Received: 7 December 2022 | Revised: 13 February 2023 | Accepted: 22 February 2023

DOI: 10.1111/soc4.13086

ORIGINAL ARTICLE

WILEY

Social welfare in the light of topic modelling

Mariusz Baranowski[1] | **Piotr Cichocki**[1] | **Jim McKinley**[2]

[1]Faculty of Sociology, Adam Mickiewicz University, Poznan, Poland

[2]University College London (UCL), UCL Institute of Education, London, UK

Correspondence

Mariusz Baranowski, Faculty of Sociology, Adam Mickiewicz University, Szamarzewskiego 89C, Poznan 60-658, Poland.

Email: mariusz.baranowski@amu.edu.pl

Funding information

Narodowe Centrum Nauki, Grant/Award Number: 2021/05/X/HS6/00067

Abstract

With an increased focus on social well-being in response to a burgeoning global economy exposing the weaknesses of social welfare policies, research output in the field has grown exponentially. Keeping track of the evolving research themes proves difficult due to the steady rise in the number of studies published in the interdisciplinary field of social welfare. Therefore, researchers need a comprehensive overview to confirm the current shape of the field based on the published research. Using a latent Dirichlet allocation algorithm as a topic modelling technique, this study identified 12 prominent themes from more than 10,000 research outputs on social welfare published from 2000 to 2020 in Scopus-indexed journals. Such an exploratory text-mining approach to literature review provides broad insights into the diversity of research and may serve as a foundation for further in-depth studies. Identifying these 12 thematic areas and their sub-themes allows us to articulate the complexity and diversity of social welfare issues, which go far beyond the field of well-established welfare economics or social work. The study shows that the topic of 'social welfare' has not only evolved over time but has significantly broadened its meaning. It can no longer be solely synonymous with institutional social security. We contend that research in this area needs to take into account a broader and more

[Correction added on 18 March 2023, after first online publication: The copyright line was changed.]

systematic range of determinants constituting the dynamic character of social welfare.

KEYWORDS
automatic literature review, latent Dirichlet allocation (LDA), machine learning, natural language processing, social welfare

1 | INTRODUCTION

Scientific activities in times of dynamic development of digital technologies (Karpf, 2012; Selwyn, 2011) or technologically networked reality (Baranowski, 2021) involve new and diverse challenges. Social sciences are no exception, as areas of social life are also subject to fundamental transformations resulting from the impact of new technologies on forms of communication, work and leisure. Exponential growth in academic publications and publication venues (Bornmann & Mutz, 2015; Milojević, 2015) means that keeping track of the torrent of publications proves difficult even for specialists focusing on specific research areas within particular disciplines. However, the task proves even more daunting for researchers needing an overview of the state of the art in a field outside their narrow specialisation. Keeping track of developments in any academic discipline or sub-discipline poses many problems as the number of books, articles, chapters, or research reports proliferates. It is estimated that there may be 'close to two million articles published each year' in academic journals alone (Altbach & de Wit, 2018).

With the development of digital technologies and mass dissemination of research results in high-impact academic journals, researchers have ready access to vast amounts of accumulated knowledge, yet, effective tracking of research activities through traditional literature reviews becomes increasingly challenging. Traditional content analysis methods and systematic literature reviews offer diminishing returns to scale and usually require restrictive criteria for document inclusion (Snyder, 2019). They remain instrumental when focused on specific topics or narrow domains of expertise but prove impractical for exploratory purposes or the study of large-scale research trends. When facing Big Text (i.e., large, loosely structured bodies of textual data), exploratory methods need to be affordable, scalable and nimble to provide meaningful insights into the ongoing academic discourse (Thangaraj & Sivakami, 2018).

Algorithmic text mining handily tackles the challenges of scale; typically, the more text is made available to them, the better they work. Moreover, they also provide explicit, reproducible and quantifiable results. Mayer-Schönberger and Cukier (2013) introduced the term 'datafication' to highlight the process of continually collecting data on each of us. They noted that 'to datafy a phenomenon is to put it in a quantified format so it can be tabulated and analysed' (Mayer-Schönberger & Cukier, 2013, p. 78). Such quantifying, in turn, poses a particular challenge for empirical social science, which cannot be abstracted from big data analyses (Cukier & Mayer-Schoenberger, 2013). Some researchers have already pointed out the coming crisis of empirical sociology 'at the level of data collection and analysis,' together with 'the question of the use and dissemination of research information' (Savage & Burrows, 2007, p. 887). Datafication of textual data through machine-based text mining provides a promising answer to that challenge.

Our analysis applies a machine text-mining approach to the academic discourse on social welfare. Due to its theoretical ambiguity and wide use by researchers from different disciplines, social welfare is an excellent point of investigation using the latent Dirichlet allocation (LDA) algorithm, as the numerous studies on social welfare involve diverse theoretical and methodological orientations. Our study is an exploratory attempt to analyse a reasonably comprehensive set of publications on social welfare. The study examines the main topics within research articles published in Scopus-indexed journals on 'social welfare' for 2000–2020. The study adopted the LDA algorithm, an established topic modelling technique. The LDA is used here as an exploratory tool for extracting a snapshot of the thematic structure of discourse, providing a rough mapping for further in-depth studies. The study identified 12 main themes within the area of social welfare, which was then subjected to a clustering procedure. This way, a unique picture of social welfare issues addressed by journals from different social science disciplines over the last two decades was obtained. Additionally, LDA analysis was applied to identify micro-topics within each specified macro-topics (we call this LDA within LDA) to broaden the exploration of social welfare issues.

The remainder of this paper proceeds as follows. The next section outlines the literature review on social welfare. Section 3 presents the description of the data and methods. In Section 4, we described the results of our analyses, with a particular focus on the 12 topics produced by the LDA algorithm. Section 5 collects the discussion and limitations of the method we used and the results we obtained. Finally, Section 6 concludes.

2 | LITERATURE REVIEW

The field of social welfare is particularly interdisciplinary (Bean & MacPherson, 2018; Forder et al., 2018; Goodman & Markowitz, 1952; Hadley & Hatch, 2019; Sullivan, 2018), making the definition of what constitutes social welfare explicitly difficult (Macarov, 1995, p. xv). Arguably, the term is yet to find a clear definition (*The Encyclopedia of Social Work*, 1987, cited by Macarov, 1995, p. xv), which raises several difficulties. For instance, Aravacik (2018, np) shows how 'social policy' is referred to as 'social policy' in Europe, while North American scholarship calls it 'social welfare policy' (cf. Rochefort, 1986). The matter can be further complicated by introducing—following Spicker (1988)—a distinction between individual and social welfare and equating the latter with the concept of 'well-being' (cf. Baranowski, 2019; Midgley, 2017).

