Investigation of Women's Health on Wikipedia—A Temporal Analysis of Women's Health Topic

Yanyan Wang [1,2,*] and **Jin Zhang** [3]

1   School of Information Resource Management, Renmin University of China, Beijing 100872, China
2   CIO Research Center, Renmin University of China, Beijing 100872, China
3   School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee WI 53211, USA;
    jzhang@uwm.edu
*   Correspondence: yanyanwangruc@ruc.edu.cn

check for
updates

Abstract: New health-related concepts, terms, and topics emerge, and the meanings of existing terms and topics keep changing. This study investigated and explored the evolutions of the women's health topic on Wikipedia. The creation time, page views data, page edits data, and text of historical versions of 207 women-health-related entries from 2010 to 2017 on Wikipedia were collected. Coding, subject analysis, descriptive and inferential statistical analysis, and Self-Organizing Map and n-gram approaches were employed to explore the characteristics and evolutions of the entries for the women's health topic. The results show that the number of the women-health-related entries kept increasing from 2010 to 2017, and nearly half of them were related to the supports and protection of women's health. The total number of page views of the investigated items increased from 2011 to 2013, but it decreased from 2013 to 2017, while the total number of page edits stayed stable from 2010 to 2017. Growing subjects were found during the investigated period, such as abuse and violence, and family planning and reproduction. However, the entries related to the economy and politics were diminishing. There was no association between the internal characteristic evolution and the external popularity evolution of the women's health topic.

Keywords: women's health; Wikipedia; health informatics; social media; temporal analysis; theme development; information science

1. Introduction

With the development of computer technology and Internet technology, the volume of health information keeps increasing on the Internet. According to Tu's report, the proportion of people among all the consumers who sought health information online increased from 15.9% to 32.6% from 2001 to 2010 [1]. A survey in 2013 reported that 87% of USA adults use the Internet and among them, 72% stated that they sought health information online during the past year [2].

The emergence of Web 2.0 advocated the creation of social media, which changed the method of communication between health organizations, health consumers, patients, and health professionals [3]. Dawson reported that 81% of European consumers and 63% of USA consumers trust the health information on social media applications [4]. Marar, Al-Madaney, and Almousawi found that 85% of patients and their companions sought health information via social media [5]. These statistics reveals that social media is recognized as an important channel for seeking health information in recent years.

The health-related information on social media covers a wide range of topics, including diseases and treatments, nutrition, health care and insurance, and healthy lifestyle [6]. The women's health topic is a main category among the various health-related topics on social media. Women usually face special health risks, such as pregnancy, cervical and breast cancer, and accompanying physical and

psychological issues [7–9]. Social media are currently utilized as important platforms by women to gain and share specific women-health-related information, to strengthen social supports by building social connections with others, and to self-manage health care [10–13].

As the volume of health information increases rapidly on social media, new terms, concepts, and topics emerge, which cause problems in health information seeking from both system and users' perspectives. Therefore, it is necessary to explore the temporal features of health-related concepts, terms, and topics on social media. Wikipedia is one of the largest social media platforms consisting of user-generated articles and the semantic relations between them [14,15]. Wikimedia Downloads stores all the historical versions and data of entries, including editors, viewers, content, interactions, and temporal data [16]. All of its data are open to the public. Hence, to investigate the temporal features of the women's health topic on social media, Wikipedia is a proper data source. However, few studies have focused on the women's health user-generated content on Wikipedia, and the temporal characteristics of the women's health topic on social media have not been adequately investigated, either. This study aims at discovering the themes, subjects, and entries related to women's health on Wikipedia and the relations among them and exploring how the women's health topic evolved from 2010 to 2017 on Wikipedia on both internal and external aspects.

2. Research Problem and Questions

The research problem of this study is to investigate and discover the evolution of the women's health topic derived from the social media website Wikipedia. The research questions are as follows:

- *RQ1:* What are the emergent themes and subjects of the women's health topic on Wikipedia and the relations among them?
- *RQ2:* How does the women's health topic evolve on Wikipedia in terms of its internal characteristics and external popularity during the investigated time periods?

In this study, a theme of the topic means a specific and distinctive concern of a group of Wikipedia entries. The entries collected can be assigned to several categories, and every category has its own theme. A subject of the topic means the focus of an entry. Every Wikipedia entry can have one or more than one subjects. The structure of entries, themes, and subjects for the Women's Health topic in a specific time period is shown in Figure 1.
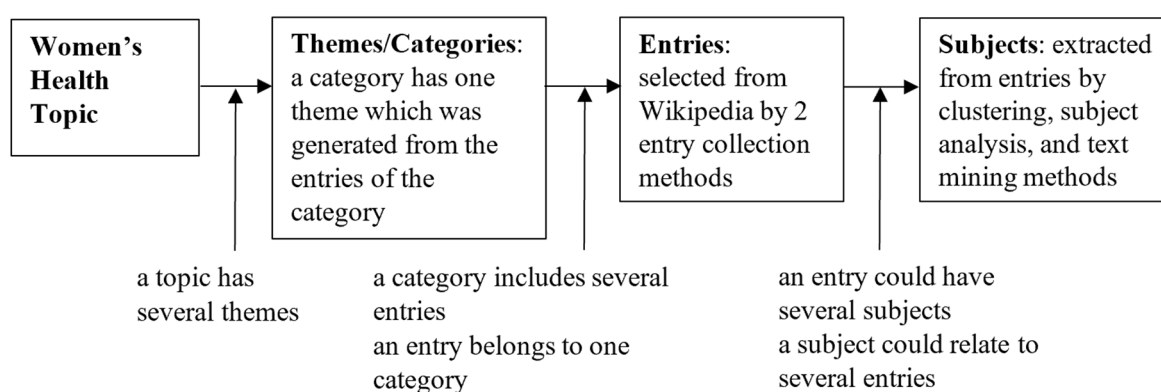

Figure 1. Concept Map.

The internal characteristics of a specific topic in different time periods show the emergences, growths, and disappearances of entries, subjects, and terms in each theme. The external popularity of an entry was measured by its number of page edits and number of page views. The number of page edits reflects the popularity of an entry among the Wikipedia editors, and the number of page views reflects the popularity of an entry among the Wikipedia viewers. This study explored the internal characteristics and external popularities of the women's health topic from 2010 to 2017. Four periods

were defined: 2010 to 2011 (Period 1), 2012 to 2013 (Period 2), 2014 to 2015 (Period 3), and 2016 to 2017 (Period 4).

3. Materials and Methods

3.1. Data Collection

The data collection processes of this study contain entries collection, text collection, and page views and edits data collection from Wikipedia. Figure 2 illustrates the entire data collection processes.
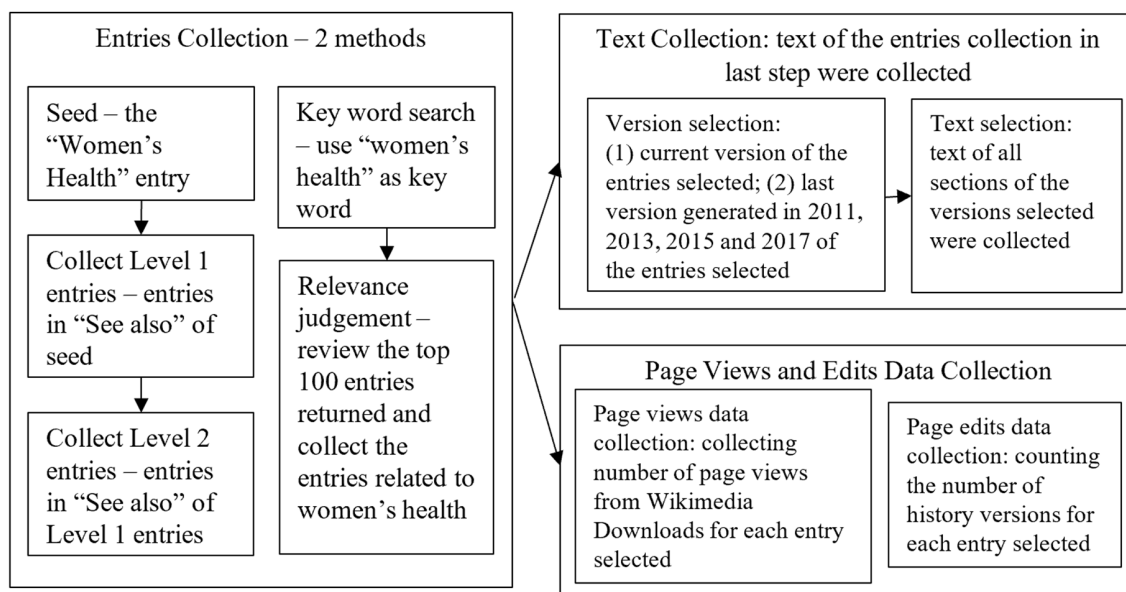
Figure 2. Data Collection Processes.

Wikipedia is regarded as a social media platform and its content is open to the public. It does not include any private data of editors and users. The investigated study does not involve any human subjects and the topic is not sensitive. Therefore, it was exempt from ethics approval.

3.1.1. Entries Collection

To explore the evolution of the women's health topic on social media, Wikipedia was selected as the data source in this study. Milne and Witten argued that Wikipedia is a rapidly growing platform containing vast interlinked information [15]. The richness of its content makes it an important resource for knowledge sharing and citation, and even for research. The history of each entry on Wikipedia is accessible to users, which means that all the historical versions of the entry are recorded by Wikipedia and could be viewed and collected by researchers.

Two methods were applied to retrieve entries related to women's health on Wikipedia. For the first method, the entries in the "See also" section of the women's health entry and the entries in the "See also" sections of these related entries were regarded as the associated entries, since the "See also" section of an entry contains the relevant entries selected by editors. For the second method, the term "women's health" was used as the search term to retrieve related entries on Wikipedia. The search results returned were ranked by relevance, and the top 100 search results returned were examined by the researcher. The associated entries, which were not the same as the entries obtained by the first method, were collected.

