


ANÁLISE DE SENTIMENTO EM AMBIENTE URBANO, EDIFICADO OU CITADINO: ANÁLISE BIBLIOMÉTRICA, 2009-2023

IURIA BETCO^{1,2} 

JORGE ROCHA^{1,2} 

RESUMO – O reconhecimento de que diversos aspetos do ambiente urbano podem afetar a saúde mental dos indivíduos tem vindo a aumentar, dado que estes são responsáveis por facilitar ou inibir comportamentos e estilos de vida que impactam o sentimento. A crescente disponibilidade de dados gerados pelos utilizadores nas redes sociais constitui uma fonte de informação sobre os utilizadores e o seu ambiente envolvente, possibilitando a análise de sentimento. Este estudo averigua a produção global de documentos sobre análise de sentimento e ambiente construído, urbano ou citadino, publicados entre 2009 e 2023. Analisaram-se indicadores bibliométricos sobre as tendências, artigos mais citados, autores, instituições, países, agências de financiamento e áreas temáticas de investigação. Foi utilizada a biblioteca bibliométrica do *software R* para a análise de desempenho e o *software VOSviewer* para o mapeamento das conexões científicas. Foram identificados 728 documentos elegíveis de 2009 a junho de 2023. A maioria das publicações eram documentos da área de ciência da computação (n = 487). Este tipo de análises fornece aos investigadores orientações sobre os possíveis pontos críticos que se podem encontrar neste campo de pesquisa, bem como informar acerca da multiplicidade de pesquisas científicas conduzidas globalmente até ao momento por autores, países, fontes, financiamentos atribuídos e instituições ativas, permitindo melhorar a produção de documentos sobre análise de sentimento com relação ao ambiente urbano.

Palavras-chave: Análise de sentimento; ambiente urbano; cidade; edificado; análise bibliométrica.

ABSTRACT – SENTIMENT ANALYSIS IN URBAN, BUILT ENVIRONMENT OR CITY: BIBLIOMETRIC ANALYSIS, 2009-2023. The recognition that different aspects of the urban environment can affect individuals' mental health is growing, as they are responsible for facilitating or inhibiting behaviors and lifestyles that affect sentiment. The increasing availability of user-generated data on social media provides information about users and their surroundings, enabling sentiment analysis. This study investigates the global production of documents on sentiment analysis and the built environment, urban or city, published between 2009 and 2023. Bibliometric indicators on trends, most cited articles, authors, institutions, countries, funding agencies, and thematic research areas were analyzed. The bibliometrix library of R software was used for the performance analysis and the VOSviewer software for mapping scientific connections. A total of 728 eligible papers were identified from 2009 to June 2023. Most publications were computer science papers (n=487). This type of analysis provides researchers with guidance on the possible critical points that can be found in this field of research, as well as informs about the multiplicity of scientific research conducted globally so far by authors, countries, sources, allocated funding, and active institutions, allowing improving the production of documents on sentiment analysis concerning the urban environment.

Keywords: Sentiment analysis; urban environment; city; built fabric; bibliometric analysis.

RESUMEN – ANÁLISIS DEL SENTIMIENTO EN EL ENTORNO URBANO, CONSTRUIDO O DE LA CIUDAD: ANÁLISIS BIBLIOMÉTRICO, 2009-2023. Cada vez se reconoce más que los diversos aspectos del entorno urbano pueden afectar a la salud mental de las personas, ya que son responsables de facilitar o inhibir comportamientos y estilos de vida que repercuten en el sentimiento. La creciente disponibilidad de datos generados por los usuarios en las redes sociales es una fuente de información sobre los usuarios y su entorno que permite analizar el sentimiento. Este estudio investiga la producción mundial de documentos sobre el análisis del sentimiento y el entorno construido, urbano o de la ciudad, publicados entre 2009 y 2023. Se analizaron indicadores bibliométricos sobre tendencias, artículos más citados, autores, instituciones, países, organismos de financiación y áreas temáticas de investigación. Se utilizó la biblioteca *bibliometrix* del programa informático R para analizar el rendimiento y el programa *VOSviewer* para cartografiar las conexiones científicas. Se identificaron 728 documentos elegibles desde 2009 hasta junio de 2023. La mayoría de las publicaciones eran documentos de informática (n=487). Este tipo de análisis orienta a los investigadores sobre los posibles puntos críticos que pueden encontrarse en este campo de investigación, además de informarles sobre la multiplicidad de investigaciones científicas realizadas globalmente hasta la fecha por autores, países, fuentes, financiación asignada e instituciones activas, lo que permite mejorar la producción de documentos sobre análisis de sentimientos en relación con el entorno urbano.

Palavras chave: Análisis de sentimientos; entorno urbano; ciudad; edificado; análisis bibliométrica.

Recebido: 29/11/2023. Aceite: 14/05/2024. Publicado: 25/05/2024.

¹ Centro de Estudos Geográficos, Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Rua Branca Edmée Marques, 1600-276, Lisboa, Portugal. E-mail: betcoiuria@edu.ulisboa.pt, jorge.rocha@edu.ulisboa.pt

² Laboratório Associado TERRA, Universidade de Lisboa, Lisboa, Portugal.

I. INTRODUÇÃO

O ambiente urbano há muito que é reconhecido como um importante determinante de saúde e do bem-estar. Apesar de inicialmente os investigadores apontarem a poluição, a doença e a sobrepopulação como principais determinantes, existem agora provas crescentes dos impactes de muitos outros elementos do ambiente urbano físico na saúde e bem-estar, tanto positivos como negativos (Pineo & Rydin, 2018).

Os planeadores urbanos e os profissionais de saúde têm demonstrado preocupação com o impacto dos ambientes urbanos (*e.g.*, espaços verdes e qualidade do ar) no bem-estar mental dos residentes em áreas urbanas (Hidaka, 2012). Embora a qualidade de vida em geral tenha melhorado em todo o mundo, os problemas de saúde mental aumentaram, o que pode estar associado ao estilo de vida urbano (Centers for Disease Control and Prevention, 2011). De facto, algumas das doenças mentais mais prevalentes parecem ser causadas pelos elevados níveis de *stress* despoletados pela vida nas cidades (Adli *et al.*, 2017; Peen *et al.*, 2010). A crescente urbanização tem um impacto significativo na saúde mental em todo o mundo. Os indivíduos que residem em áreas urbanas enfrentam um maior risco de problemas como depressão, transtornos de ansiedade generalizada, transtornos de adição, transtornos de humor e psicoses (Lederbogen *et al.*, 2011; Peen *et al.*, 2010). Alguns estudos constataram que, em moradores urbanos, o risco de desenvolver transtornos mentais é 38% maior (39% maior para transtornos do humor, 21% maior para transtornos de ansiedade) do que em pessoas que vivem em áreas rurais (Peen *et al.*, 2010). A distribuição de recursos na área da saúde nem sempre é adequada para suprir as necessidades adicionais de serviços nas áreas urbanas, resultando em listas de espera relativamente longas e uma maior pressão para manter os tratamentos e internamentos mais breves, o que compromete a qualidade dos cuidados de saúde mental. Idealmente, a dever-se-ia estabelecer uma correspondência entre a oferta de serviços e a procura por estes cuidados (Peen *et al.*, 2010).

Os estressores, o isolamento e a densidade social são determinantes independentes da saúde, que ocorrem com maior frequência nas cidades (hipótese do estresse social). Na presença de fatores de risco individuais, que diminuem a resiliência de um indivíduo (*e.g.*, genético, relacionado à personalidade, sociodemográfico, idade, pobreza e *status* migratório), o estresse social pode ser relevante para a saúde. O risco de diminuição do bem-estar mental é provavelmente agravado por gradientes socioeconómicos extremos num espaço confinado, condições de habitação inadequadas e violência (Adli *et al.*, 2017).

Existe consenso na literatura de que há uma maior prevalência de transtornos psicóticos em áreas urbanas do que em áreas não urbanas (Galea, 2011). Este crescente problema deve-se, em parte, à crescente desconexão entre as pessoas e a natureza que resulta em estilos de vida mais urbanizados e sedentários (Miller, 2005), existindo provas substanciais que demonstram que o *design* urbano e a arquitetura paisagística são ferramentas preponderantes para melhorar a condição humana e a saúde (Adli *et al.*, 2017; Olszewska & Bil, 2016).

Noutros estudos examinou-se o potencial restaurador mental da exposição ao ambiente natural, bem como dos elementos e sistemas naturais no ambiente urbano construído (Hartig, 1993; Kaplan, 1995). Por exemplo, viver em áreas urbanas com características naturais, como parques, jardins, água, árvores e pássaros, está associado a níveis mais elevados de bem-estar mental e à redução da incidência de doenças mentais crónicas (Bakolis *et al.*, 2018; Cox *et al.*, 2017; Hammoud *et al.*, 2022).

A natureza dinâmica das cidades, entendidas como sistemas complexos com uma variedade de fatores concomitantes, impõe desafios significativos à análise urbana para apoiar os processos de planeamento. Devido à disponibilidade de grandes quantidades de dados, as redes sociais (*e.g.*, X antigo *Twitter*) oferecem a possibilidade de análises espaciais e temporais com grande detalhe, especialmente no que diz respeito aos sentimentos e emoções das pessoas (Kovacs-gyori *et al.*, 2018). A transformação do espaço urbano, e simultaneamente, a digitalização da sociedade, veio possibilitar o conhecimento acerca do quanto o bem-estar individual é influenciado pelo ambiente construído envolvente (Romice *et al.*, 2016). As tendências contemporâneas mostram que a transformação digital impacta todos os ambientes onde as pessoas se conectam, se expressam e se juntam. Assim, o espaço cívico e a sociedade civil já revelam tanto aspetos positivos quanto negativos dessa influência. As tecnologias digitais viabilizaram novas formas de exercer as liberdades de associação, reunião e expressão, ao mesmo tempo que introduziram novas maneiras de limitar esses direitos. Isto faz do bem-estar um caso de estudo ideal para a análise do sentimento (Zunic *et al.*, 2020).

