Review of the Scientific Literature on the Topic of Online Dating Services in a Demographic and Social Context

German A. Klimenko

1 Lomonosov Moscow State University, Moscow, 119991, Russia

Abstract

The use of online dating services has surged dramatically in recent years. Concurrently, a substantial body of scientific literature has emerged, analyzing these services from demographic and social perspectives. This review, based on 528 English-language publications from the past 13 years, compiles a bibliographic database and employs content analysis to systematize research directions on the use of online dating services. The classification includes topics such as locations and methods of dating (both traditional and digital), the risks and drawbacks of digital dating services, user actions and interactions, dating goals, marital status, and more. This bibliographic database has helped identify the primary vectors of publication activity on the topic of online dating services over the past 13 years.

Keywords

dating sites, online dating services, tinder, services, users

JEL codes: J11, J12, C88.

Introduction

Due to the increasing pace of digitalization in society, particularly in the realm of interpersonal communications, social networks have become more popular. These platforms enable individuals with common interests, hobbies, or passions to connect, broaden their horizons, and share news and materials.

Simultaneously, online dating services have been developed, providing opportunities to build romantic relationships without the necessity of face-to-face meetings during the initial stages of dating. These services facilitate partner searches by automatically matching indi-
viduals based on desired characteristics and preferences. However, they also present several risks, including fraud, sexual harassment, and other safety concerns.

In recent years, the topic of online dating services has gained popularity in the popular science discourse in Russia. However, abroad, scientific discourse on this topic began earlier and is more active today due to the earlier adoption of online dating services. Despite many years of studying online dating services, general theoretical and methodological approaches to analyzing this social phenomenon have not yet been established. Consequently, there is a need for a comprehensive analysis of scientific publications in the English-speaking world.

The purpose of this study is to identify the main topics of English-language scientific research on online dating services.

Analyzing scientific research on the topic of online dating services conducted before 2022 is particularly valuable, as it allows for the identification of trends in the development of the topic during a period when international online dating services were available and functioning in Russia. By the end of this period, the Russian online dating market began its separate development.

Research topics in the English-language literature reflect significant social issues related to the process of finding a romantic partner in an online environment. These issues attract the attention of the scientific community, the state, and the business sector. Analyzing dating processes through the lens of online services allows for a comprehensive assessment of matrimonial attitudes. For instance, studies that focus on attitudes toward marriage and childbirth without considering the dating stage—which has undergone significant transformation due to the development of information technology—lack sufficient predictive power, in our opinion.

The identification of the main topics will be carried out using the method of content analysis of scientific English-language publications. The applicability of this method for addressing similar issues is well-documented in the works of domestic researchers (Kalabikhina & Chesnokov 2020; Gerasimenko et al. 2021).

**Method and data**

To analyze the scientific literature on the topic of online dating services, we utilized the five largest publishers of scientific works in English: Science Direct (Elsevier), Springer, Taylor & Francis, Wiley, and SAGE.

To optimize the search for academic papers in these online publishing platforms, we first performed a frequency analysis of expressions related to the research topic in the English-speaking internet environment (using search engines and systems).

We conducted trial uploads of scientific publications using the selected expressions and performed an outlier analysis to filter out irrelevant articles. After this process, we identified nine key expressions, which formed the basis for downloading scientific publications. A universal query was then compiled using these expressions:

«Online dating» OR «dating sites» OR «dating apps» OR «dating services» OR «dating platforms» OR «mobile dating» OR «dating applications» OR «dating websites» OR «Tinder.»

1 “Online dating” OR “dating sites” OR “dating apps” OR “dating services” OR “dating platforms” OR “mobile dating” OR “dating apps” OR “dating sites” OR “tinder”
The services of scientific citation and access to publications in the English-speaking environment have unified forms and principles for making a request.