3.1.2. Text Collection

The content of every history version of an entry on Wikipedia contains several sections, such as content, main text, and reference. Although not all entries consist of the same sections, certain sections are included in almost all entries. They are the title, other entries associated to this entry, a short description of the entry, content, main body, "See also", and reference. The content section includes the content table of an entry, the main text or main body of an entry, and the reference section, which consists of references and URLs of references of an entry.

For each of the associated entries, the text data of the current version and the last version generated in 2011, 2013, 2015, and 2017 were collected based on the time periods determined. For each history version, the text of all the sections of each entry was collected. The WikipediR package developed by Oliver Keyes run on R was adopted for text data collection [17]. The software R is developed by The R Foundation for Statistical Computing who is seated in Vienna and Austria [18]. It enables the researcher to retrieve and gather the text content of an entry's current and historical versions.

3.1.3. Page Views and Edits Data Collection

The page views data during 2010 to 2017 were collected from Wikimedia Downloads. This website provides the Wikipedia data dumps that store all the historical page views data of all the Wikipedia entries since January 2010. The page edits data were collected from the view history page of the associated entries by R and RStudio. The software RStudio is developed by the RStudio Team in Boston, MA, USA [19].

3.2. Data Analysis

3.2.1. Categories and Themes

The entries obtained related to the topic on different aspects. In order to explore the relations among the entries, they were grouped into several categories in terms of their content. Since there are no existing categories of these entries, the open coding method was employed to analyze the associated entries and group them into several categories. In this study, every category had only one theme.

3.2.2. Text Data Processing

The subjects of every theme were extracted from the entries belonging to it by clustering and text mining approaches. To apply these approaches, the text data obtained for the themes were cleansed and transformed first.

The open source software R and RStudio and the tm package were adopted for text data cleansing, transformation, and processing. The punctuations, stop words, meaningless words (e.g., numbers, dates, equations, and so on), and words whose frequencies were less than 4 were removed. For each theme, a document–term matrix (Equation (1)) of the vector space model was presented. The matrix has m rows and n columns. The value of the cell (a_{ij}) in the matrix represents the frequency of the term j in the entry *i*.

$$M = \begin{pmatrix} a_{11} & a_{12} & .. & .. & a_{1n} \\ a_{21} & a_{22} & .. & .. & a_{2n} \\ .. & .. & a_{ij} & .. & .. \\ & & \cdots & & \\ a_{m1} & a_{m2} & .. & .. & a_{mn} \end{pmatrix} \tag{1}$$

Then, the document–term matrices were transformed to Term Frequency × Inverse Document Frequency (TF-IDF) matrices. Each value of the TF-IDF matrices (w_{ij}) was calculated based on Equation (2). In this equation, m is the number of the entries of a matrix, and e_j represents the number of the entries containing the term *j* in the matrix. The TF-IDF matrices obtained were the input matrices for the following clustering process.

$$w_{ij} = a_{ij} \times Log\left(m/e_j\right) \tag{2}$$

3.2.3. SOM Approach

To cluster the entries of the categories, the Self-Organizing Map (SOM) approach was employed in this study. It is a widely used neural network method that measures similarities among items of input data so as to form similarity graphs. The whole procedure of this approach is a recursive regression process [20]. The input matrices are the TF-IDF matrices created for the categories.

The output of the SOM approach is an output display map. Similar entries were assigned to the same node on the output display map. A U-matrix was used for projecting the clustering results to SOM displays. Every entry was projected to the SOM display as a number. Numbers with shorter distances among them were more similar than those with longer distances. Moreover, the similarity among entries was indicated by the color of an SOM output. The color projected to the SOM display background was determined by a U-matrix [21]. Higher values of the U-matrix stood for cluster borders, while lower values represented clusters.

According to the distances between numbers and the background colors, the entries of each matrix were clustered. The criteria for clustering the numbers are as follows: (1) the numbers located in the same SOM node were grouped into one cluster; and (2) if the numbers are located in two or more nodes, and the nodes are adjacent, or separated by only one empty node, and at the same time, the numbers are located in the same area where the U-matrix values are lower than half of the highest U-matrix value of the matrix, then these numbers were grouped into one cluster. In this way, the entries of each category were assigned into several clusters.

3.2.4. Subject Analysis

To identify the subjects of the clusters and categories, the n-gram approach was employed. The n-gram package offered by R extracts the n-word phrases in unstructured text files.

The historical revisions of the entries in one cluster were merged into one document. For each category, the historical revisions in a specific period of its entries were also merged into one document. The most used 2-word, 3-word, and 4-word phrases in each document were extracted by the n-gram package. The set phrases and meaningless phrases (e.g., "of the" and "the study is a") were removed from the dataset. If a phrase was a part of another one and the two phrases had the same meaning, then they were regarded as one phrase, and their frequencies were added together (e.g., "child development index" and "the child development index"). After data processing, a list containing phrases and frequencies was obtained for each document.

The researcher manually reviewed the lists to summarize the subjects of each document. One phrase could relate to more than one subject, and different phrases could relate to the same subjects. In this way, the subjects of each cluster were generated, and the subjects in each period of each theme were generated.

To find the increasing and decreasing phrases in each category, the differences of the frequencies obtained from the adjunct periods for a phrase was calculated. After all the frequency differences were obtained for each category, the researcher reviewed the most increasing and most decreasing phrases to generate the subjects.

3.2.5. Inferential Analysis

Inferential statistical tests allow the researcher to gain insights into the differences among the objects. In addition to descriptive statistical methods, inferential statistical analysis was applied to test the differences among the determined periods for the women's health topic. The hypotheses are:

- H01: There were no significant differences among the investigated time periods in terms of the number of views of the entries relevant to women's health.

- H02: There were no significant differences among the investigated time periods in terms of the number of edits of the entries relevant to women's health.

Since the independent variable (time period) was categorical, the dependent variables (number of views and number of edits) were continuous with repeated measures, and the distributions of the dependent variables did not follow the normal distribution; meanwhile, the Friedman's Test was applied to test the differences among the periods. To explore the difference among every two periods, a series of pairwise comparisons were conducted. Since the distribution of the differences among every two periods was not symmetrical, the Sign Test was used. The significant level of the inferential statistical tests was 0.05.

4. Results and Discussion

4.1. Descriptive Results

4.1.1. Entries and Themes

According to the data collection and analysis strategy, 207 associated entries were obtained, and four themes were generated from them. Table 1 lists the themes of the women's health topic, the number of entries related to each theme, and the description of each theme. The Support and protection (WH-SP) theme had the most relevant entries (99 entries) among the four themes, which reflects that the general public cared more about the protection of women's health than the other themes.

Table 1. Themes of Women's Health Topic.

Themes	No. of Entries	Descriptions
Discrimination, violence, harm, and subordination (DVHS)	37	Entries related to mental and physical violence and harm, discrimination, and subordination to women
Health problems and risks (WH-HPR)	25	Entries related to health problems and risks
Medical and interdisciplinary subjects (MIS)	46	Entries related to medical subjects and health-related interdisciplinary subjects
Support and protection (WH-SP)	99	Entries related to policies, laws, research studies, literary and artistic work, treatments, people, and organizations that support and protect women, and improve women's health

Table 2 presents the numbers of the entries created from 2010 to 2017 of each theme. The number of the entries had a steady rise. The WH-SP theme contributed to the entry increase the most among the four themes every year from 2010 to 2017. In 2010 and 2015, more new entries were created for this theme compared with the other years. Another special case is that for the MIS theme, 4 new entries were generated in 2013, which was larger than the other years.

Table 2. Number of Entries Created during the Investigated Time Periods.

| Themes	2010	2011	2012	2013	2014	2015	2016	2017
Discrimination, violence, harm, and subordination	1	1	1	1	2	1	0	1
Health problems and risks	0	1	0	2	0	3	0	1
Medical and interdisciplinary subjects	1	0	1	4	2	1	0	1
Support and protection	7	4	5	4	3	8	3	3
Total	9	6	7	11	7	13	3	6

4.1.2. Page Views and Edits

Figure 3 illustrates the Numbers of Yearly Page Edits (NYPEs) of the four themes of the *women's health* topic, and the NYPE of the topic as well. This figure reveals that the general trend of the NYPE of the *women's health* topic decreased from 2010 to 2017. The *Support and protection* theme received the largest NYPEs for six years (2010 to 2012 and 2015 to 2017) among the investigated eight years. In 2013 and 2014, the *Discrimination, violence, harm, and subordination* theme surpassed *Support and protection* and occupied the first position. The NYPEs of these two themes and *Medical and interdisciplinary subjects* fluctuated from 2010 to 2017, and no obvious ascending or descending trend was found for them. The NYPEs of the *Health problems and risks* theme rose from 2010 and reached its peak in 2012; then, it began to drop and reached its trough in 2017.
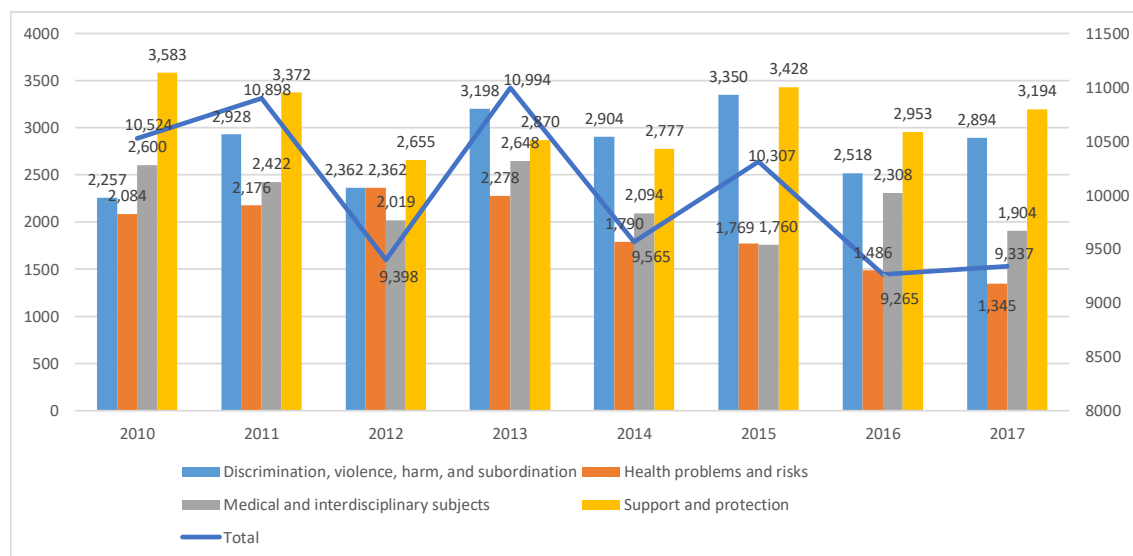
Figure 3. Numbers of Yearly Page Edits for Each Theme of Women's Health.

