

MAPEAMENTO CIENTOMÉTRICO DA ATIVIDADE ACADÊMICA:

pesquisa em Big Data e Humanidades Digitais

Rogério Mugnaini¹
Suzane Albino da Silva Guimarães²
Ednéia Silva Santos Rocha¹

Resumo: Realizou-se um mapeamento cientométrico da atividade acadêmica dos pesquisadores das Universidades Estaduais Paulistas dedicados aos temas Big Data e Humanidades Digitais. Diferentes fontes de dados foram utilizadas: Web of Science, para mapeamento da produção científica; Biblioteca Virtual da FAPESP e Diretório de Grupos de Pesquisa do CNPq, para levantamento de auxílios e grupos de pesquisa; e Plataforma Lattes, para outras atividades acadêmicas. Concluiu-se que estudos cientométricos que contemplem dados de pesquisa de diferentes fontes e atividades se mostraram eficientes para analisar como está se desenvolvendo a pesquisa em tais temáticas nas Universidades Estaduais Paulistas.

Palavras-Chave: Cientometria. Atividade Acadêmica. Big Data. Humanidades Digitais. Universidades Estaduais Paulistas.

1 INTRODUÇÃO

O fenômeno Big Data (BD) vem atraindo a atenção de pesquisadores de diversas áreas, cujo foco visa lidar com volumes massivos de dados que não são manipulados por ferramentas típicas de bancos de dados. Sua complexidade é alvo dos pesquisadores das áreas de Matemática e Ciência da Computação – dado o número de autores que publicaram na Web of Science (WoS) até 2017, com crescimento expressivo a partir de 2013 – e reunido o maior percentual de autores considerados por Lyu e Costas (2021) como especialistas – as áreas de Ciências Sociais & Humanidades (CSHs) vem em segundo lugar.

A presença de pesquisadores de CSHs na produção científica em BD pode estar relacionada à antecedência do campo de “computação das humanidades” (*computing humanities*) cuja origem remonta à metade do século XX, e na virada do século é considerado sob o termo guarda-chuva Humanidades Digitais (HDs). Pode-se então afirmar que BD e HDs são

¹ Universidade de São Paulo (USP)

² Instituto de Pesquisas Energéticas e Nucleares (IPEN)

contemporâneos (SVENSSON, 2009), resultando nos primeiros artigos na WoS, respectivamente, de 2004³ e 1998 (HU; ZHANG, 2017; SU; ZHANG; IMMEL, 2020).

Como observa Smit (2021) sobre as HDs, trata-se de mais do que a justaposição entre áreas (computação e humanidades) – que caracterizaria uma visão utilitarista da primeira, da qual não se espera ter simplesmente a otimização do trabalho da segunda –, mas de uma simbiose que a autora simplifica como a alteração de um modo de produção caracterizado pela retórica para sua subordinação à lógica. Nesse sentido, Sousa (2011) destaca o caráter transdisciplinar das HDs, que incorpora os saberes tradicionais às tecnologias computacionais.

Dessa forma, a presença das CSHs – assim como da Ciência da Informação (CI) – é mais evidente nas HDs do que no BD, conforme observaram, respectivamente, Su, Zhang e Immel (2020) e Hu e Zhang (2017). A CI, BD e as HDs apresentam similaridades que expressam a oportunidade que pesquisadores da CI têm de atuar nesse campo sobretudo nos processos de produção, mediação, acesso e recuperação da informação digital compartilhando ambientes como os bancos e bases de dados, repositórios e metodologias de representação da informação com o uso de metadados (PIMENTA, 2020).

A emergência de tais campos de pesquisa e sua proximidade com a CI justificam o estudo das atividades acadêmicas da comunidade científica com foco nas Universidades Estaduais Paulistas (UEPs). A atualidade do tema sugere um olhar mais amplo, não restrito às publicações, ou “pesquisa demonstrada” – conforme afirmam Davyt e Velho (2000) –, mas também à “pesquisa potencial”, que estariam relacionadas às bolsas e projetos de pesquisa financiados por agências de fomento. O objetivo deste estudo foi mapear a atividade acadêmica dos pesquisadores das UEPs dedicados aos temas BD e HDs.

