***An approach to identifying social protest potential based on a city-to-city similarity metric***

***Introduction.*** Nowadays, the evaluation of city-to-city similarity is a relevant approach that can be applied to various tasks, including infrastructural, architectural, and other areas [1, 2, 3, 4]. Interest in this research direction is mainly driven by the possibility of using similarity assessment to address urban planning problems [4]. At the same time, social conflicts that arise during the implementation of construction projects are of great importance in the field of urban planning [5, 6]. Therefore, forecasting social risks (such as protests and conflicts) is a relevant task for urban planning. Undoubtedly, this problem is even more relevant for socio-economic policymaking [7–11].

Motives and factors of modern social protests are under careful investigation due to their significant impact on the socio-political situation, which, in turn, is closely connected with socio-economic processes [7–11]. The relevance of this research area is beyond doubt, since certain mechanisms of protests remain debatable or even unknown [7, 8]. The growing number of recent studies on this topic also serves as evidence of strong interest.

The main goal of this research is to develop a method for assessing the potential for social conflict in a given municipality (city, town, or village). The core of the approach is estimating the similarity not only of the socio-economic environment, but also the demographic structure (human capital) to characteristics that could show the predisposition to certain types of social conflict. In other words, similarity between two municipalities, where one of which has experienced a particular social conflict, can serve as an indicator of a likely comparable social response to the same type of stimulus.

***Related works.*** There are various approaches to investigating social conflicts. So far, the most common approach has been to examine the problem within the context of sociological and psychological research [7, 8]. However, it should be noted that forecasts based only on sociological data do not demonstrate high accuracy [7].

Also one of the common approaches is based on mathematical modeling of collective behavior [8]. It mainly involves the development of various agent-based systems that allow the analysis of protest dynamics through imitation processes. For example, an epidemiological model is often used as a representation of protest movements [8].

In recent years, machine learning approaches have also gained popularity in this field of research [9, 10, 11]. These studies are generally based on analyzing public sentiment through the monitoring of social networks and mass media [9, 10, 11].

In study [9], protests are investigated through the analysis of the stylistic features of mass media articles covering labor protests that took place in Romania (2018–2025). Within the context of the present research, particular interest lies in the evidence supporting the spatio-temporal hypothesis of protest diffusion, meaning that a protest in one location increases the occurrences of protests in neighboring areas [9]. This conclusion resonates with the effectiveness of the epidemiological approach to analyzing protest dynamics. However, the present research proposes a method that considers at least the “spatial” component in terms of the socio-economic and demographic space, which is often influenced by geographical closeness but is not determined by location alone.

The study [10] used social network data to forecast civil unrest events. A modified neural network was proposed, trained on Twitter data (now X.com) using the example of the 2019 protests in Hong Kong. The effectiveness of this approach was demonstrated by comparing it with various other machine learning algorithms [10].

In [11], protests in Latin America (2012–2014) were examined using a logistic regression model that was trained on data not only from the social network Twitter (now X.com), but also from other sources, including news feeds, political event databases, the TOR network, and economic indicators. It is important to note that only currency exchange rates were used as the economic indicator [11]. While this is a valuable factor, it is not sufficient to capture the full socio-economic condition of a territory within a country.

***Data.*** The analyzed dataset includes socio-economic and demographic data for over 1800 municipal entities in Russia (districts, cities, and villages) for the period 2014 to 2022, collected from the Federal State Statistics Service (ROSSTAT). Most records (around 80%) correspond to municipal districts, as separate data for small cities and villages are generally unavailable in open access. In total, the dataset includes more than 9000 examples, since each municipality–year combination is treated as an independent observation.

Table 1 presents the selected socio-economic indicators. These factors were chosen based on their demonstrated effectiveness in migration forecasting study []. Therefore, they provide a reliable basis for assessing the socio-economic conditions of a municipality.