Suppose we add to these many sociological or socio-political meanings of the term social welfare (Butler & Drakeford, 2005; Dover, 2016; Williamson & Fleming, 1977) the 'classical' psychological (Stewart & Viney, 1975), political (Wickwar, 1946), legal (Friedman, 1969) or economic (Arrow, 1950) interpretations. In that case, the matter becomes even more complicated. Notably, in this broad and diverse field, the number of publications within each discipline is growing by leaps and bounds, thanks partly to digitalisation and changing patterns of academic careers. Traditional literature review methods (Callahan, 2014), geared towards *in-depth knowledge*, face the challenge of incorporating increasing research findings (*knowledge across*). Sacrificing depth for breadth, machine text mining utilises technologically advanced algorithms capable of structuring large corpora of heterogeneous data (Roberts et al., 2016).

The definitional difficulties with the concept of social welfare have a long academic pedigree (Arrow, 1950), and successive generations of researchers working within the various social sciences and humanities on this phenomenon have addressed particular slices of it (sometimes referred to as happiness, prosperity, wealth, quality of life, life satisfaction, subjective or societal well-being). Those interested in the role of the state (government) in meeting human needs have emphasised the institutional aspects of achieving social goals (e.g., Esping-Andersen, 1990; Gough, 2000, 2019; Kipo-Sunyehzi, 2021; Madison, 2019; Wilson et al., 2020). They used terms such as social welfare services, social welfare policy or national welfare system. This perspective focused on particular dimensions of the welfare state and social policy linked to specific thematic areas (cf. Esping-Andersen, 1999; Greve, 2020; Titmuss, 2018). Others have developed the subjective component of social welfare, most often referred to as (subjective/individual) well-being (e.g., Adler, 2019; Ellison, 1991; Ngamaba et al., 2018). There is no shortage of approaches unifying or harmonising these two levels of analysis, for example, in the form of 'personal' and 'social' welfare (Esping-Andersen, 1999, p. 59) or proposing to consider well-being as a component of the broader phenomenon of social welfare (Baranowski, 2019) or mutually constitutive (Taylor, 2011).

From a content analysis perspective, it is also important to note that individual studies on social welfare are published in a variety of journals, such as the *British Journal of Sociology*, *International Journal of Social Welfare*, *Journal of Economic Theory*, *Journal of Applied Psychology*, and *International Social Work*, to name a few. Individual journals prefer specific theoretical approaches, empirical methods or geographical areas, which further influences the formula for presenting research findings on social welfare.

Aware of the differences mentioned above within social welfare and the theoretical approaches and measures for its study, Baldock (2007, p. 21) noted that social welfare 'gains little form being defined very tightly'. This observation provides a starting point for seeking ways to analyse the highly heterogeneous notion of social welfare, especially in a period of digital proliferation and general accessibility of research findings (Alvarez, 2016; Karpf, 2012; Ruppert et al., 2013; Salganik, 2019; Savage & Burrows, 2007). We believe that a LDA algorithm as a topic modelling technique is an appropriate tool for the exploratory study of the diverse phenomenon of social welfare. Additionally, we would like to emphasise that although we use the automatic literature review technique, we disagree with Asmussen and Møller's (2019, p. 1) position that 'manual exploratory literature reviews are soon to be outdated'. The research problems determine the choice of analysis method, and each method has advantages and disadvantages depending on the intended objectives.

3 | DATA AND METHODS

This study employed content analysis of an initial 10,050 article abstracts from Scopus-indexed journals between 2000 and 2020. The selection was limited to English-language scientific articles from the areas of (i) Social Sciences, (ii) Economics, Econometrics and Finance, (iii) Environmental Science, (iv) Psychology, and (v) Arts and Humanities. We used the LDA algorithm to generate topics within the collected database. The LDA is a probabilistic topic model for 'performing unsupervised analysis of large document collections and requires no manual construction of the training data' (Sukhija et al., 2016, p. 1198). Setting up the analysis workflow is work-intensive, as the original Scopus abstracts required substantial cleaning and processing before feeding into the LDA algorithm. However, once the code is written, the pre-processing procedures may be easily implemented in other similar literature reviews. The code requires only minor manual tweaks when used on abstracts from any other research field.

The data processing workflow (Figure 1) comprised a sequence of steps: (1) import of .csv files downloaded from Scopus into a data frame comprising 10,050 records, (2) basic cleaning reducing the database size to 9522 records by filtering out those (a) without an abstract, (b) with an abstract not written in English, (c) with an abstract shorter than 70 words, (3) NLP processing, including (a) tokenisation and part-of-speech recognition retaining only noun phrases (the procedure uses the Spacy language model—the widely used algorithm for parsing large bodies of text), (b) deep cleaning of the tokenised output including (c) stemming of tokens towards common forms, (d) removal of frequently occurring stop-words, (e) n-gramisation of commonly occurring phrases (e.g., 'social' and 'policy' become 'social_policy'). Once the reasonably clean database was forged, it was turned into a document-term matrix: a data structure with documents assigned to rows and tokens to columns, with every cell indicating how many times a particular token occurs in any given document. The original DTM is far too sparse (9522 rows × 14,472 tokens, with

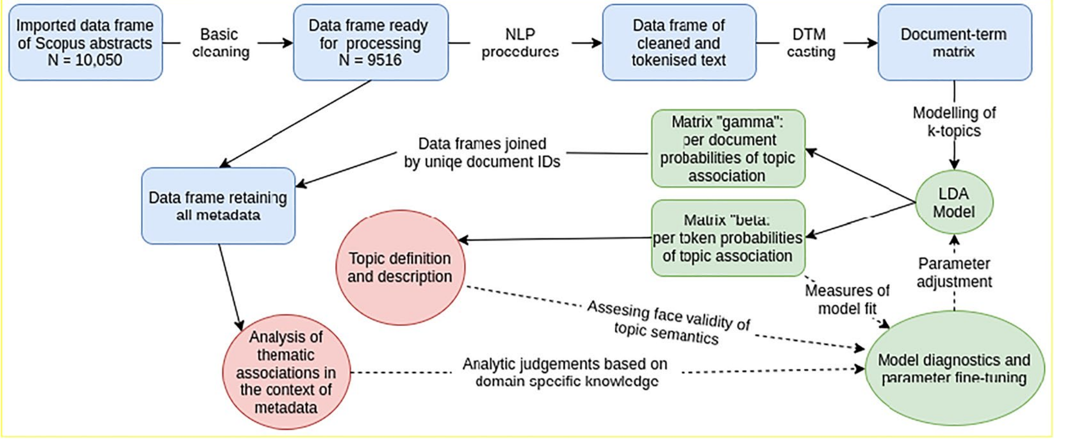

FIGURE 1 Data processing workflow scheme.

99.76% of the cells with 0s), so a sparsity reduction was performed with the minimum token frequency set at 0.1% of documents resulting in a matrix of 9522 rows × 3205 tokens.

