



# Network Entropy as a Measure of Socioeconomic Segregation in Residential and Employment Landscapes

Nandini Iyer<sup>1</sup>(✉) , Ronaldo Menezes<sup>1,2</sup> , and Hugo Barbosa<sup>1</sup>

<sup>1</sup> BioComplex Laboratory, Computer Science, University of Exeter, Exeter, UK  
`n.iyer@biocomplex.org`, `{r.menezes,h.barbosa}@exeter.ac.uk`

<sup>2</sup> Computer Science, Federal University of Ceará, Fortaleza, Brazil

**Abstract.** Cities create potential for individuals from different backgrounds to interact with one another. It is often the case, however, that urban infrastructure obfuscates this potential, creating dense pockets of affluence and poverty throughout a region. The spatial distribution of job opportunities, and how it intersects with the residential landscape, is one of many such obstacles. In this paper, we apply global and local measures of entropy to the commuting networks of 25 US cities to capture structural diversity in residential and work patterns. We identify significant relationships between the heterogeneity of commuting origins and destinations with levels of employment and residential segregation, respectively. Finally, by comparing the local entropy values of low and high-income networks, we highlight how disparities in entropy are indicative of both employment segregation and residential inhomogeneities. Ultimately, this work motivates the application of network entropy to understand segregation not just from a residential perspective, but an experiential one as well.

**Keywords:** Commuting networks · Network entropy · Socioeconomic inequality

## 1 Introduction

Often, the prosperity of urban areas is understood as a function of economic input and output [12]. However, experiential variables such as social inclusion have been shown to fuel the productivity of cities [5, 15]. Accordingly, investing in improving the diversity of cities can lead to positive socioeconomic outcomes, fostering innovation and entrepreneurship [13]. The urban experience, in terms of mobility, is largely comprised of trips to work [8], underscoring the importance of considering social inclusion not only from the residential dimension, but the employment perspective as well. In this sense, mobility, specifically commuting patterns, can be viewed as a potential way to improve diversity and integration by creating points of social inclusion in employment areas [7].

Networks are particularly useful for analysing commuting behaviours, as they capture structural patterns that other approaches may overlook [10]. Specifically,

network entropy can capture the concentration of labour supply and demand as well as the level of diversity of where workers are commuting to or from [11]. Entropy has been used in commuting networks to explain economic growth [6], identify spatial inequalities [9], and measure social assortativity [1]. However, the majority of previous research on this topic analyses the commuting networks of an entire population. Thus, we consider not only the commuting networks of an entire population, but also commuting networks comprised of individuals from particular socioeconomic groups. Disaggregating commuting networks by demographics allows us to study whether disparities in structural diversity could serve as an indicator of social exclusion.

In this paper, we explore the intersection of socioeconomic segregation in cities and structural diversity in commuting patterns by applying an information-theoretic approach to mobility networks. Specifically, we leverage measures of global and local entropy of commuting networks to clarify whether demographic disparities in reliance on commuting origins and destinations correspond with levels of segregation on a residential and employment level. Our analyses of the commuting networks of 25 US cities indicate that, for high-income workers, higher levels of residential concentration correlate with higher employment segregation. Employment segregation strays from residential segregation, in that it measures the level of segregation based on the individuals that work in a region, rather than individuals that live there. Low-income workers, on the other hand, tend to commute to a larger range of workplace areas. Finally, our network approach to segregation reveals facets of structural socioeconomic inhomogeneities in commuting patterns beyond what traditional measures of segregation can capture.