Figure 4 displays the Numbers of Yearly Page Views (NYPVs) of the four themes of the *women's health* topic, and the NYPV of the topic as well. The trend of the total page views decreased from 2010 to 2011, increased from 2011 to 2013, but then decreased again after that. The trends of the four themes were similar to the trend of the entire *women's health* topic. The *Support and protection* theme and the *Medical and interdisciplinary subjects* theme ranked in the top two places among the four themes from 2010 to 2017. The *Health problems and risks* theme occupied the third place from 2010 to 2012 but fell to the last place from 2013. The trend of the *Discrimination, violence, harm, and subordination* theme was slightly different from the other three themes, because its NYPV increased rapidly from 2010 to 2013. However, the decreasing of its NYPV from 2013 to 2017 was similar to the other themes and the *Women's Health* topic.

For each theme and the entire topic, its NYPE trend differed a lot from its NYPV trend from 2010 to 2017. No association was found between the NYPEs and the NYPVs. It indicates that the user groups who created the page edits and the page views had different interests in the investigated time periods. The Wikipedia editors were more interested in *Support and protection*, and *Discrimination, violence, harm, and subordination,* while the Wikipedia viewers were more interested in the *Support and protection* and *Medical and interdisciplinary subjects.*
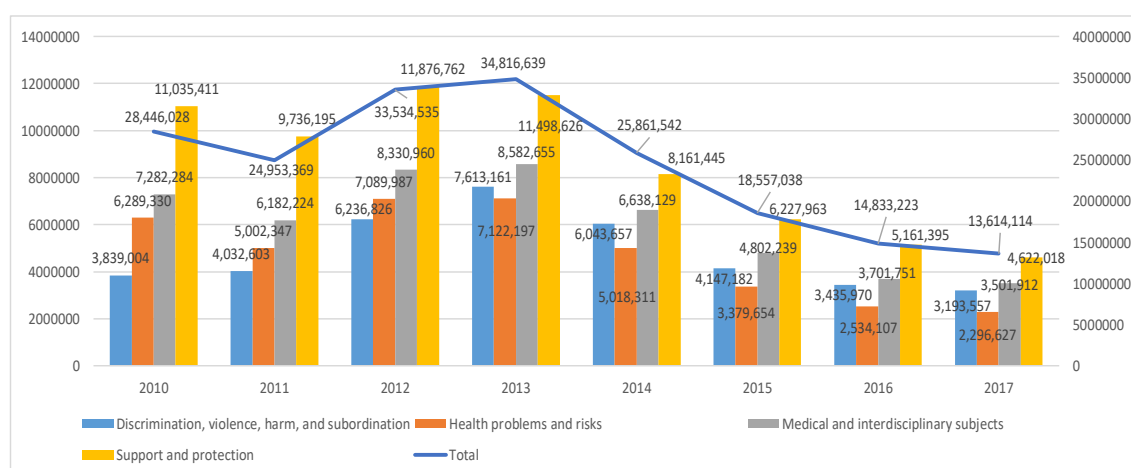
Figure 4. Numbers of Yearly Page Views for Each Theme of Women's Health.

4.2. Subject Analysis Results and Discussion

Figures 5–8 are the SOM displays for the four identified themes. The color bars on the right side of the figures represent different values of the U-matrix. A lower value means higher similarity. In the displays, every number stands for an entry, and the corresponding entry of each number is presented in Appendix A. Every rectangle or polygon represents a cluster and the numbers in the same rectangle/polygon represent the entries belonging to the same cluster. The numbers not included in any rectangle/polygon stand for the isolated entries, which were not grouped to any clusters. The clusters with more than three entries were recognized as large clusters, while those with three or less entries were the small clusters. The large clusters were represented by purple rectangles or polygons. The small clusters were represented by red rectangles.

Tables 3–6 display the high-frequency terms and phrases, as well as the subjects discovered within each large cluster. The high-frequency terms/phrases were extracted from the entries by the n-gram approach. The high-frequency terms and phrases are displayed in the second column of each table, and the frequency of each term/phrase is included in the brackets following the term/phrase. The researcher proposed the subjects of each large cluster by examining the high-frequency terms and phrases of it. The small clusters and isolated entries were not included in these tables.

4.2.1. The Discrimination, Violence, Harm, and Subordination (DVHS) Theme

Figure 5 presents four large clusters and two small clusters discovered for the DVHS theme. Clusters C3 and C4 were all located in the same blue area. Therefore, the entries in these two clusters were relevant to one another. Table 3 lists the clusters and their high-frequency terms, phrases, and subjects.

The *minority group* subject occurred in all the four clusters of this theme, so it was the most dominant subject of the DVHS theme. The minority groups mentioned in the entries associated to this subject contained LGBT people, women, and African Americans (black).

The *inequality and discrimination* subject and the *abuse and violence* subject appeared in three clusters. The former subject had three lower-level subjects, which were *health care inequality and discrimination*, *inequality and discrimination in work*, and *inequality and discrimination in research*. *Health care inequality and discrimination* was reflected by the "missing women" phenomenon. As it was demonstrated in the "Missing women" entry, this phenomenon indicated that the number of women in a region was smaller than the expected number of women, which was caused by sex-selective abortion, female infanticide, and inadequate health care and nutrition for female children.
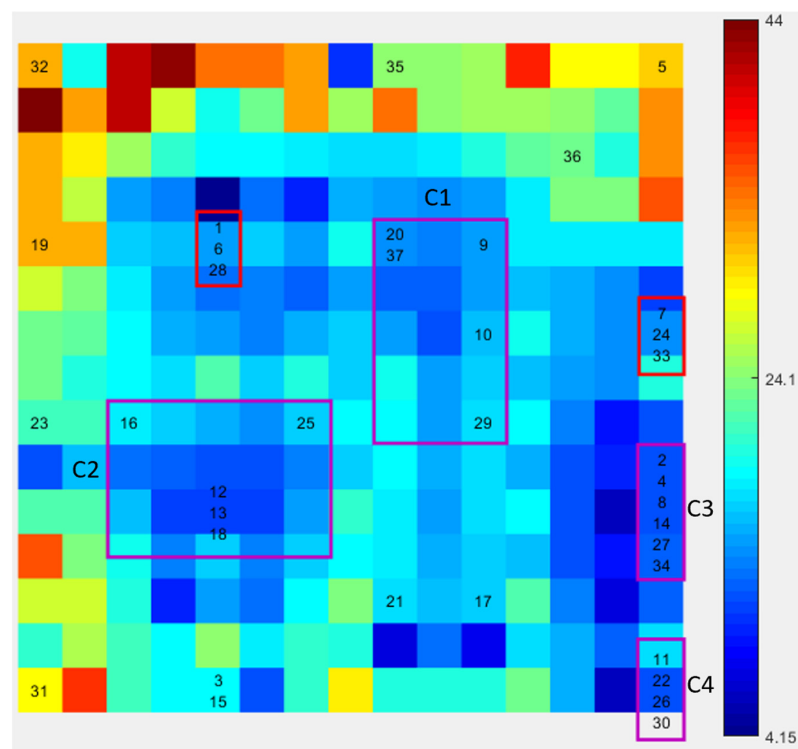
Figure 5. Self-Organizing Map (SOM) Display of DVHS.

Table 3. Subject Analysis of DVHS.

