ORIGINAL PAPER

SN

Check for updates

Social heterogeneity in urban India: a case study on five selected metropolitan cities

Arghya Saha[1] (ORCID) **· Sudipto Kumar Goswami[2] · Swarnali Saha[3]**

Abstract

Economic, social, ethnic, racial, cultural, linguistic, occupational diversity is an inseparable feature of the city. For several years, urban sociologists recognized heterogeneity as one of the city's defining and important characteristics. Heterogeneity in population composition is connected with globalization, so increased connectivity between people and territories is the goal of this process. Globalization also makes a closer cultural, economic and political tie between different regions. Cities act as magnets for the flow of people by their capital, goods and information. This paper considers three aspects of city's heterogeneity, namely ethnic, linguistic and religious characteristics of five metropolitan cities in India. The calculation regarding diversity is computed by Shannon's Wiener Diversity Index, Herfindahl–Hirschman Index, Simpson's Diversity Index, Pielou Index and Czekanowski coefficient for similarities between places. Using city level data from Census 2011, detailed description which deals with much smaller data components is absent. Finally, the article finds out that ethnicity, religion and language are significant predictor of cultural diversity in these five Indian metropolitan cities despite they constitute a homogenous and economically advantaged groups.

✉ Arghya Saha
arghyasaha21@gmail.com

Swarnali Saha
swarnalisaha025@gmail.com

[1] Department of Geography, Bidhan Chandra College, Hooghly 712248, India

[2] Visva Bharati University, Santiniketan, India

[3] Department of Geography, Mahitosh Nandy Mahavidyalaya, Hooghly 712404, India

Introduction

A heterogeneous group of objects is the polar opposite of a homogeneous group of objects, which is defined as "a group of units or samples fundamentally similar". A homogeneous society is a group of people with identical properties, consisting of a similar composition in the society. Heterogeneity matters because discovering strong regularities among classes of phenomena is one of the central goals of positivist science. Heterogeneity in societies is of many types, the first point to consider is the diversity of social causes and influences. The second factor is heterogeneity within social groups such as cities, religions, electoral democracies, and social movements. The third point to consider is the heterogeneity that can be found both across and within social groups. Social motives, emotions, moral frameworks and modes of agency are all examples of a fourth type of heterogeneity. Instead of seeking legal regulations for whole categories of events and things, heterogeneity suggests an explanatory strategy. The investigation is about finding the fact of heterogeneity and about different processes and things involved. Social diversity is a characteristic quality of most urban spaces. According to Sennett (1992), 'differences are an overwhelming sociological fact' of cities. The concept adopted in this study relates to social heterogeneity and includes many forms of differences between individuals within the same culture. From early accounts, the issue of social diversity has been apparent. With their human ecological approach, the Chicago School focused their work on the relationship between social differences and functional divisions in space. (Tonkiss 2005). According to some academics, diversity can promote tolerance, creativity, and make city officials aware of lifestyles that would otherwise go unnoticed. (Sennett 1996; Florida 2002; Fainstein 2005). The counter-argument contends that diversity is overstated because, despite this, there is a lack of interaction between diverse individuals and groups (Wessel 2009). In the analysis of social heterogeneity, structural stability over space is unrealistic. Many scholars replaced structural stability with regression models with spatial structural change. A focus on local patterns of association for local instabilities has been more appropriate perspective. A small approach to urban inequality is a starting point for additional research into its impact on urban quality. (Lelo et al. 2018). This approach is further complementing with downscaling method, which would facilitate decision-making for spatial land use (Sikder et al. 2017).

Soviet ethnographers conducted ethno-linguistic research in the 1960s and published Atlas Narodov Mira. Researchers in the Soviet Union primarily used language to define groups, but they occasionally included groups that appeared to be differentiated by race rather than language. Most empirical works on ethnic diversity are the works of Easterly and Levine's work on economic growth (Easterly and Levine 1997). Following this, researchers include an ethnic diversity measure in their cross-country growth regressions (Alesina et al. 2003; Brock and Durlauf 2001; Collier and Gunning 1999; Easterly 2002; Englebert 2000; Hall and Jones 1999; Rodrik 1999). Until recently, most of the measures of ethnic fractionalization called Ethno-Linguistic Fractionalization. The data compiled by

the Soviet team of ethnographers in the early 1960s and published in the Atlas Narodov Mira are calculated using the Herfindahl concentration index (1964). Another work on linguistic and racial diversity is a list of "minorities at risk" in 115 countries (Gurr 1996). Alesina et al. (2003) attempt to differentiate between ethnic, linguistic, and religious groups in a sample of approximately 190 countries, and then use the list to create measures of ethnic, linguistic, and religious fractionalization. (Alesina et al. 2003). In India, localities were founded on the identities like exclusion and othering despite identity is an unstable relation of difference (Gupta and Ferguson 1997). This phenomenon describes the mechanism by which identity is constantly shaped and reshaped whereas socio spatial forms have affected identity formation. Early Indian cities replicated the spatial form of the village in terms of caste segregation and working specialities. (Rowe 1973). Socio spatial changes in urban forms were happened on twentieth century developments, like emerging housing market, shifts in the traditional economy into service-based economy, diversification out of caste based or family-based occupations, etc. Large metropolitan cities, which were selected for this study were the first cities that experienced the change. However, there was considerable urban migration across India following the end of British colonial control in 1947 and the subsequent partition of India and Pakistan (Evenson 1989). Millions of people moved to urban areas, in the big cities like Delhi and Kolkata (Davis 2007). In this period, slum settlements are claimed to be growing at 250 times faster than the overall population growth rate (Mehta 2004). People shapes heir social interaction and social networks, health outcomes and sense of self. The residential locations of individuals and groups reflect the hierarchies of advantage in a society. Gupta et al (2009) found that historically disadvantaged castes disproportionately live in slums in eight Indian cities. Previous quantitative studies in Pune and Delhi were based on segregation by socio-economic status, religion and caste (Dupont 2004; Mehta 1968, 1969).