The LDA algorithm included several inputs, principally: (1) hyperparameter delta—how likely it is for a token to belong to more than one topic (here we set delta at 0.01, which is tilted towards token exclusivity but not excessively so); (2) hyperparameter alpha—how likely it is for a document to be a mixture of more than one topic (here the initial alpha value has a very low setting of 1.5, but we allow the model to estimate alpha further as it learns), (3) k-topic number—the algorithm was commanded to find 12 topics. These particular settings were developed through multiple explorations to fine-tune the original model. The most controversial issue seems to come in the setting of the number of topics—the algorithm cannot estimate it on its own. Although measures of fit can be calculated (e.g., log-likelihood, perplexity) in order to see at which point adding an additional topic does not make the overall model much better, it remains ultimately a judgement call on the part of the researchers (for a different approach, cf. Jacobs & Tschötschel, 2019). Our purpose was to look for general themes in discourse rather than highly specific topics; hence, we chose to look for 12 macro-topics (see Table A1), which would contain hundreds of documents each and be internally differentiated into multiple sub-topics.

4 | RESULTS AND ANALYSIS

Twelve main topics were generated from the cleaned database with a list of the most frequent words (Figure 2). These themes are associated with specific documents, defined by tokenised words and common phrases. Given the available publication metadata, the distribution of topics can be studied relative to other factors, such as source journals. Additionally, this allows researchers to explore the extracted topics further, taking into account the affiliation of documents not only to journals but also to individual authors (including their affiliation or sources of research funding).

However, the most important result of applying the LDA algorithm to a collection of several thousand abstracts of scientific articles is, of course, the identification of the 12 topics alone within social welfare based on a generative statistical model that 'assumes that the observed documents are produced from a mixture of latent topics' (Lindstedt, 2019, p. 308). The topics identified provide valuable information for social welfare researchers, both from the side of historically shaped discourses, which are subject to (i) specific fashions (certain problem areas are more readily taken up in certain periods and some are not), (ii) divisions within particular disciplines (institutional struggle in the field of science, using Pierre Bourdieu's (1986) terminology), or (iii) assessment in terms of their impact on the discipline(s) as measured by bibliometric indicators. However, since they were generated without being influenced by the researchers' prior assumptions (leaving aside strictly methodological issues that had to be addressed from the research procedure point of view), these topics have a highly objectified character from an epistemological standpoint (cf. the polemic position in Pääkkönen & Ylikoski, 2021), not to mention the value of overview knowledge (cf. Savin & van den Bergh, 2021).

It is worth making a secondary division within the 12 topics due to the specificity of the individual social sciences. Such a division fundamentally cannot be sharp, as the boundaries between disciplines are fluid, and they themselves are multi-paradigmatic (Kiser, 1999; Lamont & Molnár, 2002; Wallerstein, 2001). However, the heuristic value of such a secondary division, with its corollary of specific topics and even specialised journals, cannot be overestimated. Researchers working on the issue of the ambiguous and variously defined phenomenon of social welfare do not have a broad and framed perspective over two decades, as they focus on its fragment. As a result of applying cluster analysis to the gamma matrix (document-topic probabilities) (see Figure A2), several topic clusters can be clearly distinguished.

The 12 topics presented in Figure 3 represent the popularity patterns regarding different facets of social welfare. Take, for instance, the following three topics: T3 Community health care and education, T9 Intervention and assistance support and T12 Family and children support services. All of them fall within the purview of social policy,

FIGURE 2 Labels for 12 main topics with top associated words/tokens.

which, according to Taylor-Gooby (2019, np), 'is a field of study rather than a discipline'. The most frequent tokens (see Figure 2) also indicate a strong relationship with social work and its associated health, education, homelessness, and employment components. Applying cluster analysis to the topic-document association matrix reveals that the T4 Risk of abuse and violence remains related to the three abovementioned themes (Figure A2). To a degree, this is also true with respect to T1 Political government and moral discourse, which seems more strongly involved with the questions of governance and politics. This group of topics (T1, T3, T4, T9 and T12) designates an area of interest identified strictly with the 'classical' or 'traditional' understanding of social policy as social welfare (Drover & Woodsworth, 1978; Hughes & Lewis, 1998; Walsh et al., 2000).

Similarly, the second broad understanding of social welfare is typically associated with economics, or more precisely, with one of its branches, welfare economics (cf. Albert & Hahnel, 2017; Hicks, 1939; Pigou & Aslanbeigui, 2002). This perspective is represented in the following three topics: T2 Market economy and competition, T8 Utility functions and preferences and T6 Financial management. Cluster analysis (see Figure A2) indicates a close alignment of T6 and T8, with T2 remaining an obvious outsider. Regarding T6 and T8, the former is associated

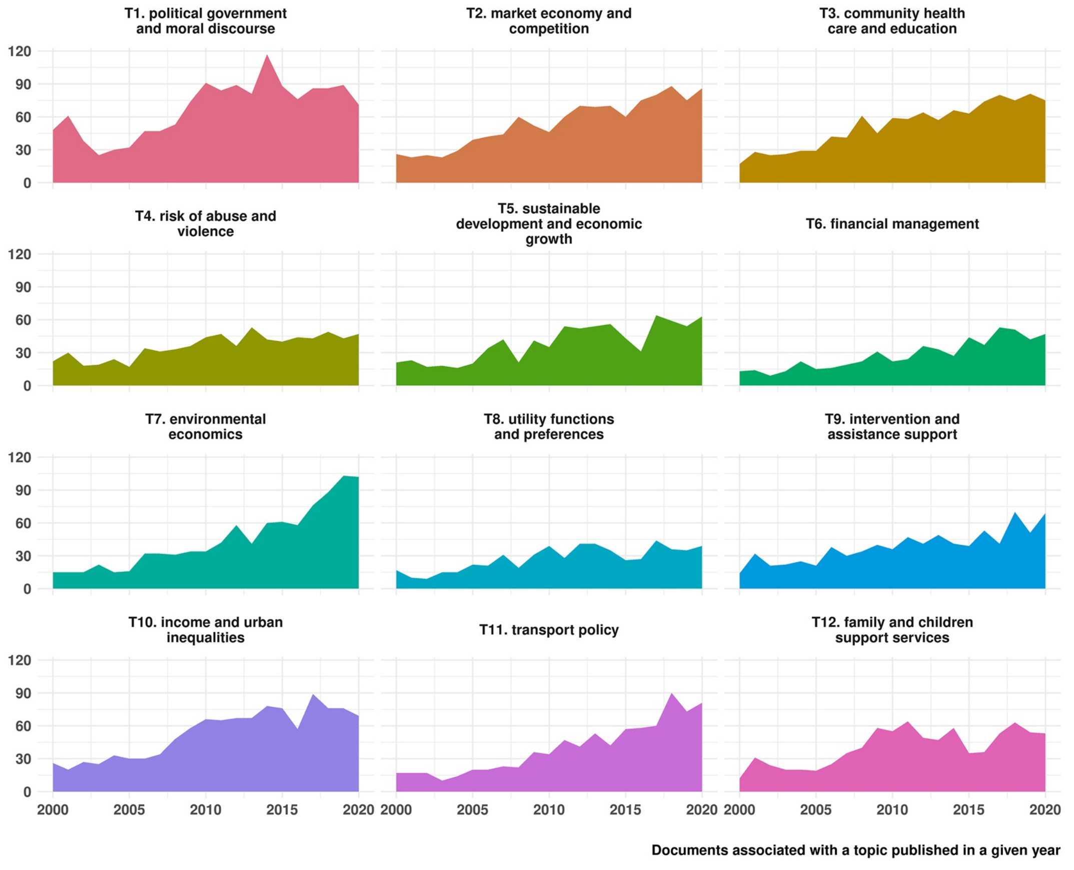

FIGURE 3 Prevalence of topics over time.

with management and finance, while the latter is with theoretical economics and its emphasis on agent preferences. Even though T2 is not clustered closely with T6 and T8, from the point of view of domain knowledge, it constitutes their obvious complement and, therefore, should be analysed as a component of welfare economics.