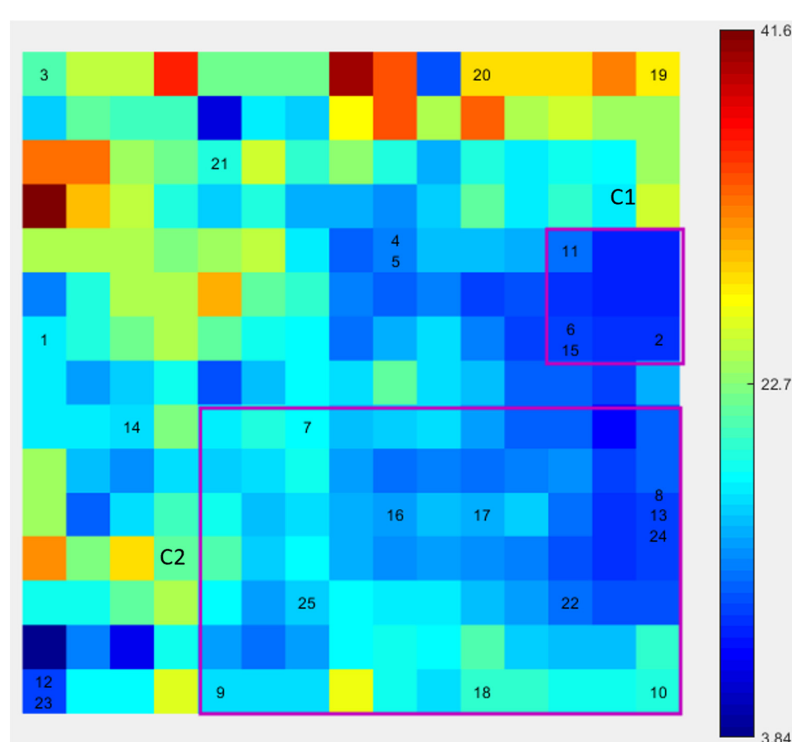| Clusters	High-Frequency Terms and Phrases	Subjects
C1	sex ratio (81), missing women (65), gender inequality (47), gender gap (45), transgender people (43), men and women (40), United States (39), Gender Gap Report (25), sexual harassment (21), labor force (16), gender identity (16), female children (15), wage gap (15), World Economic Forum (15), birth sex (14)	Inequality and discrimination (health care, work), minority group (LGBT, woman), abuse and violence (sexual violence)
C2	rape culture (73), hegemonic masculinity (65), gay men (34), horror films (23), sexual violence (22), sexual assault (20), rape victims (17), sexual orientation (15), rape myths (15), South Africa (15), violence against women (11), slasher films (10), United States (9), victim blaming (9), mass media (9), LGBT people (9)	Minority group (LGBT, woman), abuse and violence (sexual violence, heterosexist violence)
C3	gender gap (22), law enforcement (20), female child (20), police officers (14), Wikimedia foundation (13), New York Times (13), Wikipedia editors (12), Silicon Valley (11), gender bias (9), British Airways (8), New Zealand (8), black women (8), technology industry (7), female officers (6), sexual harassment (6)	Inequality and discrimination (work, research), minority group (woman), abuse and violence (sexual violence)
C4	glass cliff (28), triple oppression (20), black women (15), Communist Party (9), leadership positions (7), women executives (6), Michelle K (5), United States (5), black feminist (5), gender roles (4), occupational sexism (4), Claudia Jones (4), Socialist Party (4), reverse sexism (4), glass ceiling (4), Supreme Court (4), men and women (4)	Inequality and discrimination (work), minority group (black, woman), woman protection (organization, research)

The *abuse and violence* subject had two lower-level subjects: sexual violence and heterosexist violence. The findings imply that these two types of violence were the most discussed violence-related

subjects in the *women's health* topic. The associated entries of the *abuse and violence* subject mentioned not only different types of violence, but also the causes of the violence. For instance, rape culture was one of the main causes of high rape rates in certain countries, such as India.

Women protection was another subject of the DVHS theme. Its associated entries were about the *organizations* (e.g., the Supreme Court of the United States) and *research* (e.g., the triple oppression theory) of women protection.

4.2.2. The Health Problems and Risks (WH-HPR) Theme

Figure 6 presents that two clusters were generated for the WH-HPR theme. These two clusters were close to each other and in the same blue area, which indicates that the entries in the two clusters shared some similarities. Table 4 presents the high-frequency terms/phrases and subjects of the two clusters.

Figure 6. SOM Display of WH-HPR.

Table 4. Subject Analysis of WH-HPR.

Clusters	High-Frequency Terms and Phrases	Subjects
C1	blood pressure (20), high blood pressure (11), bacterial vaginosis (7), passive partner (7), active partner (6), chronic hypertension (5), diseases of affluence (5), sexually transmitted infections (4)	Health issue (problem)
C2	mental health (70), health care (53), United States (42), medical anthropology (39), public health (30), African Americans (23), heart disease (22), World Health Organization (19), gender equality (17), gender polarization (17), mental illness (16), sub-Saharan Africa (16), social class (15), men and women (15), women's health (14)	Health issue (service, research, organization, problem), inequality and discrimination (research)

The WH-HPR theme only had two subjects: *health issues* and *inequality and discrimination*. Cluster C1 mainly concentrated on *health problems*, such as high blood pressure and sexually transmitted

infections. In this cluster, a frequently used synonym of high blood pressure was found, which was hypertension. The term "hypertension" was often used by health professionals and the term "high blood pressure" was usually used by lay people. Since Wikipedia is a user-generated platform, these two expressions were both utilized in the Wikipedia entries.

The *health issue* subject of Cluster C2 had four lower-level subjects, including *health service* (e.g., health care), *research* (e.g., medical anthropology), *organization* (e.g., the World Health Organization), and *problem* (e.g., heart disease). Another subject of this cluster was *inequality and discrimination,* and this subject had a lower-level subject, *research*. For example, an entry of this subject was about the "Gender polarization" concept proposed by American psychologist Sandra Bem [22].

4.2.3. The Medical and Interdisciplinary Subjects (MIS) Theme

Figure 7 presents that four large clusters and two small clusters emerged from all the entries of the MIS theme. Two of the four clusters were either not close to each other or had green areas between them. These results show that the entries of the four clusters were not quite relevant. Table 5 lists the clusters, high-frequency terms, phrases, and subjects.
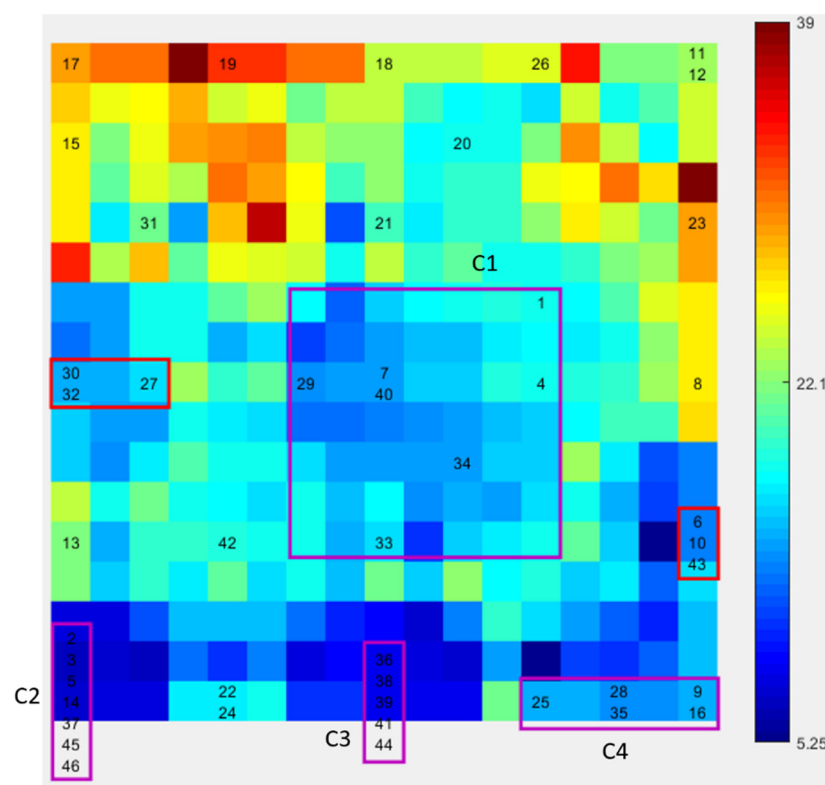

Figure 7. SOM Display of MIS.

Table 5. Subject Analysis of MIS.

Clusters	High-Frequency Terms and Phrases	Subjects
C1	health care (68), family planning (68), United States (37), public health (32), health disparities (28), rural areas (24), health outcomes (24), structural violence (22), social determinants of health (21), World Health Organization (19), health equity (18), living conditions (16), spirit level (15), life expectancy (15), socioeconomic status (14), medical care (14), birth control (14)	Health issue (service, problem, research, organization), family planning and reproduction, abuse and violence (structural violence)
C2	United States (22), sexually transmitted disease (14), reproductive health (13), facial prominence (9), social epidemiology (8), intrauterine contraception (8), birth control (8), women's health (13), health and human (8), family medicine patients (8), hormonal contraceptives (7), CDC exploratory research (7), family planning (6), human services (5), reproductive age (5), bisexual women (5), sexual health (5), health and human services (5)	Health issue (problem, research, service), family planning and reproduction (method), minority group (LGBT)
C3	Whitehall Study (25), pelvic floor (23), reproductive health (21), Whitehall II (16), health care (16), reproductive rights (13), heart disease (13), Russian women (13), coronary heart disease (10), reproductive law and policy (8), Center for Reproductive Law (8), women's health (7), civil servants (6), social class (6), mortality rate (6), live births (6), risk factors (6), social determinants (6), blood pressure (6), pelvic floor muscles (6)	Health issue (research, problem, service), family planning and reproduction (law, organization), social factor
C4	social determinants of health (44), health care (43), population health (35), reproductive health (33), maternal mortality (30), oral health (29), World Health Organization (23), family planning (23), maternal health (20), public health (16), maternal deaths (14), health services (14), prenatal care (13), developing countries (12), United Nations (12), United States (12)	Health issue (problem, service, organization), population issue (problem), family planning and reproduction

The *health issue* subject and the *family planning and reproduction* subject appeared in all the four clusters, but each cluster of the MIS theme had their own unique subject: C1 had the *abuse and violence* subject, C2 had the *minority group* subject, C3 had the *social factor* subject, and C4 had the *population issue* subject. For the *abuse and violence* subject, a new lower-level subject emerged from this theme, which was the *structural violence* subject. Different from the previous types of violence, structural violence was caused by social structure or social institution.

4.2.4. The Support and Protection (WH-SP) Theme

Figure 8 presents that eight large clusters and five small clusters were discovered for the WH-SP theme. Clusters C1 to C7 were all located in the same blue area, which means that their entries had similarities to some extent. Cluster C8 stayed in another blue area, and the yellow and green areas separated it from the other clusters, which means that its entries had no strong connections with the entries of the other clusters. The eight large clusters and their high-frequency terms/phrases and subjects are displayed in Table 6.