A análise de sentimento também – erradamente – conhecida como *opinion mining*, tem como objetivo classificar automaticamente o sentimento expresso num texto (Zunic *et al.*, 2020). Enquanto, o *opinion mining* extrai e analisa a opinião das pessoas sobre uma entidade, a análise de sentimento procura encontrar opiniões, identificar os sentimentos que expressam e, ainda, classificar a sua polaridade (Medhat *et al.*, 2014). A polaridade de um determinado texto pode ser classificada como positiva, negativa ou neutra, embora existam abordagens com mais níveis de classificação (*e.g.*, os projetos *OpeNER* 2014, que utilizam as classificações positiva, negativa, neutra e fortemente positiva e fortemente negativa; e o *EmoLex*, que utiliza as classificações não associado, fracamente, moderadamente e fortemente associado a um sentimento positivo ou negativo) (Bollen *et al.*, 2011; Mohammad & Turney, 2015).

Apesar da compreensão dos sentimentos e emoções humanas ser um tema de investigação que remonta a várias décadas, o termo "análise de sentimento" e a sua aplicação na extração de informações subjetivas de textos obteve principal destaque a partir do início dos anos 2000, com o surgimento das abordagens computacionais para aferição do afeto (*e.g.*, *opinion mining*, deteção de subjetividade, análise de sentimentos e emoções), que se focam maioritariamente na identificação de opiniões, emoções, sentimentos, avaliações, crenças e especulações (Balahur *et al.*, 2014; Medhat *et al.*, 2014). Nesta década, surgiram investigações marcantes, como a de Bo Pang e Lillian Lee, em 2004, denominada *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*, que introduziu uma nova abordagem para a classificação de sentimentos utilizando técnicas de *machine learning* (Pang & Lee, 2004).

Inicialmente, a análise de sentimentos era particularmente baseada em abordagens menos complexas, como o uso de léxicos de sentimento, que consistem em listas de palavras com polaridades de sentimento associadas. Os léxicos de sentimento são criados manualmente, contendo listas de palavras manualmente rotuladas com polaridade positiva e negativa, permitindo assim, calcular a pontuação geral do sentimento de um determinado texto (Shayaa *et al.*, 2018). As técnicas de Classificação de Sentimento dividem-se deste modo em abordagens de *machine learning*, abordagens Baseadas no Léxico e abordagens Híbridas (Maynard & Funk, 2012).

O rápido crescimento da análise de sentimento está correlacionado com o advento da *Web 2.0* e a crescente disponibilidade de dados gerados pelo utilizador nas plataformas *online* que fornecem serviços de redes sociais (Zunic *et al.*, 2020). A abundância de dados gerados pelo utilizador torna desafiante a análise e obtenção de percepções a partir dos mesmos, levando a um aumento da procura pela análise de sentimento (Afriliana *et al.*, 2022). O incremento das interações dos cidadãos com as redes sociais abriu caminho para a realização de análises de sentimento, permitindo uma maior compreensão da geografia humana nas áreas urbanas (Niu & Silva, 2020a).

Assim, a análise dos sentimentos permite a compreensão dos impactes das características do ambiente no bem-estar individual. Este estudo possibilita uma visão geral bibliométrica dos resultados da investigação sobre análise dos sentimentos em ambiente construído, urbano ou cidade entre 2000 e 2023. A análise centra-se na exploração das tendências e publicações, autores mais impactantes, instituições, fundos, países, áreas temáticas, e possíveis futuros *hotspots* no campo da análise de sentimento.

A análise bibliométrica tem como objetivo resumir grandes quantidades de dados bibliométricos de forma a apresentar o estado da estrutura intelectual e as tendências emergentes de um tópico ou campo de pesquisa (Donthu *et al.*, 2021). Tem sido amplamente utilizada na investigação científica, independentemente do campo de aplicação (Musa *et al.*, 2021), para encontrar tendências emergentes de artigos e revistas, padrões de colaboração e componentes de investigação (*e.g.*, documentos, palavras-chave, autores, revistas e países/territórios), e para explorar a estrutura de um específico domínio na literatura existente (Donthu *et al.*, 2021; Verma & Gustafsson, 2020).

As técnicas de análise bibliométrica dividem-se em duas categorias: (1) análise de desempenho e (2) mapeamento científico. A análise de desempenho é responsável por examinar as contribuições dos componentes da investigação para um determinado campo (Donthu *et al.*, 2021). Existem diversas medidas de análise de desempenho, sendo as mais proeminentes o número de publicações e citações por ano ou por componente de investigação, em que a publicação é um indicador de produtividade, enquanto a citação é uma medida de impacte e influência (Donthu, *et al.*, 2021). Outras medidas, como a citação por publicação, o índice-*h* e o índice-*g* combinam tanto as citações como as publicações para medir o desempenho dos constituintes da investigação. Por exemplo, um índice-*h* de cinco significa que ao autor publicou pelo menos cinco artigos que receberam pelo menos cinco citações cada. Por sua vez, o índice-*g* é calculado com base na distribuição das citações recebidas pelas publicações, ou

seja, um índice-*g* de cinco significa que o autor publicou, pelo menos cinco artigos que, em conjunto, receberam pelo menos 25 citações (g^2). A análise de desempenho, apesar de descritiva, permite reconhecer a importância de diferentes constituintes num campo de investigação.

Por sua vez, o mapeamento científico analisa as relações entre os constituintes da investigação dentro de uma determinada temática (Baker *et al.*, 2021; Cobo *et al.*, 2011; Ramos-Rodríguez & Ruíz-Navarro, 2004). Esta análise diz respeito às interações intelectuais e estruturais entre os constituintes da investigação. As técnicas de mapeamento científico incluem análise de citações, análise de co-citações, acoplamento bibliográfico, análise de co-palavras, e análise de co-autoria. Estas técnicas, quando combinadas com a análise de redes, são instrumentais na apresentação da estrutura bibliométrica e da estrutura intelectual do campo de investigação (Baker *et al.*, 2020; Donthu *et al.*, 2021; Tunger & Eulerich, 2018).

A bibliometria é uma ferramenta amplamente utilizada na produção científica em diversas áreas de pesquisa (Musa *et al.*, 2021). Por exemplo, um estudo bibliométrico realizado por Niu e Silva, em 2020, examinou a utilização de dados de *crowdsourcing* em atividades urbanas, explorando áreas temáticas, fontes de publicações, referências mais citadas, autores e palavras-chave. Os autores sintetizaram as pesquisas encontradas em três secções, sendo que uma delas engloba a análise de sentimento (Niu & Silva, 2020b). Nessa secção, observou-se uma predominância do uso de dados provenientes do *Twitter* (atual *X*), e a utilização de métodos de *machine learning* (e.g., *Maximum entropy classifier*, *Multinomial Naïve Bayes classifier*) e lexicais (e.g., *AFINN dictionary*) nas análises de sentimento.

Com este estudo pretende-se obter uma visão geral da investigação realizada até o momento sobre análise de sentimento em ambiente urbano, identificar lacunas de conhecimento, gerar ideias inovadoras e definir a contribuição que se pretende deixar na área. Para tal, aplicou-se métodos bibliométricos à base de dados eletrónica *Scopus*, para avaliar a tendência da literatura publicada sobre análise de sentimento relacionada com o ambiente constituído, urbano ou cidadão. O artigo encontra-se estruturado em 5 secções. Excluindo a primeira e quinta secções. Introdução e conclusão respetivamente, na segunda secção apresentam-se os materiais e métodos, faz-se referência à fonte dos dados, estratégia de pesquisa e análise estatística adotada. A terceira secção assenta sobre os resultados dos dados extraídos, inclui tendências, análise de autores, instituições, países ou regiões, índice de colaboração e análise de redes, entre outros fatores. A quarta secção apresenta uma breve discussão dos resultados obtidos.

II. METODOLOGIA

1. Fonte dos dados

O presente estudo é sustentado na base de dados eletrónica *Scopus* (<https://www.scopus.com/>) (powered by Elsevier), a 15 de junho de 2023, por apresentar uma cobertura de investigação mais abrangente face a outras bases de dados (e.g., *Google Scholar* e *Web of Science*) (Bakkalbasi *et al.*, 2006). As abordagens metodológicas foram adotadas a partir de estudos anteriores realizados por Donthu *et al.* (2021) e Musa *et al.* (2021).

2. Estratégia de investigação

Utilizou-se a base de dados eletrónica *Scopus* para obter as publicações pretendidas, aplicando uma consulta baseada na utilização das seguintes palavras ("*sentiment analysis*" AND *urban* OR *city* OR "*built environment*"). O operador booleano 'AND' foi utilizado para obter documentos de investigação que relacionassem a análise de sentimento, com o ambiente edificado, urbano ou cidadão. A consulta à base *Scopus* foi aberta a todos os domínios da ciência. Desta pesquisa resultou uma base de dados composta por 762 documentos, os resultados foram posteriormente restringidos apenas a documentos publicados em inglês, com estado de publicação final posterior a 2000. Foram obtidos um total de 728 documentos. A escolha de utilizar apenas o *Scopus* baseou-se no fato de ser a maior base de dados de literatura revista por pares e amplamente utilizado para criar conjuntos de dados para revisões. A base de dados *Elsevier-Scopus* também é considerada mais abrangente na representação de pesquisas em ciências sociais (Burnham, 2006). Além disso, tem havido uma sobreposição crescente entre o *Scopus* e o *Thompson Reuter Web of Science* (Aria & Cuccurullo, 2017).

A figura 1 ilustra o fluxograma que detalha as características do processo de extração da documentação.

