

УДК 004.9:316.4

А.О. САВЕЛЬЕВ,

кандидат технических наук, доцент

Национальный исследовательский Томский политехнический университет

А.Ю. КАРПОВА,

доктор социологических наук, профессор

Национальный исследовательский Томский политехнический университет

С.А. КУЗНЕЦОВ,

инженер

Национальный исследовательский Томский политехнический университет

ПОДХОД К ОЦЕНКЕ ВЗАИМОСВЯЗЕЙ СООБЩЕСТВ СОЦИАЛЬНЫХ МЕДИА НА БАЗЕ ТЕМАТИЧЕСКОГО СХОДСТВА ТЕКСТОВОГО КОНТЕНТА (НА ПРИМЕРЕ СООБЩЕСТВ СОЦИАЛЬНО АКТИВНЫХ ОТЦОВ И СООБЩЕСТВ С ПРИЗНАКАМИ РАДИКАЛИЗАЦИИ)

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-011-00644 «Генеративное поведение российских мужчин в условиях демографического кризиса: Приволжский и Северо-Западный округа РФ (сравнительный анализ)». Руководитель гранта – Ильдарханова Ч.И., доктор социологических наук, профессор, директор Центра семьи и демографии Академии наук Республики Татарстан.

Аннотация. Автоматизация исследований онлайн-радикализации относится к числу актуальных задач современной социологии. С одной стороны, доступность интернет-данных, в том числе данных социальных сетей, предоставляет уникальные возможности с точки зрения полноты охвата исследуемых явлений и процессов. С другой стороны, несмотря на возросшую доступность инструментов Data Science, эффективность их использования зависит в равной степени как от соответствующих навыков, так и от компетентности в прикладной предметной области. В текущей работе представлен опыт междисциплинарной исследовательской группы в адаптации и применении инструментов автоматизированного сбора и предварительной обработки данных об онлайн-радикализации. Рассмотрен инструментарий оценки значимости слов в постах на основе статистической меры TF-IDF применительно к задаче оценки взаимосвязей сообществ социально активных отцов, аффилированных с организациями «Союз Отцов» и «Совет отцов», и сообществ с признаками радикализации, относящихся к различным идеологическим платформам. Исследование выполнено на базе открытых данных социальной сети «ВКонтакте». Сбор данных реализован с использованием API социальной сети. Сбор и обработка персональных данных не осуществлялись.

Результаты экспериментальной апробации показали наличие в отдельных случаях устойчивых взаимосвязей между рассматриваемыми группами сообществ. Результаты могут быть использованы в процессах поддержки принятия решений при планировании региональных политик в отношении семьи и детства, а также при планировании мероприятий по профилактике деструктивного информационно-психологического воздействия.

Ключевые слова: social media mining, онлайн-радикализация, TF-IDF, обработка естественного языка, тематическое сходство.

Введение

Использование современных интернет-технологий представителями радикальных идеологических платформ является сегодня одной из наиболее обсуждаемых проблем на уровне

политиков, органов безопасности, ученых и общества в целом. Ультрарадикальные сообщества применяют различные таргетинговые технологии для продвижения своих целей и осуществляют радикализацию онлайн. От-

слеживание наличия такого контента в социальных сетях в реальном времени важно для аналитиков безопасности. С точки зрения М. Hall, М.К. Logan, G.S. Ligon и D.C. Derrick, авторов работы “Do Machines Replicate Humans? Toward a Unified Understanding of Radicalizing Content on the Open Social Web”, легкость распространения деструктивного контента, а также разнообразие способов его доставки обеспечивают радикальным сообществам доступ к колоссальной аудитории, позволяя решать задачи поиска и вербовки сторонников, распространения идеологически значимой информации, привлечения ресурсов и общего управления [1]. При этом Интернет сам по себе не является причиной радикализации, а лишь инструментальной платформой, обеспечивающей недостижимые ранее плотность и скорость социальных коммуникаций. В качестве ответа на возникшие угрозы вместе с развитием и удешевлением использования методов интеллектуального анализа данных развивается междисциплинарное направление Social Media Mining. Исследования деструктивного контента, при помощи которого продвигаются радикальные идеи, являются примером области, в которой социальные медиаданные оказывают огромное влияние. Собранные данные социальных сетей могут быть проанализированы с целью изучения «цифровых следов», поведения пользователей социальных сетей, мотивации и технологий продвижения ультрарадикального контента.

