanismo de empresas que prometem direcionamentos pessoais e profissionais. O papel de desenvolvimento do indivíduo é fundamentalmente da educação básica, que precisa de uma visão holística para caminhar junto com a nova realidade.

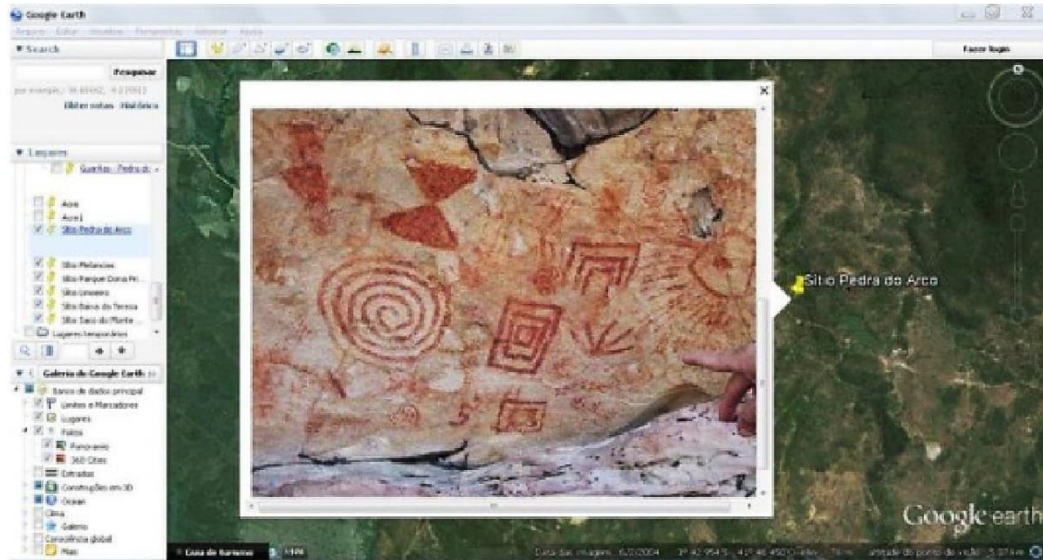
Para entendimento geral, devemos pensar individualmente os termos que formam a Ciência de Dados Sociais. Retornando à definição de ciência, pode-se entender por que o tipo de conhecimento que busca compreender as verdades para explicar o funcionamento do objeto em estudo deve passar por uma série de testes para validação do conhecimento obtido. Já o conceito de dados traz uma definição mais ampla devido ao fato de que dados não são necessariamente digitais, são informações obtidas através de alguma forma de pesquisa ou extração. É através dos dados que se obtêm as informações pertinentes ao objeto de estudo.

Refinando a definição para dados digitais, delimita-se que são representações da informação de forma abstrata que podem ser manipuladas por meios digitais. Conforme Telles (2017), o digital é produto da ação humana, construído a partir das decisões humanas. Os dados digitais também estão englobados nessa compreensão, visto que são objetos que passaram por um processo de codificação e transmissão digital. Deve-se refletir sobre o eixo tecnológico da atual sociedade como a forma que direciona a troca de informação e conhecimento, mas não aprisionar a reflexão e negligenciar o passado humano. A tecnologia sempre esteve presente no desenvolvimento das sociedades, atribuindo novas formas de se pensar soluções e encontrar caminhos mais amplos, originando, assim, transformações a partir das quais baseiam-se os marcos históricos e a direção seguida no transcorrer do tempo e da construção do conhecimento.

Ao direcionar o exemplo às pesquisas arqueológicas feitas utilizando o geoprocessamento de imagens por satélite, encontram-se modelos onde sítios arqueológicos são explorados por meio das tecnologias digitais, categorizando-os como dados digitais, porém o componente específico do estudo faz parte de uma sociedade do passado, anterior ao processo de “digitalização” da cultura. Na figura 7 é possível acompanhar as marcas do passado e as marcas do presente através da codificação da mesma informação em diferentes tecnologias, separadas

pela ação do tempo cronológico, porém com a possível ligação da necessidade humana de transmitir conhecimento.

Figura 7 - Sítio arqueológico identificado através de geolocalização (dado codificado e transmitido) digitalmente



Fonte: Lanna Oliveira (2015).