2 METODOLOGIA

As fontes de dados do estudo foram: *Web of Science*, para mapeamento da produção científica; *Biblioteca Virtual da FAPESP (BV-FAPESP)* e *Diretório de Grupos de Pesquisa do CNPq (DGP/CNPq)*, para levantamento de bolsas e projetos de pesquisa; e *Plataforma Lattes*, para mapeamento de outras atividades acadêmicas como disciplinas, linhas de

³ Diferente de Hu e Zhang (2017), Lyu e Costas (2021) identificam 1993 como o ano mais recente. Contudo uma busca na WoS evidencia que os artigos de anos anteriores a 2004 apenas mencionavam termos como “big data set”. Hu e Zhang (2017) revisaram os resultados e decidiram restringir a busca do termo somente ao título ou palavras-chave do artigo (ignorando o resumo), garantindo que a pesquisa tratasse efetivamente de Big Data.

pesquisa, projetos de pesquisa, participação em bancas, produção técnica e produção científica nas diversas tipologias documentais.

A coleta de dados na WoS, BV-FAPESP e DGP/CNPq realizada no mês de junho de 2020, utilizou os seguintes termos de busca: "big data" or "humanidades digitais" or "digital humanities" or "humanitas digitalis" or "humanidades computacionais" or "computing humanities". Não houve qualquer restrição de ano de publicação ou tipo de documento, mas apenas a presença de ao menos um autor de alguma das UEPs: Universidade de São Paulo (USP), Universidade Estadual de Campinas (UNICAMP) e/ou Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Por fim, realizou-se uma padronização dos nomes dos autores e de instituições evitando duplicidades.

Na WoS identificou-se 171 registros, com participação de 319 autores afiliados a alguma das UEPs. Na BV-FAPESP recuperou-se 94 pesquisadores com auxílio (bolsa ou projeto) nas temáticas. E no DGP/CNPq 20 pesquisadores responsáveis por grupos de pesquisa. Com 6 pesquisadores constantes de ambas fontes de agências de fomento, resultaram 108 pesquisadores. Então procedeu-se a busca do currículo Lattes de cada um, resultando na exclusão de Nicholas Spyridon Vonortas, cujo currículo não foi encontrado. Então os 107 currículos restantes foram exportados no formato XML e convertidos em arquivos CSV, para identificação das atividades acadêmicas nas temáticas.

O total de pesquisadores do estudo é de 408, sendo que apenas 25 (6,1%) foram localizados nas agências de fomento e publicaram na WoS, 294 (72,1%) somente publicaram na WoS, e 89 (21,8%) constavam das agências mas não publicaram na WoS.

As análises se dividiram em duas etapas, a primeira relacionada à produtividade científica e tipo de colaboração dos 319 pesquisadores na WoS, diferenciando UEPs, temas e presença nas fontes das agências de fomento. E a segunda analisou a atividade acadêmica dos 114 pesquisadores registrados nas fontes das agências de fomento, diferenciando UEPs, temas e o fato de haverem ou não publicado na WoS.

3 RESULTADOS

O conjunto de 319 autores que publicaram sobre os temas BD e HDs na WoS pode ser analisado na Tabela 1, considerando primeiramente a instituição de afiliação e a temática. Os totais variam pelo fato de dois autores haverem assinado alguns de seus artigos com dupla instituição (USP e UNESP), e por haver 6 autores que publicaram em cada um dos temas.

A USP apresenta o maior número de autores (186), cuja presença na BV-FAPESP evidencia que a temática vem sendo foco em auxílios à pesquisa, assim como o maior número de líderes de grupo no DGP/CNPq. Em seguida, a UNICAMP com 93 autores, e a UNESP com 42. Por outro lado, a UNESP é a que apresenta o maior percentual de autores com mais de um artigo publicado (26,2%), seguida da USP (19,4%) e UNICAMP (17,2%).