Целью нашего исследования является экспериментальная апробация программно-алгоритмического инструментария для установления наличия и оценки силы связей между сообществами социальной сети, относящимися, с одной стороны, к сообществам социально активных отцов, с другой – к сообществам с признаками радикализации, принадлежащим к различным идеологическим платформам.

Исследования процессов онлайн-радикализации, включая задачи поиска целевых сообществ и контента, активно развиваются в последнее время. Однако отсутствие единой формализованной концепции радикализации не позволяют изучить в полной степени и установить четкие взаимосвязи между онлайн- и офлайн-тенденциями на уровне отдельных

пользователей, сообществ и групп. По типу решаемых задач современные исследования можно разделить на следующие категории:

1. Обнаружение целевых сообществ, противоправного и деструктивного контента [2–7], в том числе fake news, языка вражды и контента, распространяемого экстремистскими и террористическими организациями.

2. Анализ и интерпретация данных социальных медиа: статистический анализ распространения деструктивного контента, анализ сетевой топологии с целью поиска наиболее влиятельных узлов сети [8–11].

3. Формирование баз данных и знаний, описывающих значимые характеристики деструктивного контента и сообществ; формализация механизмов радикализации, позволяющая перейти к этапу алгоритмизации решения отдельных задач противодействия распространению деструктивного контента и изучения онлайн-радикализации [12–16].

Такие ограничения, как масштабируемость, неопределенность, парадокс больших данных [17–18] и смежные с ними, при отсутствии проработанных, компенсирующих их влияние механизмов значительно сказываются на общем качестве получаемых результатов и трудоемкости построения и сложности соответствующих программных решений.

В рамках настоящей работы предложен инструментарий оценки взаимосвязей сообществ социальных медиа, основанный на применении статистической меры значимости слов TF-IDF. Предложенный подход является альтернативой применению метрики «Общие подписчики», прикладная значимость которой ограничена, с одной стороны, потенциально неоднородным характером связи между пользователями и сообществами, а с другой – необходимостью очищать исходные данные от нерелевантных учетных записей, например, ботов.

Методология

Обобщенная методология исследования описывается следующей последовательностью действий:

1. Экспертный отбор двух групп сообществ социальной сети «ВКонтакте». Первая группа представлена сообществами, явно аффили-

рованными с организациями «Союз отцов» и «Совет отцов». Во второй группе представлены сообщества с признаками радикализации (например, призыв или оправдание насильственных действий в отношении иной социальной группы), принадлежащие к различным идеологическим платформам.

2. Автоматизированная оценка значимости используемых в постах сообществ слов:

- автоматизированное извлечение информации о 100 последних опубликованных постах;

- формирование единого корпуса текстов онлайн-сообществ и их предварительная обработка (нормализация слов, удаление знаков препинания, удаление стоп-слов);

- оценка значимости каждого слова отдельного поста на основе статистической меры TF-IDF;

- формирование перечня наиболее значимых слов для каждого из сообществ обеих групп.

3. Поиск пересекающихся значимых для различных сообществ слов (для оценки сходства обсуждаемых тем).

4. Автоматизированный поиск пересечений аудитории рассматриваемых онлайн-сообществ:

- извлечение информации о подписках каждого сообщества;

- поиск пересечений уникальных подписчиков;

- обобщение и оценка уровня пересечений.

5. Интерпретация полученных результатов.

На рисунке 1 представлен фрагмент социального графа – сетевой топологии исследуемых сообществ, построенной по принципу общих подписчиков. Узлы графа обозначают отдельные сообщества. Сообщества первой группы (социально активные отцы) отмечены вертикальной штриховкой, сообщества второй группы (с признаками радикализации) – сплошной штриховкой. Всего два сообщества первой группы имеют значимое число общих подписчиков с сообществами с признаками радикализации, представленными одной идеологической платформой – MGTOW (men going their own way). В самом общем смысле объединяющим фактором для MGTOW является идея «враждебного сексизма», связанная с их убеждением в наличии угроз мужчинам со стороны женщин, даже если угрозы эти в основном косвенные и политические, а не явные и физические.