The aim of this article is to finds out how the pattern of social heterogeneity is constructed in five selected metropolises, based on ethnic, linguistic and religious diversity using four indicators. The research question finds out the nature of social heterogeneity and associated phenomena in five cities. The argument is divided into two parts. First, a short review of the article on Indian urban system and its changing pattern which affects social heterogeneity. It will be demonstrated in the next section that how globalization reshapes the economy and society, especially in urban context. The second section deals with the same problem in a different light, looking at the results from the analysis, finds out a pattern which eventually supports the research question. Preliminary explanations have been given on the observed differences of social diversity by language, religion and ethnicity across Indian megacities. This section consists a brief review of the five different methods for social diversity, the measures are applied to the census data of population in India. The conclusion finds that whereas there is disparity between cities in terms of social heterogeneity which ultimately gives a chance to find new assumptions on social diversity. This article explores a pathway to exploration of various heterogeneous character of urban India. In future different other modified indexes are also used to determine complex heterogeneity in urban places.

Ethnicity, language and religion are three important pillars in this study. The social construction of ethnic categories impacts identity development as well as personal and societal understandings of group differences and similarities. Collective ethnic identification contributes to the preservation of group interests. Only religion distinguishes the group ethnically, but culture is playing the role of social bonding among similar ethnic groups. Ethnic identity creation is a developmental process affected by the social environment in which it happens, and it differs for members of minority and majority groups. Ethnic identities happened to be fluid, changing across the life course and each shift bring changes in cultural expression and behavior. Language is a means of communication, argument, learning, negotiation, documentation, legislation and celebration. A better understanding of linguistic diversity and of their speakers helps to better understand how society functions, is in the domain of social heterogeneity. Like ethnicity, language is also fluid, evolutionary and never static. Language is the principal vehicle for the transmission of cultural practices, which enhances diversity in social scenario. Multicultural society is vibrant society by which contents from other cultures gained access. Religion refers to a variety of belief and practice systems that determine what people regard as holy or spiritual. It is found in every known culture in various forms and include festivals, marriage and funeral, music and art and other aspects of culture. Religion is dependent on society for its survival, value, and importance, and, as a result, religion fosters social interaction and group formation. It offers social support and networking opportunities, as well as a place to meet others who share similar beliefs. Finally, religion enforces social norms such as appropriate dressing, law and order and regulating sexual behavior. These three institutions maintain patterns of social structure.

Contextual background

Indian cities have become more dynamic through time to time, especially the metropolitan cities. The answer is lying beyond the boundaries of large cities and it has impact on the most dynamic cities of the region—the Indo Gangetic Plain or North Indian River Plain. This extends from far north (Punjab state) to far south Bay of Bengal (West Bengal) and encompassing most of the northern and eastern India. This is the home of half Indian population and include many major cities. Clearly, it includes two of India's five 10 million plus cities and the fortunes of the cities are influenced by the agricultural, industrial, mining and service sector resources of Indo- Gangetic Plain. Four among all of these five metropolitan cities are served as state capitals while New Delhi is India's capital city. However, the growth of these cities during the last fifty years has been changed significantly by transnational processes and changing their role in terms of policies related to commercial and industrial sectors. After economic liberalization, initiated in 1991, Indian economy grew increasingly market and service oriented, with private and international investment playing a larger role. There were some specific changes that had been taken including the reduction of import tariffs, deregulation of markets, reduction of taxes and greater foreign investment. The economic changes have occurred during the last thirty years are due to globalization and urbanization. India's metropolitan

areas have thrived due to the expansion of the service sector, particularly in business and financial services, advertising and media, medical, law, tourism, education, and high technology. These postindustrial cities are distinguished by the rise of both well paid, well-educated, and generally secure elites as well as low-paid, poorly educated, and insecure laborers. Many people employed in these dynamic areas have both indigenous and migrant backgrounds and are obvious in the large globalizing cities such as Mumbai, Bengaluru, Delhi, Chennai and Kolkata. However, these globalizing cities contain declining areas, localities with industrial decline, high rates of unemployment and poorly educated workers. However, the impact of globalization and rising service sector have attracted migrant workers.

Some of the migrant communities in metropolitan cities, like- Marwari, Parsi, Gujarati (Shetty) have been able to generate huge contribution to the local economy as they have earned the reputation for popular business model. Both previous and contemporary research in India indicate that urbanization and economic development are inextricably linked. (Sovani 1966; Bhagat 2012). About 65% of GDP acquired from urban areas that compromise of one-third of India's total population (31% urban population according to 2011 census) (Bhagat and Mohanty 2009; Bhagat 2012). As a result of increased contributions of urban resources, large metropolitan cities have seen fast urban population expansion (Map 1).