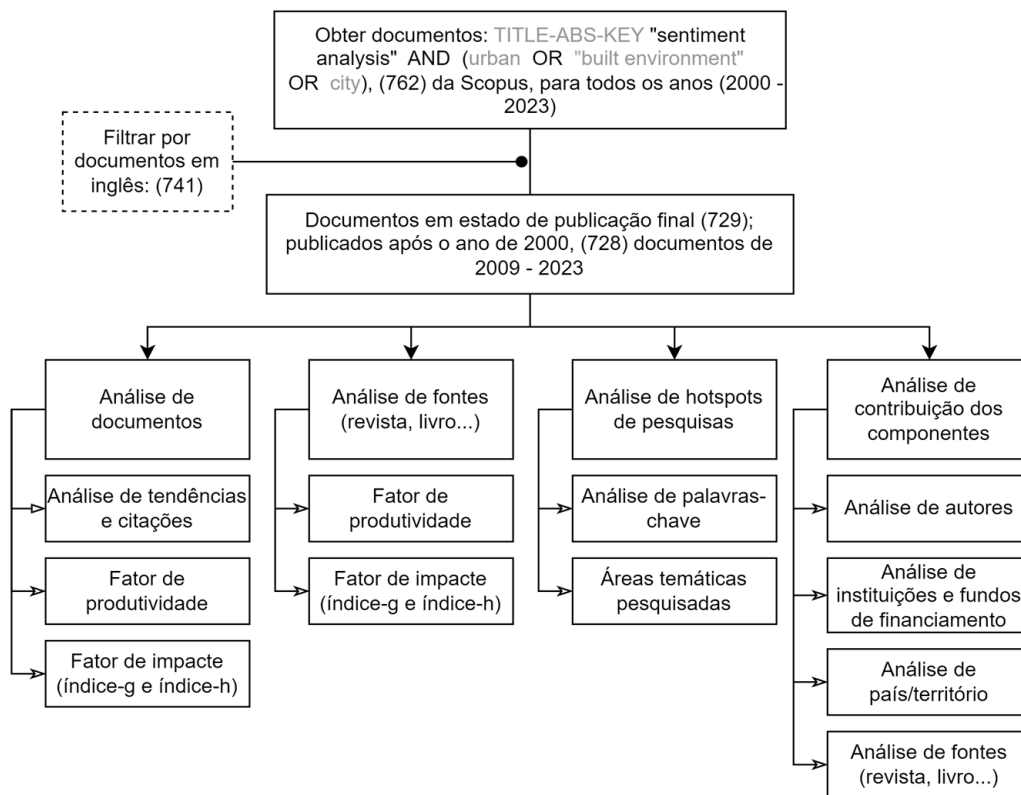


Fig. 1 – Fluxograma do processo de extração da documentação.
 Fig. 1 – Flowchart of the documentation extraction process.

3. Análise estatística

Para a análise dos dados utilizou-se a biblioteca *bibliometrix* (<https://CRAN.R-project.org/package=bibliometrix>) do *software R*, desenvolvida por Aria e Cuccurullo (2017), por permitir realizar análises de desempenho de grupos de componentes, tais como documentos, palavras-chave, autores, revistas e países/territórios, etc., e averiguar o seu impacto no domínio da investigação (Aria & Cuccurullo, 2017). Já o *VOSviewer*, o outro *software* utilizado, alia a qualidade das saídas gráficas à simplicidade, flexibilidade e capacidade de resposta às exigências do utilizador, exigindo, no entanto, análises repetidas devido à sua incapacidade de combinar dados de diferentes fontes. O *Bibliometrix* é mais robusto e versátil, sendo capaz de maior customização pelos utilizadores e de realizar análises utilizando arquivos de múltiplas bases de dados, oferecendo análises mais avançadas, tendo, no entanto, uma curva de aprendizagem mais acentuada que inclui programação (Arruda *et al.*, 2022).

Foram utilizadas estatísticas descritivas para determinar a frequência, percentagem, soma, média, índice-*h* e índice-*g*. O índice-*h* mede o número de artigos (ou outras formas de publicações académicas) que um investigador publicou e que receberam pelo menos *h* citações cada. Por outro lado, o índice-*g* tem em consideração o número total de citações recebidas pelo investigador, distribuídas uniformemente entre todas as suas publicações, em vez de focar apenas nas mais citadas. O primeiro tende a ser menos sensível a publicações extremamente citadas do que o segundo. Isto significa que o índice-*h* pode não ser tão afetado por uma única publicação altamente citada quanto o índice-*g*. Desta forma, o índice-*h* é frequentemente considerado uma medida mais conservadora do impacto académico, enquanto o índice-*g* pode refletir de forma mais precisa a distribuição de impacto entre as publicações.

Para o mapeamento científico, foi ainda utilizado o *software VOSviewer* versão 1.6.19 que permite a análise de redes sociais. Este explora a coautoria, cocorrência e citações, agrupamento

bibliográfico e links de co-citações através de três representações possíveis: visualização de rede, sobreposição ou densidade.

III. RESULTADOS

1. Características dos documentos e análise de tendências e citações

A partir da base de dados eletrônica *Scopus* foram identificados 728 artigos publicados entre 2000 e 2023 relacionados com a análise de sentimento e ambiente edificado, urbano ou cidadão (fig. 2), que incluem 281 (38,60%) artigos de conferência, 288 (39,56%) artigos em revista indexada, 127 (17,45%) resumos de conferências, 18 (2,47%) capítulos de livros, nove (1,24%) revisões, três (0,41%) erratas, um (0,14%) livro, um (0,14%) editorial, a 15 de junho de 2023. Estes documentos foram redigidos por 2068 autores, publicados em 307 fontes (*e.g.*, revistas, livros, etc.), de 70 países. Os documentos receberam 142 citações localmente e 5642 citações globalmente, com uma média de 7,75 citações por documento. O número de contribuidores foi de 2068 autores, aproximadamente 2,84 coautores por documento com um índice de colaboração de 3,4.

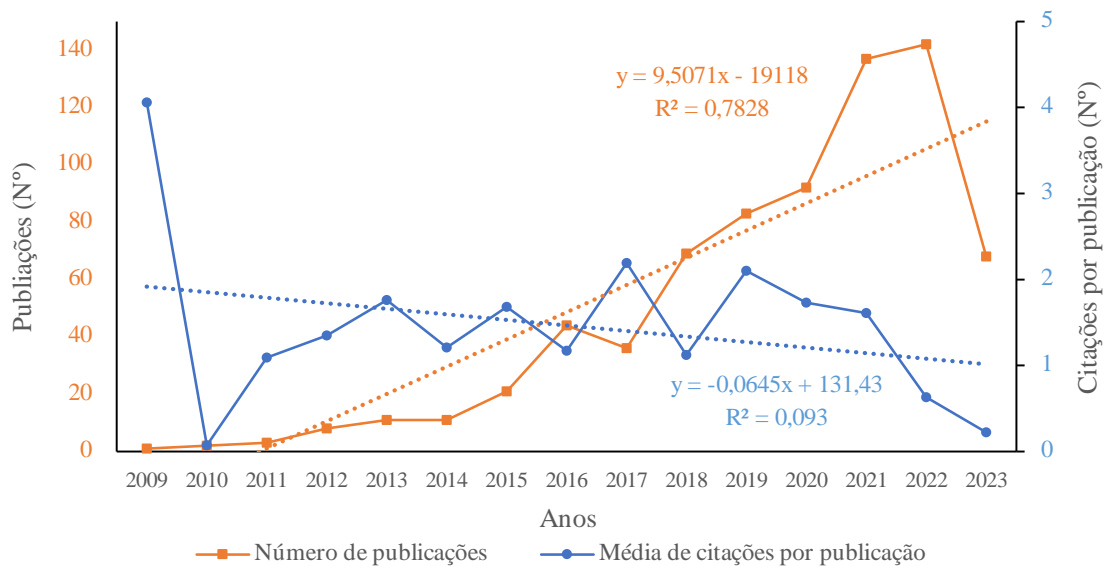


Fig. 2 – Tendências anuais e média total de citações por ano de publicações sobre análise de sentimento relacionado com ambiente edificado, urbano e cidadão. Figura a cores disponível online.

Fig. 2 – Annual trends and total average citations per year of publications on sentiment analysis related to built, urban and city environment. Colour figure available online.

Os documentos obtidos abrangeram oito tipos de publicações, dos quais 288 são artigos em revistas indexadas, compreendendo 39,56% da produção total, seguido por 281 artigos de conferência. Os restantes tipos de documentos incluem resumos de conferências, capítulos de livros, revisões, erratas, livros e editoriais.

O primeiro documento sobre sentimento com relação ao ambiente contruído, urbano ou cidadão foi publicado em 1999 por Busted: *Little islands of Erin: Irish settlement and identity in mid-nineteenth-century Manchester*. O autor estudou o sentimento durante a colonização irlandesa nas cidades britânicas do século XIX, analisando a opinião popular irlandesa através das baladas (músicas) que circulavam amplamente nas cidades (Busted, 2010).

No entanto, seguiu-se um hiato de quase dez anos e o primeiro documento sobre análise de sentimento e a sua relação com o ambiente posterior a 2000 surgiu em 2009 (fig. 2). Depois disso, verificou-se um crescente interesse pelo tema ao longo dos anos (reta com declive de 9,5; fig. 2), através do constante aumento de publicações ($R^2=0,7828$), à exceção do ano de 2017 que sofreu um decréscimo, como se pode verificar na figura 2. Entre 2009 e 2023 a média de publicação foi de 47 documentos ano. Em 2020, foram publicados 92 documentos, seguindo-se 137 documentos em 2021

(+48,9% face a 2020), 142 documentos em 2022 e 68 documentos até junho de 2023. Dado o padrão de desenvolvimento observado nos últimos anos, presume-se que esta tendência se manterá na próxima década. Verificou-se ainda uma aparente diminuição da média de citações (declive de -0,06), com principal incidência a partir do ano de 2019.

2. Publicações mais citadas e influentes

Dos 728 documentos que compõe a base de dados sobre análise de sentimento com relação com o ambiente urbano, os dez artigos mais citados globalmente (TGC), encontram-se representados no quadro I. O número total de citações local e global é diferente, 142 e 5642 respetivamente, uma vez que é determinado pelo conjunto de dados, enquanto as citações globais têm em consideração todos os documentos existentes na base de dados eletrónica *Scopus*, as citações locais apenas consideram os 728 documentos selecionados como elegíveis. O total de citações normalizado (TC) de um documento é calculado dividindo a contagem real de itens citados pela taxa de citação esperada para documentos com o mesmo ano de publicação.

O número de citações globais (TLC) no Top 10, varia de 90 a 170. O artigo mais bem classificado é *What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook*, da autoria de Sara Hofmann *et al* (2013), com 170 citações (quadro I). No entanto, numa perspetiva temporal percebe-se que há artigos mais recentes, mas com uma média de citações por ano mais elevada (*e.g.*, T = 2, 3, 5,7 e 8), podendo nos próximos anos virem a destacar-se como os mais proeminentes. Neste contexto sobressaem três bastante recentes, *Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data* e *COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model*, ambos de 2020 e com 21,20 (106 no total) e 20,20 (101 no total) citações por ano, respetivamente, e *COVID-19 open source data sets: a comprehensive survey* de 2021, com 27,5 citações por ano (110 no total). Este último é ainda o artigo que apresenta o maior TC normalizado, com um valor de 17,9. O TC normalizado de um documento é calculado dividindo o número de citações pela taxa de citação esperada para documentos com o mesmo ano de publicação.