1. Average number of employers in organizations - avgemployers (ppl.)

2. Average salary - avgsalary (rub.)

3. Shopping area - shoparea (sq.m.)

4. Number of seats in café, bars and restaurants - foodseats (num.)

5. Retail turnover - retailturnover (thnd. rub.)

6. Live area per capita - livarea (sq.m.)

7. Number of sporting venues - sportsvenue (num.)

8. Number of services (barbershops, repairs) - servicesnum (num.)

9. Length of roads - roadslen (km.)

10. Number of livestock - livestock (num.)

11. Productivity of land (vegetables) - harvest (centners)

12. Agricultural production - agrprod (thnd. rub.)

13. Number of healthcare organizations - hospitals (num.)

14. Number of places in preschool organizations - preschool (num.)

15. Volume of self-produced goods - factoriescap (thnd. rub.)

All socio-economic indicators (except average salary) were normalized on a per capita basis to enable fairer similarity evaluation across municipalities and to provide a better assessment in terms of population well-being (see Table 1).

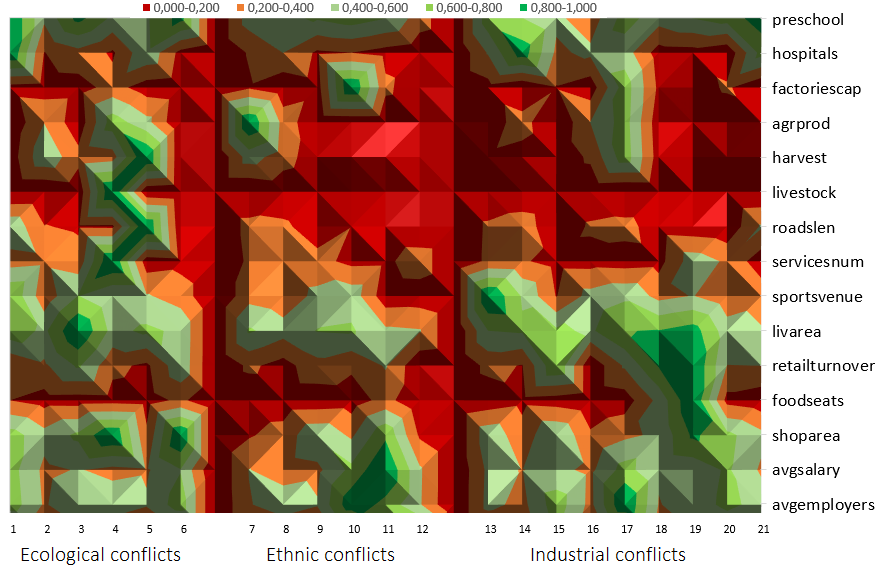
To represent the population structure of the municipalities, demographic data were collected for 14 cohorts (from ‘0–4’ to ‘65–69’), disaggregated by gender for each year. Unfortunately, data for cohorts older than ‘65–69’ were missing for many municipalities. It is important to note that the demographic data were also normalized as the proportion of each cohort (age group) in the total population of a given municipality, which makes it possible to evaluate human capital in terms of structure rather than quantity.

***Set of Social Conflicts.*** The socio-economic and demographic characteristics of a municipality at the time it experiences a protest could be used to identify a predisposition to certain social conflicts or general social tension. Thus, the set of municipalities (or conflict cases) against which similarity is assessed forms one of the crucial components of the proposed approach.

A set of the most notable social conflicts in Russia from 2013 to 2023 has been collected. It includes 21 cases of social protests in different municipalities, each involving mass gatherings of people (over 100 participants) and receiving wide media coverage. These cases can be grouped into three categories: six ecological conflicts (Shies 2018, Poltavskaya village 2022, etc.); six ethnic conflicts (Chemodanovka 2019, Makhachkala 2023, etc.); and nine industrial conflicts (Kushtau 2020, Anapa 2021, etc.).

The focus on high-profile conflicts is motivated by their potential to highlight the socio-economic and demographic indicators that predispose municipalities to social tension. In other words, these cases can be viewed as “local extrema” of an abstract function representing social conflict.