BD é representado por 91,4% dos autores, sendo ainda a temática com maior número de autores que publicaram mais do que um artigo. Na temática BD os 43,5% dos autores mapeados nas fontes das agências de fomento publicaram mais do que um artigo, enquanto que para os demais autores este percentual é de apenas 18,8%. Os dados revelam que os oito autores com algum auxílio à pesquisa (FAPESP) representam o maior percentual.

Nas HDs encontram-se 28 autores, sendo 5 deles mapeados nas fontes das agências, dos quais um publicou dois artigos, e outro autor não mapeado que publicou igualmente dois artigos. A menor quantidade de publicações na WoS é uma característica esperada em áreas de CSHs, que não têm os periódicos *mainstream* como principal canal de publicação.

Tabela 1 - Distribuição dos pesquisadores das UEPs que publicaram na WoS, segundo instituição, tema, agência de fomento e número de documentos publicados

| UEP / Tema | Agência de fomento | Número de documentos por autor | | | | | | Total |
|----------------------|--------------------|--------------------------------|----|----|---|---|---|-------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | |
| UNESP | CNPq | 1 | | | | | | 1 |
| | FAPESP e CNPq | | | 1 | | | | 1 |
| | - | 30 | 7 | 2 | | | 1 | 40 |
| UNICAMP | CNPq | | 1 | | 1 | | | 2 |
| | FAPESP | 4 | | | | 1 | | 5 |
| | - | 73 | 8 | 2 | | 1 | 2 | 86 |
| USP | CNPq | 1 | | | | | | 1 |
| | FAPESP | 8 | 3 | | 1 | | | 12 |
| | FAPESP e CNPq | | | | 1 | | 1 | 2 |
| | - | 141 | 15 | 7 | 4 | 3 | 1 | 171 |
| Total UEPs | | 258 | 34 | 12 | 7 | 5 | 5 | 321 |
| Big Data | CNPq | 2 | 2 | | | | | 4 |
| | FAPESP | 11 | 3 | | 1 | 1 | | 16 |
| | FAPESP e CNPq | | 1 | | 1 | 1 | | 3 |
| | - | 224 | 29 | 11 | 4 | 5 | 3 | 276 |
| Humanidades Digitais | CNPq | | 1 | | | | | 1 |
| | FAPESP | 2 | | | | | | 2 |
| | FAPESP e CNPq | 2 | | | | | | 2 |
| | - | 22 | 1 | | | | | 23 |
| Total Temas | | 263 | 37 | 11 | 6 | 7 | 3 | 327 |