Os 10 artigos mais citados, dos 21 676 referenciados pelos 728 documentos sobre análise de sentimento com relação com o ambiente, encontram-se representados no quadro II. O número total de citações a referências locais é de 22 808. O número total de citações (TC) no Top 10, varia de oito a 21. O artigo melhor classificado é *Opinion mining and sentiment analysis*, da autoria de Pang e Lee (2008) com 21 citações. Contrariamente ao verificado relativamente aos documentos mais citados globalmente, neste caso o mais citado também é aquele que apresenta um melhor rácio de citações/ano (1,24).

Quadro I – TOP 10 dos documentos mais citados globalmente (T: top; TC: total de citações).

Table I – TOP 10 most global cited documents (T: top; TC: total citations).

| T | Autor | Artigo | Ano | TC | TC por ano | TC normalizado | Fonte |
|---|--|--|------|-----|------------|----------------|--|
| 1 | Sara Hofmann Daniel Beverungen Michael Räckers Jörg Becker | What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook | 2013 | 170 | 14,17 | 8,03 | Government Information Quarterly |
| 2 | Minwoo Lee Miyoung Jeong Jongseo Lee | Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach | 2017 | 146 | 18,25 | 8,33 | International Journal of Contemporary Hospitality Management |
| 3 | Rodrigo Barbado Oscar Araque Carlos A. Iglesias | A framework for fake review detection in online consumer electronics retailers | 2019 | 130 | 21,67 | 10,33 | Information Processing & Management |
| 4 | M. Rosario González-Rodríguez Rocio Martínez-Torres Sergio Toral | Post-visit and pre-visit tourist destination image through eWOM sentiment analysis and perceived helpfulness | 2016 | 118 | 13,11 | 11,17 | International Journal of Contemporary Hospitality Management |

Quadro II – TOP 10 dos documentos mais citados globalmente (T: top; TC: total de citações) (continuação).

Table I – TOP 10 most global cited documents (T: top; TC: total citations) (continuation).

| T | Autor | Artigo | Ano | TC | TC por ano | TC normalizado | Fonte |
|----|--|---|------|-----|------------|----------------|---|
| 5 | Junaid Shuja Eisa Alanazi Waleed Alasmary Abdulaziz Alashaikh | COVID-19 open source data sets: a comprehensive survey | 2021 | 110 | 27,50 | 17,09 | Applied Intelligence |
| 6 | Staci M. Zavattaro P. Edward French Somya D. Mohanty | A sentiment analysis of U.S. local government tweets: The connection between tone and citizen involvement | 2015 | 109 | 10,90 | 6,48 | Government Information Quarterly |
| 7 | Danny Valdez Marijn ten Thij Krishna Bathina Lauren A Rutter Johan Bollen | Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data | 2020 | 106 | 21,20 | 12,28 | Journal Of Medical Internet Research |
| 8 | Tianyi Wang Ke Lu Kam Pui Chow Qing Zhu | COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model | 2020 | 101 | 20,20 | 11,70 | IEEE Access |
| 9 | Farman Ali Daehan Kwak Pervez Khan S.M. Riazul Islam Kye Hyun Kim K.S. Kwak | Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling | 2017 | 101 | 12,63 | 5,76 | Transportation Research Part C |
| 10 | Xinyu Chen Youngwoon Cho Suk young Jang | Crime Prediction Using Twitter Sentiment and Weather | 2015 | 90 | 9,00 | 5,35 | 2015 Systems & Information Engineering Design Symposium |

Quadro III – TOP 10 das referências mais citadas localmente (T: top; TC: total de citações).

Table II – TOP 10 most local cited references (T: top; TC: total citations).

| T | Autor | Artigo | Ano | TC | TC por ano | Publicação |
|---|---|---|------|----|------------|---|
| 1 | Bo Pang Lillian Lee | Opinion mining and sentiment analysis | 2008 | 21 | 1,24 | Foundations and Trends in Information Retrieval |
| 2 | Lewis Mitchell Morgan R. Frank Kameron Decker Harris Peter Sheridan Dodds Christopher M. Danforth | The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place | 2013 | 14 | 1,17 | PLOS ONE |
| 3 | David M. Blei Andrew Y. Ng Michael I. Jordan | Latent dirichlet allocation | 2003 | 13 | 0,59 | Journal of Machine Learning Research |
| 4 | Johan Bollen Huina Mao Xiaojun Zeng | Twitter mood predicts the stock market | 2011 | 11 | 0,79 | Journal of Computational Science |
| 5 | Bing Liu | Sentiment analysis and opinion mining | 2012 | 11 | 0,85 | Synthesis Lectures on Human Language Technologies |
| 6 | Walaa Medhat Ahmed Hassan Hoda Korashy Zheng Xiang | Sentiment analysis algorithms and applications: A survey | 2014 | 10 | 0,91 | Ain Shams Engineering Journal |
| 7 | Qianzhou Du Yufeng Ma Weiguo Fan | A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism | 2017 | 9 | 1,13 | Tourism Management |
| 8 | Stephen W. Litvin Ronald E. Goldsmith Bing Pan | Electronic word-of-mouth in hospitality and tourism management | 2008 | 8 | 0,47 | Tourism Management |

Quadro IV – TOP 10 das referências mais citadas localmente (T: top; TC: total de citações) (continuação).

Table II – TOP 10 most local cited references (T: top; TC: total citations) (continuation).

| T | Autor | Artigo | Ano | TC | TC por ano | Publicação |
|----|---------------------------------|--|------|----|------------|--|
| 9 | Liu B. Zhang L. Sakaki T. | A Survey of Opinion Mining and Sentiment Analysis | 2012 | 8 | 0,62 | Mining Text Data |
| 10 | Okazaki M. Matsuo Y. | Earthquake shakes Twitter users: real-time event detection by social sensors | 2010 | 8 | 0,53 | Proceedings of the 19th international conference on World wide web |

3. Análise de autores

O quadro III contém os dez autores mais produtivos e impactantes localmente. Os autores foram classificados com base no índice-*g*, índice-*h*, no total de citações (TC), no número de publicações (NP) e no ano de começo de atividade do autor (PY_start) na área temática e período em análise. Com base nos indicadores avaliados, Manar Alkhatib é considerado o autor mais produtivo e impactante, com um total de sete documentos publicados e um índice-*g* de seis (*i.e.*, publicou pelo menos seis artigos que, em conjunto, receberam pelo menos 36 citações). Seguiu-se-lhe Yan Wang, com índice-*g* de quatro e com quatro publicações.

Analisando as instituições de afiliação dos autores foram identificados com recurso ao Bibliometrix 70 países diferentes. De acordo com os resultados, a China é considerada o país mais produtivo em análise de sentimento com relação ao ambiente edificado, urbano, ou cidadão, com 391 artigos, seguida pelos EUA com 337, pela Índia com 174 e pela Indonésia com 125, entre outros países (quadro IV).

Foram ainda reconhecidos 58 países de autores mais produtivos e citados (quadro V). De acordo com os resultados, a China é considerada o país cujos autores são mais produtivos, com 80 artigos (11%), seguida pelos EUA com 55 (7,6%), pela Índia com 26 (3,6%) e pelo Reino Unido com 15 (2,1%). Por outro lado, aqueles com maior número de publicações com autores de diferentes países (MCP = *Multiple country publications*) foram a China (n = 23), os EUA (n = 16) e o Reino Unido (n = 7), como se pode verificar no quadro V.

Quadro V – TOP 10 dos autores com mais impacto local (T: top; TC: total de citações; NP: número de publicações; PY_start: ano de começo de atividade).

Table III – TOP 10 authors' local impact (T: top; TC: total citations; NP: number of publications; PY_start: start publication year).

| T | Autor | g-index | h-index | TC | NP | PY_start |
|----|----------------------|---------|---------|-----|----|----------|
| 1 | Alkhatib, Manar | 6 | 3 | 45 | 7 | 2019 |
| 2 | Wang, Yan | 4 | 4 | 93 | 4 | 2019 |
| 3 | El Barachi, May | 4 | 3 | 41 | 4 | 2020 |
| 4 | Resch, Bernd | 4 | 3 | 119 | 4 | 2018 |
| 5 | Hollander, Justin B. | 4 | 2 | 20 | 5 | 2017 |
| 6 | Mathew, Sujith | 4 | 2 | 28 | 4 | 2020 |
| 7 | Oroumchian, Farhad | 4 | 2 | 40 | 4 | 2019 |
| 8 | Sykora, Martin | 3 | 3 | 97 | 3 | 2017 |
| 9 | Varde, Aparna S. | 3 | 3 | 44 | 3 | 2018 |
| 10 | Shankardass, Ketan | 3 | 3 | 97 | 3 | 2017 |

Quadro VI - TOP 10 da produção científica por país.

Table IV – TOP 10 countries' scientific production.

| T | País (n = 70) | Número de documentos |
|----|----------------|----------------------|
| 1 | China | 391 |
| 2 | EUA | 337 |
| 3 | Índia | 174 |
| 4 | Indonésia | 125 |
| 5 | Itália | 115 |
| 6 | Austrália | 84 |
| 7 | RU | 77 |
| 8 | Espanha | 61 |
| 9 | Canada | 57 |
| 10 | Arábia Saudita | 53 |

Quadro VII – TOP 10 de países dos autores mais produtivos e citados (T: top; NA: número de artigos; SCP: publicação num único país; MCP: publicação em vários países; MCP_Ratio; TC: total de citações; AAC: média de citações por artigo).