Table 6. Subject Analysis of WH-SP.

| Clusters	High-Frequency Terms and Phrases	Subjects
C1	healthy people (17), Department of Health (12), black women (9), health and human services (9), women's health (8), disease prevention (6), health promotion (5), Human Services Office (5), Office of Disease Prevention (5)	Health issue (organization, problem, service), minority group (black)
C2	women's health (42), Center for Women's Health (14), health sciences (13), Health Centre (10), women's hospital (10), AnMed Health (7), health services (7), OHSU Center (6), women and newborns (6), Christie Street (5), Cancer Centre (4), World War (4), health care (4), St. Michael's Hospital (4), Honest Body (4), maternity hospital (4), The Huffington Post (4), Sunnybrook Health Sciences Centre (4), obstetric and neonatal nurses (4)	Health issue (organization, service)
C3	women's health (38), health care (11), women's studies (10), Journal Citation Reports (8), Care for Women (6), health care journal (5), impact factor (5), women's studies journals (4)	Health issue (service, research)
C4	health education (138), health literacy (105), health disparities (83), public health (69), health system (44), reproductive health (43), University of Michigan (33), health care (32), Medical School (31), health promotion (28), United States (25), New York (24), Performance Indicators for Grades (24), school health (23), health supplies (23)	Health issue (education, problem, organization, service)
C5	United Nations (60), women's health (59), public health (41), health care (29), Sutter Health (22), UN Foundation (18), New York (15), health research (14), health network (12), University of Pittsburgh (11), San Francisco (10), Women's Health Research (10), Society for Women's Health (10), Medical Center (9), United States (9), Pittsburgh Graduate School (9), School of Public Health (9)	Health issue (service, organization, research, education), population issue (organization)
C6	women's health (98), United States (57), public health (40), health care (39), health education (35), Women's Health Center (34), Health and Human Services (26), social security (22), College Hospital (20), feminist health centers (20), United Nations (17), Department of Health Education (16), Federal Security (13), Our Bodies Ourselves (13), Health Education and Welfare (11), Boston Women's Health Book (11)	Health issue (service, education, organization, law)
C7	women's health (24), Planned Parenthood (21), Heart Truth (20), Red Dress Collection (17), Conference on Planned Parenthood (10), American Medical Women's Association (10), United Nations (7), United States (7), New York (5), National Organization (5), First Ladies (5), sexual health (5), health policy (5), heart disease (5), National Organization for Men (5), Institute of Women's Health (5), Women's Health and Development (5)	Health issue (organization, law, problem, research), family planning and reproduction (organization)
C8	equity feminism (9), gender feminism (7), tobacco companies (5), United States (4), female physicians (4), Hoff Sommers (4), medical school (4)	Woman protection (activity, research)
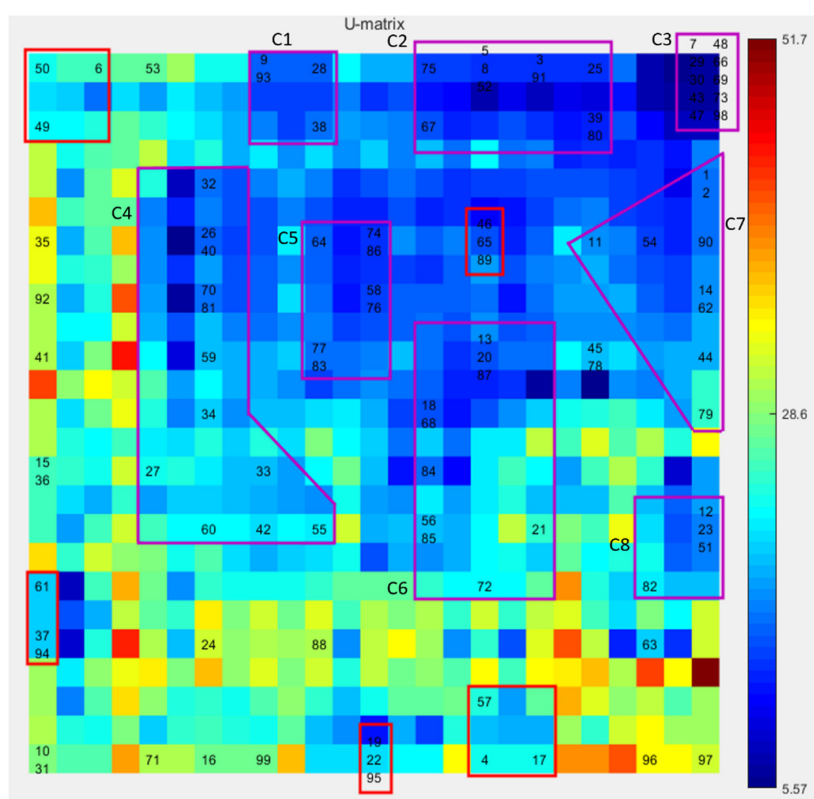
Figure 8. SOM Display of WH-SP.

Table 6 shows that the *health issue* subject appeared in the first seven clusters (Clusters C1 to C7) of the WH-SP theme, more than any other subjects of this theme. Therefore, the *health issue* subject was the salient subject of the WH-SP theme. This subject had several lower-level subjects, including *health organizations* (e.g., OHSU Center for Women's Health), *services* (e.g., obstetric and neonatal nurses), *research* (e.g., women's studies journals), *problems* (e.g., heart disease), *education* (e.g., Performance Indicators), and *laws* (e.g., Social Security Act). The *health education* subject of this theme covered the content about the performance indicators that were used for student assessment. The *health law* subject, which only occurred in this theme, referred to the entries of health-related laws and policies, such as the Social Security Act and the policies developed by the European Institute of Women's Health.

The *minority group* subject, the *population issue* subject, the *family planning and reproduction* subject, and the *woman protection* subject each only appeared in one cluster, respectively. Different from the other three subjects, the *woman protection* subject did not occur together with the *health issue* subject, which indicates that there was no strong connection between C8 and the other seven clusters. The entries in C8 were relevant to *woman protection activities* and *research*. For example, many theorists proposed a series of feminism theories (e.g., liberal feminism and gender feminism) so as to fight against gender inequality. A certain instance was the history of women fighting for equal smoking rights.

Table 7 lists the subjects of the identified themes and shows the relations between the themes and subjects. It reveals that the DVHS theme, the MIS theme, and the WH-SP theme had more diverse subjects compared with the WH-HPR theme. In other words, the entries' subjects of the WH-HPR theme were more centralized than those of the other themes. Among the four themes, the MIS theme and the WH-SP theme had more common subjects. Meanwhile, every two of the four themes had one or more subjects in common with each other, which indicates that these themes were relevant to each other.

Table 7. Subjects of Women's Health. The check mark (√) shows that a certain subject appears in a category.

Themes	Subjects							
	Inequality and Discrimination	Minority Group	Abuse and Violence	Woman Protection	Health Issue	Family Planning and Reproduction	Social Factor	Population Issue
Discrimination, violence, harm, and subordination	√	√	√	√				
Health problems and risks	√				√			
Medical and interdisciplinary subjects		√	√		√	√	√	√
Support and protection		√		√	√	√		√

4.3. Evolution of the Women's Health Topic

4.3.1. Entry Growth

New entries created in a certain period reflect the Wikipedia editors' new interests and focuses during the period. According to Table 2, among the four themes of *women's health*, the WH-SP theme had much more new entries in the four investigated periods than the other three themes.

After reviewing all the new generated entries from 2010 to 2017, it reveals that the new entries in the DVHS theme were related to sexism, such as sexism in the workplace (e.g., Women in law enforcement) and sexism in specific regions (e.g., Discrimination against girls in India). The entries in the WH-HPR theme and the MIS theme focused on women's health issues, including the research of women's health issues (e.g., Women's health issues), the determinants of health issues (e.g., Social determinants of health in poverty), and women's health status in specific regions (e.g., Women's reproductive health in Russia).

The new entries in the WH-SP theme concentrated on the techniques and methods (e.g., Gynography), research (e.g., Black Women's Health Study), organizations (e.g., EuroHealthNet), training and education (e.g., Oregon Health and Science University Center for Women's Health), services (e.g., Midwife), and works of art (e.g., The Honest Body Project) that aimed to support and protect women and improve women's health.

4.3.2. Changes of Subjects

To explore the internal characteristic evolution of each selected topic, the changes of the subjects from one period to another were explored. For each theme of a selected topic, the frequency difference of each term/phrase from one period to the next period was calculated. The terms/phrases of each theme were ranked according to their frequency differences and the terms/phrases whose frequencies increased or decreased the most from one period to the next were extracted from the rankings. Tables 8–11 display the top 20 terms/phrases of the rankings, and only the terms/phrases whose frequencies increased or decreased by more than 4 are included in the 11 tables. The subjects relevant to the terms/phrases were also included in these tables. The numbers in each table show the frequency differences. If a term's frequency decreased from Periods 1 to 2, its frequency difference would be negative, and vice versa.

1. The Discrimination, violence, harm, and subordination (DVHS) theme

Table 8 presents the terms/phrases whose frequencies changed the most from one period to the next in the DVHS theme. During all the periods, the terms about *abuse and violence*, *inequality and discrimination*, and *minority group* kept growing. One lower-level subject of *abuse and violence*, *sexual violence*, increased in all the periods, and another lower-level subject, *domestic violence*, increased from Periods 1 to 2, and Periods 3 to 4. The increase of the terms relevant to female genital mutilation mainly caused the growth of *domestic violence*.