When it comes to the less obvious areas in the domain of social welfare, the remaining topics are also important. These comprise T5 Sustainable development and economic growth, T10 Income and urban inequalities, T11 Transport policy and T7 Environmental economics. All of them are directly or indirectly connected with the problems of the natural environment, which have gained prominence in the debates within social sciences (Frank et al., 2000; Miller et al., 2012; Woodgate & Redclift, 1998). It should be noted in this context that, as remarked by Shaw (2008, p. 13): 'social welfare theory has historically ignored the natural environment'. These four main topics, as well as their subtopics (cf. Figure 4), are involved with issues traditionally associated with social welfare, but in a way which develops and modifies them in the face of prescient challenges of environmental degradation and climate change. Hence, researchers seem to capture social welfare's meanings, correlates and connotations within a broader framework. For instance, in the case of T10, which relates to issues of public spending, tax system, redistribution, inflation, consumption involving health, insurance, risks, and the situation of migrants or inhabitants of rural areas (Figure 4). When it comes to T11 Transport Policy, we are dealing with issues not usually associated with social welfare. These include the questions of the quality of various transport services (railways, urban transport, aviation), as well as the ever more prominent issues of sources and costs of energy. Within this topic, there are also prominent issues of urban congestion and congestion charges, as well as the availability of parking spaces. All those subtopics fall within the evolving category of social welfare, which can be attested by the growing body of publications over the last 20 years

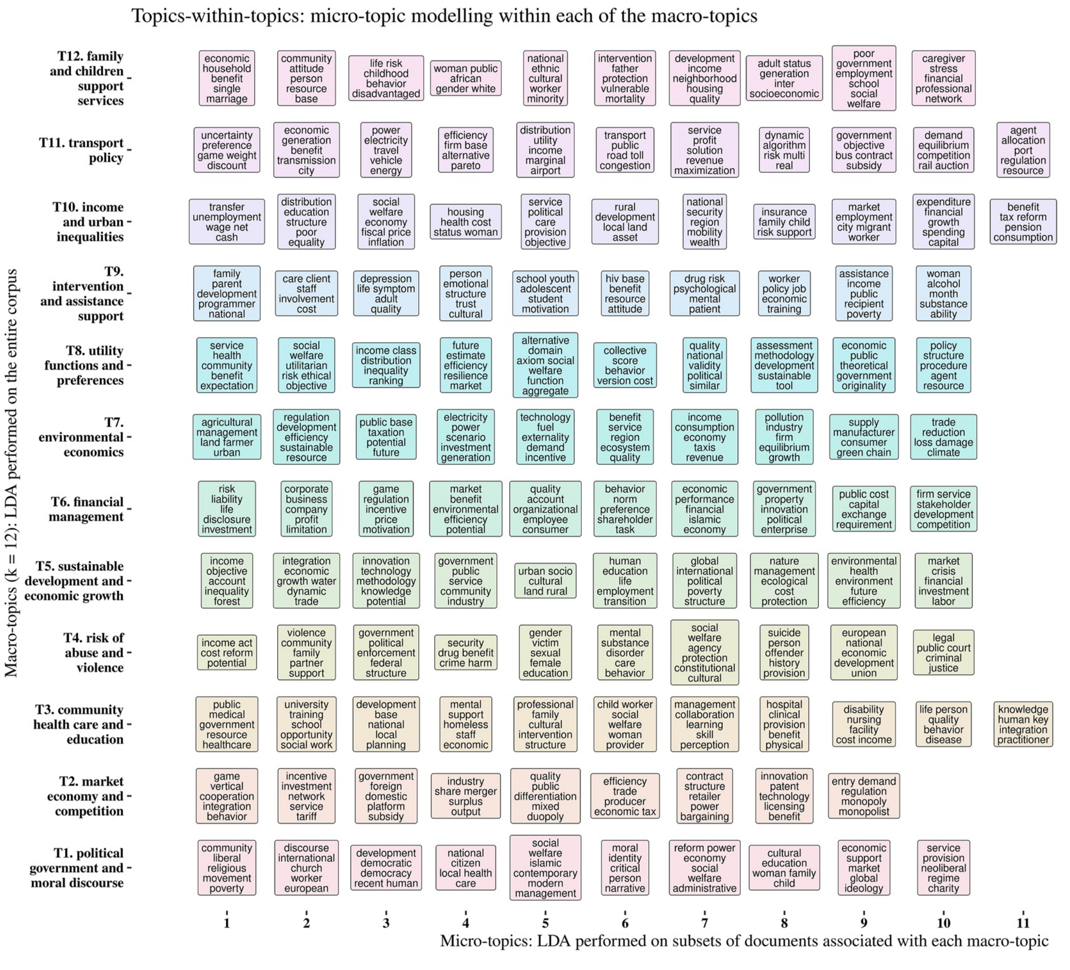

FIGURE 4 Secondary LDA: micro-topics within macro-topics. LDA, latent Dirichlet allocation.

(Baranowski, 2019; Burroughs & Rindfleisch, 2012). The innovative use of LDA within LDA allows for surveying the internal differentiation of the main 12 macro-themes within the same methodology of analysis. In T7 Environment economics, the use of LDA within LDA allows for an overview of themes present in the dynamically growing heterodox economic orientation concerning social welfare. This provides valuable insights which allow identifying themes associated with the pollution of the natural environment, research into regional ecosystems or consumption and agricultural practices, which remain of focal interest for researchers publishing in such journals as *Journal of Cleaner Production, Climate Policy, or Journal of Environmental Management*. It opens new avenues for research into nature-based social welfare (Baranowski, 2022; Gough, 2022; Linares & Cabaña, 2022), a novel approach to the study of social welfare, which tends to be different from research focusing on social policy or welfare economics and is more akin to natural science analyses.