Data and methods

This section describes the data sources as well as the methodologies that were employed, the measurement details and the challenges during the calculation of social heterogeneity by language, ethnicity and religion in India's megacities and how is the variation.

The data utilized in this study are from 2011 Census, a decennial exercise by Census of India that aims to compile information on every household in the country. The enumerators collect individual level demographic data (for example, male female members of the family, economic activities, caste, literacy status, age, migration history, etc.) for each member from the household. In the census, there is no explicit question about income or consumption levels. The religious, linguistic and ethnicity related data are not available ward wise. In this analysis, data have been collected from five selected urban municipal corporations. The boundaries of this corporation correspond to the city limits, within which local governments offer public services. Municipalities are divided into numerous wards. Each ward has an elected municipal councilor, and therefore this geographic unit has both administrative and political identity. In this analysis, aggregated individual level data provided for each municipal corporation. Table 1 lists the total population, areal extension of the cities and population density for the five largest Indian cities.

The reason of selecting the five cities are the availability of data in a large spatial unit, both in terms of population size and geographic area. Population from different social backgrounds are living with each other in closed spatial units or city wards, which gave a unique character to the social heterogeneity. According to White (1983), groups with a large social distance may reside in the same area but do not
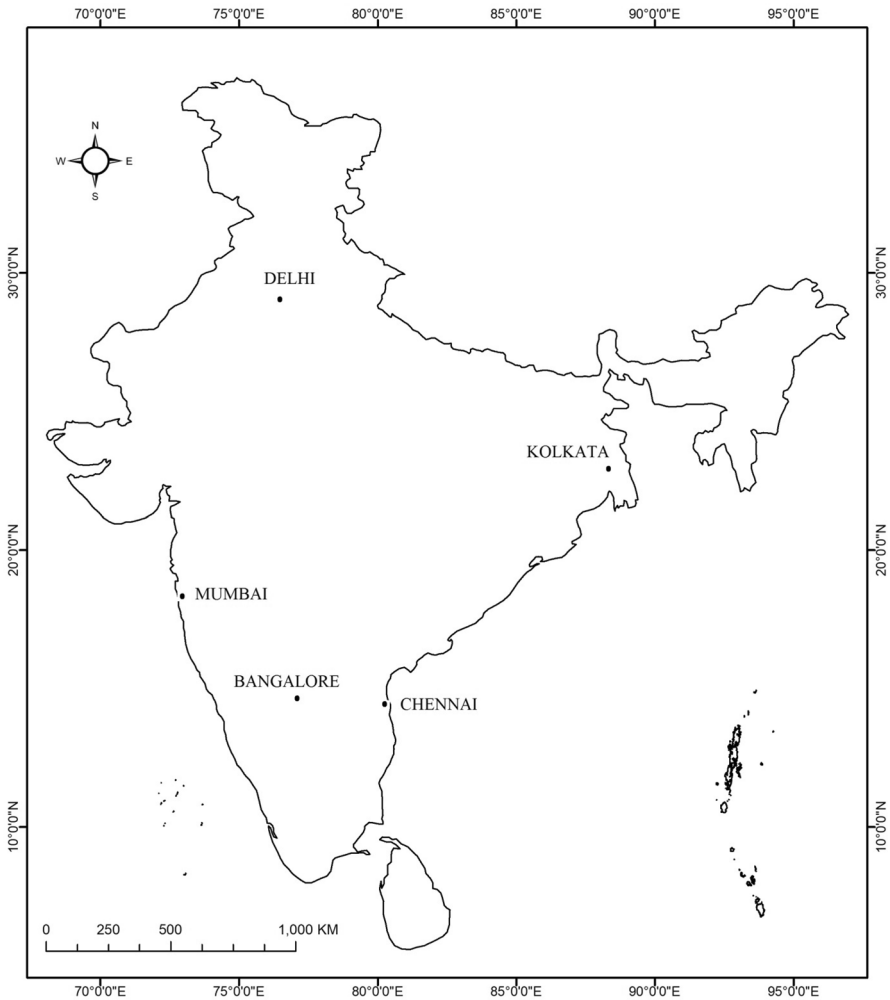
Map no. 1 India (location of Delhi, Mumbai, Chennai, Bangalore and Kolkata). *Source* Prepared by Authors, 2019

Table 1 Total population, areal extension and population density in India's five largest cities in 2011. *Source* Census of India (2011)

City	Total population	Area (Km^2)	Population density (Km^2)
Mumbai	12,442,373	603	20,634
Delhi	11,034,555	561.3	19,660
Bengaluru	8,495,492	741	11,465
Chennai	4,646,732	175	26,553
Kolkata	4,496,694	185	24,306

interact with one another. In this regard, although this paper is based on that social identity is important and shapes the nature of one's social life, it is important that shared residence in an urban ward represents a range of diverse relationships and degrees of interaction. However, the focus of this investigation is on India's macro patterns of religion, language, and ethnicity.

To explore whether there is social heterogeneity in India's megacities, three dimensions of each city has been taken into account- language spoken by the people, religion adhered by them and their ethnic setup. There are 22 scheduled languages in India, that are selected for the study, along with 8 categories of religion and 442 ethnic groups that are primarily resides over India (Census of India 2011; Joshua Project People Group data 2011). Table 1 provides a demographic description of Indian cities and the cities for this study are selected based on population.