The query is centered around the aggregator «OR,» meaning it is not necessary for all the selected expressions to appear in a publication. During the manual selection of publications, it was found that some relevant works might not contain the expressions from the query in the abstract or keywords. Therefore, the query was applied to the entire text of the publication.

The proportion of relevant articles was less than 10% of all articles. Due to the high concentration of irrelevant results in the generated sets of publications, automatic retrieval would have been ineffective and led to significant biases in the frequency analysis. Consequently, the selection, downloading, and formation of the final array of publications were performed manually in each service.

The sample of scientific publications was limited to a time period starting from 2010, when online dating services began to actively gain popularity and differed in functionality and toolsets from social networks (e.g., match.com).

In total, 528 scientific publications were included in the analysis (see the distribution by publishing houses in Table 1). Among them, the largest portion – 434 publications (82.2%) – are articles, with the remainder being chapters or excerpts from books.

Table 1. Distribution of selected publications by place of publication

<table>
<thead>
<tr>
<th>Publishing house</th>
<th>Number of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Springer</td>
<td>179</td>
</tr>
<tr>
<td>SAGE</td>
<td>163</td>
</tr>
<tr>
<td>Science direct</td>
<td>98</td>
</tr>
<tr>
<td>Wiley</td>
<td>59</td>
</tr>
<tr>
<td>Taylor &amp; Francis</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>528</td>
</tr>
</tbody>
</table>

Source: compiled by the author based on the analysis of full-text publications.

The distribution of publications by year is shown in Figure 1. The main publication activity occurred in the second half of the period under review, starting in 2018. The number of publications in 2022 is the largest for the entire period under review, even though the data collection, conducted in mid-November 2022, does not cover the entire possible volume of publications on the subject for the full calendar year.

The most complete data on publications are presented in the dataset, which will be included in Supplementary material 1 (Information on publications). A detailed analysis of the set of publications is provided in Supplementary material 2 (Description of the dataset).

In the next stage, a lexical analysis of the array of texts from the scientific papers was performed using the specialized program for qualitative data analysis, MAXQDA. Word clouds highlighting the most frequently found words in the selected publications (top 100 and 500 words) were generated (Figures 2 and 3).

The most common words found characterize online communication directly, such as «online,» «dating,» «internet,» and «participants.» Additionally, specific aspects of communication within dating apps or services are indicated by words like «physical,»
«match,» «pair,» «profile,» «young» (referring to the target audience of the services), and «status.»

Furthermore, a set of words denotes the goals of dating, including «romantic,» «sex,» «love,» and others.

Supplementary material 3 (Word frequency data) provides detailed information on the frequency of word usage within the set of scientific publications under analysis.

Figure 1. Distribution of analyzed publications by year (from 2010 to 2022), units. Source: compiled by the author based on the analysis of full-text publications.

Figure 2. Cloud of the 100 most used words in the set of analyzed documents. Source: compiled by the author based on the analysis of full-text publications.
On average, the most common words like «online,» «dating,» «internet,» and «participants» are used in more than 80% of the documents. However, certain words warrant separate attention. For instance, «app(s)» and «Tinder» are present in almost half of the publications. This prevalence is due to specific studies focusing on these particular online dating services, reflecting their significant impact and distinct characteristics within the literature.

**Encoding**

The process of encoding involved using the top 500 most frequently used words from the set of publications to create codes for subsequent quantitative analysis. During this process, certain words and constructions were excluded based on predefined criteria:

1. Prepositions, conjunctions, and other constructions without semantic load, including some adverbs like «always» and «likely.»
2. Words or phrases that are common but do not directly reflect the essence of the publication, such as «author,» «published,» «page,» and «example.»

In total, over 400 words and constructions were excluded from the encoding process. The complete list of excluded words and constructions is provided in Supplementary material 4 (Blacklist of words).

In the process of encoding, forms and words sharing the same root can indeed form unique codes, which can then be grouped together. Given that the articles under consideration are directly or indirectly related to demography (population science), it is anticipated that encoded words will typically be used within a limited range of meanings.