The content about *inequality and discrimination* focused on different aspects in different time periods. For instance, the interests about inequality in *society*, *economy*, and *work* increased from Periods 1 to 2, while from Periods 2 to 4, the interests about inequality in *health care* grew. For the *minority group* subject, the content about *LGBT* people increased in all the investigated periods. When examining the high-frequency terms/phrases about *LGBT*, it shows that "transgender people" occurred the most. In other words, from Periods 1 to 4, the Wikipedia editors paid increasing attention to the LGBT group, especially transgender people.

Table 8. Changes of Subjects in the Four Periods in the DVHS Theme.

Time Period	High-Frequency Terms and Phrases		Subjects
	Frequency Decreasing Terms	a history of women (−9), US Department (−5)	Inequality and discrimination (work), minority group (woman)
Period 1–Period 2	Frequency Increasing Terms	female genital (92), female genital mutilation (73), female circumcision (40), hegemonic masculinity (39), human rights (38), New York (34), United States (29), gender apartheid (29), United Nations (27), gender identity (23), men and women (21), transgender people (20), age discrimination (18), rape culture (18), Glick P (16), World Health Organization (16), genital cutting (14), Oxford University (14), South Africa (14), gender roles (13), sexual assault (13), Islamic law (13)	Abuse and violence (domestic violence, sexual violence), inequality and discrimination (society, economy, work), minority group (LGBT), woman protection (organization), health issue (organization)
	Frequency Decreasing Terms	genital mutilation (−32), female circumcision (−30), female genital (−29), Islamic law (−13), South Carolina (−9), gender identity (−8), Type III (−7), Hosken Report (−6), Glick P (−5), Oxford University (−5), World Health Organization (−5), medical journal (−5), Agrarian System (−5)	Abuse and violence (domestic violence), inequality and discrimination (research), health issue (organization, research)
Period 2–Period 3	Frequency Increasing Terms	New York (50), gender gap (39), sex ratio (31), first female (25), rape culture (23), gender inequality (22), age discrimination (19), missing women (19), New York Times (18), women and girls (18), United States (16), men and women (16), hegemonic masculinity (15), glass cliff (15), victim blaming (11), male privilege (11), sexual violence (10), gender equality (10), gender bias (10), transgender people (9), women and children (8)	Inequality and discrimination (health care, work), abuse and violence (sexual violence), minority group (LGBT, woman)
	Frequency Decreasing Terms	New Zealand (−5)	Inequality and discrimination (work)
Period 3–Period 4	Frequency Increasing Terms	men and women (42), missing women (37), United States (30), gender gap (23), sexual assault (20), genital mutilation (20), transgender people (17), triple oppression (16), sexual violence (15), South Africa (13), women's rights (11), New York (9), gender bias (9), rape victims (9), global gender gap (9), gender roles (9), violence against women (9), labor force (9), sex differences (9), pay gap (9), gender gap report (9)	Inequality and discrimination (health care, work), abuse and violence (sexual violence, domestic violence), minority group (LGBT, woman), woman protection

Table 9. Changes of Subjects in the Four Periods in the WH-HPR Theme.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1–Period 2	Frequency Decreasing Terms	Disease Control and Prevention (−43), medical dictionary (−5)	Health issue (organization)
	Frequency Increasing Terms	infant mortality (62), United States (40), mortality rate (38), health organization (36), health care (34), World Health Organization (29), public health (25), reproductive health (23), developing countries (18), infant mortality rate (17), Sub-Saharan Africa (14), family planning (12), urban development (10), men and women (10), ovarian cancer (9), New York (9), drinking water (9), determinants of health (9), heart disease (8), infant death (8), United Nations (8)	Health issue (problem, organization, service, cause), family planning and reproduction (organization)
Period 2–Period 3	Frequency Decreasing Terms	Today's Evidence (−15), Tomorrow's Agenda (−15)	Health issue (organization), woman protection (organization)
	Frequency Increasing Terms	ovarian cancer (102), epithelial ovarian cancer (18), infant mortality (16), United States (13), gender polarization (12), risk of ovarian cancer (12), birth control (10), women's health (10), mental health (7), public health (6), live birth (6), health issues (6), side effects (6), health care (5), mortality rate (5), New York (5), infant mortality rate (5), infant death (5), induced abortion (5), million people (5), substance abuse (5), preterm birth (5), lymph node (5)	Health issue (problem, service), family planning and reproduction (method), inequality and discrimination (society)
Period 3–Period 4	Frequency Decreasing Terms	embryo or fetus (−11), gynecologic oncology (−7), poor health (−6)	Health issue (research, problem), family planning and reproduction (method)
	Frequency Increasing Terms	mental health (67), health care (29), germ cell (26), blood pressure (21), United States (20), cell tumor (25), birth control (14), African Americans (12), socioeconomic status (12), infant mortality (11), mortality rate (8), health and human (8), health organization (7), burden of disease study (7), causes of death (7), ovarian cancer (11), infant mortality rate (6), systematic analysis (6), Stage I (6), health problems (5), health services (5), developed countries (5), medical abortion (5), systematic review (5), Cochrane Database (5)	Health issue (service, problem, organization, research, cause), family planning and reproduction (method)

Table 10. Changes of Subjects in the Four Periods in the MIS Theme.

Time Period		High-Frequency Terms and Phrases	Subjects
Period 1–Period 2	Frequency Decreasing Terms	population health (−12), Oxford University (−5), health status (−5), maternal deaths (−5), political economy (−5)	Population issue, health issue (problem), economy, politics
	Frequency Increasing Terms	public health (83), health care (72), sex segregation (70), world health (40), mental health (39), health organization (39), United States (37), men and women (28), social determinants (27), health outcomes (26), World Health Organization (26), history of medicine (25), rural areas (24), health services (22), determinants of health (21), global health (19), 19th century (19), family planning (18), maternal mortality (17), living conditions (17), sex differences (16)	Health issue (service, organization, cause, research, problem), family planning and reproduction
Period 2–Period 3	Frequency Decreasing Terms	African American (−40), health disparities (−17), global health (−13), heart disease (−11), risk factors (−8), coronary heart disease (−8), Health care Research and Quality (−7), ethnic disparities (−6), Community Health (−5), physical activity (−5), racial and ethnic disparities (−5), racial differences (−5), universal health (−5)	Health issue (service, problem, organization, research), inequality and discrimination (health care)
	Frequency Increasing Terms	reproductive health (35), mental health (33), family planning (21), medical sociology (20), New York (19), health care (18), women's health (14), public health (13), sexually transmitted diseases (13), women's health (13), molecular pathology (13), United Nations (12), social science (12), reproductive age (12), health issues (11), live births (10), rural areas (9), determinants of health (9), mental illness (9), Social Science & Medicine (9), women of reproductive age (9)	Health issue (research, service, problem, cause), family planning and reproduction (organization), population issue (organization)
Period 3–Period 4	Frequency Decreasing Terms	feminist theory (−10), women's studies (−8), Western medicine (−8), sexually transmitted diseases (−6), Chinese medicine (−6), Law Review (−5), myth of matriarchal prehistory (−5), Charlotte Perkins (−5), John Knox (−5), medical care (−5)	Woman protection (research, law), health issue (research, treatment, problem)
	Frequency Increasing Terms	mental health (72), women's health (54), United States (38), public health (34), mental illness (26), health care (25), social work (24), world health (24), female genital (24), family planning (23), United Nations (22), developing countries (22), reproductive health (20), sustainable development (20), violence against women (20), maternal mortality (16), developed countries (16), Millennium Development (16), health issues (15), social determinants (13), World Health Organization (13), health research (13), health disparities (13), intimate partner (13), cervical cancer (13)	Health issue (problem, organization, cause, research), abuse and violence (domestic violence), family planning and reproduction (organization), population issue (organization), inequality and discrimination (health care)

Table 11. Changes of Subjects in the Four Periods in the WH-SP Theme.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 1–Period 2	Frequency Decreasing Terms	health insurance (−42), Health Affairs (−13), care services (−12), women's college (−12), medical care (−10), health care services (−10), health care costs (−10), insurance coverage (−8), sex discrimination (−8), Medicaid Services (−8), Women's College Hospital (−8), Centers for Medicare (−7), Washington DC (−6), medical treatment (−6), National Organization for Women (−6), civil rights (−5), universal health (−5)	Health issue (insurance, research, service, organization, treatment), inequality and discrimination, woman protection (organization)
	Frequency Increasing Terms	women's health (163), health center (59), United States (50), reproductive health (50), women's suffrage (37), health initiative (28), public health (27), health centers (27), breast cancer (26), women's rights (25), men and women (22), Michigan Health System (22), Women's Health Initiative (20), hormone therapy (19), planned parenthood (18), health education (16), female condom (16), red dress (16), postmenopausal women (15), colorectal cancer (14), University of Michigan (14), red dress collection (14)	Health issue (organization, problem, treatment, education), inequality and discrimination (politics), woman protection, family planning and reproduction (method)
Period 2–Period 3	Frequency Decreasing Terms	reproductive health (−8), ancient Rome (−5), University of Pittsburgh (−5), Journal of Obstetrics (−5), Stefanick ML (−5)	Family planning and reproduction (organization, research, treatment), health issue (organization, education)
	Frequency Increasing Terms	women's health (82), public health (40), health literacy (35), health education (33), human rights (33), New York (30), violence against women (30), gender equality (29), domestic violence (28), United Nations (27), health care (25), Department of Health (23), Oxford University (21), women's suffrage (20), breast cancer (20), right to vote (20), reproductive rights (20), social security (18), men and women (17), United States (15), health system (15), National Institute (15)	Health issue (education, service, organization), abuse and violence (domestic violence), inequality and discrimination, woman protection (politics, health)

Table 11. *Cont*.