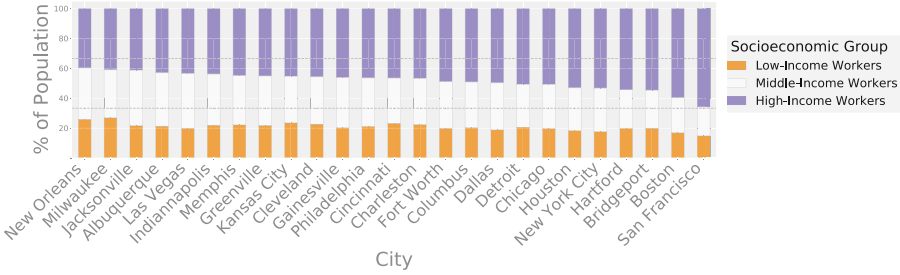
## 2 Methods and Data

### 2.1 Data

We use the LEHD Origin-Destination Employment Statistics (LODES) dataset from the United States Census’s 2019 Longitudinal Employer-Household Dynamics (LEHD) program [4]. The dataset captures the residential patterns of the surveyed workforce by measuring the number of individuals commuting from one census block group to another. We evaluate residential-employment trends on a census tract level. Census tracts are statistical partitions of counties that contain anywhere from 1,200 to 8,000 residents. We use the LODES data to construct commuting networks for 25 cities in the United States, where every node reflects a census tract and directed, weighted edges depict the number of individuals commuting from one tract to another. The 25 cities we selected cover a wide range of population sizes and socioeconomic characteristics. The LODES dataset also provides information about how the total commutes from a pair of census tracts are distributed across lower, middle, and higher-income demographics. The low-income group consists of individuals earning less than \$1,250 per month, while the minimum monthly income for the high-income group is \$3,333. In addition to building the entire commuting network of a city, we also

build a low, middle, and high-income network, which have the same nodes as the network for the entire city. However, the networks for each socioeconomic group, which we refer to as disaggregated networks, have different edge weights depending on the number of people in a socioeconomic group that commute between a pair of tracts.

## Socioeconomic Profiles of U.S. Cities



**Fig. 1.** Workforce distribution by income level split into three socioeconomic groups (low, middle and high-income) across 25 US cities.

Figure 1 captures the socioeconomic makeup of each city, according to the LODS data. The dashed lines indicate what an even distribution across demographics would look like. Thus, we observe that, within the context of the LODS dataset, cities such as Boston and San Francisco have skewed representation of socioeconomic groups. To measure levels of residential segregation, we use Table B19001 from the 2019 American Community Survey 5-Year Estimates, which measures the population distribution across income brackets for each census tract [3].

## 2.2 Network Entropy in Commuting Networks

Throughout our analyses, we use the term *commuting destinations* to refer to the workplaces that a residential population commutes to, while *commuting origins* describe the residential areas from which an employment area’s workforce commutes. Furthermore, *labour supply* refers to employment areas that supply jobs, while *labour demand* reflects the employable population of a region. We use Shannon’s entropy, which captures the level of information that can be extracted given a probability distribution, as a measure of diversity in commuting destinations and origins [14]. Moreover, network entropy can be applied at different network resolutions and can focus on the commuting in-flow to a work area or the commuting out-flow from a residential region. Global entropy of in-flow and out-flow captures the urban concentration of labour supply and demand, respectively, by characterising the degree of monocentricity. That is, when a city has one area that supplies most of the labour opportunities, there is less uncertainty in predicting commuting destinations, corresponding with a lower global in-flow

entropy value for the entire city. Thus, global in-flow entropy ( $H_{GN}^{in}$ ) leverages the node strength,  $\sum_i p_{ij}$ , of all incoming commutes to each tract,  $j$  in a city:

$$H_{GN}^{in} = \frac{-\sum_{\forall j} (\sum_{\forall i} p_{ij}) \log (\sum_{\forall i} p_{ij})}{\log(n)}, \quad (1)$$

where  $n$  is the total number of tracts in a city. Global out-flow entropy captures the distribution of labour demand, such that low entropy values depict a scenario in which most workers live in a few areas and high entropy values indicate where residential locations of workers are more evenly distributed across nodes. In contrast to  $H_{GN}^{in}$ , global out-flow entropy ( $H_{GN}^{out}$ ) calculates the out-degree node strength,  $\sum_j p_{ij}$ , for all census tracts,  $i$  in a city.