For instance, consider the word «black.» In the context of the set of publications, this word is likely used to refer to African Americans, following the standard statistical designation in
the United States. The probability of encountering this word in other contexts within these publications is low, as it would primarily serve this specific demographic-related meaning.

After analyzing the frequently used words in the set of publications, nine main vectors of the development of the analyzed topic were identified, presented in Table 2:

<table>
<thead>
<tr>
<th>№</th>
<th>Name of the code group</th>
<th>Set of codes</th>
<th>Number of marked up words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Digital resources</td>
<td>Facebook*, site*, website*, mobile, tinder*, grindr*, application*, platform*, service*, network*, app (apps, application*)</td>
<td>67341</td>
</tr>
<tr>
<td>2</td>
<td>Actions and interaction of users of digital services</td>
<td>like*, match*, messag*, search, profile, self-presentation, response*, rate*, chat*, sexting</td>
<td>41819</td>
</tr>
<tr>
<td>3</td>
<td>Dating goals</td>
<td>dating, relationship*, love, find, marriage, seek*, roman*, intim*, feel, mating, dates, long-term, emotion*, expectation*, child*, encounters, ideal*, impression, reciprocal, sexual, meet*, family, met</td>
<td>113099</td>
</tr>
<tr>
<td>4</td>
<td>Marital status and status-related words</td>
<td>single, couples, daters, married</td>
<td>7898</td>
</tr>
<tr>
<td>5</td>
<td>Thin market (limited marriage markets)</td>
<td>gay, race/racial, white, queer, msm, black, asian, bisexual, lesbi*, ethnic*</td>
<td>20777</td>
</tr>
<tr>
<td>6</td>
<td>Physical characteristics of users of digital online dating services</td>
<td>age/ages, young, attractive*, body/bodies, face/facial, adult*, student*, photo*, image*, pictur*, mascul*, old, youth, physical</td>
<td>36796</td>
</tr>
<tr>
<td>7</td>
<td>Socio-psychological features of users of digital online dating services</td>
<td>education, cultur*, hetero* traits, orientation, income, interest*, self-esteem, gender</td>
<td>28299</td>
</tr>
<tr>
<td>8</td>
<td>Places and ways of dating (both traditional and digital)</td>
<td>friend/friends, offline, college, real, place, school, face-to-face, location*</td>
<td>15854</td>
</tr>
<tr>
<td>9</td>
<td>Risks and disadvantages of digital online dating services</td>
<td>risk, hiv, violence, trust, deception, victim*, fear, safe*, anxiety, rejection, privacy/private, discrim-*</td>
<td>21846</td>
</tr>
</tbody>
</table>

Source: compiled by the author based on the analysis of full-text publications.

Note: * – words with any endings instead of asterisks are considered: word forms (website and websites) and single-root words (attractive and attractiveness).

In total, 353729 codes were marked up in all articles. More detailed information on the frequency of distribution of codes in the array of publications is provided in Supplementary material 5 (Codes) and Supplementary material 6 (Code Matrix).

Let’s take a closer look at the division of codes into selected categories, the meaning and characteristics of each group of codes.
Based on the provided information, it seems you are outlining specific categories or themes within the research on online dating services. Here's a structured approach based on the details provided:

1. **Digital Resources**
   This category encompasses online communication resources facilitating the meeting of potential partners and making acquaintances. It includes prominent services and applications such as Tinder, Grindr, and others.

   Research in this group is increasingly centered on studying online platforms and the specific characteristics of individual online dating services. Some publications focus directly on analyzing the user demographics and behaviors of these services, while others provide comprehensive analyses (Miguel 2016, 2018; Sobieraj & Humphreys 2021; Halversen et al. 2022).