As a baseline measure to serve the diversity within the society, Shannon's Diversity Index was first calculated to measure the degree of diversity by religion, language and ethnicity and provide a point of comparison with other measures of substantive interest; in this case social heterogeneity. Diversity Indices expressed three components of community structure, i.e. richness (number of species), evenness (the distribution of individuals among species) and abundance (total number of organisms' present). The most widely used diversity index has been the Shannon-Weiner Index (abbreviated as H) (Washington 1984). Shannon's diversity index is a mathematical measure of species diversity within a given community and describe the disorder or measuring uncertainty of individual species, in this case social groups. For geographical approach, human communities are selected as particular species within a human settlement (city). The index is based on species richness (the number of peoples from various communities) and species abundance (the number of individuals from each human community). The degree of uncertainty in forecasting a large number of species or groups can enhance variety; likewise, improving the uniformity of individual distribution among species can increase diversity. As the aim was to study social heterogeneity, diversity within groups of interest is assigned by this method. Shannon index is an information statistic index, which implies it presupposes all species are represented in a sample and are sampled at random.

$$\text{Shannon's Diversity Index } (H) = \sum_{i=1}^{S} (P_i \times \ln P_i)$$

where H = the Shannon diversity index, P_i = fraction of the entire population made up of species i, S = numbers of species encountered, $\sum$ = sum from species 1 to species S.

Note: The power to which the base e ($e = 2.718281828\dots$) is raised to obtain a number which is called the **natural logarithm** (ln) of the number. To calculate diversity index, first the number of individuals or species found in sample divided by the total number of individuals of all species and the result is P_i. After that, multiply the fraction by its natural log, which is ($P_i \times \ln P_i$). This process is repeated for all the different species until the last species. Lastly, the sum is calculated for all the ($P_i \times \ln P_i$) products. The sum of all the values considered as H (Shannon diversity index).

The Herfindahl–Hirschman index (HHI) is a statistical measure of concentration. The HHI was first used by Hirschman in 1940s in a trade study where the square root of the sum of squared market shares were calculated (Hirschman 1964). Herfindahl used the version of the calculation in a study of the international copper industry (Herfindahl, 1959). The index was popularized as the Herfindahl index after studies by Rosenbluth (1955, 1957). This index is also used to analyze the competitive effects of diversity. The HHI index accounts for the number of religious groups in Kolkata city and their concentration by incorporating the relative size of all religious groups in the city. It is calculated by squaring the percentage of religious groups in the city and then summing up the squares, as follows:

$$\mathrm{HHI}_{1\ldots n} = S_1^2 + S_2^2 + S_3^2 \ldots S_n^2$$

where S_i represents the market share of the ith firm.

Simpson's Diversity index is a measure of diversity which takes into account the number of species present, as well as the relative abundance of each species. As the richness and evenness increase, so diversity within an area or community increases. The Simpson index is a dominance index because it gives more weight to a common or dominant species. In this index, a few species with only a few members in the community will not affect the diversity.

$$D = 1 - \left\{ \sum n_i \, (n - 1) \, / N_i \, (N - 1) \right\}$$

where n_i = the total number of organisms of a particular species, N = the total number of organisms of all species.

Simpson's index is calculated as a weighted arithmetic mean of proportional abundance and measures the probability of two individuals randomly selected from a sample will belong to the same species. Since the mean of the proportional abundance of the species increases with decreasing number of species and increasing abundance of the most abundant species, the value of D with small values shows high diversity and large values show low density. The value of Simpson's D ranges from 0 to 1, where o representing infinite diversity and 1 representing no diversity.

Pielou's evenness Index (1966) is used in measuring species evenness. It shows how evenly the individuals are distributed among different species. It is calculated as:

$$J' = H'/\ln S \text{ where } \ln S = H' \max$$

In H' max (the maximum value of Shannon diversity), H' would be if all the species in the community had an equal number of individuals and S is the number of species. J' is constrained between 0 and 1. The less evenness in communities between the species, the lower the J' is and more evenness between communities, higher the J'.

Czekanowski Index is a measure of similarity between two categorical distributions. It was proposed by Czekanowski (1909) in the form of coefficient. It is calculated as:

$$Cz_{A,B} = \sum_i \min \left(q_i^A, q_i^B \right)$$

where q_i^A and q_i^B are relative frequencies of category i in the distribution of A and B, respectively. This index was transformed several times in economics. The Czekanowski index is regarded as "quantitative" (i.e., abundance dependent) index.

Result and discussion

For the baseline measurement in each city, degree of diversity in society by language, religion and ethnicity has been calculated.