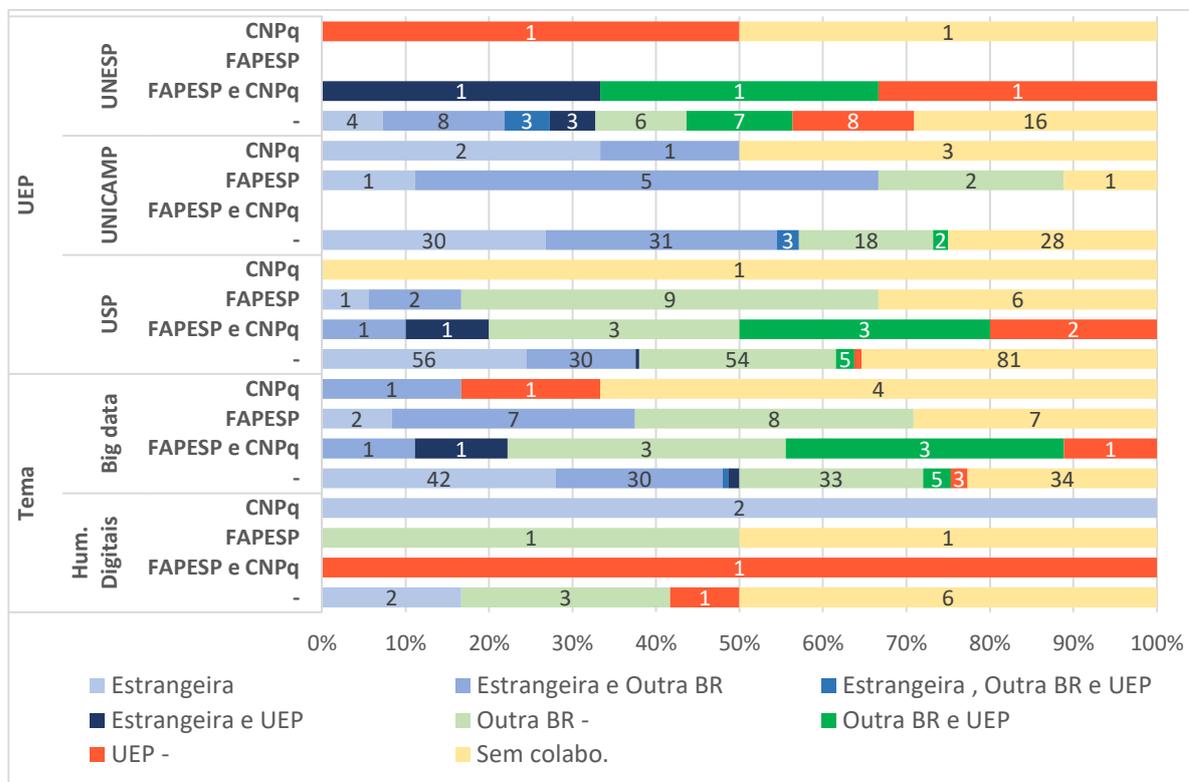
Fonte: Elaborado pelos autores.

No Gráfico 1 a distribuição dos pesquisadores é analisada em relação às colaborações institucionais estabelecidas por meio da co-autoria nos artigos. As três instituições de origem

dos autores são combinadas com demais tipos de instituições (outra UEP, outra brasileira e/ou estrangeira), sendo representados pelas cores das barras. Pelo fato dos perfis de colaboração institucional terem sido identificados para cada um dos 319 autores, somam-se 445 perfis. Nas temáticas, os perfis foram considerados, independente da IEP dos autores, fazendo com que o total de perfis dos 171 artigos chegasse a 206.

Observa-se indicativo de colaboração com instituições estrangeiras entre autores não mapeados nas agências, o que fica claro devido ao volume observado nessa categoria. Ainda na categoria dos não mapeados, a UNESP é a que apresenta perfil mais diversificado, com todos os tipos de colaboração observados, seguida da mesma categoria na temática BD. A USP apresenta três autores que receberam auxílio FAPESP e lideram grupo registrado no CNPQ, que colaboraram com outra UEP e outra brasileira (verde escuro), indicando a execução de atividades acadêmicas, institucionalização e nucleação nacional. UNICAMP e UNESP compartilham três pesquisadores que apresentam colaboração com perfil completo de instituições (envolvendo as duas UEPs, outra brasileira e estrangeira) em um de seus artigos.

Gráfico 1 - Distribuição dos pesquisadores das UEPs que publicaram na WoS, segundo instituição, tema, agência de fomento e tipo de colaboração institucional nos documentos