65

Cabe aqui explicitar o entendimento de Humanidades Digitais, sendo ela o grande contexto para análises de dados sociais, mas antes de esboçar tal definição, deve-se refletir sobre a cultura em seu aspecto geral. Segundo Eagleton (2005:184), “a cultura não é unicamente aquilo de que vivemos. Ela também é, em grande medida, aquilo para o que vivemos. Afeto, relacionamento, memória, parentesco, lugar, comunidade, satisfação emocional, prazer intelectual, um sentido de significado último”.

Sendo assim, é possível afirmar que o processo de vivência cultural de uma sociedade está ligado ao seu modo de produção, sendo este a maneira como uma sociedade produz seus bens e serviços, como os utiliza e os distribui.

O modo de produção constitui um objeto abstrato formal que, no sentido rigoroso do termo, não existe na realidade. Os modos de produção capitalista, feudal, escravagista, constituem igualmente objetos abstratos formais, visto também não possuírem essa existência. De fato, existe apenas uma formação social historicamente determinada, isto é, um todo social – no sentido mais vasto – num dado momento de sua existência histórica (Balibar, 1980:170).

Vislumbrar-se então que, simultaneamente ao modo de produção feudal, a humanidade já inclinava sua vivência cultural ao modo de produção capitalista, da mesma forma que a nossa atual sociedade vivencia o “capitalismo informacional”.

O capitalismo informacional, cognitivo ou do conhecimento corresponde ao desenvolvimento do capitalismo. O termo foi definido primeiramente pelo sociólogo espanhol Manuel Castells, em sua obra “A Sociedade em Rede”, escrita em 1996 e publicada em 2006. O capitalismo informacional teve início com a quebra da bolsa de valores de Nova York, em 1929, ganhando força na virada do século. No entanto, há controvérsias sobre sua data de origem. Para alguns estudiosos, ele começou no período do pós-guerra e, para outros, a partir da década de 80.

O atual modo de produção está centralizado na tecnologia, principalmente no que diz respeito à informação. Vivemos uma cultura digital dentro do modo de produção do capitalismo informacional. Isso quer dizer que toda cultura produzida atualmente provém do meio digital? Essa questão inicia o desenvolvimento do conceito de Humanidades Digitais.

As Humanidades Digitais têm seu marco inicial com o documento “Manifesto das Humanidades Digitais” que, conforme Castro (2020), é o instrumento que cunha oficialmente a nomenclatura. Para o autor, a tendência irreversível de criação de fontes digitais colocou às Ciências Humanas o desafio de incorporar novos métodos à sua tradicional metodologia de pesquisa. Esse desafio implica uma nova forma de trabalhar, tornando essencial sua reflexão pelos envolvidos nela.

Dentro da reflexão sobre o campo, Telles (2017) nos propõem que a estrutura das Humanidades Digitais pode ser pensada com uma interdisciplinaridade do conhecimento interessado na reflexão sobre produção, apropriação e usos das tecnologias digitais no meio acadêmico. “As humanidades digitais estão mais relacionadas à reinvenção da universidade e de seu papel na sociedade contemporânea” (Telles, 2017:13). O autor ainda reforça que, para além da produção de conhecimento, há também a questão do acesso e da difusão do conhecimento nas plataformas digitais.

Conclusão

Analisando o processo da Ciência de Dados e a metodologia em desenvolvimento para as Ciências de Dados Sociais, é preciso definir a aplicação no mundo real, a interação entre mercado e meio acadêmico. Faz-se necessário o entendimento de alguns conceitos mercadológicos como “*gap* analítico”, que é a diferença entre a capacidade de análise e a capacidade de armazenamento, e desta com a produção atual de dados. Esse fenômeno é chamado de “dilúvio de dados”. A produção de dados é muito maior do que a capacidade de armazenamento que, por sua vez, é muito maior do que a capacidade de análise humana. É nesse ponto que o interesse comercial tem um foco maior e as produções científicas não estão

acompanhando. O fato é inverso ao ocorrido na história da computação, que se desenvolveu nas academias, quando não existia valor comercial para a sua utilização. Porém, com o advento do *big data*, as informações dele derivadas se tornaram essenciais, mas são utilizadas sem a estrutura de reflexões propostas pela transdisciplinaridade que ocorre nos meios acadêmicos.

A definição de implicação social é ainda mais ampla e urgente, pois envolve tudo que é produzido e interligado nas sociedades humanas, inclusive as relações com passado e futuro. Ao delimitar o nosso entendimento ao atual nexos tecnológico - quando nosso comportamento se constrói por meio das redes sociais, nossos acessos são feitos por dispositivos móveis, todo o processamento pode ser feito em nuvem e o contexto geral é o *big data* -, constata-se que todo esse processo é anotado em formato digital e o nosso nexos tecnológico cria cadeias de transações digitais associadas à nossa identidade física.