The degree of heterogeneity in five Indian cities are measured by Shannon–Wiener Index. It is the measure of richness or evenness in society (Table 2). Typical values are generally ranging between 0.67–1.05 in case of religion, 0.70–1.72 in case of language and 3.58–4.01 in case of ethnicity. The Shannon index increases as both the richness and the evenness of the community increase. The religious richness and evenness are comparatively low in terms of language and ethnicity. There is a small number of major religions in India compared to large number of languages and ethnicity. In Kolkata, there are low richness and evenness compared to Mumbai, where religious richness and evenness is high. That means the demography of Kolkata are dominated by few numbers of religion. After 1947, refugees coming from East Pakistan (now Bangladesh) spread all over West Bengal and in other parts of India. These people are concentrated in 'colonies' in Greater Kolkata. Most of the refugees were from upper- or middle-class groups and non-agricultural background who are the Hindu people. This is the reason for the current religious structure of Kolkata. There are 6 major religion in Kolkata, but 96 percent of religious population is from two major religion. Linguistic diversity is highest in Mumbai and Bengaluru and it is lowest in Delhi. Over the past 150 years, Mumbai has been an ever-changing metropolis of global importance. It has made a quick economic shift from trade and services in the previous 25 years and has expanded its national and cross-border functions. It is by far the most globalized city in South Asia, most internationalized

Table 2 Shannon's Wiener Diversity Index. *Source* Compiled by Authors (2019)

Cities	Shannon		
	Religion	Language	Ethnicity
Mumbai	1.05	1.72	3.93
Delhi	0.70	0.70	4.01
Bengaluru	0.71	1.72	3.66
Chennai	0.70	0.87	3.73
Kolkata	0.67	1.04	3.58

economy in this region, major location for corporate headquarters and the main destination for foreign investment and joint ventures. On the other hand, Bengaluru is the home of Infosys, Wipro Technologies, Mindtree, Flipkart, Tejas Networks and several other e-commerce companies. As well as aerospace and aviation, biotechnology, manufacturing, space research is also growing. This is a main reason why the linguistic diversity is high. On the contrary, its great history and affiliation as the Capital of India affected Delhi's culture. A considerable Punjabi influence may be evident in the language, dress, and cuisine. As a result, compared to other metropolitan cities, these cities have little linguistic diversity. Ethnic diversity is highest in Delhi and lowest in Kolkata. Delhi has a long tradition as a melting pot of culture. In the medieval times. Mughal and Turkic rulers were ruled the city and it was a major transition point for business. Also, the foreign rulers employ their countrymen in their service. So, there is many Ethnic groups are present in the city. The diversity was also resulted from 1947 India Partition, where people from West Pakistan (Now Pakistan) and they formed their own community and identity. Kolkata has its roots under British rulers and its history is 300 years old. Strong Bengali presence and unique cultural identity reveals in the Ethnic setup of the country. High ethnic diversity indicates a highly complex community and greater variety of ethnic groups allows for a larger array of people's interaction. High diversity becomes a subject of discussion, because some ecologists use diversity as an index of the complexity, maturity and stability of a community whereas others advocates that there is no simple relationship between diversity and stability.

Herfindahl–Hirschman index is a statistical measure of concentration. It is also used in the analysis of the effects of a heterogeneous society. The ground rule for interpreting the HHI is that, if the value of HHI increases, it means a decrease in competition and vice versa. In Table 3, it is seen that the HHI value has amplified in case of religion, followed by language and ethnicity. The religious concentration is low, as showing in Table 2 where linguistic and ethnic concentration is high. The highest religious concentrated city is Mumbai where lowest religious concentration is found on Chennai. In this method, Chennai holds the lowest rank. In Shannon Index, Kolkata holds lowest religious diversity. The city of many faiths is Mumbai. There is a high cultural level, and religious belief/secularism go hand in hand. Mumbai has a religious mix as a result of the city's extensive expansion of various communities and migrants. Chennai is religiously cosmopolitan, but more than 97% people adheres three major religion in India. The highest linguistic concentration

Table 3 Herfindahl–Hirschman Index. *Source* Compiled by Authors (2019)

Cities	Hirschman–Herfiendahl		
	Religion	Language	Ethnicity
Mumbai	4832.65	2458.46	576.59
Delhi	6620.23	7429.33	404.22
Bengaluru	6446.05	2492.50	674.38
Chennai	6668.26	6296.35	473.21
Kolkata	6280.08	4492.09	604.27

Table 4 Simpson's Diversity Index. *Source* Compiled by Authors (2019)

Cities	Simpson's		
	Religion	Language	Ethnicity
Mumbai	0.483	0.246	0.06
Delhi	0.661	0.720	0.04
Bengaluru	0.645	0.249	0.07
Chennai	0.666	0.630	0.05
Kolkata	0.628	0.449	0.06

Table 5 Pielou's Evenness Index. *Source* Compiled by Authors (2019)

Cities	Pielou's		
	Religion	Language	Ethnicity
Mumbai	0.5	0.56	0.65
Delhi	0.34	0.23	0.66
Bengaluru	0.34	0.56	0.60
Chennai	0.34	0.28	0.61
Kolkata	0.32	0.34	0.59

is found also in Mumbai where lowest concentration is Delhi. This result is similar with the Shannon Index of linguistic diversity. Ethnic composition in Delhi is highest where it is lowest in Bengaluru. The lowest result is different from Shannon Index where Kolkata is holding lowest position among five cities. Bengaluru is a growing economy which is hi-tech in nature. For this purpose, various people from different locations visit Bengaluru but in small numbers. The city is not historically important, and its popularity is based on modern technology related industries. Therefore, the Ethnic variation is low. The difference between the highest and lowest results are not much.