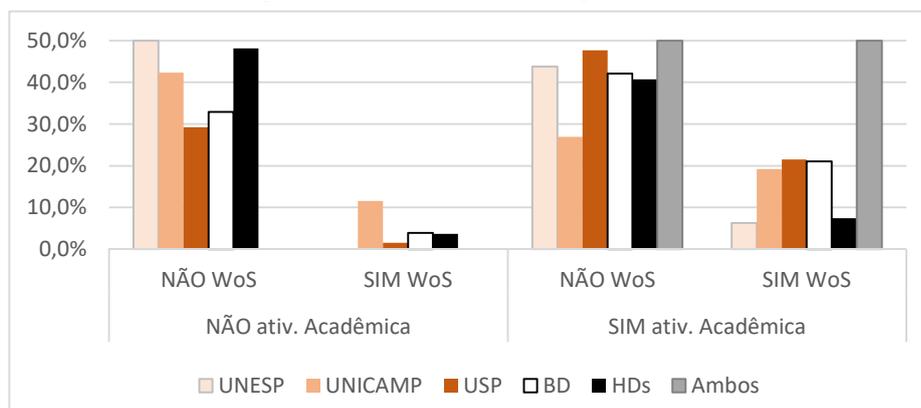


Fonte: Elaborado pelos autores.

A análise dos 107 pesquisadores mapeados nas agências (Gráfico 2) mostra que a maioria (83) que não publica na WoS se subdivide entre aqueles que apresentam ou não atividade acadêmica, com destaque USP e a temática BD, com atividades observadas. Dentre os 24 que publicaram na WoS, a observância de atividade é maior todas as UEPs e temáticas.

O Gráfico 3 reúne as atividades acadêmicas de 65 pesquisadores, apresentando o percentual de pesquisadores que executou cada tipo de atividade. Quatro pesquisadores que se dedicam a ambas as temáticas foram omitidos, sendo que três deles se destacam em diversas atividades, dos quais dois são da CI e da USP. As HDs reúnem 13 pesquisadores com atividades, sendo três da CI e também da USP, destacando-se principalmente em projetos de pesquisa.

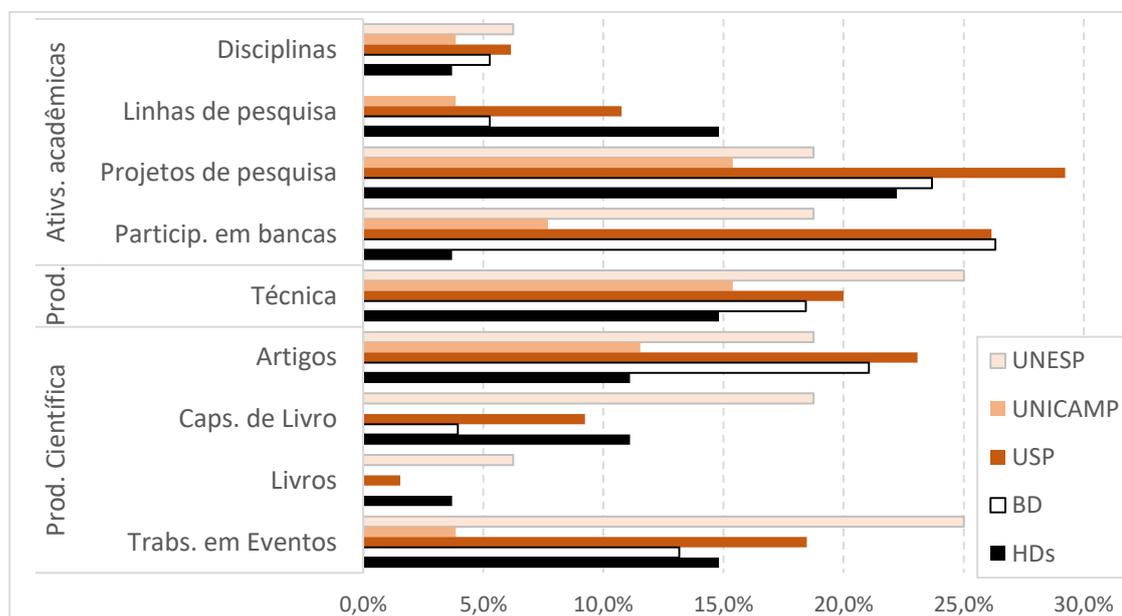
Gráfico 2 - Distribuição percentual dos pesquisadores das UEPs mapeados na BV-FAPESP e/ou DGP/CNPq, segundo instituição, tema, presença ou ausência de atividade acadêmica no currículo e presença ou ausência de publicação na WoS



Fonte: Elaborado pelos autores.