Generated macro-topics are broad and are therefore represented throughout the queried time series. Nevertheless, some broad shifts seem noticeable regarding topic prevalence over time (Figure 3). When analysing trends over time, it should also be noted that the more recent publications exert a more substantial influence over the modelling process due to the persistent inflation of academic publishing. Furthermore, overall inflation means that lack of growth constitutes a relative decline. The choice of the relatively low setting of 12 topics results in identifying broad and general content clusters, likely to have some representation over the entire time series. Had the number been set at a much larger number, identifying much narrower topics, an analysis of popularity trends over time would yield more interesting results. Our approach, however, necessitates the low setting to allow for the application of secondary LDA identifying micro-topics nested within the macro-topics. If trend analysis were the primary goal, the publication date could be made available to the modelling algorithm as a covariate, which is possible in some alternative approaches, such as Structural Topic Modelling (STM).

5 | DISCUSSIONS AND LIMITATIONS

The proposed literature review of the interdisciplinary field of social welfare involves a series of limitations, which pertain to methodological decisions when applying the LDA algorithm, the choice of data source, and the problem of social welfare itself.

The LDA approach provides the ability to deal with large bodies of textual data and distinguish latent unobserved patterns in the discourse can hardly be achieved through traditional literature review procedures due to their poor scalability. However, LDA suffers from several drawbacks and limitations. Crucially, the algorithmic solutions may seem 'objective' as the 'reading' is performed by a non-human agent, but the human researcher can introduce substantial bias into the analysis through arbitrary coding decisions. The major point of impact falls on the data pre-processing stage (i.e., the procedures transforming the unstructured text into the document-term matrix, which is fed into the LDA algorithm). Text pre-processing involves hundreds of lines of code, and at every step, researchers make decisions which impact the output (e.g., how aggressively the text should be spell-checked and stemmed, which common stopwords should be eliminated, which compound phrases should be retained as phrases rather than individual tokens, what level of sparsity in the DTM matrix is acceptable). In developing the analysis, the cleaning code is typically tweaked on multiple occasions to deal with imperfections in the output, and although all those choices are documented and can be traced in the code, they still may introduce strong biases. Notably, another approach exists, where researchers may use default package-provided pre-processing functions as a kind of black box. While this prevents intentional researcher bias, it also leads to inferior cleaning results and precludes understanding of what is actually going on in the pre-processing stage. The second major source of arbitrariness comes at the point of setting the hyper-parameters and other control options guiding the LDA algorithm, which have definite but not always apparent impacts on the resulting topic models. In our analysis, we have run through multiple iterations of the algorithm to compare the outputs in model quality metrics and the interpretability of results. While those tweaks may be compared to manipulating a lens to sharpen the picture, in topic modelling, there is no one correct solution, as the snapshot of discourse can be taken from many different vantage points depending on the researcher's interests.

The second limitation relates to the problem with the term 'social welfare', which is understood differently in different countries. As noted in the introduction, the term is sometimes used interchangeably with such words as social or collective well-being, prosperity, or happiness. The choice of the term and the area of research was not accidental, as the evolving theme of social welfare fits well with machine learning approaches to text mining due to its breadth and scale. The analysis confirmed the intuition regarding the differences in usage, even though in the presented model, it is not possible to distinguish them as sharply as human readers probably could within the framework of classical quantitative content analysis. The advantages of the LDA approach outweigh the acknowledged limitations.

Another technical limitation seems to arise from restricting the empirical base to the articles published on Scopus, which is not the only available database. However, our exploration of other databases, most notably the Web of Science, revealed that the publication records are largely duplicated. The increase would not be substantial even if the combination of records from different databases yielded more articles. At the same time, the differences in the formatting and availability of meta-data across different platforms would create serious challenges to comparability. As our research has been exploratory and innovative in the field of social welfare, the limitations arising from the choice of the Scopus database seem secondary.

Finally, so far as we know, the LDA within LDA approach has not been used thus far (cf. Baranowski & Cichocki, 2021). While the STM approach has roughly inspired it, it does not constitute its variant. The two-stage approach of selecting a small number of macro-themes and then distinguishing several micro-themes within makes it possible to identify the constituent parts of broader content clusters. Within our analysis, the micro-topic patterns seemed intuitively plausible, even though we did not attempt to tweak the analyses manually—they are all executed by the same looped procedure. Exploring the internal composition of macro-topics (Figure 4) constitutes an alternative to outright modelling dozens of topics at the first stage, as it introduces structure to the output. Implementing this procedure and evaluating its analytic efficacy requires further research.

6 | CONCLUSIONS

Social welfare is a conventional term whose meaning varies across countries and cultures (Spicker, 2012, p. 183). Our text-mining exploration of publication records was intended to make an essential and innovative contribution to understanding the diversity of research on the phenomenon of social welfare. The generation of 12 topics utilising the LDA algorithm contributed to a significant expansion of knowledge by simply extracting them from a vast corpus of data. In turn, the use of secondary LDA in LDA provided an even more precise picture of the thematic threads present in the social welfare literature of the last two decades. The knowledge gained through the application of machine learning analysis, although only beginning the discussion in this area, allows for a better understanding of the theoretical underpinnings and practical implications of concepts and policies related to social welfare (e.g. the role of the environment in welfare research, as in T5, T7 and T11). It also demonstrates the possibilities of using advanced technologies at the intersection of linguistics, computer science, and artificial intelligence in the domain areas of the social sciences. Notwithstanding its limitations, LDA provides an intersubjectively verifiable, communicable, coherent procedure for analysing large corpora of data and their latent themes. The undoubted advantages of machine learning include that 'it generates topics directly from the data and thus is unbiased by subjective judgement' (Savin & van den Bergh, 2021, p. 15). However, the results produced by the algorithm prove sensitive to parameter settings and the range of choices made by researchers in the text pre-processing stages.

Our analysis concurs with the observation that 'concepts of welfare can imply very different things to different people' (Titmuss, 2018, p. 146), as demonstrated by the 12 internally diverse themes. Equally interesting are the connections between researchers publishing on social welfare by country of affiliation (see Figure A1). The USA, China, Canada and the UK are at the centre, along with Hong Kong and South Korea. The Scandinavian countries, Sweden, Norway and Finland, are also at the centre of this collaboration on welfare topics. The positions of Germany, France and the Netherlands may seem surprisingly weak, especially in the latter case, as English publications are standard among Dutch researchers (in contrast to German and French publication cultures that prefer monographs in national languages).

Hence, the 12 themes obtained provide not only a unique overview of the thematic areas but also a deeper insight into their specificity (sub-themes), the temporal framework of the issues addressed, the impact on the discipline (studied by the number of citations), or the national specificities of the research of the phenomenon under consideration. The propaedeutic nature of this article has shown how to (a) collect and use large amounts of data, representing real challenges for the social sciences, and (b) innovatively review 'hidden' thematic areas within the dynamic social welfare research.

ACKNOWLEDGEMENTS

Mariusz Baranowski received support for this research from the National Science Centre, Poland (grant number 2021/05/X/HS6/00067).

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

DATA AVAILABILITY STATEMENT

The dataset is available from the Scopus database (or authors on reasonable request).