A virtualização do sujeito pós-moderno a partir da cibercultura traz o questionamento sobre quem é esse sujeito pós-moderno e é possível definir uma identidade para este ente que ora habita no plano físico e ora no campo virtual, todavia, no decorrer da evolução da tecnologia tem feito êxodos para o ciberespaço, dedicando mais tempo à sua vida virtual, do que a sua vida atual (Azevedo, 2014:12).

Atinge-se enfim a um ponto crucial para a definição do contexto para análise de dados sociais. O pesquisador da área de Ciência de Dados Sociais não pode ter o foco somente no nexos tecnológico atual, afinal isso limitaria a percepção das relações com o passado. Porém a metodologia precisa ser estruturada dentro dessa nova dimensão da realidade, devido aos dados que passam por um processo de digitalização estarem localizados no âmbito dos dados digitais. Lupton (2015) afirma que é necessário pensar os *big data* enquanto artefatos socioculturais formados sobre processos políticos, sociais e culturais que afetam diretamente a liberdade das pessoas e os direitos civis. Muitos projetos que ocorrem atualmente – nos quais inclusive existe um debate sobre o valor de documentos digitalizados e a possível perda de memória nacional no tocante aos arquivos públicos – são de extrema importância para a base de dados e para os projetos de “*web scraping*” (raspagem de dados) dos repositórios digitais, que organizam as estruturas de códigos e o banco de dados para facilitar as consultas e que estão em crescente desenvolvimento. Por isso, a sociologia digital crítica deve debater o impacto dessas informações (Lupton, 2015).

Deve-se manter um amplo raciocínio sobre a utilização de todo resultado de pesquisa produzido, seja acadêmico ou mercadológico, por parte dos detentores dos meios de produção. A ciência deve servir à sociedade, transformando-a em uma estrutura justa e igualitária, eliminando as bases hierárquicas de detenção de poder, seja político, cultural ou simbólico, que mantém as vestes pseudodemocráticas que servem aos interesses daqueles que dominam o atual modo de

produção capitalista informacional. Nisso a Ciência de Dados Sociais pode ajudar, pois emerge da necessidade de orientação sobre as metodologias, métodos e interesses em todo o contexto de dados sociais, então analisados sob a ótica, ainda em desenvolvimento, das Humanidades Digitais.

Referências

- ALBINO, João Pedro (2017), “Uma proposta para análise de dados abertos utilizando a linguagem R”. *Revista dos Mestrados Profissionais*, v. 6, n. 2, pp. 45-62 [Consult. 20/03/2021]. Disponível em <https://periodicos.ufpe.br/revistas/RMP/article/download/244961/35106>
- AUSUBEL, David P., NOVAK, Joseph D. y HANESIAN, Helen (1978), *Educational Psychology. A cognitive view*, 2a. ed., New York, Holt, Rinehart and Winston.
- AZEVEDO, Thiago Guimaraes (2014), “Identidade Digital: A crise das identidades no ciberespaço”. *ARTEFACTUM-Revista de estudos em Linguagens e Tecnologia*, v. 8, n. 1 [Consult. 20/11/2021] Disponível em <http://artefactum.rafrom.com.br/index.php/artefactum/article/view/225/280>
- BALIBAR, Etienne (1980), “Os conceitos fundamentais do materialismo histórico”, in L. Althusser; E. Balibar, R. Establet. *Para ler o capital Vol.02*. Rio de Janeiro, Zahar.
- CASTELLS, Manuel (2008), *A Sociedade em Rede*. 11 ed. São Paulo, Paz e Terra.
- CASTRO, Renan M. (2020), *Análise da literatura das humanidades digitais: uma proposta bibliométrica para descrição de seu escopo e congruência conceitual*. Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade Federal do Rio de Janeiro [Consult. 30/03/2021]. Disponível em <http://ridi.ibict.br/bitstream/123456789/1096/1/TESE%20VERSA%CC%83O%20DEFESA.pdf>
- CHEIN, Flávia (2019), *Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas*. Brasília, Enap.
- DE OLIVEIRA, Lucas Augusto Ferreira *et al* (2020), *Análise de texto não supervisionada. Aplicações: setores químico e elétrico*. Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química da Escola de Engenharia da Universidade Federal de Minas Gerais.
- EAGLETON, Terry (2005), *A idéia de cultura*. Trad. Sandra Castello Branco. São Paulo, Unesp.
- GIDDENS, Anthony (2005), *Sociologia*. Porto Alegre, Artmed.
- GOFFMAN, Erving (1988), *Estigma: notas sobre a manipulação da identidade deteriorada*. Rio de Janeiro.
- GRIMMER, Justin; BRANDON M. Stewart (2013), *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis*, v. 21, Issue 3, Summer, pp. 267 – 297 [Consult. 23/04/2021]. Disponível em <https://doi.org/10.1093/pan/mps028>
- HAIR, Joseph F.R.; BABIN, Babin.; MONEY, Arthur H; SAMOUEL, Phillip (2005), *Fundamentos de métodos de pesquisa em administração*. Porto Alegre, Bookman.
- HAWKINS, Douglas M. (1980), *Identification of Outliers*. Chapman and Hall, London/New York.
- LUPTON, Deborah (2015), *Digital Sociology*. [S.l.], Routledge.