   **Grindr Specific Research:**
   - Grindr is a notable online service within this category, popular among researchers studying online dating in LGBT communities. Despite its specialized audience and potentially limited coverage, Grindr remains a significant object of study. It is important to note that while Grindr is associated with the LGBT community and may be restricted or banned in certain regions (like in Russia), it serves as a subject of academic research worldwide (Blackwell et al. 2015; Stempfhuber & Liegl 2016; Anderson et al. 2018; Filice et al. 2019; Halversen et al. 2022).

   We emphasize that the word did not fall into the 5th category associated with limited marriage markets during encoding, since the code itself is more consistent with the purpose of combining codes in the first group. Publications that mention Grindr will include words from the fifth category, which should potentially not lead to distortions in accounting statistics.

2. **Actions and interaction of users of digital services**
   This category focuses on the primary actions users perform on digital dating services and applications, such as communication, profile search, likes, and chat conversations.

   This subtopic revolves around the action of liking (or swiping\(^1\)) profiles and the concept of getting matches. A match occurs when users of an online service have mutually liked each other, indicating mutual interest and a willingness to communicate. Matches are a central feature of most dating services, including Tinder, where communication can only begin after mutual likes are received (Sumter et al. 2017; Ward 2016; LeFebvre 2018). Tinder's policy of requiring mutual likes before communication aims to mitigate various potential risks, including minimizing criminal situations that may arise during male-female interactions. This policy helps reduce the likelihood of unwanted sexting, a phenomenon where sexual messages, including visual content like «dickpics,» (photos of a male genital organ sent by a file in a personal message service) are sent unsolicited (Shaw 2016; Dietzel 2022).

3. **Dating goals**
   Dating goals for using online dating services can be divided into two broad categories: short-term and long-term relationships.

---

\(^1\) Swipe - movement of a finger, swiping to the right or left of a photo of a proposed person to establish acquaintance and further communication. The function is key for Tinder.
Short-term dating goals mainly include goals with intimate or sexual overtones (Jin et al. 2019; Grøntvedt et al. 2020; Schwarz et al. 2020). First of all, it is directly sex, intimacy, sexting and so on.

Long-term relationship goals include those associated with finding a future spouse, starting a family, and having children (Potarca et al. 2015; Schwarz et al. 2020; Zinck et al. 2022).

Some codes in the group do not have a clear link to the duration of the relationship and express the main motives for using online dating services: encounters, feel, dates, etc. (Licoppe et al. 2016; Timmermans et al. 2018; Maliepaard & van Lisdonk 2019).

4. Marital status and status-related words
Description of the main marital statuses of the user of the online dating service: married, having a partner, active search status (dater).

Table 2 demonstrates the relative unpopularity of this group of codes compared to others. This lack of focus is attributed to the generally lower emphasis researchers place on marital status in the context of online dating. Most studies typically consider the situation of individuals seeking acquaintances regardless of their marital status (Murray 2020; Himawan et al. 2022).

5. Thin market (limited marriage markets)
The emergence of such markets is associated with racial, national, gender-psychological and other restrictions that reduce the number of possible partners in the marriage market. For example, this group should include codes related to representatives of the LGBT community (considered extremist on the territory of the RF and banned).

The analysis of the frequency of use of words in a set of documents shows a contradictory picture: on the one hand, a number of words forming the codes of this group are among the most common in the array of publications, on the other hand, the proportion of documents in which these words occur is much lower than that of words from other groups with a similar level of frequency of use in a set of documents. For example, the word «gay» is the 37th most frequently used and occurs in 48.5% of the analyzed documents, which is almost twice lower than the neighboring words «like», «individuals», «time» (found in 84.7%, 80.9% and 96.0% of documents).

Accordingly, the thin market theme, although relatively popular in English-language literature, is rarely mentioned in publications about online dating services.