Evenness and dominance are both compatible and their measures are complimentary. Simpson Index is denoted by DD is a measure of dominance, so as D increases, diversity decreases. D takes on values between 0 and 1. From Table 4, it is observed that Mumbai city has highest religious diversity and Chennai city experience lowest religious diversity. This result is similar to HHI index. Mumbai city has highest linguistic diversity compared to Delhi, where it is lowest. These data are similar as of Shannon's Diversity and HHI Index. Ethnic diversity is highest in Delhi city and lowest in Bengaluru city. The same trend in found on HHI index.

Pielou's Index is referred as *J*'. The range is between 0 and 1, lower value denotes less evenness and vice versa. Mumbai has most religious evenness compared to Kolkata, where it is the lowest. Mumbai and Bengaluru have most linguistic evenness where Delhi holds lowest position. Delhi has highest Ethnic evenness and Kolkata has lowest evenness. These three results are like Table 5 because Pielou's index is based on Shannon Wiener's index.

Similarity is a community parameter and the basis of classification. In this study, Czekanowski Coefficient is used as one method of objectively defining the

relationships among different city community. It is found that the degree of religious similarity is high across India's five large cities though it does vary from a similarity of 0.53 (Mumbai and Kolkata) to 0.87 (Chennai and Kolkata). Religious similarity is overall similar throughout cities because there is no sharp differentiation between religious population in these cities. But there is a sharp residential segregation is present based on religion within those cities. Religion became a sensitive issue in India for a very long period. In these five cities, Hinduism dominates as major religion followed by Islam and Christianity. These are three major religion in India with its religious affiliation. Sikhism, Buddhism and Jainism are another three religion with significant population that is found in these five cities. In Mumbai, Parsi community practiced ancient Zoroastrian religion which is considered a minority religion. The selection of cities in this study shows more similar pattern in religious affiliation, if the study area will expand, the similarity index would be different. From Table 6, it can create two comparison groups- one is Chennai and Kolkata and the other is Mumbai, Bengaluru and Delhi. Chennai is in the southern tip of India and Kolkata is in the Eastern end. For its location, Chennai did not face persecution by Mughal invaders, so the original social setup was not disturbed. Modern Chennai had its origin as a colonial city and its growth was closely related with the port facility. This city was initially ruled by Dutch, but after 1750 it became part of British Empire. Religious persecution was uncommon during these two colonial rulers. Therefore, Hinduism became most prominent and major religion in Chennai. Kolkata was founded by British traders, for business purpose, where Nawabs of Bengal were never put attention. In 1690, when East India Company started trading from the Banks of Hooghly River, there was a new era of Company rule is beginning. The religious setup was changed after Partition of India in 1947, which led to a shift in demographics in Bengal, and especially Kolkata. Hundreds of thousands of Hindus arrived to replace the Muslims who had fled to East Pakistan (now Bangladesh). Millions of refugees arrived in Kolkata alone. The war between India and Pakistan in 1971 resulted in another huge inflow of refugees from the erstwhile East Pakistan (now Bangladesh).Another comparison group consists Mumbai, Delhi and Bangalore. Delhi was the capital during Mughal period. The Mughal empire ruled Delhi for three centuries. The Mughal Empire was afflicted by a series of revolts after a series of revolts. After that, Delhi was looted by Nadir Shah. East India Company captured Delhi in 1803. The capital of British India was moved from Calcutta to New Delhi in 1911. The societal setup was changing for a long time where numbers of various groups from different religious affiliation can be found. Unlike Delhi,

Table 6 Czekanowski coefficient (religious). *Source* Compiled by Authors (2019)

Cities	Mumbai	Delhi	Bengaluru	Chennai	Kolkata
Mumbai	1	0.85	0.80	0.54	0.53
Delhi		1	0.83	0.56	0.57
Bengaluru			1	0.71	0.69
Chennai				1	0.87
Kolkata					1

history of Mumbai was founded by Portuguese invaders. They transformed Mumbai (Bombay) from a fishing village to a large megacity in India. Portuguese influence was changed religious setup of Mumbai megacity. They were actively involved in the establishment and expansion of their religious orders in Bombay. After that in 1817, British East India Company acquired Mumbai. These successive colonial periods have changed religious setup of this city. Bengaluru city was gone through successive colonial rulers like Sultanate of Bijapur and Mughal Empire and after 1799, it was becoming part of British India. But most of its growth was taken place after Independence of India and 1991 Economic Reforms. For this reason, religious setup of this city has been changed.

Table 7 shows the Ethnic Linguistic Similarity composition between the cities. There is very little similarity between the cities. It is inevitable because the formation of Indian states had been conducted based on language. Therefore, every city has most population from the state language group. Hindi is served as Lingua Franca between different states. In this table, similarity between Mumbai and Bengaluru is highest (0.48) followed by Bengaluru and Chennai (0.37); Mumbai and Delhi (0.33). Lowest similarity is found between Delhi and Chennai (0.06) and Chennai and Kolkata (0.09). Marathi is a common language in Mumbai and Bengaluru; Urdu is a common language in Mumbai and Delhi and Chennai and Bengaluru have one common language- Kannada. But Bengali and Tamil are regional languages with little presence in other states in India. Same thing is true in case of ethnic groups. There are different ethnic groups in India and the number/percentage varies from one state to another. So, it is not uncommon that there is little similarity between Ethnic Linguistic groups in India.

