

## **Bibliometric Structure of Collective Identities: Progress of Publication on Collective Identity in the Scopus Database**

**Hafizan Mohamad Naim<sup>1\*</sup>, Mohd Azizuddin Mohd Sani<sup>2</sup>**

<sup>1</sup>Ghazali Shafie Graduate School of Government  
Universiti Utara Malaysia, Sintok, Kedah, 06010, MALAYSIA

<sup>1</sup>Fakulti Sains Sosial dan Kemanusiaan  
Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, 94300, MALAYSIA

<sup>2</sup>School of International Studies  
Universiti Utara Malaysia, Sintok, Kedah, 06010, MALAYSIA

\*Corresponding Author

DOI: <https://doi.org/10.30880/ahcs.2023.04.01.005>

Received 7 October 2022; Accepted 7 April 2023; Available online 30 June 2023

**Abstract:** The formation of collective identity, at its core, is the product of people navigating through social relations. Collective identity has become one of the most prominent sub theories in social identity theory. This study examines the progress of publication on collective identity and mapping the structure of its bibliometric information on journals published in the Scopus database. The online retrieval process was done on 28 July 2020 and obtained 689 documents. The data were analysed using a standard bibliometric approach using Microsoft Excel, VOSviewer, and Harzing's Publish and Perish. The result shows an increased growth rate of literature in the year 1997. The growth particularly linked to the publication of "Who Is This "We"? Levels of Collective Identity and Self Representations" by M.B. Brewer and W. Gardner' in 1996 and "The Construction of Collective Identity." S.N. Eisenstadt, B. Giesen in 1995. This study reveals research on collective identity subject are predominantly related to constructive analysis, identity analysis, discourse analysis, and collective action analysis. There is increased attention in linking social media with collective identity research, while the social movement domain will continue to be an important domain within the corpus. This study offers an insight to researchers on current trends and the future trajectory of collective identity research.

**Keywords:** bibliometric analysis, collective identity, keywords analysis, authorship analysis, citation analysis

## 1. Introduction

Collective identity is a social construction from social collectivities such as groups, families, communities, organisations, nations, regions, or groups such as political parties or even social movements [1]. Individuals themselves cannot seclude themselves from the process of self-identification and experiencing a sense of belonging within the community at large. Perhaps, it has always been a characteristic feature of human interaction, whether among early preliterate humans or those in the modern social world [2]. As social animals, people have the innate drive to feel they belonged in their most superficial sense. As time goes by and the world becomes globalised, people tend to establish their collective identities [3], making the idea of their collective identities are increasingly becoming structurally differentiated and more fragmented. Furthermore, modern society is increasingly influenced by information and the environment they are in – bringing about new forms of social control, conformity pressures, and information processing response [4], which further expounds the dissimilarity.

Research on collective identity also goes through series of changes, from collective identity conceptualisation [5], identity construction [6], collective agency and collective action [7], and contestation [8]. In the last quarter of a century, along with the technological revolution, the collapse of capitalism, and the decline of statism, we have witnessed the widespread emergence of powerful expressions of collective identity that question globalisation and cosmopolitanism for the sake of cultural singularity and the sovereignty of people over their lives and environment [3]. Hence, it is not surprising that collective identity is placed under the subtheory of social identity theory.

If the main terms and critical concepts can be identified, capturing the various domains of the concept, the research trend can be identified. It may be reflected in the scholarly work that deals with the conceptualisation, empirical manifestations, analytical discussions, and analysis corpus of collective identity. Although the study on collective identity has steadily grown over the years, there have been relatively few attempts to report its literature trend using bibliometric analysis. Bibliometric analysis is considered an extensive quantitative indicators that considers diverse sectors (i.e., researchers, research administrators, government, etc.) on all levels (i.e., individual, institutional, national, and global) [9]. Since the main aim of science is to produce and disseminating scientific knowledge, the application of quantitative methods such as bibliometric analysis is viewed as a good technique in assessing scientific knowledge. Therefore, this study is a good starting point to systematically explore the progress of publication on collective identity and mapping it through figurative representations of its bibliometric information's relationship and structure.

## 2. Methods

This paper aims to identify the publication progress on collective identity research that previous researchers have achieved. For this study, a query was conducted on the Scopus database based on the title of the document (TITLE (“collective identity”)) on 28 July 2020. This query produced 689 documents comprising of literature produced within the time frame of 1963 to 2020. Some of the bibliometric indicators such as document and source type, language, country, subject area, year, source title, and keywords, citation, and authorship network visualisation mapping will be presented in this paper to identify the future direction of this research.

### 3. Bibliometric analysis

Bibliometric analysis is instrumental in analysing the publication progress for any specific corpus of information irrespective of the subject area or domain [10]. It can produce bibliometric maps consisting of the scientific communities' relationship and structure within a given domain topic. Bibliometric maps or maps of science are a figurative representation of the scientific field visualized by several map components showing association themes [11]. The exploration of bibliometric data will then produce a meaningful insight into research progress achieved by previous researchers and identify the potential future direction that has yet to be discovered.

### 4. Result

The data collected were analysed to identify document and source types, distribution of languages and countries of the publications, subject areas, keywords, authorship, and citation analysis. Most of the findings are presented as frequency and percentage.

#### 4.1 Document types and source types

The first analysis will identify document types and source types. As indicated in Table 1, more than half of the total publications are in the form of an article (69.67%), followed by book chapter (16.11%) and review (8.56%). The rest of the document types are representing less than 5% of the total publication.

**Table 1 - Document Types**

Document types	Number of publications	Percentage
Article	480	69.67%
Book Chapter	111	16.11%
Review	59	8.56%
Book	16	2.32%
Conference Paper	13	1.89%
Editorial	4	0.58%
Note	3	0.44%
Erratum	2	0.29%
Undefined	1	0.15%
Total	689	100.00%

Analysing source types for publication on collective identity research reveals five types of source types: journal, book, book series, conference proceedings, and trade journal. Table 2 shows out of these five source types, publication through journal is the most popular type of publication source with 529 journal publications (76.78%), followed by book publications with 125 publications (18.14%) and book series with 25 publications (3.63%). Both conference proceedings and trade journals are the least popular source types, amounting to only 1.46% of the overall publications.