The main body of publications on this topic is related to the analysis of the features of online dating of representatives of both racial and sexual minorities (Miller & Behm-Morawitz 2016; Rafalow et al. 2017; Wang 2020; Smith 2022). It is worth noting some studies on specific territories, with the premise of varying degrees of freedom of behaviour of these people in different regions. As an example, we can cite a study (Li & Chen 2021) about dating on Australian online dating services for Chinese queer women, where an important aspect is the racial context associated with negative attitudes towards the Chinese during the period of active spread of the coronavirus.

Some publications are devoted to analyzing the choice of a partner among users who prefer representatives of only certain racial or ethnic groups for dating. For example, racial minorities in the United States who do not want to mix with representatives of other races (Alhabash et al. 2014; Weser et al. 2021). However, a number of scientific publications, on
the contrary, focus on interracial communication in online services (Curington 2021; Curington et al. 2015, 2020).

6. Physical characteristics of users of digital online dating services
This group examines the physiological characteristics of users, which can be assessed through a photo or basic profile description: facial features, body structure, age, etc. The physical characteristics of users of online dating services are reflected in the photo or set of photos posted by the user in their profile (McGloin & Denes 2018). For the most part, users of online dating services choose photos that they believe showcase their physical qualities in the most favorable light, thereby increasing their attractiveness to the target audience (Todorov & Porter 2014; Miguel 2016; Jin et al. 2019; Daniels 2020; Gao et al. 2021). Some physical characteristics are further emphasized in the profile description (Brand et al. 2012). For example, the age of the user, which is a separate indicator of physical characteristics.

It should be noted that there is a high concentration of publications devoted to studying the physical characteristics and peculiarities of users of online dating services using neural analysis. Several publications focus on analyzing user choices based on visual characteristics, such as profile photos. Notably, some works highlight the automatic generation of photos using neural networks, examining both the preferences of users and the selection or editing of existing user photos to simulate impressions formed from photographs (Kalra & Peterson 2019; van der Zanden et al. 2022).

Despite the complexity of hiding physical characteristics and the implementation of advanced identification and authentication processes\(^1\), cases of deception by users occasionally come to light. These include editing profile photos, providing false information, and uploading photos of another person (Cross & Holt 2021; Jozsa et al. 2021; Zhao & Yan 2022).

7. Socio-psychological features of users of digital online dating services
This group examines the socio-psychological characteristics that are important when searching for a partner. These features include those detailed in the profile description and relate to a person’s internal state and culture: interests, hobbies, level of education, and other significant aspects (Ranzini & Lutz 2017; Xiao & Qian 2020).

The profile description is an important point of a person’s self-presentation, which underlies some publications on psychological topics (Finkel et al. 2012; Hutmacher & Appel 2023).

Individual socio-psychological traits can be emphasized by photographs. For example, the financial situation is often highlighted by male users who try to demonstrate their financial viability through photos featuring expensive objects, such as luxury watches, private cottages, and expensive cars. Such photos help users increase their appeal to and reach within the female audience (Ong & Wang 2015). However, such photos may pose risks to female

---
\(^1\) New registered users of the dating site or application must go through the identification process. To date, the most popular procedure is to compare a photo with a real person being registered by using the user to perform certain actions on the camera in online mode during registration. The necessary actions are performed by the user according to the instructions, which are displayed before identification in video format.
users, as they may be unreliable and specifically posted for deceptive or criminal purposes. The luxury items displayed may not belong to the person posting them, and the individual may have no genuine connection to these items (Cross & Lee 2022).

As with physical characteristics, users may mislead others with the socio-psychological characteristics indicated in their profiles (Cassiman 2019; Cross & Holt 2021). This type of deception is potentially more traumatic and dangerous for the affected party, as the identification of undesirable socio-psychological traits during offline interactions can lead to psychological trauma.

8. Traditional (offline) places and ways of dating
The category includes several groups of codes:

A. Possible places for offline meetings after people can meet online (for example, public places) (Licoppe et al. 2016; Hallam et al. 2019).

B. The main ways/institutions of offline dating outside of online dating services: work, college, university, through friends, etc. (Potarca 2017; Hanson 2022).