Table V – TOP 10 corresponding author's countries (T: top; NA: number of articles; SCP: single country publication; MCP: multi-country publication; MCP_Ratio; TC: total citations; AAC: average citations per article).

| T | País (n = 58) | NA | SCP | MCP | Frequência | MCP_Ratio | TC | AAC |
|----|---------------|-----|-----|-----|------------|-----------|------|------|
| | | 362 | 313 | 49 | 0,497 | 0,135 | | |
| 1 | China | 80 | 57 | 23 | 0,110 | 0,287 | 322 | 4 |
| 2 | USA | 55 | 39 | 16 | 0,076 | 0,291 | 1060 | 19.3 |
| 3 | Índia | 26 | 25 | 1 | 0,036 | 0,038 | 70 | 2.7 |
| 4 | Reino Unido | 15 | 8 | 7 | 0,021 | 0,467 | 158 | 10.5 |
| 5 | Austrália | 13 | 11 | 2 | 0,018 | 0,154 | 97 | 7.5 |
| 6 | Espanha | 13 | 7 | 6 | 0,018 | 0,462 | 374 | 28.8 |
| 7 | Itália | 11 | 11 | 0 | 0,015 | 0,000 | 73 | 6.6 |
| 8 | Canada | 10 | 4 | 6 | 0,014 | 0,600 | 91 | 9.1 |
| 9 | Indonésia | 10 | 10 | 0 | 0,014 | 0,000 | 35 | 3.5 |
| 10 | Coreia | 10 | 5 | 5 | 0,014 | 0,500 | 234 | 23.4 |

4. Citações em fontes de publicação

Os 728 documentos sobre análise de sentimento com relação ao ambiente urbano, edificado ou cidadão foram publicados em 307 fontes diferentes (quadro VI). No conjunto das 307 fontes, as dez mais destacadas publicaram 120 artigos, o que representa 16% do total de artigos amostrados (120 de 728). A fonte com um maior impacto local foi a revista *International Journal of Environmental Research and Public Health*, na qual foram publicados 15 documentos, com um índice-g de 12 e um total de 147 citações (quadro VI). Seguindo-se a *Cities* (10), *Lecture Notes in Computer Science* (10), *Sustainability* (7) e a *Sustainable Cities and Society* (6) por ordem decrescente de acordo com o número de vezes citadas segundo o índice-g. As fontes com um maior número de publicações são a *Lecture Notes in Computer Science* (44), *Communications in Computer and Information Science* (24), *The ACM International Conference Proceeding Series* (23), *Lecture Notes in Networks and Systems* (21) e *Advances in Intelligent Systems and Computing*. É possível constatar que as revistas onde mais se publica sobre este tema pertencem à área das ciências da computação, contudo estas não são necessariamente as mais citadas. Efetivamente, as ciências da saúde ambiental, saúde pública, planeamento e da política urbana apresentam maior impacto nesta temática.

Quadro VIII – Top 10 de fontes de publicações mais impactantes localmente (T: top; TC: total de citações; NP: número de publicações; PY_start: ano da primeira publicação).

Table VI – Top 10 sources' local impact (T: top; TC: total citations; NP: number of publications; PY_start: year of first publication).

| T | Fonte (n = 307) | g-index | h-index | TC | NP | PY_start |
|----|--|---------|---------|-----|----|----------|
| 1 | International Journal of Environmental Research and Public Health | 12 | 6 | 147 | 15 | 2018 |
| 2 | Cities | 10 | 6 | 146 | 10 | 2017 |
| 3 | Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) | 10 | 6 | 136 | 44 | 2012 |
| 4 | Sustainability (Switzerland) | 7 | 4 | 55 | 14 | 2019 |
| 5 | Sustainable Cities and Society | 6 | 6 | 150 | 6 | 2019 |
| 6 | Journal of Medical Internet Research | 6 | 4 | 154 | 6 | 2019 |
| 7 | IEEE Access | 6 | 3 | 182 | 6 | 2017 |
| 8 | 2018 IEEE SmartWorld | 6 | 2 | 65 | 6 | 2018 |
| 9 | CEUR Workshop Proceedings | 5 | 2 | 32 | 8 | 2013 |
| 10 | ISPRS International Journal of Geo-Information | 5 | 2 | 59 | 5 | 2018 |

5. Fundos de financiamento e afiliação de autoria

Foram identificados 159 financiadores e 160 instituições durante o período de 2009 a 2023. A *National Natural Science Foundation of China* foi a principal financiadora com 42 documentos (quadro VII). Entre os principais institutos, encontra-se a *Chinese Academy of Sciences*, sediada na China, com nove publicações (1,96%), seguida pela *British University in Dubai*, sediada nos Emirados Árabes Unidos, com 8 publicações (1,74%) (quadro VIII).

Quadro IX – TOP 10 de fundos de financiamento (NP: número de publicações).

Table VII – TOP 10 funding sponsor (NP: number of publications).

| T | Financiador (n = 159) | NP | Porcentagem (%) |
|----|---|----|-----------------|
| 1 | National Natural Science Foundation of China | 42 | 5,07 |
| 2 | National Science Foundation | 13 | 1,57 |
| 3 | Horizon 2020 Framework Programme | 9 | 1,09 |
| 4 | European Commission | 8 | 0,97 |
| 5 | Fundamental Research Funds for the Central Universities | 7 | 0,84 |
| 6 | National Office for Philosophy and Social Sciences | 7 | 0,84 |
| 7 | Coordenação de Aperfeiçoamento de Pessoal de Nível Superior | 6 | 0,72 |
| 8 | Fundação para a Ciência e a Tecnologia | 6 | 0,72 |
| 9 | China Scholarship Council | 5 | 0,60 |
| 10 | Engineering and Physical Sciences Research Council | 5 | 0,60 |

Quadro X – Top 10 de instituições mais produtivas (NP: número de publicações).

Table VIII – Top 10 affiliation (NP: number of publications).

| T | Instituição (n = 160) | País | NP | Porcentagem (%) |
|---|-----------------------------|------------------------|----|-----------------|
| 1 | Chinese Academy of Sciences | China | 9 | 1,96 |
| 2 | British University in Dubai | Emirados Árabes Unidos | 8 | 1,74 |
| 3 | University of Melbourne | Austrália | 7 | 1,53 |
| 4 | Tongji University | China | 7 | 1,53 |

| | | | | |
|----|-----------------------------------|------------------------|---|------|
| 5 | University of Florida | Estados Unidos | 6 | 1,31 |
| 6 | University of Toronto | Canada | 6 | 1,31 |
| 7 | Wuhan University | China | 6 | 1,31 |
| 8 | University of Wollongong in Dubai | Emirados Árabes Unidos | 6 | 1,31 |
| 9 | Ministry of Education China | China | 5 | 1,09 |
| 10 | The University of Hong Kong | China | 5 | 1,09 |

6. Área temática mais estudada

Apenas cinco áreas temáticas indexadas na base de dados *Scopus* apresentam mais de 100 documentos. A maioria das publicações pertence à área das ciências da computação (487), seguindo-se a engenharia (225) e ciências sociais (182) (fig. 3).

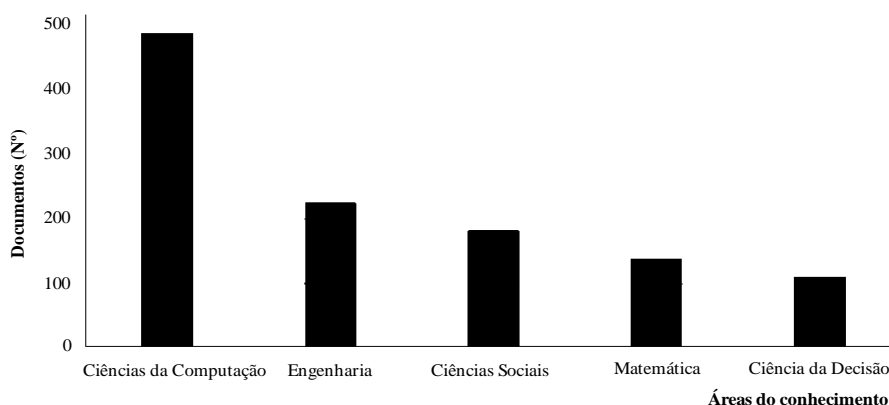


Fig. 3 – TOP 5 das áreas temáticas mais pesquisadas com mais de 100 documentos.

Fig. 3 – TOP 5 of the most searched thematic areas with more than 100 documents.

7. Palavras-chave e termos mais frequentes

A nuvem de palavras consiste numa representação visual da frequência e da importância das palavras. Estas, revelam os temas de tendência na investigação nos últimos anos. Na figura 4.a), encontram-se representadas as *KeyWords Plus*, que são palavras ou frases que aparecem frequentemente nos títulos das referências de um artigo, mas não aparecem no próprio título do artigo. A nuvem de palavras foi gerada a partir de títulos citados que se encontram relacionados com os documentos da base de dados, sobre análise de sentimento com relação ao ambiente edificado, urbano ou cidade. A figura 4.b), compreende as palavras-chave frequentemente utilizadas pelos autores. Segundo Zhang *et al.*, o *KeyWords Plus* é tão eficaz quanto as palavras-chave do autor na análise bibliométrica para mapear a estrutura conceitual de um campo de investigação. A análise *KeyWords Plus* pode captar artigos com maior variedade e identificar tendências em diversas áreas científicas (Zhang *et al.*, 2016).

As dez principais palavras-chave da análise *KeyWords Plus* foram *sentiment analysis* (461), seguida de *social media* (186), *social networking (online)* (184), *data mining* (128), *smart city* (94), *big data* (58), *China* (54), *learning systems* (50), *covid-19* (45) e *semantics* (44). Por sua vez, a nuvem de palavras-chave do autor encontra-se representada na figura 4.b), sendo as dez mais frequentes, *sentiment analysis* (357), seguida por *Twitter* (90), *social media* (89), *machine learning* (64), *natural language processing* (40), *text mining* (40), *covid-19* (37), *smart cities* (33), *opinion mining* (29) e *big-data* (26).



a)



b)

Fig. 4 – Análise de frequência de *keywords plus* (a) e palavras-chave do autor (b). Figura a cores disponível online.

Fig. 4 – Frequency analysis of keywords plus (a) and author's keywords (b). Colour figure available online.

8. Análise de redes

No caso das ligações de coautoria, o atributo *Links* indica o número de ligações de coautoria de um determinado componente com outros componentes, enquanto o atributo Força Total de Ligações indica a força total das ligações de coautoria de um dado componente com outros componentes (van Eck & Waltman, 2023). No caso das ligações de coocorrência (*i.e.*, co-palavras), cada linha representa a relação entre duas palavras-chave (*i.e.*, quando aparecem juntas no mesmo documento). A dimensão dos nós, reflete a frequência das palavras-chave (*i.e.*, quanto maior a frequência de uma palavra-chave, maior o tamanho de um nó). Os *clusters* (identificados por cores) são gerados automaticamente pelo *software VOSViewer*, e distinguem as diferentes redes umas das outras (Knani *et al.*, 2022).