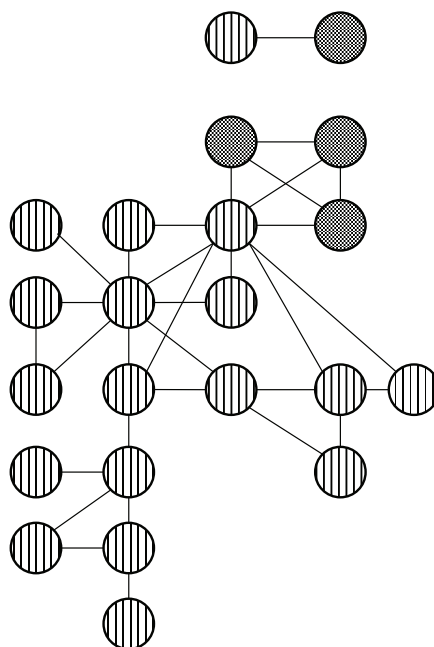


Рис. 1. Фрагмент графа взаимосвязей сообществ, выявленных на основе общности подписчиков

Оценка тематического сходства сообществ осуществляется на основе применения метрики статистической значимости слов TF-IDF. TF (term frequency) – это частотность термина, которая измеряет, насколько часто термин встречается в документе. IDF (inverse document frequency) – инверсия частоты, с которой некоторое слово встречается в документах коллекции. TF-IDF эффективно работает только на больших корпусах текста, без чего истинное значение IDF слова вычислить нельзя. В своем исследовании мы придерживались рекомендации использовать корпус документов, не более чем на 10 % состоящий из документов интересующих нас сообществ, а остальное – это набор разнообразных документов, который позволил бы алгоритму определить статистику использования слов в русском языке в целом. Для составления 90 % корпуса использовались датасеты из lenta.ru, сообщения в Twitter, публикации на политические темы в «ВКонтакте».

Посты каждого из сообществ за один календарный год были объединены в единый документ. Для всех слов каждого из документов была определена мера значимости.

Автоматизированная оценка значимости используемых в постах сообществ слов выполнена следующим образом. Все извлеченные

посты каждого из сообществ были объединены в единый корпус. Для всех слов каждого из постов была определена мера значимости слова для поста с использованием статистической меры TF-IDF:

$$\text{weight}(w, d) = \frac{n_w}{\sum_k n_k} \times \log \frac{|D|}{|\{d_i \in D \mid w \in d_i\}|}, \quad (1)$$

где n_w – число вхождений слова w в документ; $\sum_k n_k$ – общее число слов в документе; w – слово в документе; d – документ; $|D|$ – корпус всех документов; $|\{d_i \in D \mid w \in d_i\}|$ – число документов в корпусе D , где встречается w .

Перечень наиболее значимых слов для документа вычисляется на основе абсолютной значимости:

$$\text{weight}_{abs}(w, C) = \frac{\sum_1^n (\text{weight}(w, d_i))}{n_d}, \quad (2)$$

где C – корпус постов радикального сообщества ($C \in P$); n_d – количество постов в C , где встречается w .

Визуализация результатов оценки тематического сходства рассматриваемых сообществ представлена на рис. 2.

Оценка тематического сходства рассматриваемых сообществ позволила выявить в отдельных случаях устойчивые связи сообществ первой группы с сообществами, относящимся к таким радикальным идеологическим платформам как MGTOW, националисты и неоязычники.

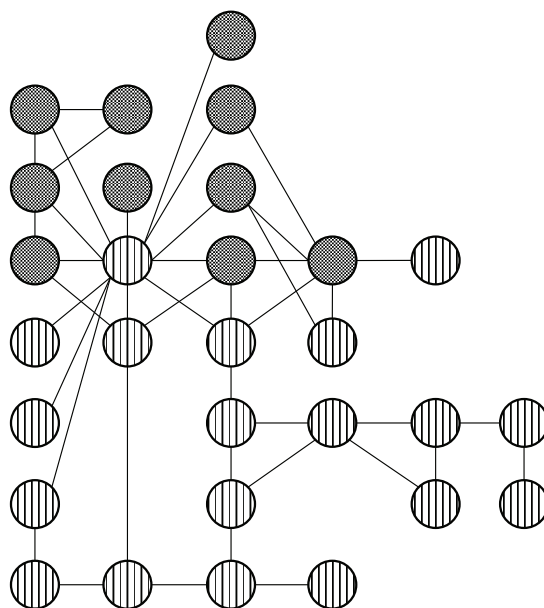


Рис. 2. Фрагмент графа взаимосвязей сообществ, выявленных на основе тематического сходства

Результаты

1. Извлечение новых знаний из сетевой топологии, сформированной на базе общих подписчиков, является затруднительным. Полученный результат очевиден: учетные записи пользователей с высокой долей вероятности привязаны к нескольким однотипным сообществам и образуют тем самым плотные однородные кластеры сообществ. По сути именно такой часто встречаемый в российском исследовательском поле подход влечет за собой не получение нового знания о сообществах и механизмах онлайн-радикализации, а, наоборот, искажает картину, фактически «запирая» пользователей в эхо-камеры. Следовательно, выводы из такого рода данных заведомо искажают реальное положение исследуемых объектов.