Um total de 48 pesquisadores realizam atividades em BD, sendo 22 de Ciência da Computação, com destaque para participação em bancas, projetos de pesquisa, artigos em periódicos e produção técnica. A USP apresenta um perfil similar, já que responde por aproximadamente 70% dos pesquisadores da temática, com um percentual maior apenas em linhas de pesquisa e trabalhos em anais de eventos. Neste último aspecto a UNESP se destaca em relação à USP, assim como em produção técnica.

Gráfico 3 - Distribuição percentual dos pesquisadores das UEPs mapeados na BV-FAPESP e/ou DGP/CNPq, segundo instituição, tema, tipo de atividade acadêmica e produções técnica e científica



Fonte: Elaborado pelos autores.

4 CONSIDERAÇÕES FINAIS

O desenvolvimento de estudos cientométricos que contemplem dados de pesquisa de diferentes fontes, cotejando tanto a produção científica como as outras atividades científicas como os auxílios de pesquisa, se mostraram eficientes para analisar como está se desenvolvendo as temáticas Big Data e Humanidades Digitais nas Universidades Estaduais Paulistas. A atualidade das publicações na WoS em âmbito mundial evidencia a importância de se recorrer a fontes que permitam observar a criação de infraestrutura de pesquisa cujos resultados em publicações deverão se evidenciar posteriormente.

REFERÊNCIAS

- DAVYT, Amilcar; VELHO, Léa. A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro? **História, Ciências, Saúde-Manguinhos**, [s.l.], v.7, n.1, p.93-116, 2000. DOI 10.1590/S0104-59702000000200005
- HU, Jiming; ZHANG, Yin. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. **Scientometrics**, Budapest, v. 112, n. 1, p. 91-109, 2017. DOI 10.1007/s11192-017-2383-1
- LYU, Xiaozan; COSTAS, Rodrigo. Studying the characteristics of scientific communities using individual-level bibliometrics: the case of Big Data research. **Scientometrics**, Budapest, v. 126, p. 6965–6987, 2021. DOI 10.1007/s11192-021-04034-6

PIMENTA, Ricardo Medeiros. Por que humanidades digitais na ciência da informação? perspectivas pregressas e futuras de uma prática transdisciplinar comum. **Informação & Sociedade: Estudos**, João Pessoa, v. 30, n. 2, 2020. DOI 10.22478/ufpb.1809-4783.2020v30n2.52122

SMIT, Johanna Wilhelmina. A organização do conhecimento e as humanidades digitais: pontos de interseção. *In*: SMIT, Johanna Wilhelmina *et al.* **Humanidades digitais, big data e pesquisa científica**. São Paulo: Fundação Fernando Henrique Cardoso (FFHC), 2021. p. 125-150. Disponível em: <https://fundacaofhc.org.br/files/Humanidades%20Digitais.pdf>. Acesso em: 12 jan. 2022.

SOUSA, Maria Clara Paixão de. **Humanidades digitais: um breve panorama**. [São Paulo]: Universidade de São Paulo, 2011. Disponível em: <https://humanidadesdigitais.org/breve-panorama/>. Acesso em: 12 jan. 2022.

SU, Fangli; ZHANG, Yin; IMMEL, Zachary. Digital humanities research: interdisciplinary collaborations, themes and implications to library and information science. **Journal of Documentation**, Bingley, v. 77, n. 1, p. 143-161, 2020. DOI 10.1108/JD-05-2020-0072

SVENSSON, Patrik. Humanities computing as digital humanities. **Digital Humanities Quarterly**, [s.l.], v. 3, n. 3, 2009. Disponível em: <http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html>. Acesso em: 11 jan. 2022.