Os 728 documentos identificados apresentam ainda 1615 palavras-chave diferentes selecionadas pelos autores. Na figura 5 encontram-se representadas as 52 palavras-chave, com frequência superior a 5. Quanto maior a colaboração menor a distância entre as palavras. No geral, são visualizados 8 *clusters* de palavras-chave, com 341 ligações (*Links* - L) e uma Força Total de ligações (*Total Links Strength* - TLS) de 1059. As principais palavras-chave por cluster são *machine learning* (*cluster* 3 - vermelho), *Twitter* (*cluster* 1), *text mining* (*cluster* amarelo), *sentiment analysis* (*cluster* 2), *deep learning* (*cluster* 4), *support vector machine* (*cluster* azul-claro), *nlp* (*natural language processing*) (*cluster* laranja) e *social network* (*cluster* castanho). A palavra-chave *sentiment analysis* apresenta 356 ocorrências de, seguida pela palavra *Twitter* (89) e *social media* (88).

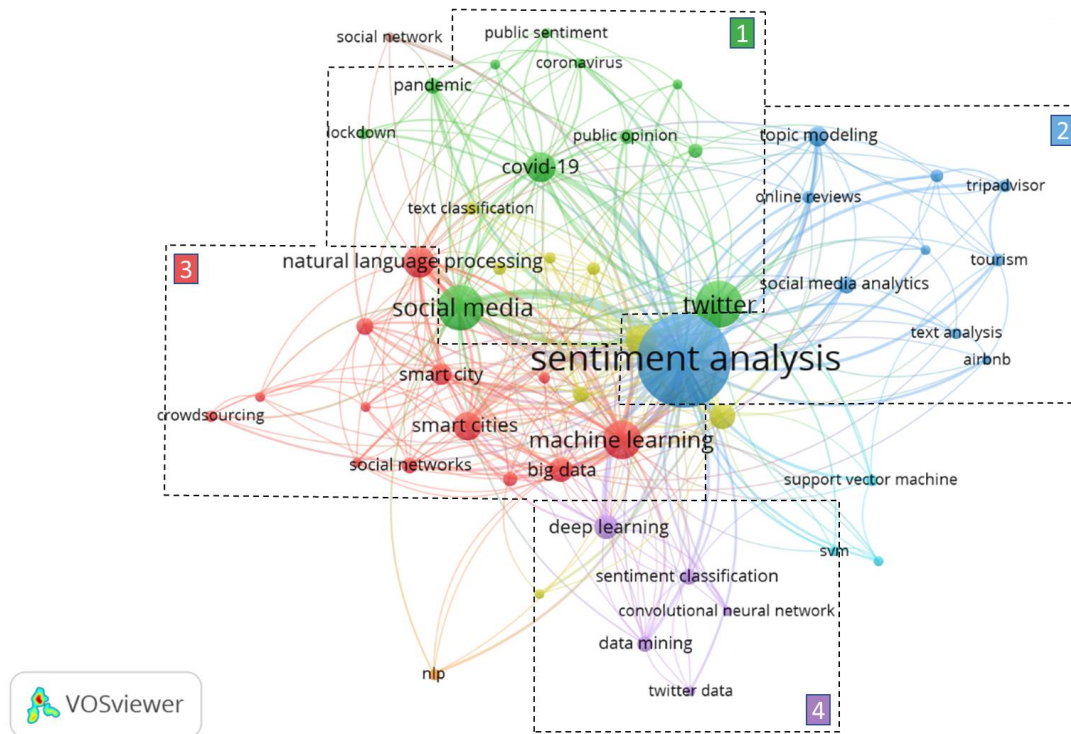


Fig. 5 – Análise de coocorrência de palavras-chave. Figura a cores disponível online.

Fig. 5 – Keyword co-occurrence analysis. Colour figure available online.

O *cluster 3* (vermelho) ilustra conexões entre *machine learning*, *natural language processing*, *smart cities*, *big data* e *smart city*. Este cluster revela a frequente utilização de *big data* e métodos de *machine learning* em pesquisas relacionadas com *smart cities*. Estes estudos desempenham um papel fundamental na tomada de decisões mais informadas e no desenvolvimento de soluções mais eficientes face aos desafios urbanos. O *cluster 1* (verde) apresenta conexões entre palavras, como *Twitter*, *social media*, *covid-19* e *pandemic*. A pandemia de *covid-19* teve um impacto na sociedade que se refletiu nas interações dos utilizadores nas redes sociais, destacando-se a importância da análise de sentimento, utilizando dados do *Twitter* para compreender as emoções e opiniões das pessoas, durante este período. O *cluster 2* – (azul escuro) revela conexões entre *sentiment analysis*, *topic modeling*, *social media analytics*, *tripadvisor* e *tourism*. O *TripAdvisor* tem sido cada vez mais utilizado em análises de sentimento devido à crescente quantidade de dados que disponibiliza, como avaliações e opiniões de utilizadores. Esses dados possuem uma natureza semiestruturada que permite extrair informações relevantes para o setor do turismo, tais como oportunidades de mercado. Verificam-se ainda outras conexões evidentes a roxo (*cluster 4*), entre *deep learning*, *data mining* e *sentiment classification*. Estas conexões ilustram a recente utilização de técnicas de *deep learning* em análise de sentimento, devido à sua eficácia em lidar com problemas complexos no processamento de linguagem natural.

A Figura 6 mostra as tendências dos tópicos de investigação ao longo do tempo. A partir da mesma é possível verificar uma elevada presença das redes sociais ao longo dos últimos anos em análises de sentimento, com um maior destaque para o *Twitter*, utilizado em estudos relacionados com o planeamento urbano (2017 a 2020) e *smart cities* (2018 a 2021). Mais recentemente surgiu a incorporação de dados do *TripAdvisor* em estudos relacionados com o turismo (Raja & Juliet, 2023), devido às opiniões deixadas pelos utilizadores que apresentam informações detalhadas e subjetivas sobre o local. Efetivamente, dada a proliferação de avaliações em *sites* de viagens *online* e o impacto resultante no consumidor, muitos investigadores têm feito esforços para explorar a relação entre as avaliações de viagens *online* e o comportamento do consumidor, e até que ponto essas avaliações influenciam as decisões e escolhas desse consumidor (Chu *et al.*, 2022).

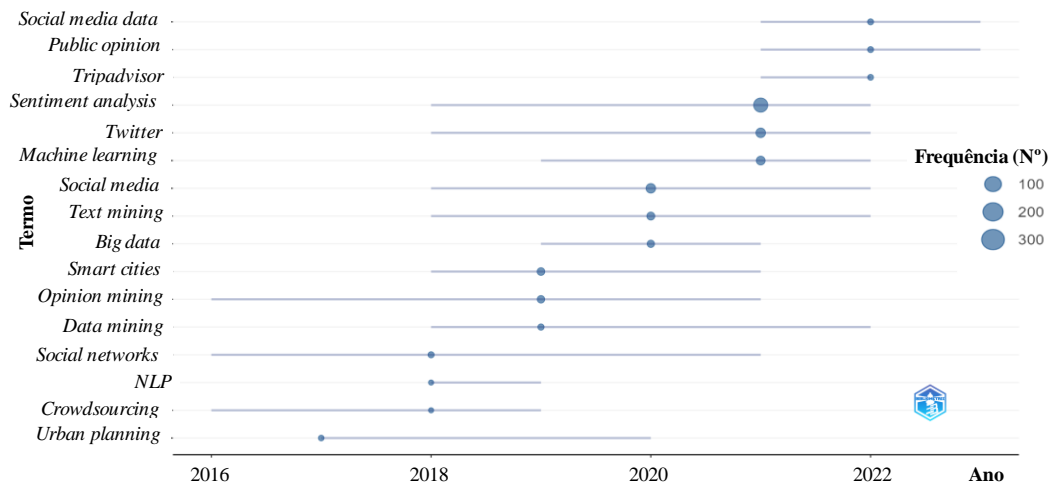


Fig. 6 – Tendências dos tópicos com base na frequência das palavras-chave dos autores ao longo do tempo. Figura a cores disponível online.

Fig. 6 – Trending topics based on the frequency of authors' keywords over time. Colour figure available online.

Muitos dos desenvolvimentos ocorreram nos últimos sete anos (2016-2023) com o aumento dos *big data*. As palavras-chave, *big data*, *text mining* e *social media* atingiram o seu ponto mais expressivo em 2019, enquanto *machine learning*, *Twitter*, e *sentiment analysis* atingiram o máximo de relevância em 2021. O *urban planning* perdeu alguma importância nos últimos anos, tendo-se desviado o foco para questões relacionadas com o turismo, o que poderá revelar uma falta de estudos em planeamento urbano com o uso de técnicas mais recentes.

De acordo com a análise dos dados obtidos da *Scopus* usando o filtro do *VosViewer*, os documentos identificados foram redigidos em instituições de 85 países diferentes. Na Figura 7 encontram-se representados por círculos 33 países, de um total de 36 países com mais de cinco publicações. Quanto maior a colaboração menor a distância entre eles. No geral, são visualizados 6 clusters de países, com 127 ligações (L) e uma Força Total de Ligação (TLS) de 226. O país que mais documentos produziu foi os EUA (123), seguindo-se a China (109), Índia (65), Itália (38) e Reino Unido (33). Em termos de agregação destaca-se o cluster 7 (vermelho) correspondente a um eixo asiático. Menos produtivos, mas igualmente com ligações fortes, observam-se os clusters 2 (verde – países mediterrânicos e Brasil) e 6 (amarelo – países muçulmanos).

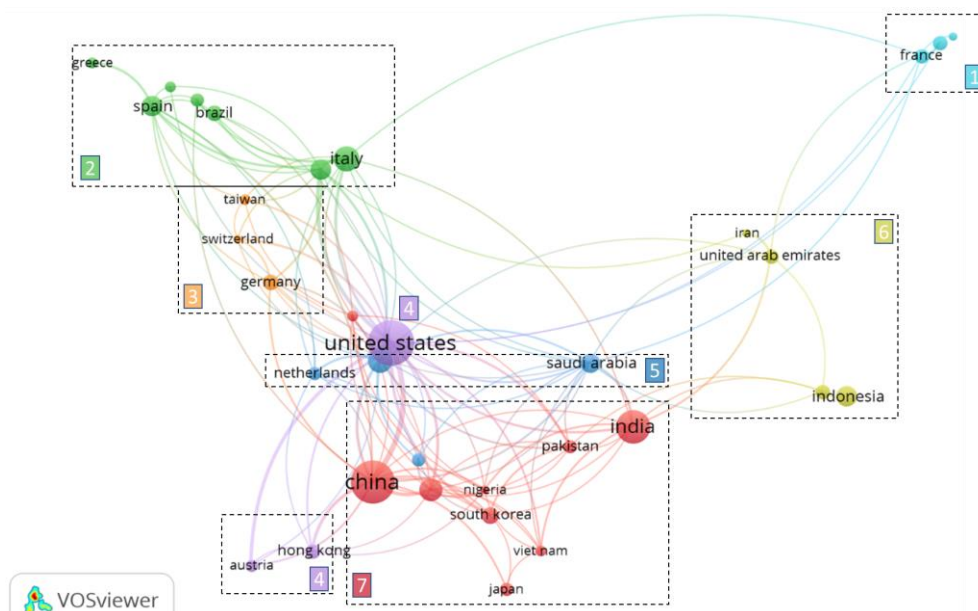


Fig. 7 – Análise de coautoria (países). Figura a cores disponível online.