2. Мера сходства «Общность подписчиков» также требует дополнительной предварительной обработки, поскольку общие подписчики могут быть у сообществ, не относящихся к целевым (в нашем случае – с признаками радикализации). Например, сообщества средств массовой информации, банков, территориальных образований должны быть исключены из анализируемой выборки, так как объединяют подписчиков, которые могут также составлять значимую аудиторию целевых сообществ.

3. Следует отметить, что TF-IDF-мера выявляет значимые слова в текстах сообщений. Таким образом, граф тематического сходства показывает фактическое пересечение используемых в текстах ключевых слов. Сетевая топология TF-IDF заметно отличается от топологии «Общие подписчики». Обнаружено большее по сравнению с общностью подписчиков число взаимосвязей сообществ активных отцов и сообществ с признаками радикализации, в том числе относящихся к различным идеологическим платформам.

4. Оценка сходства на базе TF-IDF позволяет отразить связь между сообществами, выявленную через общность обсуждаемых в постах значимых тем. Применительно к задаче изучения онлайн-радикализации TF-IDF способен расширять знание о ее механизмах в части влияния радикальных сообществ одной идеологической платформы на сообщества других

радикальных платформ через обсуждение, продвижение и пропаганду значимой тематической повестки, что также является инструментом рекрутинга.

Заключение

Использование метрик семантического сходства текстового контента, в том числе векторизации на базе статистической меры TF-IDF, позволяет обнаруживать неявные взаимосвязи сообществ социальных медиа и является эффективным инструментом поиска целевых сообществ, применимым в сетях со слабыми фактическими связями, выраженными через общность подписчиков, но со значительным уровнем гомофилии – тенденции схожих сообществ и индивидов к образованию связей.

Применимость подхода к оценке взаимосвязей сообществ социальных медиа на основе тематического сходства контента апробирована на примере задачи установления наличия связей между сообществами социально активных отцов и сообществ с признаками радикализации. Результаты апробации оформлены в виде базы данных и зарегистрированы [19]. Выявлены устойчивые связи между отдельными сообществами социально активных отцов и различными радикальными идеологическими платформами. Количество выявленных связей не позволяет сделать вывод об общей тенденции к радикализации сообществ социально активных отцов в настоящий момент, при этом присутствует ярко выраженный соответствующий риск. В силу наличия плотных связей сообществ «отцов» внутри своей группы распространение и поддержка радикальных идей в отдельных из них потенциально может привести как к дискредитации всего движения в целом, так и к общей трансформации социальных практик ответственного отцовства в девиантные формы социальной активности. Указанный риск должен учитываться при проработке мер поддержки семьи и детства, а также при реализации профилактических мероприятий в области противодействия распространению деструктивного информационно-психологического воздействия.