Time Period	High-Frequency Terms and Phrases		Subjects
Period 3–Period 4	Frequency Decreasing Terms	hormone replacement (−20), health system (−16), hormone replacement therapy (−11), Michigan Health System (−10), red dress collection (−9), suffrage referendum (−5), equine estrogen (−5)	Health issue (treatment, organization), inequality and discrimination (politics)
	Frequency Increasing Terms	New York (84), United States (55), women's health (36), violence against women (35), United Nations (26), health care (25), reproductive health (25), gender equality (24), health organization (21), Medical Association (21), women's rights (20), women's suffrage (16), birth control (16), World Health Organization (16), sexual and reproductive health (15), Department of Health (14), health and human (14), United States Department (14), human rights (13), reproductive rights (13), family planning (13), women's education (13)	Health issue (service, organization), abuse and violence, women protection (politics, education), family planning and reproduction

2. The Health problems and risks (WH-HPR) theme

Table 9 shows that the terms about the *health issue* subject and the *family planning and reproduction* subject increased in all the periods. From Periods 1 to 2, the increasing terms about *family planning and reproduction* were related to *family planning and reproduction organizations* (e.g., United Nations Population Fund), while in the following periods, the terms were related to *family planning and reproduction methods* (e.g., induced abortion and medical abortion).

The increasing terms of the health issue subject were related to different aspects in different periods. From Periods 1 to 2, the increasing terms covered various lower-level subjects, including *health problems*, *health organizations*, *health services*, and *causes of health problems*. Among these lower-level subjects, only *health problems* and *health services* attracted more attention than before in the next period. From Periods 3 to 4, a new lower-level subject emerged, which was *health research*.

3. The Medical and interdisciplinary subjects (MIS) theme

Table 10 shows that the frequency of increasing terms covered more and more subjects as time went by in the MIS theme. From Periods 1 to 2, the terms were relevant to *health issue* and *family planning and reproduction*. It means that the Wikipedia editors' interests focused on these two subjects. In Period 3, a new interest about the *population issue* emerged. In addition to the previous subjects, in Period 4, the Wikipedia editors had two more interests, *violence* and *inequality and discrimination*.

4. The Support and protection (WH-SP) theme

Table 11 illustrates that the terms about *health issue* and *woman protection* kept increasing from Periods 1 to 2, although in different periods, these terms focused on different aspects of the two subjects. For instance, the terms about *treatment* only increased from Periods 1 to 2, while the terms about *health education* increased from Periods 1 to 3.

The woman protection subject had three lower-level subjects, which were *politics, health, and education*. The terms about *politics* increased from Periods 2 to 4, which indicates that the Wikipedia editors had increasing interests in this subject in recent years. Furthermore, examination of the terms about *politics* demonstrates that the Wikipedia editors' interests increased the most in women's suffrage.

Table 12 summarizes the growing, diminishing, and fluctuating subjects of the Women's Health topic from 2010 to 2017. The growing/diminishing subjects were the subjects whose associated terms and phrases kept increasing/decreasing during the investigated periods. In other words, the growing/diminishing subjects attracted increasing/decreasing attention during the investigated periods. The fluctuating subjects were the subjects whose associated terms and phrases increased in some periods but decreased in other periods.

Table 12. Growing, Diminishing, and Fluctuating Subjects.

Types	Subjects
Growing subjects	Abuse and violence, family planning and reproduction, health issue, inequality and discrimination, minority group, woman protection
Diminishing subjects	Economy, politics
Fluctuating subjects	Population issue

The *minority group* subject's associated terms/phrases kept increasing from Periods 1 to 4. It became more and more important from Periods 1 to 4. In other words, the Wikipedia editors paid increasing attention to the minority groups from 2010 to 2017.

4.3.3. Changes of External Popularities

The external popularity of a topic/theme was defined as the numbers of the page edits and the numbers of the page views of its associated entries. The Friedman's Test was applied to test for the differences among the periods. Table 13 presents the results.

Table 13. Hypothesis Testing Results of H01 and H02.

Hypotheses	Chi-Square Value	*p*-Value
H01	$\chi^2(3) = 73.384$	0.000
H02	$\chi^2(3) = 0.383$	0.944

The results show that *H01* was rejected. It means that: (1) there *were significant differences* among the four periods in terms of the number of the page views; (2) there *were no significant differences* among the four periods in terms of the number of the page edits.

The Sign Test was used to explore the differences between every two periods. The comparisons intended to reveal the differences from one period to the next in order to show the temporal changes of external popularities. Hence, only the adjacent periods were compared. Since the result of *H02* was not significant, no follow-up test was conducted for this hypothesis. The results of the follow-up tests for *H01* are presented in Table 14.

Table 14. Pairwise Comparison Results of H01 and H02.

Values	Period 1 vs. Period 2	Period 2 vs. Period 3	Period 3 vs. Period 4
Z-value	−7.363	−1.331	−6.134
p-value	0.000	0.183	0.000

Table 14 displays that there were significant differences among Periods 1 and 2, and Periods 3 and 4, but no significant difference was found among Periods 2 and 3 in terms of the number of the page views. When investigating the detailed results obtained from the pairwise comparisons, it shows that the number of the page views of Period 2 was larger than that of Period 1 (129 positive signs versus 34 negative signs) and the number of the page views of Period 4 was smaller than that of Period 3 (53 positive signs versus 139 negative signs). Therefore, the number of the page views of the associated entries in *women's health* grew from Periods 1 to 2, remained stable from Periods 2 to 3, and dropped from Periods 3 to 4. These findings reveal that the Wikipedia editors' interests in women's health did not change quickly from 2010 to 2017, while the Wikipedia viewers' interests in this topic grew rapidly from Periods 1 to 2 but dropped quickly from Periods 3 to 4, which indicates that these groups were built by different people.

5. Conclusions

This study discovers the evolution characteristics of the women's health topic on Wikipedia. Two hundred and seven associated entries of *women's health* were retrieved on Wikipedia, and four themes emerged from these entries, which were (2) *Discrimination, violence, harm, and subordination*; (2) *Health problems and risks*; (3) *Medical and interdisciplinary subjects*; and (4) *Support and protection*. It indicates that the Wikipedia editors focused on these four aspects of women's health.

From the internal characteristic's aspect, the women's health content on Wikipedia kept increasing from 2010 to 2017. The subjects became increasingly diverse as time went by. The editors paid more and more attention to *abuse and violence*, *family planning and reproduction*, *health issue*, *inequality and discrimination*, *minority group*, and *woman protection*, while their interests in *economy* and *politics* decreased. If a subject was quickly changed in certain periods, it was usually caused by social events or social issues.

From the external popularity's aspect, the overall popularity of the women's health topic declined from 2010 to 2017, contrary to the growth of their content and the growth of extensive online health information seeking. The themes identified in this study had similar trends of popularities among the Wikipedia viewers. Their popularity all grew rapidly from Periods 1 to 2, remained stable from Periods 2 to 3, and fell dramatically from Periods 3 to 4. However, the trends of the popularities among the viewers were not consistent with those among the editors. Therefore, the two groups were not composed of the same members.

The results show that no association was found between the internal characteristic evolution and the external popularity evolution of the women's health topic on Wikipedia. The content generation or change of Wikipedia entries had no impact on the Wikipedia users.

The findings can enable health professionals, health care givers, and general users to get a more comprehensive understanding of women's health information on social media by illustrating and discovering the entries, subjects, and themes of women's health discussed on Wikipedia and the relationships among them. Exploring the women-health-related themes and subjects will contribute to the developments of health ontologies and consumer health vocabularies and assist Website designers in organizing online women's health information. Revealing the temporal features of the women's health topic can support the temporal information retrieval of women-health-related information.

There are plenty of health-related topics on social media, such as women's health, men's health, and children's health, which are worthy research topics. However, because of the limitations of time and paper length, it is difficult to investigate all the related topics on different social media platforms in one study. In future research, the researchers will explore the characteristics of more health topics on Wikipedia and other social media platforms, and compare different health topics and health information on different platforms.

Author Contributions: Conceptualization, J.Z. and Y.W.; methodology, Y.W. and J.Z.; software, Y.W.; validation, Y.W.; formal analysis, Y.W.; investigation, Y.W.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, J.Z. and Y.W..; visualization, Y.W.; supervision, J.Z.; project administration, J.Z. and Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Appendix A

Table A1. Themes and Investigated Entries of Women's Health Topic.

Themes	Entries
Discrimination, violence, harm, and subordination (DVHS): 37 entries	(1) Ageism, (2) Airline seating sex discrimination controversy, (3) Ambivalent sexism, (4) Discrimination against girls in India, (5) Female genital mutilation, (6) Femicide, (7) Gender apartheid, (8) Gender bias on Wikipedia, (9) Gender inequality in India, (10) Gender inequality, (11) Glass cliff, (12) Hegemonic masculinity, (13) Heterosexism, (14) Husband stitch, (15) Hypermasculinity, (16) LGBT stereotypes, (17) Male privilege, (18) Misogyny in horror films, (19) Misogyny, (20) Missing women, (21) Occupational segregation, (22) Occupational sexism, (23) Patriarchy, (24) Pink-collar worker, (25) Rape culture, (26) Reverse sexism, (27) Sexism in the technology industry, (28) Sexism, (29) Transphobia, (30) Triple oppression, (31) Victim blaming, (32) Wife selling, (33) Women in firefighting, (34) Women in law enforcement, (35) Women in medicine, (36) Women in Pakistan, (37) Women in the workforce
Health problems and risks (WH-HPR): 25 entries	(1) Abortion, (2) Anilingus, (3) Birth control, (4) Complications of pregnancy, (5) Disease, (6) Diseases of affluence, (7) Diseases of poverty, (8) Drift hypothesis, (9) Gender disparities in health, (10) Gender polarization, (11) Hypertensive disease of pregnancy, (12) Incarceration of women in the United States, (13) Inequality in disease, (14) Infant mortality, (15) List of bacterial vaginosis microbiota, (16) Medical anthropology, (17) Mental health inequality, (18) Misandry, (19) Molar pregnancy, (20) Ovarian cancer, (21) Schistosomiasis, (22) Unnatural Causes: Is Inequality Making Us Sick?, (23) Water supply and sanitation in India, (24) Women's Health Issues (journal), (25) Women and smoking

Table A1. *Cont*.