During the comparison of the socio-economic conditions of these local extrema (conflict cases), certain trends can be observed. Figure 1 shows that in the entertainment (“servicesnum”, “sportsvenue”, “retailturnover”, “foodseats”) and in the production sectors (“factoriescap”, “agrprod”, “harvest”, “livestock”) municipalities that experienced ethnic conflicts exhibit the lowest values (many red and few green zones). This indicates that municipalities with ethnic conflicts were in a more depressed socio-economic environment than those experiencing industrial or ecological conflicts. In particular, municipalities with ecological conflicts appear to be the most balanced (see Fig. 1).



Some trends and dependencies were also identified in the analysis of the demographic structure of these municipalities. Figure 2 illustrates certain differences between categories of conflicts. For example, municipalities that experienced ethnic conflicts have the highest proportion of young people in the population (cohorts “20–24,” “25–29,” and “30–34”; see Fig. 2). In the “25–29” cohort, the proportion of both males and females is nearly double that in municipalities with ecological conflicts.

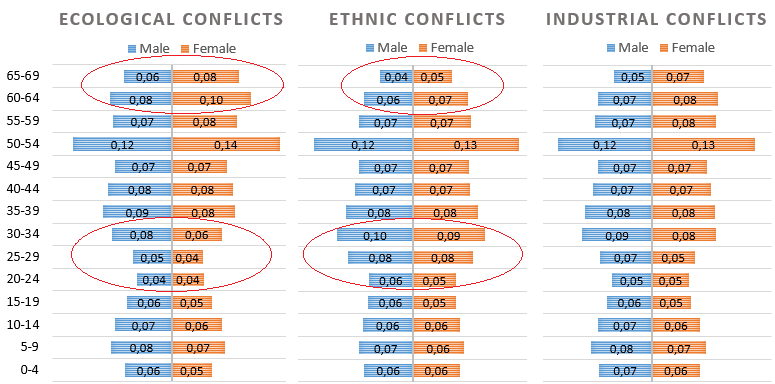


Рис. 2. Усредненные значения демографических структур МО по каждому типу конфликта

Municipalities with ecological conflicts show the lowest proportion of young people, but at the same time have the highest proportion of older population (cohorts “60–64” and “65–69”; see Fig. 2). In contrast, the demographic structure of municipalities with ethnic conflicts demonstrates the lowest proportion of people in these cohorts. For example, compared to ecological conflicts, in the “60–64” cohort municipalities with ethnic conflicts have about 25% fewer men and 30% fewer women, while in the “65–69” cohort the difference is even greater – about 33% fewer men and nearly 38% fewer women. The population structure of municipalities with industrial conflicts falls somewhere in between.

Thus, notable differences between municipalities with certain types of conflicts can be observed both in socio-economic conditions and in demographic structure. This provides strong evidence for the potential use of these similarities as a measure of the risk of social conflict.

***Similarity metric.*** Each municipal entity (municipal district, city, village etc.) in the data defined by the socio-economic and demographic components. The socio-economy condition could be represented as a tuple:

where is the value of the specific socio-economic indicator (see Section 3, Table 1)

The demographic structure can be represented by two separate tuples, indicating the proportions of females and males in the population:

here is the number of females in a given cohort, is the total number of females in the population and *k* is the number of cohorts. The division enables calculation of the proportion of each cohort within the population, allowing assessment of the demographic structure. The same approach is used to represent the male population structure:

where is the number of male in the cohort, is the total number of male in the population.

For evaluating the similarity between two municipal entities x and y, it is proposed to use the combined sum of the mean squared deviations of socio-economic and demographic indicators:

where – is the value of socio-economic feature of municipality x*, n* is the number of socio-economic features, and is the proportion of female/male in the certain cohort of municipality x*, k* isthe number of cohorts and is the weighting coefficient that could balance the socio-economic and demographic component.