Fig. 7 – Co-authorship analysis (countries). Colour figure available online.

IV. DISCUSSÃO

Este estudo assenta numa análise bibliométrica dos documentos sobre análise de sentimento com relação ao ambiente urbano, edificado ou cidadão, indexados na base de dados eletrônica *Scopus* nos últimos 14 anos. A análise incorporou técnicas bibliométricas de duas categorias vigentes, análise de desempenho e mapeamento científico. Nomeadamente, análise sobre os autores mais impactantes e os países, fontes, instituições e fundos de financiamento mais produtivos compreendidos nesta área de investigação. Além disso, o estudo explora a relação entre os tópicos identificados pelos *KeyWords Plus* e as "palavras-chave do autor", com recurso a ferramentas visuais, como o gráfico de nuvem de palavras e o gráfico de tendências dos tópicos ao longo do tempo.

A partir dos resultados das publicações é possível verificar que a análise de sentimento relacionada com o ambiente urbano tem tido um interesse crescente na última década. Espera-se que o aumento da produção científica se mantenha nos próximos anos. As citações recebidas anualmente sofreram um aumento significativo com início em 2010. Todavia, a partir de 2019 tem-se observado um decréscimo acentuado. Contudo os valores apurados do total de citações por ano nos artigos mais recentes e principalmente os valores normalizados de total de citações deixam antever uma inversão desta tendência mais recente.

O autor mais produtivo nesta temática é Manar Alkhatib, professora assistente de inteligência artificial na *The British University in Dubai* (BUiD) ao contrário do que se verificava até há uns anos atrás em que era Cambria Erik, da *Nanyang Technological University*, Singapura (Musa *et al.*, 2021). O acompanhamento das revistas e livros ou as citações das mesmas constitui um outro meio de avaliar a qualidade das publicações. O maior número de documentos, sobre análise de sentimento com relação ao ambiente urbano, foi publicado na revista *International Journal of Environmental Research and Public Health*, seguindo-se a *Cities* e a *Lecture Notes in Computer Science* (LNCS), incluindo as subséries *Lecture Notes in Artificial Intelligence* (LNAI) e *Lecture Notes in Bioinformatics* (LNBI). Os países que contribuíram com um maior número de documentos e citações foram a China, EUA e Índia. Estes países predominam não só nas publicações com autores de um único país, mas também naquelas que envolvem colaborações de múltiplas nacionalidades. Da mesma forma, a análise de autoria indicou que os autores chineses (seguidos dos norte americanos e indianos) também são os principais contribuintes para a investigação neste domínio. Estes resultados consubstanciam os resultados dos estudos desenvolvidos por Keramatfar e Amirkhani (2019), e Musa *et al.* (2021).

A nuvem de palavras obtida a partir das *KeyWords Plus* resulta das 50 palavras-chave mais frequentes, de um universo de 2985. Sendo que, as cinco palavras-chave mais frequentes foram *sentiment analysis*, *social media*, *social networking* (online), *data mining* e *smart city*. Estes tópicos permitem obter uma visão geral dos temas com maior destaque nos últimos 14 anos. As *KeyWords Plus* apresentam a vantagem de ampliar o conjunto de palavras-chave, fornecendo uma maior variedade (Zhang *et al.*, 2015). Por sua vez, a nuvem de palavra-chave do autor resultou de 50 palavras-chave de um total de 1612, em que as cinco mais frequentes foram: *sentiment analysis*, *Twitter*, *social media*, *machine learning*, *natural language processing*. Estas palavras-chave podem ser influenciadas pelas preferências, interesses específicos ou estratégias de publicação do autor.

Os resultados são consistentes com outros estudos bibliométricos, realizados anteriormente, sobre *opinion mining* ou análise de sentimento por (Keramatfar & Amirkhani, 2019; Musa *et al.*, 2021). As investigações bibliométricas sobre estudos de análise de sentimento realizadas por Keramatfar e Amirkhani (2019), e Musa *et al.* (2021), exploram também fatores, como autores e instituições mais influentes, documentos mais citados e evolução das palavras-chave. Os autores verificaram que a abordagem mais adotada para a classificação de sentimento foi o *machine learning*, e o *Twitter* a rede social mais adotada para análise de sentimento. Esta situação ainda se mantém muito embora em termos de dados se venha a observar um aumento de importância do *TripAdvisor*, muito em função das recentes restrições de acesso aos dados implementadas pela *X* (antigo *Twitter*).

Através das tendências dos tópicos foi possível constatar que os trabalhos relacionados com o planeamento urbano decresceram nos últimos anos, tendo sido dado um maior destaque a temas relacionados com as *smart cities* e o turismo, demonstrado assim a pertinência de retomar tais trabalhos devido à desatualização de que os mesmos podem estar a ser alvo.

V. CONCLUSÃO E LIMITAÇÕES

Este estudo fornece uma visão geral bibliométrica dos resultados da investigação sobre análise dos sentimentos em ambiente construído, urbano ou cidade. Nos últimos anos verificou-se uma maior incidência da análise de sentimento em tópicos como as *smart cities*, pandemia de covid-19 e turismo. Dada a importância da análise dos sentimentos na compreensão dos impactos das características do ambiente no bem-estar individual, é pertinente desenvolver futuramente estudos orientados para esta temática, através da incorporação de métodos mais recentes, como o *deep learning*, que permite reconhecer padrões complexos em textos, processando os dados de uma forma inspirada no cérebro humano.

Efetivamente, o uso de arquiteturas de redes neuronais profundas, como as redes neuronais convolucionais (*Convolutional Neural Network – CNN*) e redes de neurónais recorrentes (*Recurrent Neural Network – RNN*), tem mostrado resultados promissores na análise de sentimentos a partir de textos. Alguns modelos mais avançados, como as redes neuronais transformadoras (*Transformer*) desenvolvida pela Google®, também têm sido explorados devido à sua capacidade de capturar relações semânticas complexas. O *Transfer Learning*, por seu turno, permite que modelos pré-treinados em grandes conjuntos de dados sejam ajustados para a análise de sentimentos em redes sociais. Isso permite melhorar o desempenho dos modelos, especialmente quando os dados disponíveis para treino são limitados. No entanto, em muitos casos, pode ser difícil obter grandes conjuntos de dados reclassificados para treinar modelos de análise de sentimentos. Nestas situações pode-se recorrer à aprendizagem semi-supervisionada que combina dados classificados e não classificados para melhorar o desempenho do modelo, permitindo que este aprenda com exemplos não classificados durante o treino.

Outras metodologias a explorar futuramente serão os modelos multimodais que podem analisar simultaneamente múltiplos tipos de dados (as sociais não consistem apenas em texto, mas também incluem imagens, vídeos e áudio) e estão a adquirir cada vez mais importância na captura de nuances na expressão de sentimentos. Isto leva-nos precisamente ao ambiente urbano e à análise de contexto. O facto de se considerar o contexto em que as mensagens são postadas pode melhorar significativamente a precisão da análise de sentimentos. Os métodos que levam em conta o contexto social, temporal e situacional podem ajudar a interpretar de forma mais precisa o sentimento expresso.

Este tipo de análises fornece aos futuros investigadores orientações/indicações sobre os possíveis pontos críticos/limitações que se podem encontrar no campo de investigação, bem como informar acerca da multiplicidade de pesquisas científicas conduzidas globalmente até ao momento por autores, países, fontes, financiamentos atribuídos e instituições ativas, permitindo melhorar a produção de documentos sobre análise de sentimento com relação ao ambiente urbano.

A análise bibliométrica efetuada apresentada algumas limitações, nomeadamente na utilização de dados relativos a um indexador, a Scopus. Em trabalhos futuros, poderão ser consideradas mais bases de dados para além das eletrónicas (*e.g., Google Académico, Web of Science, PubMed, Embase*), tais como, bibliotecas e repositórios físicos, publicações impressas, etc. Para além disso, foram considerados para análise apenas os artigos publicados em inglês e posteriores ao ano de 2000.

Apesar da pertinência das conexões entre os autores, não foi possível a realização de análises de coautoria e copulação de autores devido à fraca colaboração e acoplamento entre os mesmos.

AGRADECIMENTOS

Iuria Betco foi financiada pela Fundação para a Ciência e Tecnologia (FCT) através da Bolsa de Doutoramento [2022.11665.BD]. Este trabalho foi financiado pelo Centro de Estudos Geográficos, Universidade de Lisboa e FCT através da bolsa [UIDB/00295/2020 + UIDP/00295/2020]. Gostaríamos de agradecer ao GEOMODLAB – Laboratório de Detecção Remota, Análise Geográfica e Modelação – do Centro de Estudos Geográficos/IGOT pelo fornecimento dos equipamentos e software necessários.

CONTRIBUTOS DOS/AS AUTORES/AS

Iuria Betco: Conceptualização; Software; Investigação; Recursos; Curadoria dos dados; Escrita – preparação do esboço original. **Jorge Rocha:** Redação – revisão e edição; Visualização; Supervisão; Administração do projeto; Aquisição de financiamento.