Themes	Entries
Medical and interdisciplinary subjects (MIS): 46 entries	(1) Epidemiology, (2) Etiology, (3) Face-ism, (4) Family planning, (5) Gender-blind, (6) Global health, (7) Health equity, (8) Health in China, (9) Health in India, (10) Health, (11) History of medicine, (12) History of nursing, (13) Immigrant paradox, (14) International Conference on Population and Development, (15) Intersectionality, (16) Maternal health, (17) Matriarchy, (18) Medical sociology, (19) Menstruation, (20) Mental health, (21) Molecular pathological epidemiology, (22) Pathogenesis, (23) Pathology, (24) Population Health Forum, (25) Population health, (26) Public health, (27) Race and health, (28) Reproductive health, (29) Richard G. Wilkinson, (30) Sex differences in humans, (31) Sex segregation, (32) Sexual division of labour, (33) Social determinants of health in Mexico, (34) Social determinants of health in poverty, (35) Social determinants of health, (36) Social determinants of obesity, (37) Social epidemiology, (38) Vaginal tightening, (39) Whitehall Study, (40) Women's health in China, (41) Women's health in Ethiopia, (42) Women's health in India, (43) Women's health, (44) Women's reproductive health in Russia, (45) Women's reproductive health in the United States, (46) Women who have sex with women
Support and protection (WH-SP): 99 entries	(1) Alexandria Regional Center for Women's Health and Development, (2) American Medical Women's Association, (3) AnMed Health Women's & Children's Hospital, (4) Antifeminism, (5) Association of Women's Health, Obstetric and Neonatal Nurses, (6) Australian Longitudinal Study on Women's Health, (7) Australian Women's Health Network, (8) B.C. Women's Hospital & Health Centre, (9) Black Women's Health Study, (10) Condom, (11) Dennis Raphael, (12) Equity feminism, (13) EuroHealthNet, (14) European Institute of Women's Health, (15) Female condom, (16) Female education, (17) Feminism, (18) Feminist health centers, (19) Feminist movement, (20) Feminist Women's Health Center (Atlanta, Georgia), (21) Florence Hartley, (22) Gender equality, (23) Gender feminism, (24) Gender neutrality, (25) Global Library of Women's Medicine, (26) Global Task Force on Expanded Access to Cancer Care and Control in Developing Countries, (27) Gynaecology, (28) Gynography, (29) Health (magazine), (30) Health Care for Women International, (31) Health care in the United States, (32) Health Disparities Center, (33) Health education, (34) Health literacy, (35) Health professional, (36) Healthcare and the LGBT community, (37) Healthcare in Canada, (38) Healthy People program, (39) HealthyWomen, (40) Hopkins Center for Health Disparities Solutions, (41) Hormone replacement therapy (menopause), (42) Howard Atwood Kelly, (43) International Journal of Women's Health, (44) International Planned Parenthood Federation, (45) International Women's Health Coalition, (46) Ipas (organization), (47) Journal of Midwifery & Women's Health, (48) Journal of Women's Health, (49) Kegel exercise, (50) Laura W. Bush Institute for Women's Health, (51) List of first female physicians by country, (52) List of health and fitness magazines, (53) List of medical journals, (54) List of women's studies journals, (55) Madsen v. Women's Health Center, Inc., (56) Martha Ballard, (57) Men and feminism, (58) Michael Marmot, (59) Michigan Medicine, (60) Midwife, (61) Midwifery, (62) National Organization for Men Against Sexism, (63) National Organization for Women, (64) National Women's Health Network, (65) New Space for Women's Health, (66) Office on Women's Health, (67) Oregon Health and Science University Center for Women's Health, (68) Our Bodies, Ourselves, (69) Psychology of Women Quarterly, (70) Reproductive Health Supplies Coalition, (71) Reproductive rights, (72) Separatist feminism, (73) Sex Roles (journal), (74) Society for Women's Health Research, (75) Sunnybrook Health Sciences Centre, (76) Sutter Health, (77) Sybil Shainwald, (78) Tamika D. Mallory, (79) The Heart Truth, (80) The Honest Body Project, (81) The NeuroGenderings Network, (82) Torches of Freedom, (83) United Nations Foundation, (84) United Nations Population Fund, (85) United States Department of Health and Human Services, (86) University of Pittsburgh Graduate School of Public Health, (87) Women's College Hospital, (88) Women's empowerment, (89) Women's Health (magazine), (90) Women's Health Action and Mobilization, (91) Women's Health Care Nurse Practitioner-Board Certified, (92) Women's Health Initiative, (93) Women's health nurse practitioner, (94) Women's medicine in antiquity, (95) Women's rights in Iran, (96) Women's rights, (97) Women's suffrage, (98) Women & Health, (99) Women in India

References

1. Tu, H.T. Surprising decline in consumers seeking health information. *Track. Rep.* **2011**, *26*, 1–6.
2. Fox, S.; Duggan, M. Health Online. 2013. Available online: http://www.pewinternet.org/2013/01/15/health-online-2013/ (accessed on 20 June 2020).
3. Moorhead, S.A.; Hazlett, D.E.; Harrison, L.; Carroll, J.K.; Irwin, A.; Hoving, C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *J. Med. Internet Res.* **2013**, *15*, e85. [CrossRef] [PubMed]
4. Dawson, J. Doctors Join Patients in Going Online for Health Information. Available online: http://connection.ebscohost.com/c/opinions/49259197/doctors-join-patients-going-online-health-information (accessed on 25 March 2010).
5. Marar, S.D.; Al-Madaney, M.M.; Almousawi, F.H. Health information on social media. perceptions, attitudes, and practices of patients and their companions. *Saudi Med. J.* **2019**, *40*, 1294–1299. [CrossRef] [PubMed]
6. Zhang, J.; Zhao, Y. A user term visualization analysis based on a social question and answer log. *Inf. Process. Manag.* **2013**, *49*, 1019–1048. [CrossRef]

7. Brasil, P.; Pereira, J.P.J.; Moreira, M.E.; Ribeiro Nogueira, R.M.; Damasceno, L.; Wakimoto, M.; Nielsen-Saines, K. Zika Virus Infection in Pregnant Women in Rio de Janeiro. *N. Engl. J. Med.* **2016**, *375*, 2321–2334. [CrossRef] [PubMed]

8. Oteng-Ntim, E.; Tezcan, B.; Seed, P.; Poston, L.; Doyle, P. Lifestyle interventions for obese and overweight pregnant women to improve pregnancy outcome: A systematic review and meta-analysis. *Lancet* **2015**, *386*, S61. [CrossRef]

9. Subramaniam, M.; Prasad, R.O.; Abdin, E.; Vaingankar, J.A.; Chong, S.A. Single mothers have a higher risk of mood disorders. *Ann. Acad. Med. Singap.* **2014**, *43*, 145–151. [PubMed]

10. Asiodu, I.V.; Waters, C.M.; Dailey, D.E.; Lee, K.A.; Lyndon, A. Breastfeeding and use of social media among first-time African American mothers. *J. Obstet. Gynecol. Neonatal Nurs.* **2015**, *44*, 268–278. [CrossRef] [PubMed]

11. Gleeson, D.M.; Craswell, A.; Jones, C.M. Women's use of social networking sites related to childbearing: An integrative review. *Women Birth* **2019**, *32*, 294–302. [CrossRef] [PubMed]

12. Holtz, B.; Smock, A.; Reyes-Gastelum, D. Connected motherhood: Social support for moms and moms-to-be on Facebook. *Telemed. J. E-Health* **2015**, *21*, 415–421. [CrossRef] [PubMed]

13. Ure, C.; Cooper-Ryan, A.M.; Condie, J.; Galpin, A. Exploring Strategies for Using Social Media to Self-Manage Health Care When Living with and Beyond Breast Cancer: In-Depth Qualitative Study. *J. Med. Internet Res.* **2020**, *22*, e16902. [CrossRef] [PubMed]

14. Milne, D.; Witten, I.H. Learning to Link with Wikipedia. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 26–30 October 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 509–518.

15. Milne, D.; Witten, I.H. An open-source toolkit for mining Wikipedia. *Artif. Intell.* **2013**, *194*, 222–239. [CrossRef]

16. Wikimedia Downloads. Available online: https://dumps.wikimedia.org/ (accessed on 9 July 2020).

17. Keyes, O. Package 'WikipediR'. Available online: http://www.stats.bris.ac.uk/R/web/packages/WikipediR/WikipediR.pdf (accessed on 20 June 2020).

18. Kohonen, T.; Kaski, S.; Lagus, K.; Salojarvi, J.; Honkela, J.; Paatero, V.; Saarela, A. Self organization of a massive document collection. *IEEE Trans. Neural Netw.* **2000**, *11*, 574–585. [CrossRef] [PubMed]

19. The R Foundation for Statistical Computing. The R Project for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 17 July 2020).

20. RStudio Team. RStudio. Available online: https://rstudio.com/ (accessed on 17 July 2020).

21. Ultsch, A.; Siemon, H. Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis. In Proceedings of the INNC'90, International Neural Networks Conference, Palais des Congres, Paris, France, 9–13 July 1990; Kluwer Academic: Dordrecht, The Netherlands; Boston, MA, USA, 1990; pp. 305–308.

22. Bem, S.L. Dismantling gender polarization and compulsory heterosexuality: Should we turn the volume down or up? *J. Sex Res.* **1995**, *32*, 329–334. [CrossRef]