- MANIFESTO das humanidades digitais (2011) [Consult. 20/11/2020]. Disponível em <https://humanidadesdigitais.org/manifesto-das-humanidades-digitais/>
- METZ, Jean (2006), *Interpretação de clusters gerados por algoritmos de clustering hierárquico*. Tese de Doutorado. Universidade de São Paulo.
- MOREIRA, Marco Antonio (1999), *Aprendizagem significativa*. Brasília, editora da UnB.
- NASCIMENTO, Leonardo F. (2016), “A sociologia digital: um desafio para o século XXI”. *Sociologias*, v. 18, p. 216-241 [Consult. 20/11/2020]. Disponível em <https://doi.org/10.1590/15174522-018004111>
- OLIVEIRA, Lanna (2015), “Estudo prático sobre a geolocalização de sítios arqueológicos no Google Earth”. *Revista Arqueologia Pública*, n. 9, pp. 35-46 [Consult. 28/04/2021]. Disponível em https://www.researchgate.net/publication/322904669_Estudo_pratico_sobre_a_geolocalizacao_de_sitios_arqueologicos_no_Google_Earth
- PETERNELLI, Luiz Alexandre (2005), *INF 162 - Estatística I* [Consult. 20/11/2021]. Disponível em em: <http://www.dpi.ufv.br/~peterneli/inf162.www.16032004/materiais/CAPITULO9.pdf>
- PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de (2007), *Conhecendo o R: uma visão estatística*. Viçosa, Editora UFV.
- TELLES, Helyom Viana (2017), “História digital, Sociologia digital e humanidades digitais: Algumas questões metodológicas”. *Revista Observatório*, v. 3, n. 4, pp. 74-101 [Consult. 05/05/2021]. Disponível em <https://sistemas.uft.edu.br/periodicos/index.php/observatorio/article/view/3810/11270>

Abstract

This work aims to locate and contextualize some scientific and methodological definitions for the emerging field of Social Data Science. As it is a highly new area, the search encompasses guidelines to organize this new look at the Social Sciences, using authors and methods from the constitutive areas of Computational Social Sciences and Digital Humanities, including Digital Sociology. Following the process proposed by Data Science, the research brings reflections on the making and use of graphic products as effectors of extracting information and conducting human decision making or machine learning. Human development linked to technology is relatively new, and this work relates to this moment of humanity.

Keywords: Digital Humanities; Digital Sociology; Social Data Science; Computational Social Sciences; Information.

Resumen

El presente trabajo tiene como objetivo localizar y contextualizar algunas definiciones científicas y metodológicas para el campo emergente de la Ciencia de Datos Sociales. Por tratarse de un área sumamente nueva, la búsqueda engloba pautas para organizar esta nueva mirada a las Ciencias Sociales, utilizando autores y métodos de las áreas constitutivas de las Ciencias Sociales Computacionales y Humanidades Digitales, incluida la Sociología Digital. Siguiendo el proceso propuesto por Data Science, la investigación trae reflexiones sobre la elaboración y uso de productos gráficos como efectores del proceso de extracción de información y realización de decisiones humanas o aprendizaje automático. El desarrollo humano vinculado a la tecnología es relativamente nuevo y este trabajo se relaciona con este momento de la humanidad.

Palabras-clave: Humanidades digitales; Sociología digital; Ciencia de datos sociales; Ciencias Sociales Computacionales.