First, it is necessary to evaluate the applicability of the similarity metric (4). This involves estimating the similarity of a given municipality to all other municipal entities in the dataset. Table illustrates the case of Michurinsk City in 2022. As expected, in the majority of cases, the proposed metric (4) identifies the same municipality observed in different time periods as the most similar to a given municipality. This confirms the reliability of the similarity metric. It should be mentioned that in the case of Michurinsk City the set of top similar municipalities includes not only its geographical neighbor Morshansk City (both located in Tambov Oblast), but also the geographically distant Vyshnevolotsky Dist. (Tver Oblast), demonstrating similarity despite the lack of spatial closeness.

The next step is to examine the potential of using the similarity metric in the context of the main idea, that a municipality similar to another, which has experienced a specific type of “social turbulence”, may demonstrate a comparable social response to a more or less similar stimulus.

As an example, the similarity metric is applied to identify the most similar municipal entities to Krasnoarmeysky Dist., where Poltavskaya Village experienced large-scale protests against a landfill in 2022 [link to media sources???]. The results of the similarity metric evaluation are shown in Figure 4 (table in the image), where all municipal entities are ranked according to their similarity (4) to Krasnoarmeysky District. The metric values are normalized to a range from 0 to 1.

The Figure 4 (in the right) also shows a graph that illustrates all values of metric (4) from the table (in the left). The graph demonstrates that the similarity value from 0 to 0.1 is reached by approximately 300-th example, while the range from 0.1 to 0.2 is crossed only around ~5000-th example. This result shows that certain homogeneity is observed in Russia, at least in terms of per capita characteristics. In other words, there are a large number of municipal entities that do not differ significantly from each other, which is generally normal within a single country.

***Reference***

[1] A Similarity Approach to Cities and Features. 2022. URL: <https://www.researchgate.net/publication/358646626_A_Similarity_Approach_to_Cities_and_Features>

[2] Analysis of the Uniqueness and Similarity of City Landscapes Based on Deep Style Learning. 2021. URL: <https://doi.org/10.3390/ijgi10110734>

[3] Exploring venue-based city-to-city similarity measures. 2013. URL: <https://www.researchgate.net/publication/257212810_Exploring_venue-based_city-to-city_similarity_measures>

[4] The scenario method in urban planning. 2022. URL: <https://www.researchgate.net/publication/276321581_The_scenario_method_in_urban_planning>

[5] Are land use conflicts a barrier to sustainable city development? Evidence from Chattogram District of Bangladesh. 2025. URL: <https://www.researchgate.net/publication/388926768_Are_land_use_conflicts_a_barrier_to_sustainable_city_development_Evidence_from_Chattogram_District_of_Bangladesh>

[6] Conflicts in urban peripheries in Europe. 2023. URL: <https://www.researchgate.net/publication/373236871_Conflicts_in_urban_peripheries_in_Europe>

[7] Davydov D. A. (2022) Dynamics of Mass Protest Actions in Modern Russia: An Event Study. Monitoring of Public Opinion: Economic and Social Changes. No. 5. P. 72–93. https:// doi.org/10.14515/monitoring.2022.5.2199 (In Russ.)

[8] Petrovskii S, Shishlenin M, GlukhovA (2025) Understanding street protests: from amathematical model to protest management. PLoS ONE 20(4): e0319837. <https://doi.org/10.1371/journal.pone.0319837>

[9] Nicula, Alexandru-Sabin and Cretan, Remus and Simionescu, Mihaela and Oancea, Bogdan and Dragan, Alexandru, The City as Stage: Labour Protest, Sentiment, and Machine Learning. DOI: <http://dx.doi.org/10.2139/ssrn.5357390>

[10] Iyda, J.J., Geetha, P. An improved deep belief neural network based civil unrest event forecasting in twitter. Appl Intell 53, 5714–5731 (2023). https://doi.org/10.1007/s10489-022-03746-3

[11] Korkmaz, G., Cadena, J., Kuhlman, C.J. et al. Multi-source models for civil unrest forecasting. Soc. Netw. Anal. Min. 6, 50 (2016). https://doi.org/10.1007/s13278-016-0355-8