ORCID

Mariusz Baranowski https://orcid.org/0000-0001-6755-9368
Piotr Cichocki https://orcid.org/0000-0002-6501-9082
Jim McKinley https://orcid.org/0000-0002-9949-8368

REFERENCES

Adler, M. D. (2019). *Measuring social welfare: An introduction*. Oxford University Press.

Albert, M., & Hahnel, R. (2017). *Quiet revolution in welfare economics*. Princeton University Press.

Altbach, P. G., & de Wit, H. (2018). Too much academic research is being published. *University World News*. 7 Sept. 2018 https://www.universityworldnews.com/post.php?story=20180905095203579

Alvarez, R. M. (2016). *Computational social science: Discovery and prediction*. Cambridge University Press.

Aravacik, E. D. (2018). Social policy and the welfare state. In B. Acikgoz (Ed.), *Public economics and finance* (pp. 1–20). IntechOpen. https://doi.org/10.5772/intechopen.82372

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*(4), 328–346. https://doi.org/10.1086/256963

Asmussen, C. B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(93), 1–18. https://doi.org/10.1186/s40537-019-0255-7

Baldock, J. (2007). Social policy, social welfare and the welfare state. In J. Baldock, N. Manning, & S. Vickerstaff (Eds.), *Social policy* (3rd ed., pp. 5–30). Oxford University Press.

Baranowski, M. (2019). The struggle for social welfare: Towards an emerging welfare sociology. *Society Register*, *3*(2), 7–19. https://doi.org/10.14746/sr.2019.3.2.01

Baranowski, M. (2021). The sharing economy: Social welfare in a technologically networked economy. *Bulletin of Science, Technology & Society*, *41*(1), 20–30. https://doi.org/10.1177/02704676211010723

Baranowski, M. (2022). Nature-based social welfare and socially responsible consumption: Is circular economy a viable solution? In J. Bhattacharyya (Ed.), *Dealing with socially responsible consumers* (pp. 33–50). Palgrave Macmillan.

Baranowski, M., & Cichocki, P. (2021). Good and bad sociology: Does topic modelling make a difference? *Society Register*, *5*(4), 7–22. https://doi.org/10.14746/sr.2021.5.4.01

Bean, P. & MacPherson, S. (Eds.). (2018). *Approaches to welfare*. Routledge.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. https://doi.org/10.1002/asi.23329

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood.

Burroughs, J. E., & Rindfleisch, A. (2012). What welfare? On the definition and domain of transformative consumer research and the foundational role of materialism. In D. G. Mick, S. Pettigrew, C. Pechmann, & J. L. Ozanne (Eds.), *Transformative consumer research for personal and collective well-being* (pp. 249–266). Routledge.

Butler, I., & Drakeford, M. (2005). *Scandal, social policy and social welfare*. The Policy Press.

Callahan, J. L. (2014). Writing literature reviews: A reprise and update. *Human Resource Development Review*, *13*(3), 271–275. https://doi.org/10.1177/1534484314536

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*, *92*(3), 28–40. http://www.jstor.org/stable/23526834

Dover, M. A. (2016). Human needs: Overview. In *Encyclopedia of social work* (pp. 1–32). Oxford University Press. https://doi.org/10.1093/acrefore/9780199975839.013.554

Drover, G., & Woodsworth, D. (1978). Social welfare theory and social policy. *Canadian Journal of Social Work Education/Revue canadienne d'éducation en service social*, *4*(2–3), 19–41.

Ellison, C. G. (1991). Religious involvement and subjective well-being. *Journal of Health and Social Behavior*, *32*(1), 80–99. https://doi.org/10.2307/2136801

Encyclopedia of social work. (1987). National Association of Social Workers.

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Polity Press.

Esping-Andersen, G. (1999). *Social foundations of postindustrial economics*. Oxford University Press.

Forder, A., Caslin, T., Ponton, G., et al. (2018). *Theories of welfare*. Routledge.

Frank, D. J., Hironaka, A., & Schofer, E. (2000). The nation-state and the natural environment over the twentieth century. *American Sociological Review*, *65*(1), 96–116. https://doi.org/10.2307/2657291

Friedman, L. M. (1969). Social welfare legislation: An introduction. *Stanford Law Review*, *21*(2), 217–247. https://doi.org/10.2307/1227652
Goodman, L. A., & Markowitz, H. (1952). Social welfare functions based on individual rankings. *American Journal of Sociology*, *58*(3), 257–262. https://doi.org/10.1086/221147
Gough, I. (2000). *Global capital, human needs and social policies: Selected essays, 1994-99*. Macmillan Press.
Gough, I. (2019). Universal basic services: A theoretical and moral framework. *The Political Quarterly*, *90*(3), 534–542. https://doi.org/10.1111/1467-923X.12706
Gough, I. (2022). Two scenarios for sustainable welfare: A framework for an eco-social contract. *Social Policy and Society*, *21*(3), 460–472. https://doi.org/10.1017/S1474746421000701
Greve, B. (2020). *Myths, narratives and welfare states: The impact of stories on welfare state development*. Edward Elgar Publishing.
Hadley, R., & Hatch, S. (2019). *Social welfare and the failure of the state*. Routledge.
Hicks, J. R. (1939). The foundations of welfare economics. *The Economic Journal*, *49*(196), 696–712. https://doi.org/10.2307/2225023
Hughes, G. & Lewis, G. (Eds.). (1998). *Unsettling Welfare: The reconstruction of social policy*. Routledge.
Jacobs, T., & Tschötschel, R. (2019). Topic models meet discourse analysis: A quantitative tool for a qualitative approach. *International Journal of Social Research Methodology*, *22*(5), 469–485. https://doi.org/10.1080/13645579.2019.1576317
Karpf, D. (2012). Social science research methods in Internet time. *Information, Communication & Society*, *15*(5), 639–661. https://doi.org/10.1080/1369118x.2012.665468
Kipo-Sunyehzi, D. D. (2021). Global social welfare and social policy debates: Ghana's health insurance scheme promotion of the well-being of vulnerable groups. *Journal of Social Service Research*, *47*(1), 73–87. https://doi.org/10.1080/01488376.2020.1716919
Kiser, E. (1999). Comparing varieties of agency theory in economics, political science, and sociology: An illustration from state policy implementation. *Sociological Theory*, *17*(2), 146–170. https://doi.org/10.1111/0735-2751.00073
Lamont, M., & Molnár, V. (2002). The study of boundaries in the social sciences. *Annual Review of Sociology*, *28*(1), 167–195. https://doi.org/10.1146/annurev.soc.28.110601.141107
Linares, J., & Cabaña, G. (2022). Towards an ecology of care: Basic Income after the nation-state. *Society Register*, *6*(3), 29–56. https://doi.org/10.14746/sr.2022.6.3.03
Lindstedt, N. C. (2019). Structural topic modeling for social scientists: A brief case study with social movement studies literature, 2005-2017. *Social Currents*, *6*(4), 307–318. https://doi.org/10.1177/2329496519846505
Macarov, D. (1995). *Social welfare: Structure and practice*. SAGE Publications.
Madison, B. Q. (2019). *The meaning of social policy: The comparative dimension in Social welfare*. Routledge.
Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. John Murray.
Midgley, J. (2017). *Social welfare for a global era: International perspectives on policy and practice*. SAGE.
Miller, S. E., Hayward, R. A., & Shaw, T. V. (2012). Environmental shifts for social work: A principles approach. *International Journal of Social Welfare*, *21*(3), 270–277. https://doi.org/10.1111/j.1468-2397.2011.00848.x
Milojević, S. (2015). Quantifying the cognitive extent of science. *Journal of Informetrics*, *9*(4), 962–973. https://doi.org/10.1016/j.joi.2015.10.005
Ngamaba, K. H., Panagioti, M., & Armitage, C. J. (2018). Income inequality and subjective well-being: A systematic review and meta-analysis. *Quality of Life Research*, *27*(3), 577–596. https://doi.org/10.1007/s11136-017-1719-x
Pääkkönen, J., & Ylikoski, P. (2021). Humanistic interpretation and machine learning. *Synthese*, *199*(1–2), 1461–1497. https://doi.org/10.1007/s11229-020-02806-w
Pigou, A. C., & Aslanbeigui, N. (2002). *The economics of welfare* (1st ed.). Routledge.
Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684
Rochefort, D. A. (1986). *American social welfare policy: Dynamics of formulation and change*. Routledge.
Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, *30*(4), 22–46. https://doi.org/10.1177/0263276413484941
Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, *41*(5), 885–899. https://doi.org/10.1177/0038038507080443
Savin, I., & van den Bergh, J. (2021). Main topics in EIST during its first decade: A computational-linguistic analysis. *Environmental Innovation and Societal Transitions*, *41*, 10–17. https://doi.org/10.1016/j.eist.2021.06.006
Selwyn, N. (2011). Making sense of young people, education and digital technology: The role of sociological theory. *Oxford Review of Education*, *38*(1), 81–96. https://doi.org/10.1080/03054985.2011.577949
Shaw, T. V. (2008). An ecological contribution to social welfare theory. *Social Development Issues*, *30*(3), 13–26.
Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Spicker, P. (1988). *Principles of social welfare: An introduction to thinking about the welfare state*. Routledge.