Spatial variations in human community pattern and diversity measures are analyzed for the five cities investigated. The indices reflecting the changes that occurred in the urban community structure which is different from one place to another. To calculate the level of social heterogeneity for each of the three dimensions discussed earlier, three diversity indices and one similarity indices have been selected. Diversity indices gave the idea that how religious, linguistic and ethnic groups in each city are creating a cluster in the demography of the city. The number and presence of dominant groups often results to disharmony. Communities are an organized system of organism which is responding in a related manner to changes. Due to the concomitant and continuous interaction taking place between individuals and other socio-economic factors, so the community remains dynamic. Other than Shannon–Wiener and Pearson Diversity Index, Herfindahl–Hirschman index explore the

Table 7 Czekanowski coefficient (ethno-linguistic). *Source* Compiled by Authors (2019)

Cities	Mumbai	Delhi	Bengaluru	Chennai	Kolkata
Mumbai	1	0.33	0.48	0.13	0.27
Delhi		1	0.14	0.06	0.19
Bengaluru			1	0.37	0.19
Chennai				1	0.09
Kolkata					1

competitiveness of the society, where the society is perfectly competitive or there is monopoly of a single group. The results reflect the structure and function of the community. The conventional types of diversity indices that are used in this work do not consider whether the assemblage within the same community which are closely related to each other or whether they are only distantly related. These indices are heavily dependent on population size and they do not reflect population behavior. The term Diversity is used as Heterogeneity in this study for easily measured. Heterogeneity contains two separate ideas—species richness and evenness and Pielou's Evenness Index is used to measure the evenness component. It is the fact that most communities contain a few dominant groups and many groups that are relatively uncommon. Pielou's Evenness Index attempted to quantify this unequal representation against a hypothetical community in which all groups of people are equally common. Index of Similarity (Czekanowski Coefficient) is another measure of evenness has been used. The word 'Similarity' is the logical complement because similarity indices indicates how 'close' two samples are to one another. There are two reasons- First, this index used five city comparison as the variables that are used in this study. The second one is it is defined by the pairwise comparison of the values for each community attributes in the five sample cities. This index uses intermediate similarity scores defined on the pairwise comparisons of the values for each community attributes in the five sample cities will be discussed. This is also a measure of community structure (Pinkham and Pearson 1976) and an index of sample diversity also.

Because of the structure of the data, there are two methodological problems in analyzing social heterogeneity. First, within each city, there are varying urban area size in terms of population number, composition and areal extension. Second, there is also a substantial variation in the urban area size in terms of population number, composition and areal extension across cities. To meet the first difficulty of a city's changing population structure, the degree of residential segregation across several demographic measures, and the divergence in residential segregation depending on religion, language, and ethnicity. This within city comparison is only assessed by species diversity indices that helps to make sense of the similarity level for religion, language and ethnicity in each city.

The second challenge of varying city sizes raises the question of whether differences in similarity across cities are mere reflection of similarities in city size. Since the index of similarity is sensitive to the size of underlying area units, it raises a concern that as the city size increases, similarity may increase. Given the small sample of cities, testing for the relationship between city sizes and similarity level could not provide convincing results. But from this study, it can be said that there is convincing result between city size and similarity index. So, the study is restricted the comparison between cities with similar sizes with a conservative approach. By comparing cities with same areal extension, the problem of any biased comparison could be avoided.

In addition to these issues, problems with sample diversity (species richness) estimation also exist. The richness estimator, in particular, exhibits sample size dependence. (Smith and Van Belle 1984). Simple abundance or richness estimators, such as the sum of observed species in a sample underestimate richness at

all sample sizes and this underestimation decreases with increases in sample size. The most widely used measures of diversity in ecology are species richness- the Shannon Index, Pielou's evenness index and the Simpson Index. Each index compares different communities to each other, but diversity values of the different indices cannot be directly compared to each other. This is due to the fact that each index has a distinct meaning in terms of variety. As a result, diversity measures obtained using various classical indices are not directly comparable, which is a disadvantage for scientific research. Another drawback suffered by classical indices is the problem of non-linearity. However, this problem is not statistically solved, but the spatial and temporal discussion on five sample cities could reduce non-linearity in an effective manner. Another measure from economics, which is called Hirschman-Herfindahl Index similarly have an erroneous conflation of market power (in this case population diversity). In this measure, number is the only important matter and completely discard network building and functional relationship gave no importance. So, the way social groups interact with each other with concentration is not explained by Hirschman-Herfindahl index.

Conclusion

After conceptually grounding, operationalizing and constructing indicators, it is found out these social identities historically shaped the association of residential space, especially at the village level, and it appears to continue the same in contemporary urban India. At the start of the twenty-first century, it is inevitable to find that India's five largest metro cities have residential segregation by religion, language or ethnicity and language that is sizably larger than the level of segregation by other factors. The relationship between these social communities have always been complicated, however intermarriages and cultural adaptation is common too. In terms of opportunities, mainstream educational system provided by Indian Government could promote multiculturalism to enhance intermixing. These five metropolises in this article experience some of the bloodiest sectarian violence from time to time (Bombay Riot, 1992, Religious; Anti Hindi Agitation in Chennai, 1968, 1986, 2014, Language; Naxalite Movement in Kolkata, 1969, Political revolt; Attacks on North Eastern Indians in Delhi, Ethnic, 2017–2019). Therefore poor, unorganized, unemployed and vulnerable minority groups can experience certain degree of upward and outward mobility through the educational systems by community organizations and NGOs. The high level of ethnic, religious and linguistic conflict over the last twenty years is the result of economic issues, lack of general prosperity through India as well as political and cultural factors which led to the alienation of minority groups. In this regard, a comparative social study on five largest cities is better in understanding the mechanisms that segregation within cities. Finally, it is concluded that disparity in social heterogeneity is the effect of economy and opportunity among space. It is hoped that this study has been able to understand the social relations in India across large urban scales in India.