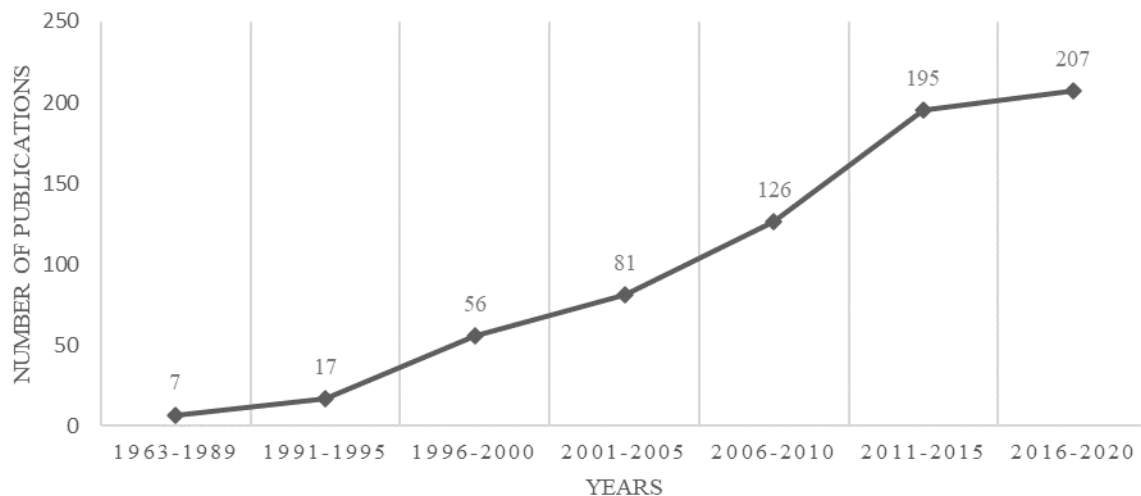
Analysing source types for publication on collective identity research reveals five types of source types: journal, book, book series, conference proceedings, and trade journal. Table 2 shows out of these five source types, publication through journal is the most popular type of publication source with 529 journal publications (76.78%), followed by book publications with 125 publications (18.14%) and book series with 25 publications (3.63%). Both conference

proceedings and trade journals are the least popular source types, amounting to only 1.46% of the overall publications.

**Table 2 - Source Types**

Source types	Number of Publications	Percentage
Journal	529	76.78%
Book	125	18.14%
Book Series	25	3.63%
Conference Proceeding	9	1.31%
Trade Journal	1	0.15%
Total	689	100.00%

Figure 1 shows the statistic of annual publications on collective identity from 1963 to 2020. The first published research on collective identity recorded in the Scopus database was a review paper by Butwell [12] entitled “Individual and Collective Identity and Nation-Building”, published by Cambridge University Press. From 1963 until 1989, only seven publications on collective identity were published. The number somewhat improves from 1990 to 1995 to 17 publications. The growth of publications only starts picking up from 1996-2000 and consistently rising. During the last ten years, the publication on the topic is averaging 40.2 publications per year compared to 20.7 publications per year during 2001-2010 and 7.3 publications on average during 1991-2000.



**Fig. 1 - Annual growth of publications**

#### 4.2 Languages distribution and countries of publications

Based on Table 3, English is the most common language for most publications for this research domain (89.63%). Other common languages used are French (2.98%), Spanish (2.98%), German (1.42%), and Italian (1.14%). The rest of the nine languages published less than two publications each.

**Table 3 - Languages Used for Publication**

<b>Language</b>	<b>Number of Publications</b>	<b>Percentage</b>
English	631	89.63%
French	21	2.98%
Spanish	21	2.98%
German	10	1.42%
Italian	8	1.14%
Polish	2	0.28%
Russian	2	0.28%
Slovak	2	0.28%
Slovenian	2	0.28%
Afrikaans	1	0.14%
Bosnian	1	0.14%
Czech	1	0.14%
Portuguese	1	0.14%
Catalan	1	0.14%
Total	704	100.00%

For the geographical distribution of publications for this research domain, only 13 countries out of the overall 61 countries contribute to more than ten publications per country. The publication from these 13 countries encompasses more than half of the overall publication at about 68.82%. Out of these 13 countries, the most significant contributing country is the United States of America (USA), contributing a total of 215 publications (27.35%) from the overall 689 publications produced (refer to Table 4). It should be noted that even though the USA publishes a large amount of the published work, the publications from European countries are contributing to more than half of the overall publications produced.

**Table 4 - Most Active Countries with a Minimum of 10 Publications**

<b>Country</b>	<b>Number of publications</b>	<b>Percentage</b>
United States	215	27.35%
United Kingdom	67	8.52%
Germany	56	7.12%
Israel	40	5.09%
Canada	31	3.94%
Spain	26	3.31%
Italy	19	2.42%
Australia	18	2.29%
Netherlands	18	2.29%
France	16	2.04%
Sweden	13	1.65%
South Africa	11	1.40%
Switzerland	11	1.40%
Total	541	68.82%

### 4.3 Subject area

This study also classifies the published documents based on their subject areas. More than 70 per cent of the collective identity publications are produced under social sciences and arts and humanities. Besides being published under social sciences and humanities, studies on collective identity have also been published in diverse subject areas ranging from psychology, business, management and accounting, computer science, medicine, nursing, and even physics and astronomy.

**Table 5 - Subject Area**

<b>Subject area</b>	<b>Number of publications</b>	<b>Percentage</b>
Social Sciences	526	50.14%
Arts and Humanities	256	24.40%
Psychology	78	7.44%
Business, Management, and Accounting	63	6.01%
Economics, Econometrics, and Finance	32	3.05%
Computer Science	17	1.62%
Environmental Science	17	1.62%
Medicine	15	1.43%
Earth and Planetary Sciences	12	1.14%
Decision Sciences	8	0.76%
Agricultural and Biological Sciences	7	0.67%
Engineering	7	0.67%
Agricultural and Biological Sciences	7	0.67%
Health Professions	2	0.19%
Mathematics	2	0.19%
Multidisciplinary	2	0.19%
Biochemistry, Genetics and Molecular Biology	1	0.10%
Energy	1	0.10%
Materials Science	1	0.10%
Nursing	1	0.10%
Physics and Astronomy	1	0.10%
<b>Total</b>	<b>1049</b>	<b>100.00%</b>

### 4.4 Most active source titles

Table 6 presents the most active source title that has five or more publications on collective identity. The most active source title is Mobilisation, while the most active publisher is Routledge. Out of the list of the most active source titles provided in Table 7, Routledge alone published three source titles, in which the total publications produced by the source titles are 17 publications. The three Routledge's source titles are Information Communication and Society, Social Movement Studies, and Social Identities.