Spicker, P. (2012). Social welfare, provision and finance. In R. Chadwick (Ed.), *Encyclopedia of applied ethics* (2nd ed., pp. 182–188). Academic Press.

Stewart, N. P., & Viney, L. L. (1975). Role of the psychologist in social welfare. *Australian Psychologist*, *10*(2), 163–178. https://doi.org/10.1080/00050067508256454

Sukhija, N., Tatineni, M., Brown, N., Van Moer, M., Rodriguez, P., & Callicott, S. (2016). Topic modeling and visualisation for big data in social sciences. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)* (pp. 1198–1205). https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0183

Sullivan, M. (2018). *Sociology and social welfare*. Routledge.

Taylor, D. (2011). Well-being and welfare: A psychosocial analysis of being well and doing well enough. *Journal of Social Policy*, *40*(4), 777–794. https://doi.org/10.1017/S0047279411000249

Taylor-Gooby, P. (2019). What is social policy? *The British Academy*. https://www.thebritishacademy.ac.uk/blog/what-is-social-policy/

Thangaraj, M., & Sivakami, M. (2018). Text classification techniques: A literature review. *Interdisciplinary Journal of Information, Knowledge, and Management*, *13*, 117–135. https://doi.org/10.28945/4066

Titmuss, R. M. (2018). Welfare state and welfare society. In N. Timms & D. Watson (Eds.), *Talking about welfare* (pp. 145–161). Routledge.

Wallerstein, I. (2001). *Unthinking social science: The limits of nineteenth-century paradigms*. Temple University Press.

Walsh, M., Stephens, P., & Moore, S. (2000). *Social policy and welfare*. Stanley Thornes.

Wickwar, W. (1946). Social welfare and political science. *American Political Science Review*, *40*(3), 563–571. https://doi.org/10.2307/1949330

Williamson, J. B., & Fleming, J. J. (1977). Convergence theory and the social welfare sector. *International Journal of Comparative Sociology*, *18*(3–4), 242–253. https://doi.org/10.1163/156854277x00168

Wilson, D. B., Solomon, T. A., & McLane-Davison, D. (2020). Ethics and racial equity in social welfare policy: Social work's response to the COVID-19 Pandemic. *Social Work in Public Health*, *35*(7), 617–632. https://doi.org/10.1080/19371918.2020.1808145

Woodgate, G., & Redclift, M. (1998). From a 'sociology of nature' to environmental sociology: Beyond social construction. *Environmental Values*, *7*(1), 3–24. https://doi.org/10.3197/096327198129341447

AUTHOR BIOGRAPHIES

Mariusz Baranowski, Ph.D., is an adjunct researcher at the Faculty of Sociology Adam Mickiewicz University, Poznan. His research interest includes political and economic sociology, social welfare issues, and the critique of neoliberalism. His most recent projects are concerned with topic modelling (principal investigator of a grant of the National Science Centre, Poland, titled *Social welfare in the light of topic modelling: a preliminary study*) and improvement of higher education (coordinator and project manager of a grant of the Erasmus + KA220-HED—Cooperation partnerships in higher education). He is an Editor-in-Chief of the journal Society Register.

Piotr Cichocki, Ph.D., is an adjunct researcher at the Faculty of Sociology, Adam Mickiewicz University, Poznan. His research interest includes monitoring social and political attitudes by cross-country surveys, attitudes towards European integration, and research on European Identity.

Jim McKinley (SFHEA) is Associate Professor of Applied Linguistics and TESOL at University College London. He has taught in higher education in the UK, Japan, Australia, and Uganda, as well as US schools. His research targets implications of globalisation for L2 writing, language education, and higher education studies, particularly the teaching-research nexus. Jim is co-author and co-editor of several books on research methods in applied linguistics. He is an Editor-in-Chief of the journal System, and Co-Editor of the Cambridge Elements series Language Teaching (Cambridge University Press).

How to cite this article: Baranowski, M., Cichocki, P., & McKinley, J. (2023). Social welfare in the light of topic modelling. *Sociology Compass*, e13086. https://doi.org/10.1111/soc4.13086

APPENDIX

See Table A1 here.