C. Words and collocations that are synonymous with offline meetings: face-to-face meetings, real meetings, etc.

9. Risks and disadvantages of digital online dating services
This group includes codes that demonstrate the main negative consequences that can occur during or after online dating interactions. These consequences encompass unpleasant communication experiences, fears, blackmail, insecurity, and privacy violations (Choi et al. 2022; Filice et al. 2022). In addition, potential dangers associated with meeting online contacts in person, including violence and the risk of HIV infection (Mazanderani 2012; Heijman et al. 2016; Wayal et al. 2019).

Most of the group's research is related to codes that identify offline dating sites. Most risks correlate with communication during offline meetings (Gillett 2018; Almond et al. 2020; Qu et al. 2022). However, risks can also arise in online communications, which can also be violent and criminal in nature (Gillett 2018; Hirayama 2019; Almond et al. 2020). And there are also risks of leakage of personal data and confidential information: personal photos, profile data, correspondence and other attributes that can be used for blackmail or other purposes (Pujason-Zazik & Park 2010; Jozsa et al. 2021).

HIV-related publications are popular in scientific discourse. Online dating in the thin market group is considered separately (Luo et al. 2019; Macapagal et al. 2019; Choi et al. 2021). However, the HIV problem is one of the key threats for all categories of users of online dating services (Mazanderani 2012; Tsai et al. 2019).

Conclusion

Content analysis of English-language scientific publications on online dating and dating applications reveals the development of at least nine major thematic groups within this field. Key trends and vectors of publication activity include:

1. The growing popularity of the topic in scientific discourse, especially in recent years.

   Presumably, this trend is partly related to the COVID pandemic. The pandemic
has spurred the development of the topic due to the growing popularity of online communication.

2. Research on limited marriage markets, including those involving the LGBT community (considered extremist and banned in the RF) and various racial and ethnic groups, is prominent. This popularity is partly due to the predominance of authors from Western countries where these topics are highly relevant.

3. Focus on risks and problems. In the scientific discourse, many publications address the risks associated with online communication, both online and during offline meetings. This includes potential dangers such as fraud, sexual harassment, and other safety concerns.

4. In the last few years, many publications appeared, the authors of which use neural networks in their research to analyze the features of accessing online dating services. Recent studies employ neural networks to analyze user interactions and preferences on online dating services. These technologies help in generating ideal user profiles and adjusting images based on user feedback.

5. Although it was anticipated that alternative dating services like Badoo and Bumble would be widely studied, over 80% of the publications focus predominantly on Tinder. This highlights Tinder’s significant impact and popularity within the research community.

The popularity of online dating services is expected to continue growing for several reasons:

1. Online dating services require less material investment, effort, and time compared to offline spaces.

2. Features such as geolocation-based search and mobility (the ability to use services anywhere) make online dating services highly attractive.

There is a growing trend of people consciously avoiding offline interactions, preferring digital communications such as messaging over phone calls or face-to-face meetings.

The continuous transformation of online dating processes underscores the relevance of ongoing research. Future studies should focus on understanding the values and guidelines of users of online dating services, providing deeper insights into user behavior and preferences.

**Financing**

The study was carried out within the framework of the research project «Population reproduction in socio-economic development».

**Reference list**


Gerasimenko VV, Chesnokov AA, Romanova YuA (2021) Izuchenie klientskogo opyta s ispol’zovaniem kontent-analiza dlja povyshenija privlekatel’nosti brenda rabotodatelja [Studying customer...
experience using content analysis to increase the attractiveness of the employer's brand]. Marketing and Marketing Research 1: 48–58. https://doi.org/10.36627/2074-5095-2021-1-1-48-58 (in Russian)


Supplementary material 1

Information about publications
Explanatory note: Set of data on scientific publications. The data is presented in CSV format (UTF-8 encoding). Description of table headers:

- № - Number of the scientific publication.
- Type of publication - Article in a scientific journal or a book chapter
- Publication year - The numerical value is from 2010 to 2022 inclusively.
- Authors - The entire list of authors of the publication is presented.
- Title - The title of the publication is presented in English. For book chapters, the title is the title of the chapter/part of the book presented in the dataset.
- Publication Title - The name of the journal (for articles) or book (for chapters).
- ISBN - International Standard Book number. It is used only for book chapters, and the code is assigned directly to the book. Consists of thirteen digits, two codes are listed.
- ISSN - International Standard Serial number. It is used only for publications in the form of an article in a scientific journal. It consists of eight digits, two codes are listed (unique for each publication).
- DOI - Digital indicator of an object used for publication in the form of an article in a journal.
- URL - Unified resource index from the systems of unified addresses of digital resources. Internet link to the digital version of the document.
- Abstract - The abstract of the article is presented in English. It is worth noting that not all articles have an abstract, since the journal in which the work is published does not provide for an abstract.
- Date - The date of publication on the online publishing house's digital service. For the most part, due to simplification, this is the first day of the month of publication of the article.
- Pages - The range of page numbers of the issue of the journal in which this article is published. It is worth noting that in many cases the page numbers do not have a classic look: ..., but are presented in other formats that the publication in the international database has under its DOI.
- Issue - Number of issues in which the work was published.
- Volume - Volume of the journal in which this issue was published.
- Journal Abbreviation - Short abbreviation of the journal name, if present in the system.
- Short Title - Short title of the article, if available in an international database.
- Publisher - Publishing house where the book was published. Present only for chapters/parts of books.
- Place - City where the publishing house is located.
- Editor - List of editors of the book.

Link: https://doi.org/10.3897/popecon.8.e104663.suppl1

Supplementary material 2

Description of the dataset
The primary analytics of the publication dataset includes data on the distribution of publications by year, publishing houses, journals, and authors. The detailed data is provided in PDF format.

Link: https://doi.org/10.3897/popecon.8.e104663.suppl2
Supplementary material 3

Word frequency data
Data for all words that occur in a set of publications. The data is presented in CSV format (UTF-8 encoding).
Link: https://doi.org/10.3897/popecon.8.e104663.suppl3

Supplementary material 4

Blacklist of words
A set of words, introductory structures, letters and other lexical structures that occur in a set of publications with a high level of frequency, but are excluded from consideration in the analysis. The data is presented in CSV format (UTF-8 encoding). Description of table headers:
- Word - Word from a set of texts.
- Word length - Number of letters in a word.
- Frequency - Frequency of use of a word in a set of documents.
- Share - Percentage of word usage among all words in the set of documents (in %).
- Rank - Position of the word in frequency.
- Number of documents - The number of documents containing this code.
- Share of documents - Proportion of documents in which the word occurs.
Link: https://doi.org/10.3897/popecon.8.e104663.suppl4

Supplementary material 5

Codes
Data on the selected codes. The data is presented in CSV format (UTF-8 encoding).
Link: https://doi.org/10.3897/popecon.8.e104663.suppl5

Supplementary material 6

Matrix of codes
A summary table where each value indicates the number of publications in which the codes occur in the publication. Maximum value – 528. The data is presented in CSV format (UTF-8 encoding). Description of table headers:
- Color - Color used for the markup code in the documents.
- Code - Code used (in many codes there is no ending due to the need to take into account the forms of the word and words of the same root).
- Number of coded segments (of all documents) - Number of encoded segments in the document set.
- Encoded segments of all documents - The percentage of encoded segments with this code from the entire set of encoded segments in the document set.
- Number of documents - The number of documents containing this code.
Link: https://doi.org/10.3897/popecon.8.e104663.suppl6

Information about the author

- Klimenko German Andreevich - graduate student, Moscow State University, Moscow, 119991, Russia, german89000@mail.ru