**Table 6 - Most Active Source Title**

Source title	Publisher	Source type	Total	Percentage
Mobilization	San Diego State University	Article	10	3.37%
Information Communication and Society	Routledge	Article	6	2.02%
Social Movement Studies	Routledge	Article	6	2.02%
European Journal of Social Theory	SAGE Publications Ltd	Article	5	1.68%
Journal of Common Market Studies	Blackwell Publishing Ltd	Article	5	1.68%
Nationalities Papers	Cambridge University Press	Article	5	1.68%
Research in Social Movements Conflicts and Change	Emerald Group Publishing Ltd.	Article	5	1.68%
Social Identities	Routledge	Article	5	1.68%

#### 4.5 Keyword analysis

Keyword analysis is often accompanied by mapping the network visualisation of significant terms. A keyword analysis is often done to understand a) the structure of a bibliometric network, b) the identification of key themes in a domain, c) the relationship between the subject, and d) the diffusion of ideas over time [10]. Specific for this study, author keywords have been mapped with VOSviewer to identify key themes and the diffusion of ideas over time. VOSviewer generates a complex visualisation map, processes massive volumes of data, and creates clusters where each cluster includes significant correlations between nodes [13]. To analyse the visualisation map, elements such as colour, circle size, font size, and thickness of connecting lines were used to present the relationship with other keywords [14]. Nodes size indicates the weight of the term's occurrence, while the line's thickness indicates the strength of the relationship amongst the terms [15].

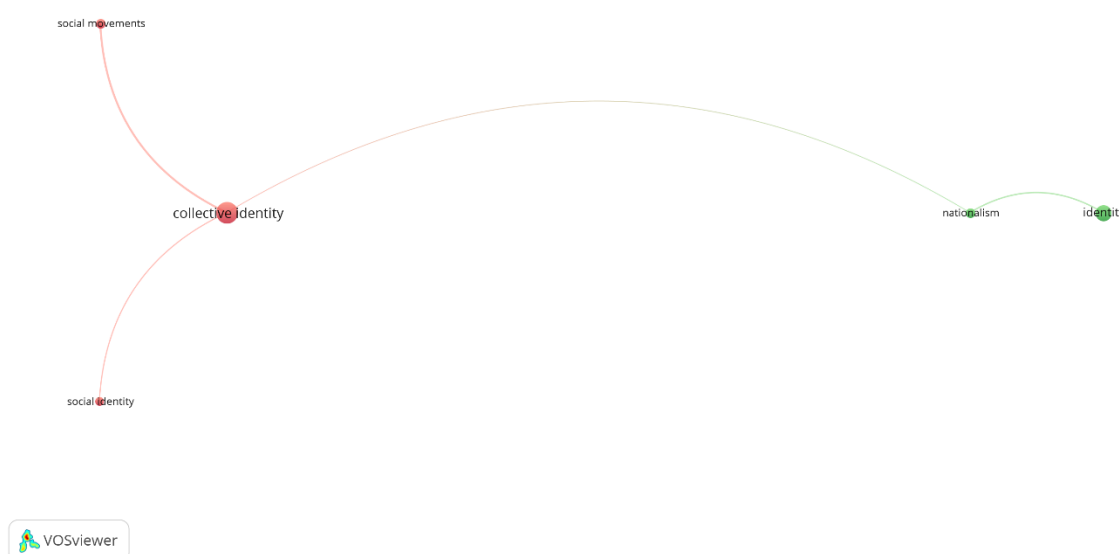
For this study, results for keyword analysis will be presented through three network visualisation maps. The first visualisation map (Figure 2) will explain the relationship of author keywords within the corpus title of published work from 1963 to 2010. It will mainly show the established baseline keywords set by published authors relating to collective identity. The second map (Figure 3) will explain the relationship of author keywords used for the subsequent years 2011 until 2020, while the third map (Figure 4) will show the overlay visualisation map of average citation in 2011-2020, indicating the latest development of collective identity research. The fourth map (Figure 5) will explain the overall relationship of general author keywords within the corpus title. Figure 3, Figure 4 and Figure 5 will be used to highlight the latest and emerging trend within the corpus title, bring to light the general idea of the progression and diffusion of ideas on collective identity literature.

##### a) Author keywords 1963-2010

Figure 2 shows the term co-occurrence network's visualisation map based on title fields with five minimum numbers of occurrences for a term between the years 1963 to 2010. Out of

397 keywords generated, only five keywords meet the required threshold. These five keywords can be identified as the baseline keywords for the corpus title collective identity.

The baseline keywords that have been identified are collective identity, identity, nationalism, social movements, and social identity. These five keywords are grouped within two significant clusters represented by the red cluster keywords and the green cluster keywords. The first cluster, which is coloured in red, is related to collective identity, social movement, and social identity cluster, while the second cluster is coloured in green, is related to nationalism and identity cluster.



**Fig. 2 - Network visualisation map of author keywords 1963 to 2010**

The publications related to the corpus during the period are mainly dealing with two main domains, collective identity (49.30%) and identity (26.76%) (refer to Table 7), while the nationalism domain acts as a connecting domain for both identity and collective identity domains (refer to Figure 2).

**Table 7 - Author Keywords 1963 to 2010**

<b>Keyword</b>	<b>Occurrences</b>	<b>Percentage</b>
collective identity	35	49.30%
identity	19	26.76%
nationalism	6	8.45%
social movements	6	8.45%
social identity	5	7.04%

Earlier literature focuses more on the construction of understanding collective identity from all levels of social collectivities, be it groups, families, communities, organisations, nations, regions or groups such as political parties or even social movements [1]. It is consistent with the findings which saw the five top-cited publications during the year 1963 to 2010 presenting varying perspectives in both approach and emphasis – from social psychological analysis, social movement to identity formation and international state but somehow constraint