TABLE A1 Model diagnostics.

t_num	t_labels	t_size	t_coherence	t_rank
1	T1. Political government and moral discourse	399	−162.1	6
2	T2. Market economy and competition	201	−124.0	1
3	T3. Community health care and education	252	−153.7	4
4	T4. Risk of abuse and violence	300	−207.9	12
5	T5. Sustainable development and economic growth	222	−142.6	2
6	T6. Financial management	276	−194.3	10
7	T7. Environmental economics	234	−147.1	3
8	T8. Utility functions and preferences	273	−202.5	11
9	T9. Intervention and assistance support	278	−177.8	8
10	T10. Income and urban inequalities	209	−168.0	7
11	T11. Transport policy	270	−179.0	9
12	T12. Family and children support services	211	−155.4	5

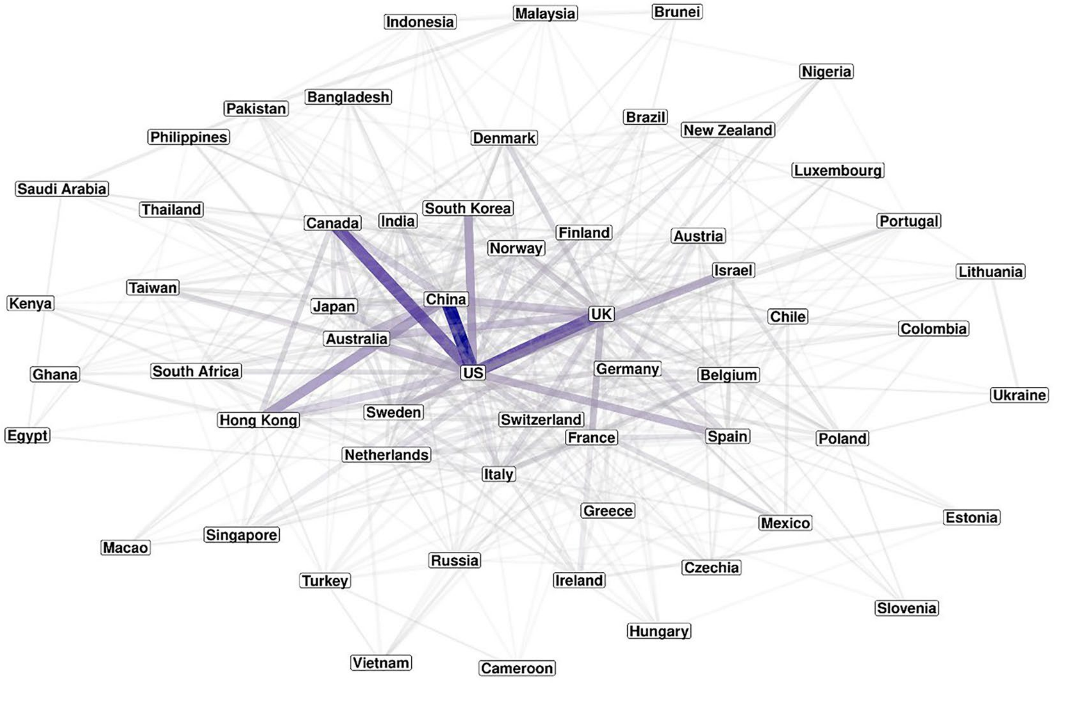

FIGURE A1 Links between countries of affiliation of authors publishing on social welfare.

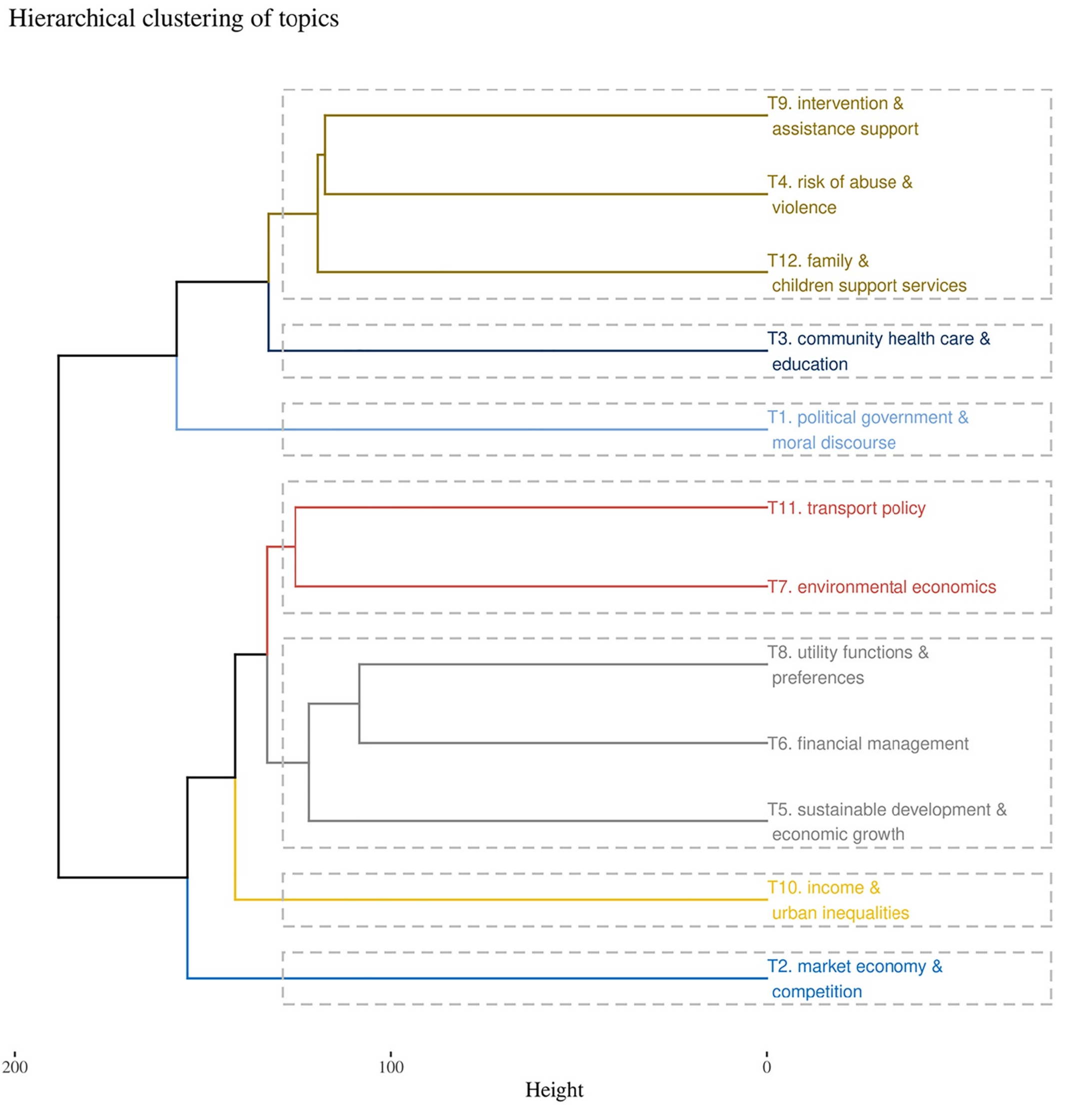

FIGURE A2 Hierarchical clustering of topics.