ORCID ID

Iuria Betco  <https://orcid.org/0000-0001-8714-686X>

Jorge Rocha  <https://orcid.org/0000-0002-7228-6330>

REFERÊNCIAS BIBLIOGRÁFICAS

- Adli, M., Berger, M., Brakemeier, E.-L., Engel, L., Fingerhut, J., Gomez-Carrillo, A. ... Stollmann, J. (2017). Neurourbanism: towards a new discipline. *The Lancet Psychiatry*, 4(3), 183-185. [https://doi.org/10.1016/S2215-0366\(16\)30371-6](https://doi.org/10.1016/S2215-0366(16)30371-6)
- Afriliana, N., Iswari, N. M. S., & Suryasari. (2022). Sentiment analysis of user-generated content: a bibliometric analysis. *Journal of System and Management Sciences*, 12(6), 583-598. <https://doi.org/10.33168/JSMS.2022.0634>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Arruda, H., Silva, E. R., Lessa, M., Júnior, D. P. & Bartholo, R. (2022). VOSviewer and Bibliometrix. *Journal of the Medical Library Association*, 110(3), 392-395. <https://doi.org/10.5195/jmla.2022.1434>
- Baker, H. K., Kumar, S., & Pandey, N. (2021). Forty years of the Journal of Futures Markets: A bibliometric overview. *Journal of Futures Markets*, 41(7), 1027-1054. <https://doi.org/10.1002/fut.22211>
- Baker, H. K., Pandey, N., Kumar, S., & Haldar, A. (2020). A bibliometric analysis of board diversity: current status, development, and future research directions. *Journal of Business Research*, 108, 232-246. <https://doi.org/10.1016/j.jbusres.2019.11.025>
- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3, 1-8. <https://doi.org/10.1186/1742-5581-3-7>
- Bakolis, I., Hammoud, R., Smythe, M., Gibbons, J., Davidson, N., Tognin, S., & Mechelli, A. (2018). Urban Mind: using smartphone technologies to investigate the impact of nature on mental well-being in real time. *BioScience*, 68(2), 134-145. <https://doi.org/10.1093/biosci/bix149>
- Balahur, A., Turchi, M., Steinberger, R., Perea-Ortega, J.-M., Jacquet, G., Küçük, D. ... El Ghali, A. (2014). Resource creation and evaluation for multilingual sentiment analysis in social media texts. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani... S. Piperidis, (Eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014* (pp. 4265-4269). European Language Resources Association. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85009226696&partnerID=40&md5=c60a943c0079c385b8fb448db8ee8621>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Burnham, J. F. (2006). Scopus database: a review. *Biomedical Digital Libraries*, 3(1), 1-8. <https://doi.org/10.1186/1742-5581-3-1>
- Busteed, M. (2010). Little Islands of Erin: Irish settlement and identity in mid-nineteenth-century Manchester. *Immigrants & Minorities*, 18(2-3), 94-127. <https://doi.org/10.1080/02619288.1999.9974970>
- Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). Language interpretation in travel guidance platform: text mining and sentiment analysis of TripAdvisor reviews. *Frontiers in Psychology*, 13, 1029945. <https://doi.org/10.3389/fpsyg.2022.1029945>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146-166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Cox, D. T. C., Shanahan, D. F., Hudson, H. L., Plummer, K. E., Siriwardena, G. M., Fuller, R. A. ... Gaston, K. J. (2017). Doses of neighborhood nature: The benefits for mental health of living with nature. *BioScience*, 67(2), 147-

155. <https://doi.org/10.1093/biosci/biw173>
- Donthu, N., Kumar, S., & Pandey, N. (2021). A retrospective evaluation of Marketing Intelligence and Planning: 1983–2019. *Marketing Intelligence & Planning*, 39(1), 48-73. <https://doi.org/10.1108/MIP-02-2020-0066>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: an overview and guidelines. *Journal of Business Research*, 133, 285-296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Galea, S. (2011). The urban brain: new directions in research exploring the relation between cities and mood-anxiety disorders. *Depression and Anxiety*, 28(10), 857-862. <https://doi.org/10.1002/da.20868>
- Hammoud, R., Tognin, S., Burgess, L., Bergou, N., Smythe, M., Gibbons, J. ... Mechelli, A. (2022). Smartphone-based ecological momentary assessment reveals mental health benefits of birdlife. *Scientific Reports*, 12(1), 17589. <https://doi.org/10.1038/s41598-022-20207-6>
- Hartig, T. (1993). Nature experience in transactional perspective. *Landscape and Urban Planning*, 25(1-2), 17-36. [https://doi.org/10.1016/0169-2046\(93\)90120-3](https://doi.org/10.1016/0169-2046(93)90120-3)
- Hidaka, B. H. (2012). Depression as a disease of modernity: explanations for increasing prevalence. *Journal of Affective Disorders*, 140(3), 205-214. <https://doi.org/10.1016/j.jad.2011.12.036>
- Kaplan, S. (1995). The restorative benefits of nature: toward an integrative framework. *Journal of Environmental Psychology*, 15(3), 169-182. [https://doi.org/10.1016/0272-4944\(95\)90001-2](https://doi.org/10.1016/0272-4944(95)90001-2)
- Keramatfar, A. (2019). Bibliometrics of sentiment analysis literature. *Journal of Information Science*, 45(1), 3-15. <https://doi.org/10.1177/0165551518761013>
- Knani, M., Echchakoui, S., & Ladhari, R. (2022). Artificial intelligence in tourism and hospitality: bibliometric analysis and research agenda. *International Journal of Hospitality Management*, 107, 103317. <https://doi.org/10.1016/j.ijhm.2022.103317>
- Kovacs-Gyori, A., Ristea, A., Havas, C., Resch, B., & Cabrera-barona, P. (2018). #London2012: towards citizen-contributed urban planning through sentiment analysis of Twitter data. *Urban Planning*, 3(1), 75-99. <https://doi.org/10.17645/up.v3i1.1287>
- Lederbogen, F., Kirsch, P., Haddad, L., Streif, F., Tost, H., Schuch, P. ... Meyer-Lindenberg, A. (2011). City living and urban upbringing affect neural social stress processing in humans. *Nature*, 474(7352), 498-501. <https://doi.org/10.1038/nature10190>
- Maynard, D., & Funk, A. (2012). Automatic Detection of Political Opinions in Tweets. In R. García-Castro, D. Fensel & G. Antoniou (Eds.), *The Semantic Web: ESWC 2011 Workshops*, ESWC 2011. Lecture Notes in Computer Science, (pp. 88–99). Springer. https://doi.org/10.1007/978-3-642-25953-1_8
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: a survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Miller, J. R. (2005). Biodiversity conservation and the extinction of experience. *Trends in Ecology and Evolution*, 20(8), 430-434. <https://doi.org/10.1016/j.tree.2005.05.013>
- Mohammad, S., & Turney, P. (2015). *NRC Word-Emotion Association Lexicon*. <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- Musa, I. H., Zamit, I., Xu, K., Boutouhami, K., & Qi, G. (2021). A comprehensive bibliometric analysis on opinion mining and sentiment analysis global research output. *Journal of Information Science*, 49(6), 1506-1516. <https://doi.org/10.1177/01655515211061866>
- Niu, H., & Silva, E. A. (2020). Crowdsourced data mining for urban activity: review of data sources, applications, and methods. *Journal of Urban Planning and Development*, 146(2), 04020007. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000566](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000566)
- Olszewska, A. A., & Bil, J. S. (2016). Therapeutic garden design for patients with neurodegenerative diseases. *Przestrzeń i Forma*, 25, 259-270. <https://doi.org/10.21005/pif.2016.25.D-01>
- Pang, B., & Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In D. Scott. (Ed.), *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, ACL-04*, (pp. 271-278). ACL. <https://doi.org/10.3115/1218955.1218990>
- Peen, J., Schoevers, R. A., Beekman, A. T., & Dekker, J. (2010). The current status of urban-rural differences in psychiatric disorders. *Acta Psychiatrica Scandinavica*, 121(2), 84-93. <https://doi.org/10.1111/j.1600-0447.2009.01438.x>
- Pineo, H., & Rydin, P. Y. (2018). *Cities, health and well-being*. Royal Institution of Chartered Surveyors.

<https://www.rics.org/news-insights/research-and-insights/cities-health-and-well-being>

- Raja, J. G. J. S., & Juliet, S. (2023). Deep learning-based sentiment analysis of Trip Advisor reviews. In *2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, (pp. 560-565). <https://doi.org/10.1109/ICAAIC56838.2023.10140848>
- Ramos-Rodríguez, A. R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the *Strategic Management Journal*, 1980-2000. *Strategic Management Journal*, 25(10), 981-1004. <https://doi.org/10.1002/smj.397>
- Centers for Disease Control and Prevention (2011). Mental illness surveillance among adults in the United States. *MMWR supplements*, 60(3), 1-32. https://www.cdc.gov/mmwr/preview/ind2011_su.html
- Romice, O., Thwaites, K., Porta, S., Greaves, M., Barbour, G., & Pasino, P. (2016). Urban design and quality of life. In G. Fleury-Bahi, E. Pol & O. Navarro, (Eds.), *The Handbook of Environmental Psychology and Quality of Life Research* (pp. 241-273). Springer. https://doi.org/10.1007/978-3-319-31416-7_14
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (WWW '10). Association for Computing Machinery, 851-860. <https://doi.org/10.1145/1772690.1772777>
- Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W. ... Al-garadi, M. A. (2018). Sentiment analysis of big data: methods, applications, and open challenges. *IEEE Access*, 6, 37807-37827. <https://doi.org/10.1109/ACCESS.2018.2851311>
- Tunger, D., & Eulerich, M. (2018). Bibliometric analysis of corporate governance research in German-speaking countries: applying bibliometrics to business research using a custom-made database. *Scientometrics*, 117(3), 2041-2059. <https://doi.org/10.1007/s11192-018-2919-z>
- van Eck, N. J., & Waltman, L. (2023). {VOSviewer} manual. Leiden: Universteit Leiden. http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.1.pdf
- Verma, S., & Gustafsson, A. (2020). Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. *Journal of Business Research*, 118, 253-261. <https://doi.org/10.1016/j.jbusres.2020.06.057>
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2015). Comparing keywords plus of WOS and authors keywords: a case study of patient adherence research. *Journal of the Association Science and Technology*, 67(4), 967-972. <https://doi.org/10.1002/asi.23437>
- Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment Analysis in Health and Well-Being: systematic review. *JMIR Medical Informatics*, 8(1), 22. <https://doi.org/10.2196/16023>

¹ Em Psicologia, a densidade social é um dos principais determinantes da aglomeração, e há evidências de que tem um efeito mais poderoso na resposta humana do que a densidade – populacional – espacial. Traduz o número de interações interpessoais que provavelmente ocorrerão num determinado espaço. Intimamente relacionada com a proxémica, *ie.*, o estudo do comportamento espacial interpessoal, que se preocupa com a territorialidade, a distância interpessoal, os arranjos espaciais, a aglomeração e outros aspetos do ambiente físico que afetam o comportamento.