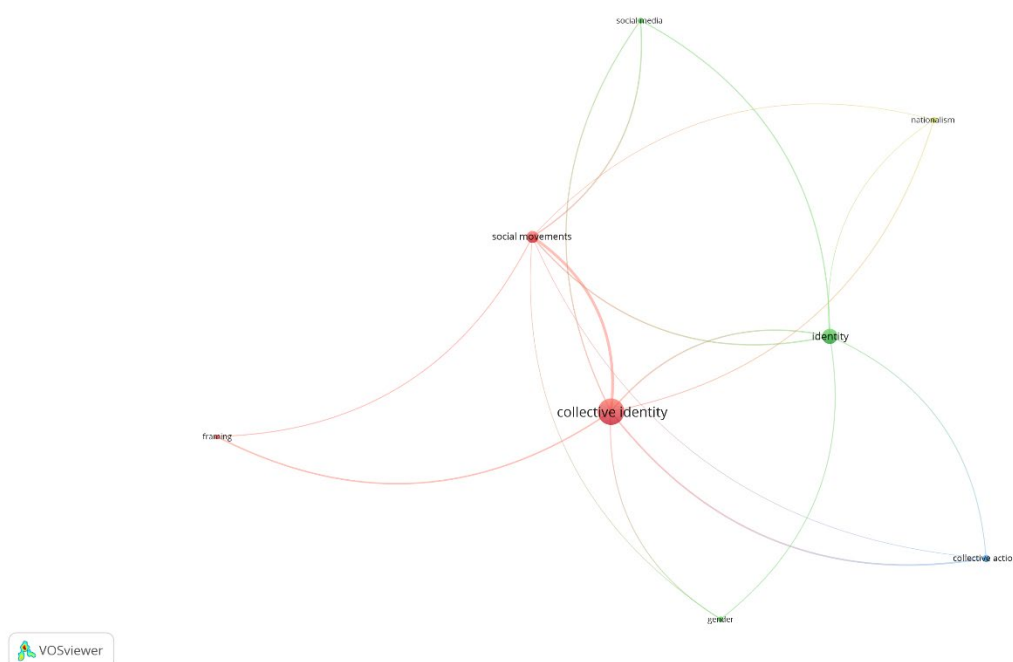
in a way that it only emphasises on the discussion of understanding collective identity and identity issues (refer to Table 8).

**Table 8 - Five Top-cited Publications Year 1963 to 2010**

Cites	Cites/ Year	Authors	Title	Year	Source	Publisher
1806	75.25	M.B. Brewer, W. Gardner	Who Is This “We”? Levels of Collective Identity and Self Representations	1996	Journal of Personality and Social Psychology	American Psychological Association Inc.
1136	59.79	F. Polletta, J.M. Jasper	Collective identity and social movements	2001	Annual Review of Sociology	Annual Reviews Inc.
928	58	R.D. Ashmore, K. Deaux, T. McLaughlin-Volpe	An Organising Framework for Collective Identity: Articulation and Significance of Multidimensionality	2004	Psychological Bulletin	American Psychological Association Inc.
512	34.62	A. Wendt	Collective identity formation and the international state	1994	American Political Science Review	Cambridge University Press
900	33.26	B. Simon, B. Klandermans	Politicised collective identity: A social psychological analysis	2001	American Psychologist	American Psychological Association Inc.

b) Author keywords 2011-2020

Figure 3 shows the visualisation of the term co-occurrence network based on title fields with five minimum numbers of occurrences of a term for published works between the year 2011 to 2020. Out of 1087 keywords, nine terms meet the required threshold. The map shows four significant clusters identified between the years 2011 to 2020 instead of only two from 1963 to 2010. The increase in the number of keywords indicates that the corpus title is accommodating for emerging trends. Interestingly, the social identity node has not made it into the captured nodes, although the identity node has been clustered together with gender and social media.



**Fig. 3 - Network visualisation map of author keywords within 2011-2020**

Comparing to clusters of author keywords for 1963- 2010 (Figure 2) with the clusters of author keywords for 2011-2020 (Figure 3), it is apparent that the 1963-2010 clusters have expanded, allowing the keywords that have previously situated under the clusters to be settled into new clusters in 2011-2020. The occurrence can be observed within the previous 1963-2010 nationalism-identity cluster (Figure 2), where both nodes expanded in 2011-2020 to form their respective cluster, with identity nodes clustered with gender and social media while nationalism is forming a new cluster on its own.

Based on Table 9, there is a distinctive pattern on how nodes are clustered together. The related terms are clustered together and indicated by the same colour, indicating that they are frequently co-occurred [15]. The terms are segregated into four clusters. The first cluster, presented in red, consists of keywords such as social movements, framing and collective identities, while the second cluster in green consists of terms such as identity, gender and social media. The third cluster, which is coloured in blue, includes the terms collective identity and collective action. The last cluster is the nationalism cluster which is given a yellow colour.

**Table 9 - Keyword Clusters 2011 to 2020**

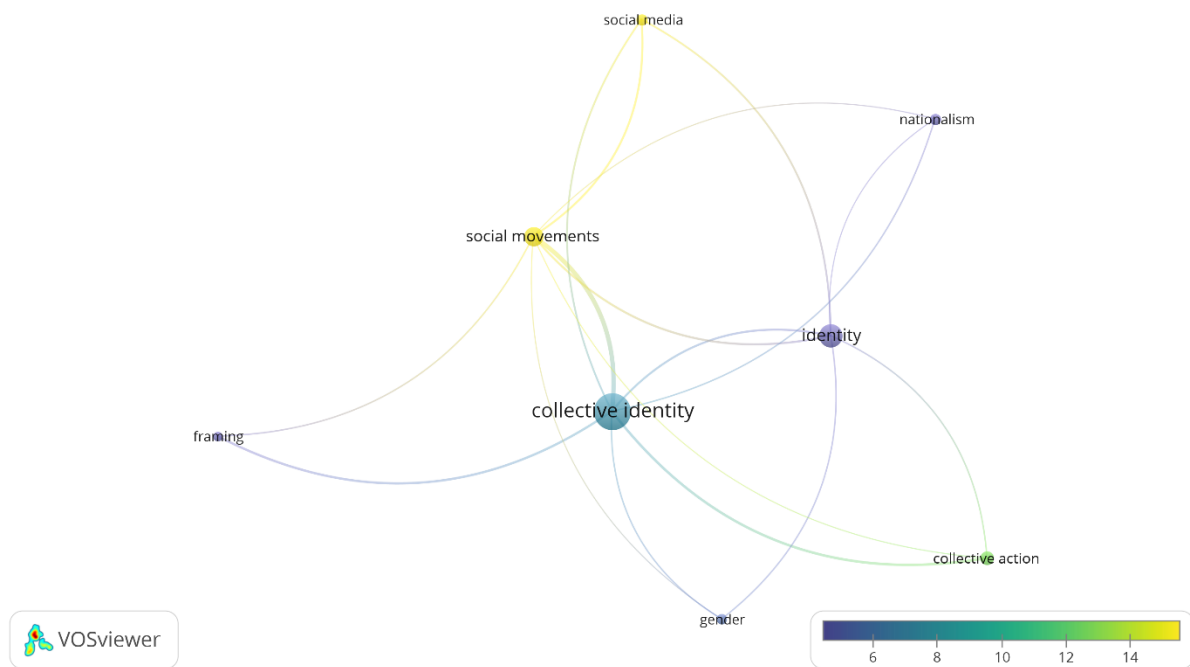
Cluster 1	Cluster 2	Cluster 3	Cluster 4
collective identity	identity	collective action	nationalism
social movements	gender		
framing	social media		

Within the timeframe of 2011 to 2020, identity and social movement domains are starting to catch on with the publication trend being that they have formed their specific clusters (refer to Table 9). It is observed that the most encountered author keywords under collective identity subject are identity (17.67%), social movement (11.64%), collective action (4.74%) (refer to Table 10), where all three keywords are prominent keywords within their respective cluster.