Литература

1. Hall M., Logan M.K., Ligon G.S., Derrick D.C. Do Machines Replicate Humans? Toward a Unified Understanding of Radicalizing Content on the Open Social Web // *Policy & Internet*. – 2019. – Vol. 12. – Is. 3. – P. 109–138.
2. Grover T., Mark G. Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit // *Proceedings of the International AAAI Conference on Web and Social Media*. – 2019. – Vol. 13. – Is. 01. – P. 193–204.
3. Deem A. The Digital Traces of #whitegenocide and Alt-Right Affective Economies of Transgression // *International Journal of Communication*. – 2019. – No. 13. – P. 3183–3202.
4. Lara-Cabrera R., Pardo A.G., Benouaret K., Faci N., Benslimane D., Camacho D. Measuring the Radicalisation Risk in Social Networks // *IEEE Access*. – 2017. – No. 5. – P. 10892–10900.
5. Alvari H., Sarkar S., Shakarian P. Detection of Violent Extremists in Social Media // *Proceedings of the 2nd International Conference on Data Intelligence and Security (ICDIS)*. – 2019. – P. 43–47.
6. Mussiraliyeva S., Bolatbek M., Omarov B., Medetbek Z., Baispay G., Ospanov R. On Detecting Online Radicalization and Extremism Using Natural Language Processing // *Proceedings of the 21st International Arab Conference on Information Technology (ACIT)*. – 2020. – P. 1–5.
7. Lee D.H., Kim Y.R., Kim H.J., Park S.M., Yang Y.J. Fake News Detection Using Deep Learning // *Journal of Information Processing Systems*. – 2019. – Vol. 15. – Is. 5. – P. 1190–1130.
8. Zareie A., Sheikahmadi A., Jalili M., Fasaie M.S.K. Finding influential nodes in social networks based on neighborhood correlation coefficient // *Knowledge-Based Systems*. – 2020. – No. 194. – P. 105580.
9. Tsapatsoulis N., Djouvas C. Opinion Mining From Social Media Short Texts: Does Collective Intelligence Beat Deep Learning? // *Frontiers in robotics and AI*. – 2019. – No. 5. – P. 138.
10. Rowe M., Saif H. Mining Pro-ISIS Radicalisation Signals from Social Media Users // *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM 2016)*. – 2016. – P. 329–338.
11. Francisco M., Castro J.L. A fuzzy model to enhance user profiles in microblogging sites using deep relations // *Fuzzy Sets and Systems*. – 2020. – No. 401. – P. 133–149.
12. Borum R. Understanding the terrorist mind-set // *FBI Law Enforcement Bulletin*. – 2003. – No. 72. – P. 7–11.
13. Hamm M., Spaaj R. Lone Wolf Terrorism in America: Using Knowledge of Radicalization Pathways to Forge Prevention Strategies: textbook. – Washington: The United States Department of Justice, 2015. – 28 p.
14. McCauley C., Moskalenko S. Mechanisms of Political Radicalization: Pathways toward Terrorism // *Terrorism and Political Violence*. – 2008. – Vol. 20. – Is. 3. – P. 415–433.
15. Thompson R. Radicalization and the Use of Social Media // *Journal of Strategic Security*. – 2011. – No. 4. – P. 167–189.
16. Borum R. Radicalization into Violent Extremism I: A Review of Social // *Journal of Strategic Security*. – 2011. – No. 4. – P. 7–35.
17. Tang L., Liu H. Community Detection and Mining in Social Media: textbook. – San Rafael: Morgan & Claypool Publishers, 2010. – 138 p.
18. Zafarani R., Abbasi M.A., Liu H. Social Media Mining: An Introduction: textbook. – Cambridge: Cambridge University Press, 2014. – 332 p.
19. Ильдарханова Ч.И., Савельев А.О., Вильнин А.Д., Карпова А.Ю., Кайда А.Ю., Кузнецов С.А., Ширькалов А.М., Максимова Н.Г. Социологическое исследование «Сетевые взаимодействия социально активных отцов: норма и отклонение (анализ социальных сетей профильных НКО)» // База данных социологического исследования (свидетельство о государственной регистрации базы данных RU 2021621866 от 03.09.2021).

Информация об авторах

Савельев Алексей Олегович, кандидат технических наук, доцент, Национальный исследовательский Томский политехнический университет.

E-mail: sava@tpu.ru

Карпова Анна Юрьевна, доктор социологических наук, профессор, Национальный исследовательский Томский политехнический университет.

E-mail: belts@tpu.ru

Кузнецов Сергей Анатольевич, инженер, Национальный исследовательский Томский политехнический университет.

E-mail: ksall@tpu.ru

A.O. SAVELEV,
PhD in Technical Sciences, Associate Professor,
National Research Tomsk Polytechnic University

A.YU. KARPOVA,
Doctor in Social Science, Professor,
National Research Tomsk Polytechnic University

S.A. KUZNETSOV,
Engineer,
National Research Tomsk Polytechnic University

**APPROACH TO ASSESSMENT OF SOCIAL MEDIA COMMUNITIES RELATIONSHIPS BASED
ON TEXT CONTENT THEMATIC SIMILARITY (ON THE EXAMPLE OF SOCIALLY ACTIVE
FATHERS COMMUNITIES AND COMMUNITIES WITH SIGNS OF RADICALIZATION)**

Abstract. Automation of research on online radicalization is one of the urgent tasks of modern sociology. On the one hand, the availability of Internet data, including data from social networks, provides unique opportunities for the completeness of coverage of the phenomena and processes under study. On the other hand, despite the increased availability of Data Science tools, their use depends equally on both the relevant skills and competence in the applied subject area. The current work presents the experience of an interdisciplinary research group in adapting and applying tools for the automated collection and preprocessing of data on online radicalization. A toolkit for assessing the importance of words in posts based on the TF-IDF statistical measure is considered concerning the problem of assessing the relationships between communities of socially active fathers affiliated with the “Union of Fathers” and “Council of Fathers” organizations and communities with signs of radicalization belonging to various ideological platforms. The research was carried out based on open data of the social network “VKontakte”. Data collection is implemented using the social network API. The collection and processing of personal data were not carried out. The results of experimental testing showed the presence in some cases of stable relationships between the considered groups of communities. The results can be used in decision support processes when planning regional policies in relation to family and childhood and when planning measures to prevent destructive information and psychological impact.