Declarations

Conflict of interest The authors have declared that they have no conflict of interest.

References

Alesina A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R (2003) Fractionalization. J Econ Growth 8:155–194

Atlas Narodov Mira (1964) Moscow: Miklukho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.

Bhagat RB (2012) Migrants (Denied) right to the city. In: National workshop on internal migration and human development: workshop compendium, vol 2, Workshop Papers, UNESCO and UNICEF, New Delhi, pp 86–99

Bhagat RB, Mohanty S (2009) Emerging pattern of urbanization and the contribution of migration in urban growth in India. Asian Popul Stud 5(1):5–20

Brock W, Durlauf S (2001) What have we learned from a decade of Empirical research on Growth? growth empirics and reality. World Bank Econ Rev 15:229–272

Census of India (2011) District Census Handbook Bardhhaman. West Bengal, Series 20, Part XII-B, pp 7–686

Collier P, Gunning JW (1999) Why has Africa grown slowly? J Econ Perspect 13:3–22

Davis M (2007) Planet of slums. Verso, London and New York, pp 1–49

Dupont V (2004) Socio-spatial differentiation and residential segregation in Delhi: a question of scale? Geoforum 35:157–175

Easterly W (2002) The elusive quest for growth: economists, adventures and misadventures in the tropics. MIT Press, Cambridge

Easterly W, Levine R (1997) Africa's growth tragedy: policies and ethnic divisions. Quart J Econ 112:1203–1250

Englebert P (2000) State legitimacy and development in Africa. Lynne Rienner, Boulder, CO

Evenson N (1989) The Indian metropolis: a view towards the West. Yale University Press, New Haven, CT

Fainstein S (2005) Cities and diversity. Should we want it? Can we plan for it? Urban Affairs Rev 41(1):3–19

Florida R (2002) The rise of the creative class. Basic Books, New York

Gupta A, Ferguson J (1997) Culture, power, place: ethnography at the end of an era. In: Gupta A, Ferguson J (eds) Culture, power and place: explorations in critical anthropology. Duke University Press, Durham, pp 1–29

Gupta K, Fred A, Lhungdim H (2009) Health and living conditions in eight Indian cities, National Family Health Survey (NFHS-3), India, 2005–2006. International Institute for Population Sciences, Mumbai; ICF Macro, Calverton, MA.

Gurr T (1996) Minorities at risk. 111 datasets: user's manual. CIDCM, University of Maryland.

Hall RE, Jones CI (1999) Why do some countries produce so much more output per worker than others? Quart J Econ 114:83–116

Herfindahl OC (1959) Copper costs and prices: 1870–1957. The John Hopkins Press, Baltimore

Hirschman AO (1964) The paternity of an index. Am Econ Rev 54:761–762

Lelo K, Monni S, Tomassi F (2018) Socio-spatial inequalities and urban transformation. The case of Rome districts, socio-economic planning sciences. Elsevier, Amsterdam

Mehta S (1968) Patterns of Residence in Poona (India) by income, education and occupation (1937–65). Am J Sociol 73(4):496–508

Mehta S (1969) Patterns of residence in Poona, India, by caste and religion: 1822–1965. Demography 6(4):473–491

Mehta S (2004) Maximum city: Bombay lost and found. Headline Review, London

Pinkham CF, Pearson JG (1976) Application of a new coefficient of similarity to pollution surveys. J WCPF 48:717–723

Rodrik D (1999) Where did all the growth go? External shocks, social conflict and growth collapses. J Econ Growth 4:385–412

Rosenbluth G (1955) Measures of concentration. Business concentration and price policy. National Bureau of Economic Research, Princeton University Press, Special conference series no 5, pp 57–89

Rosenbluth G (1957) Concentration in Canadian manufacturing industries, vol 67(268). Princeton, Princeton University Press, pp 665–691

Rowe W (1973) Caste, kinship and association in urban India. In: Southall A (ed) Urban anthropology: cross cultural studies in Urbanization. Oxford University Press, New York

Sennett R (1992) The conscience of the eye. The design and social life of cities. W.W. Norton & Company, New York, London

Sennett R (1996) The uses of disorder. Faber and Faber, London

Sikder SK, Nagarajan MS, Koetter T (2017) A geospatial approach of downscaling urban energy consumption density in mega city Dhaka, Bangladesh, Urban Climate. Elsevier, Amsterdam

Smith EP, Van Belle G (1984) Nonparametric estimation of species richness. Biometrics 40:119–129

Sovani NV (1966) Urbanization and urban India. Asia Publishing House, New York

Tonkiss F (2005) Space, the city and social theory. Polity Press, Cambridge

White MJ (1983) The measurement of spatial segregation. Am J Sociol 88(5):1008–1018

Wessel T (2009) Does diversity in urban space enhance intergroup contact and tolerance? Geografiska Ann B, Hum Geography 91(1):5–17