**Table 10 - Author Keywords 2011 to 2020**

Keyword	Occurrences	Percentage
collective identity	128	55.17%
identity	41	17.67%
social movements	27	11.64%
collective action	11	4.74%
nationalism	7	3.02%
social media	7	3.02%
gender	6	2.59%
framing	5	2.16%

As expected, a significant increment in average citation per year for the social movements domain is observed within the year 2011-2020 when an overlay visualisation map of average citation per year is presented (refer to Figure 4). The mapping shows that another exciting domain has caught on to the citation momentum: the social media domain, which goes hand in hand with the social movement domain. It shows evidence that there is currently a growing interest in discussing collective identity, social movement, and identity in association with social media as a medium among collective identity researchers.

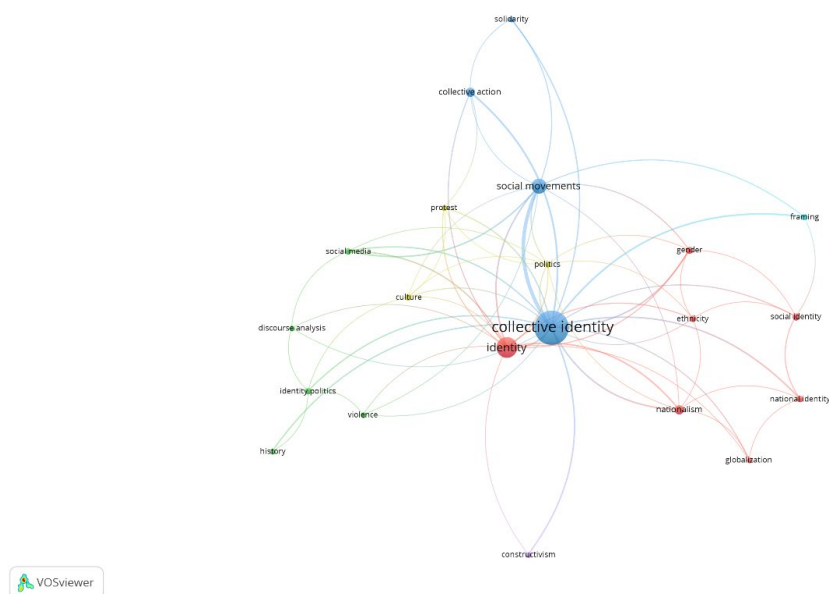


**Fig. 4 - Overlay visualisation map of average citation per year within the year 2011-2020**

Figure 5 shows the visualisation of a term co-occurrence network based on the title field with five minimum numbers of term occurrences from 1963 to 2020. Out of 1395 keywords generated, 21 keywords meet the required threshold. Based on the network visualisations map, the collective identity and identity nodes are situated near each other, indicating the relatedness between the two nodes [16]. Although the network visualisations map for 2011-2020 (refer to Figure 3) reveals that the distance between collective identity with social movements and identity nodes is relatively similar. As shown in Figure 5, the overall view indicates that studies

done on collective identity are tended to be discussed together with identity issues, showing a significant influence of the identity and identity theory discussion in shaping the corpus. The mapping also suggests publications relating to constructivism, globalisation, framing, national identity, social identity, and nationalism tend to revolve into a more specific discussion of the corpus than the rest of the clusters, showing a delineating specialisation in the corpus.

c) Author keywords 1963-2020: An overall view



**Fig. 5 - Network visualisation map of author keywords in 1963 to 2020**

Relating to the identity domain, identity researchers and theorists focus on personal and social identity, centred on the status categories and distinctiveness of the individual position within the social structure. Consequently, current research emphasises discussing common group identities such as gender, personal relationships, or social roles such as activism and social movement [17]. The same observation can be seen in this research where the connecting keywords such as ethnicity, gender, globalisation, identity, national identity, nationalism, and social identity are identified under a cluster (Table 12). It shows that the published works related to identity clusters argue that group and social bases of identity are indeed imbricated and correspond through person identity meanings. It means the group identity which constitutes membership in a community of specific others, e.g., family, school, political organisation, and social identity, which represent status categories that denote one’s position within the broader social structure, e.g., race, gender, age are imbricated and tend to be presented as a single unit [18].

**Table 12 - Keyword Clusters 1963 to 2020**

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
ethnicity	discourse analysis	collective action	culture	constructivism	framing
gender	history	collective identity	politics		
globalization	identity politics	social movements	protest		
identity national identity nationalism social identity	violence	solidarity			

Table 13 shows the overall author keywords occurrences trend suggests the five dominant main domains within the corpus. These domains are collective identity (44.18%), identity (15.87%), social movements (8.73%), nationalism (3.44%), and collective action (3.17%).

**Table 13 - Author keywords 1963 to 2020**

Keyword	Occurrences	Percentage
collective identity	167	44.18%
identity	60	15.87%
social movements	33	8.73%
nationalism	13	3.44%
collective action	12	3.17%
gender	8	2.12%
social identity	8	2.12%
social media	7	1.85%
culture	6	1.59%
ethnicity	6	1.59%
framing	6	1.59%
identity politics	6	1.59%
national identity	6	1.59%
constructivism	5	1.32%
discourse analysis	5	1.32%
globalization	5	1.32%
history	5	1.32%
politics	5	1.32%
protest	5	1.32%
solidarity	5	1.32%
violence	5	1.32%

#### 4.6 Authorship analysis

There is a total of 160 unique authors contributing to a total of 689 publications on collective identity. Table 14 listed the number of publications based on the number of authors

for each publication. Based on the table, most of the published works are single-authored (64.73%), while the rest are multi-authored. Collective identity authors tend to co-author around two to four authors per document. Of the overall 689 publications, 21.48% are co-authored by two authors, while three co-authoring and four co-authoring constitute a percentage of 8.13% and 3.34%, respectively.