Keywords: social media mining, online radicalization, tf-idf, natural language processing, thematic similarity.

References

1. Hall M., Logan M.K., Ligon G.S., Derrick D.C. Do Machines Replicate Humans? Toward a Unified Understanding of Radicalizing Content on the Open Social Web // Policy & Internet. – 2019. – No. 12(3). – P. 109–138.
2. Grover T., Mark G. Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit // Proceedings of the International AAAI Conference on Web and Social Media. – 2019. – No. 13(01). – P. 193–204.
3. Deem A. The Digital Traces of #whitegenocide and Alt-Right Affective Economies of Transgression // International Journal of Communication. – 2019. – No. 13. – P. 3183–3202.
4. Lara-Cabrera R., Pardo A.G., Benouaret K., Faci N., Benslimane D., Camacho D. Measuring the Radicalization Risk in Social Networks // IEEE Access. – 2017. – No. 5. – P. 10892–10900.
5. Alvari H., Sarkar S., Shakarian P. Detection of Violent Extremists in Social Media // 2nd International Conference on Data Intelligence and Security (ICDIS). – 2019. – P. 43–47.
6. Mussiraliyeva S., Bolatbek M., Omarov B., Medetbek Z., Baispay G., Ospanov R. On Detecting Online Radicalization and Extremism Using Natural Language Processing // 21st International Arab Conference on Information Technology (ACIT). – 2020. – P. 1–5.
7. Lee D.-H., Kim Y.-R., Kim H.-J., Park S.-M., Yang Y.-J. Fake News Detection Using Deep Learning // Journal of Information Processing Systems. – 2019. – No. 15(5). – P. 1190–1130.
8. Zareie A., Sheikahmadi A., Jalili M., Fasaie M.S.K. Finding influential nodes in social networks based on neighborhood correlation coefficient // Knowledge-Based Systems. – 2020. – No. 194.

9. Tsapatsoulis N., Djouvas C. Opinion Mining From Social Media Short Texts: Does Collective Intelligence Beat Deep Learning? // *Front. Robot. AI*. – 2019. – No. 5:138.
10. Rowe M., Saif H. Mining Pro-ISIS Radicalisation Signals from Social Media Users // *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*. – 2016. – P. 329–338.
11. Francisco M., Castro J.L. A fuzzy model to enhance user profiles in microblogging sites using deep relations // *Fuzzy Sets and Systems*. – 2020. – No. 401. – P. 133–149.
12. Borum R. Understanding the terrorist mind-set // *FBI Law Enforcement Bulletin*. – 2003. – No. 72. – P. 7–11.
13. Hamm M., Spaaj R. Lone Wolf Terrorism in America: Using Knowledge of Radicalization Pathways To Forge Prevention Strategies textbook. – U.S. Department of Justice, 2015. – 28 p.
14. McCauley C., Moskalenko S. Mechanisms of Political Radicalization: Pathways Toward Terrorism // *Terrorism and Political Violence*. – 2008. – No. 20:3. – P. 415–433.
15. Thompson R. Radicalization and the Use of Social Media // *Journal of Strategic Security*. – 2011. – No. 4. – P. 167–189.
16. Borum R. Radicalization into Violent Extremism I: A Review of Social // *Journal of Strategic Security*. – 2011. – No. 4. – P. 7–35.
17. Tang L., Liu H. Community Detection and Mining in Social Media: textbook. – Morgan & Claypool Publishers, 2010. – 138 p.
18. Zafarani R., Abbasi M.A., Liu H. Social Media Mining: An Introduction: textbook. – Cambridge University Press, 2014. – 332 p.
19. Ildarkhanova Ch.I., Savelev A.O., Vilnin A.D., Karpova A.Yu., Kaida A.Yu., Kuznetsov S.A., Shirykalov A.M., Maksimova N.G. Sociological research “Network interactions of socially active fathers: norm and deviation (analysis of social networks of specialized non-profit organizations)” // Certificate of state registration of the database 2021621866 dated 03.09.2021.