**Table 14 - Number of Author(s) Per Document**

<b>Number of authors</b>	<b>Number of publications</b>	<b>Percentage</b>
0	3	0.44%
1	446	64.73%
2	148	21.48%
3	56	8.13%
4	23	3.34%
5	3	0.44%
6	3	0.44%
7	3	0.44%
8	2	0.29%
10	1	0.15%
19	1	0.15%

This study also presents the most active authors that published the documents on collective identity. Table 15 listed the most active authors with a minimum of five publications. Based on the table, Eisenstadt, S.N., Shriver, T.E., Gongaware, T.B., and Simon, B. are among the most active authors in this research field, publishing more than five publications on collective identity.

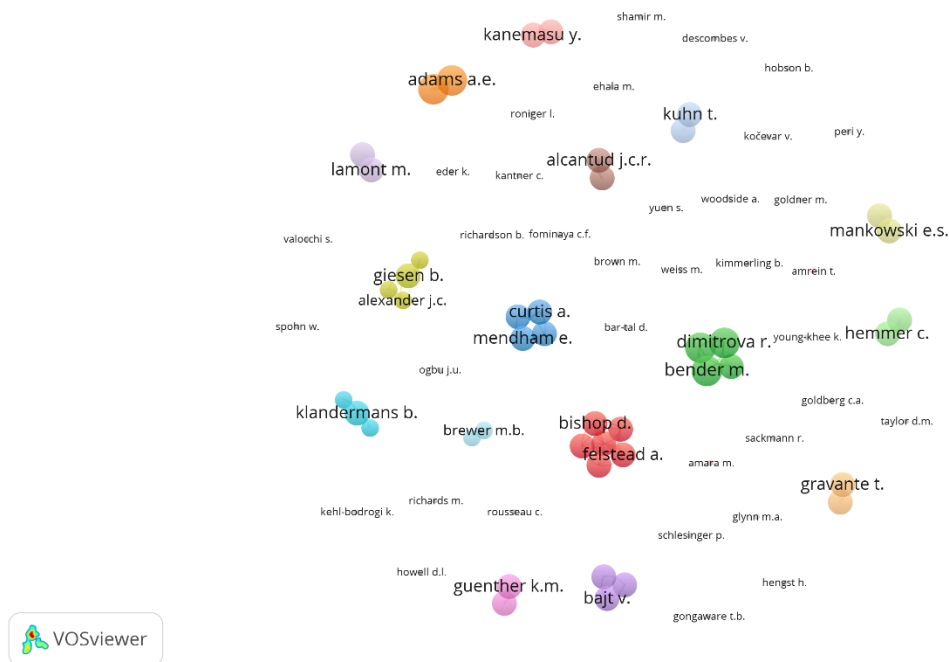
**Table 15 - Most Active Authors with a Minimum of Five Publications**

<b>Author name</b>	<b>Number of publications</b>	<b>Percentage</b>
Eisenstadt, S.N.	9	3.24%
Shriver, T.E.	7	2.52%
Gongaware, T.B.	5	1.80%
Simon, B.	5	1.80%
Descombes, V.	4	1.44%
Dimitrova, R.	4	1.44%
Fominaya, C.F.	4	1.44%
Ogbu, J.U.	4	1.44%
Roniger, L.	4	1.44%
Schlesinger, P.	4	1.44%
Adams, A.E.	3	3.24%
Alcantud, J.C.R.	3	2.52%
Amara, M.	3	1.80%
Bender, M.	3	1.80%
Chasiotis, A.	3	1.44%

Giesen, B.	3	1.44%
Kimmerling, B.	3	1.44%
Klandermans, B.	3	1.44%
Sackmann, R.	3	1.44%
Weiss, M.	3	1.44%

This paper further analyses the collaboration of the authors by conducting the co-authorship analysis using VOSviewer. The analysis is done based on the authors with more than two citations and calculated using the fractional counting method. The elements are shown in the mapping, such as colour indicating connected authors, while node size, font size and thickness of connecting lines, indicate the strength of the relationship amongst the authors [15].

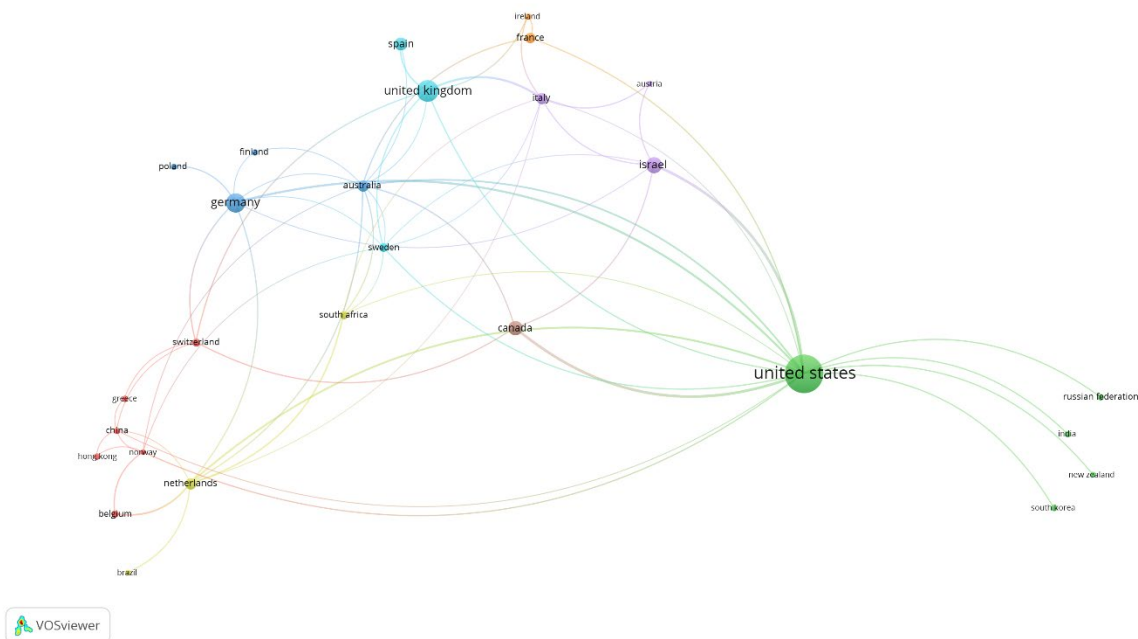
Figure 6 suggests that most authors who contribute to collective identity are not collaborating widely other than their existing collaboration cliques. There is no linkage connecting the individual cluster. For example, the mapping suggests that Bishop, D., Felstead, A., Fuller, A., Jewson, N., Kakavelakis, K. and Unwin, L., are a group of authors that collaborate closely together (clustered in red), but the cluster is not associated with the other clusters.



**Fig. 6 - Network visualisation map of the co-authorship based on authors that have a minimum of two citations (fractional counting)**

Despite the lack of collaboration across co-authorship clusters, collective identity research is highly affiliated across countries. Figure 7 shows the authors’ network visualisation map based on the countries with which they are affiliated. Only countries with more than five publications and more than five citations are considered for this analysis. Based on the fractional counting method, the United States authors have collaborated with 16 other states,

showing that the United States authors play a prominent role in collaborating with other countries.



**Figure 7 - Network visualisation map of the co-authorship based on countries that have a minimum of five documents and five citations (fractional counting)**

#### 4.7 Citation analysis

The RIS formatted file from the Scopus database data gathered from the Scopus database is imported into Harzing’s Publish or Perish software to generate the citation metrics. Table 16 shows the summaries of the citation metrics. The summary comprises information such as publication years, citations per year, citations per paper, and authors per publication. As indicated, 15,581 citations were reported from 1963 to 2020 for 689 published works on collective identity. The average number of citations per year is 268.64 citations, and the average number of citations per paper is 22.61 citations. From the 689 data retrieved, 107 publications (15.53%) receive more than the average 22 citations per paper.

<b>Metrics</b>	<b>Data</b>
Publication years	1963-2020
Citation years	58
Papers	689
Citations	15581
Citations/year	268.64
Citations/paper	22.61
Authors/paper	1.58
Hirsch h-index	52
PoP hI,norm	48
PoP hI,annual	0.83



Table 17 shows the top 10 most cited articles. The top-cited paper entitled “Who Is This “We”? Levels of Collective Identity and Self Representations” by M.B. Brewer, W. Gardner (1996) has received the highest number of citations (1,803 citations) with an average of 72.12 citations per year. Out of the top 10 cited papers, eight papers received more than 22 citations per year.

**Table 17 - Top 10 cited analysis**

<b>Authors</b>	<b>Title</b>	<b>Source</b>	<b>Cites/Paper</b>	<b>Cites/Year</b>
M.B. Brewer, W. Gardner	Who Is This “We”? Levels of Collective Identity and Self Representations	Journal of Personality and Social Psychology	1803	72.12
F. Polletta, J.M. Jasper	Collective identity and social movements	Annual Review of Sociology	1132	56.60
R.D. Ashmore, K. Deaux, T. McLaughlin-Volpe	An Organising Framework for Collective Identity: Articulation and Significance of Multidimensionality	Psychological Bulletin	927	54.53
A. Wendt	Collective identity formation and the international state	American Political Science Review	897	33.22
B. Simon, B. Klandermans	Politicised collective identity: A social psychological analysis	American Psychologist	631	31.55
J.C. Alexander, R. Eyerman, B. Giesen, N.J. Smelser, P. Sztompka	Cultural trauma and collective identity	Cultural Trauma and Collective Identity	511	30.06
J. Eccles	Who am I and what am I going to do with my life? Personal and collective identities as motivators of action	Educational Psychologist	508	42.33
C. Hardy, T.B. Lawrence, D. Grant	Discourse and collaboration: The role of conversations and collective identity	Academy of Management Review	376	23.50
C. Hemmer, P.J. Katzenstein	Why is there no NATO in Asia? Collective identity,	International Organization	283	14.89

	regionalism, and the origins of multilateralism			
A.D. Brown	A narrative approach to collective identities	Journal of Management Studies	242	16.13

## 5. Discussion

The bibliometric analysis presents a more precise insight in explaining the trend of collective identity research. Publication through journals is the most preferred type of publication source, where about 69.67% of publications on collective identity are published in an article. Research on this topic has started as early as 1963 and steadily increasing year by year since. For the last ten years, the subject's publications have recorded an average of 40.2 publications per year. Despite its early publication in 1963, the interest in the subject has only drastically risen in 1997 after the publication of two influential papers by M.B. Brewer, W. Gardner in 1996 entitled "Who Is This "We"? Levels of Collective Identity and Self Representations" and S.N. Eisenstadt, B. Giesen in 1995 entitled "The Construction of Collective Identity." Both papers continue to become prominent papers cited until these days.

During the early research period from 1963 to 2010, the baseline keywords for the collective identity corpus are collective identity, identity, nationalism, social movements, and social identity. The corpus mainly dealt with two domains: collective identity and identity. The earlier literature focuses more on constructing understanding collective identity, where five top-cited publications from 1963 to 2010 show varying perspectives in both approach and emphasis in understanding collective identity and identity issues. Despite the differing views, they are limited only to the understanding of collective identity and identity issues.

The current collective identity trend challenges the identity theory model, often linking emotional attachments with collectivities and social movements. It often tackles group identity and activism issues, where the connection between collectivity and social movement is theorised by the identity theory model [18]. Consequently, the identity domain is recognised as one of the most widely studied subjects across social sciences, making it one "elusive and ubiquitous" subject [19], hence its high keyword occurrences percentage. In contrast, social movement garners interest among scholars due to the enlightening work done by Melucci in the 1980s, which introduces the model of collective identity [20]. Therefore, it is reasonable to expect the social movement domain will continue to garner its momentum within the corpus for years to come.

The corpus later expanded, accommodating emerging research trends. The expansion of ideas has taken place, allowing identity nodes to be clustered with gender and social media nodes, signalling new interest among researchers on the topics. The growing interest in these two subjects goes hand in hand with the current collective identity research trend, which tackles group identity and online activism issues [18]. With the increased attention given by researchers in linking social media with collective identity research, it is expected that the social movement domain will continue to be an important domain within the corpus. Furthermore, the network visualisations map for the year 2011-2020 reveals that the distance between collective identity with social movements and identity nodes is relatively similar, showing the significant influence of social movement and identity domains towards the collective identity corpus.

Nevertheless, relating to the identity domain, identity researchers focus on personal and social identity, particularly examining the status categories and distinctiveness of the individual position within the social structure. This study reveals that connecting keywords such as ethnicity, gender, globalisation, identity, national identity, nationalism, and social identity are identified under a cluster, showing that the published works under the identity cluster tend to support the argument that group and social bases of the group identity are indeed imbricated. Consequently, current research related to identity clusters emphasises discussions related to common group identities or social roles.

It is important to note that this study finds some areas of the corpus to be shrinking, suggesting delineating specialisation within the corpus where publications relating to constructivism, globalisation, framing, national identity, social identity, and nationalism tend to revolve into more specific discussions than the rest of the clusters within collective identity research.

This study reveals research on collective identity subject are predominantly related to constructive analysis, identity analysis, discourse analysis, and collective action analysis. Regardless of the various discussion on the subject, most collective identity authors are confined within a small collaboration grouping. About 64.73% of authors published their work as a single author. However, despite having a narrow co-authorship landscape, collective identity researchers are highly affiliated across countries. The study shows that United States authors play a prominent role in collaborating with authors from other countries.

## 6. Conclusion

Collective identity has become one of the most prominent sub theories in social identity theory. Due to the increased interest among collective identity researchers to link emotional attachments with collectivities and social movements, it is expected that the identity domain will continue to be significant for the corpus. Bibliometric evidence shows the published works under the identity cluster support the argument that group and social bases of identity are imbricated. The collective identity corpus is experiencing an expansion as it has allowed the current identity domain to be diversified, revolving around the group and online activism issues, particularly touching on gender and social media issues. Nonetheless, this study also finds some corpus areas appear to be shrinking, suggesting delineating specialisation within the corpus. Most collective identity authors are more inclined to single author their work. However, regarding the collaboration aspect, collaborations are often done within small groupings. Despite having a narrow co-authorship landscape, collective identity researchers are highly affiliated across countries, with the United States authors play a prominent role in collaborating with authors from other countries.

There are a few limitations identified for this research. The first limitation of this study is mainly on the aspect of data coverage. Since this study primarily is data from the Scopus database, it is anticipated that the data used for this research do not cover all scientific papers on collective identity. Using complementary bibliometric sources such as Web of Science and Google Scholar will reduce omission in the analysis. Second, this study shows the bibliometric structure and thematic evolution for collective identity research. However, since research is a continuous process, the total number of publications and citations is only correct at the search time. We may expect the research's scientific landscape to change gradually. Therefore, it is suggested for future studies to track the changes that take place over time. Despite all these limitations, this study is among the first to analyse the detailed bibliometric indicators of the published literature.

### **Acknowledgement**

The author(s) wish to express sincere gratitude to the Malaysian Ministry of Higher Education for providing the Bumiputera Academic Training Scheme (SLAB) scholarship, Universiti Malaysia Sarawak (UNIMAS), and Universiti Utara Malaysia (UUM) for supporting the conduct of this study.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Fundamental Research Grant Scheme (FRGS) [SO Code: 13810].

### **Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

### **References**

- [1] K. Eder, "A theory of collective identity making sense of the debate on a 'European identity,'" *Eur. J. Soc. Theory*, vol. 12, no. 4, pp. 427–447, 2009, doi: 10.1177/1368431009345050.
- [2] D. A. Snow, "Collective Identity and Expressive Forms," in *International Encyclopedia of the Social & Behavioral Sciences*, 2001.
- [3] M. Castells, "Globalisation and identity. A comparative perspective," *Transfer*, 2006, doi: 10.1080/02533950008458681.
- [4] S. M. Buechler, "New social movement theories," *Sociol. Q.*, vol. 36, no. 3, pp. 441–464, 1995, doi: 10.1111/j.1533-8525.1995.tb00447.x.
- [5] A. Wendt, "Collective Identity Formation and the International State," *Am. Polit. Sci. Rev.*, 1994, doi: 10.2307/2944711.
- [6] S. N. Eisenstadt and B. Giesen, "The construction of collective identity," *Eur. J. Sociol.*, vol. 36, no. 1, pp. 72–102, 1995, doi: 10.1017/S0003975600007116.
- [7] A. Melucci, *Challenging codes: collective action in the information age*. Cambridge: Cambridge University Press, 1996.
- [8] R. Ranta and N. Nancheva, "Unsettled: Brexit and European Union nationals' sense of belonging," *Popul. Space Place*, vol. 25, no. 1, pp. 1–10, 2019, doi: 10.1002/psp.2199.
- [9] S. Asadi, F. A. Deligani, B. Rasuli, and M. S. Majd, "Bibliometrics in practice in developing nations: A study on the development of scientometrics and bibliometrics careers in Iran," *J. Inf. Sci. Theory Pract.*, vol. 6, no. 2, pp. 24–35, 2018, doi: 10.1633/JISTaP.2018.6.2.2.
- [10] Nurul Mardhiah Azura Md Nadzar, A. Bakri, and R. Ibrahim, "A bibliometric mapping of malaysian publication using co-word analysis," *Int. J. Adv. Soft Comput.*

*its Appl.*, 2017.

- [11] F. Muñoz-Leiva, M. I. Viedma-del-Jesús, J. Sánchez-Fernández, and A. G. López-Herrera, "An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective," *Qual. Quant.*, vol. 46, no. 4, pp. 1077–1095, 2012, doi: 10.1007/s11135-011-9565-3.
- [12] R. Butwell, "Individual and Collective Identity and Nation-Building," *World Polit.*, vol. 15, no. 3, pp. 488–494, 1963, doi: 10.2307/2009475.
- [13] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, 2010, doi: 10.1007/s11192-009-0146-3.
- [14] W. M. Sweileh, S. W. Al-Jabi, A. S. AbuTaha, S. H. Zyoud, F. M. A. Anayah, and A. F. Sawalha, "Bibliometric analysis of worldwide scientific literature in mobile - health: 2006-2016," *BMC Med. Inform. Decis. Mak.*, 2017, doi: 10.1186/s12911-017-0476-7.
- [15] A. Ahmi and M. H. Mohd Nasir, "Examining the trend of the research on extensible business reporting language (xbrl): A bibliometric review," *Int. J. Innov. Creat. Chang.*, 2019.
- [16] N. J. van Eck and L. Waltman, "Visualizing Bibliometric Networks," in *Measuring Scholarly Impact*, Y. Ding, R. Rousseau, and D. Wolfram, Eds. Springer, 2014, pp. 285–320.
- [17] J. Franco-Zamudio and H. Dorton, "Collective identity," in *Encyclopedia of Critical Psychology*, 2014.
- [18] J. L. Davis, T. P. Love, and P. Fares, "Collective Social Identity: Synthesizing Identity Theory and Social Identity Theory Using Digital Data," *Soc. Psychol. Q.*, vol. 82, no. 3, pp. 254–273, 2019, doi: 10.1177/0190272519851025.
- [19] P. Gleason, "Identifying identity: A semantic history," *J. Am. Hist.*, 1983, doi: 10.2307/1901196.
- [20] A. Bartholomew and M. Mayer, "Nomads of the Present: Melucci's Contribution to 'New Social Movement' Theory," *Theory, Cult. Soc.*, vol. 9, no. 4, pp. 141–159, 1992, doi: 10.1177/026327692009004007.