Global entropy defines a city with respect to how labour supply or demand is distributed across all census tracts. Meanwhile, local entropy defines each tract based on how evenly distributed all its incoming or outgoing commutes are. A high local in-entropy for a census tract implies that the commuting origins for individuals who work in that tract are evenly distributed across all the potential origins. Local in-flow entropy ( $H_L^{in}$ ) for a census tract  $j$  accounts for the probability,  $p_{(i|j)}$ , of tract  $j$  receiving commutes from a census tract  $i$ , for all possible commuting origins,  $i$ :

$$H_L^{in} = \frac{-\sum_{\forall i} \frac{p_{ij}}{p_j} \log \frac{p_{ij}}{p_j}}{\log(n-1)}. \quad (2)$$

Local out-flow entropy ( $H_L^{out}$ ) can be defined similarly, except rather than consider the probability  $p_{(i|j)}$  for incoming commutes, it calculates the probability,  $p_{(j|i)}$  of tract  $i$  sending commuters to a census tract  $j$ , for all possible workplaces  $j$ . In this manner, we leverage network entropy measures on a global and local scale to understand how commuting characteristics of low versus high income workers indicate levels of segregation not only in terms of where each demographic group lives, but also in the context of the areas where they work. The denominators in Eqs. 1 and 2 serve to normalise the entropy values for cities of varying sizes, based on the number of tracts in a city (Eq. 1), and the focal node's maximal possible degree (Eq. 2).

### 2.3 Characterising Segregation in Urban Landscapes

We analyse features of the urban space to better understand the mechanisms that drive the identified differences in commuting networks across demographics. For such, we characterise urban areas in terms of the concentration of socioeconomic groups, using the Index of Concentration at the Extremes (ICE), proposed by Douglas Massey in an attempt to describe a region using both spatial attributes and the characteristics of polarised demographics [2]. While most segregation metrics account for how segregated the minority group is in relation to the entire population, ICE incorporates both ends of the demographic spectrum:

$$ICE_i = \frac{A_i - P_i}{T_i}, \quad (3)$$

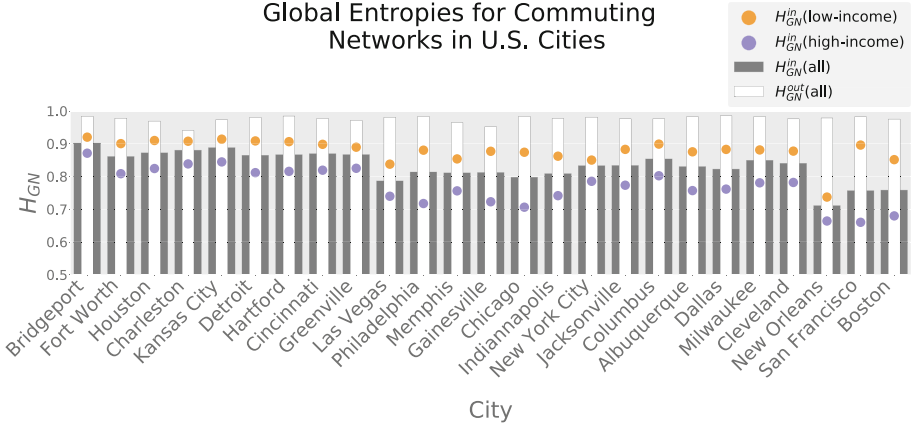
where the ICE for a census tract  $i$  is defined as the difference between the number of affluent residents,  $A_i$ , and the residents under the poverty line,  $P_i$ , over the entire population,  $T_i$ . While the numerator captures the imbalance between the extremes, the denominator expresses the degree of imbalance in relation to the entire population of tract  $i$ . Thus, ICE aims to measure the imbalance between affluence and poverty by measuring the concentration of both the extremely disadvantaged and advantaged in a given population. In a similar vein, we can use Eq. 3 to measure the level of segregation for a census tract,  $i$ , from an employment perspective, defining  $A_i$  and  $P_i$  as the number of workers commuting to tract  $i$  from the high and low-income group, respectively. Thus,  $T_i$  captures the total number of individuals commuting to  $i$ , regardless of demographics. Accordingly, we can use ICE to capture segregation levels at both a residential ( $\text{ICE}_{\text{res}}$ ) and employment ( $\text{ICE}_{\text{emp}}$ ) scale. We use the LODES dataset to calculate  $\text{ICE}_{\text{emp}}$  as it provides unique information regarding demographic characteristics of a workforce. Meanwhile, We use the ACS median household income distributions to calculate residential segregation as it provides a high granularity of income-levels to characterise tracts by its residents.

### 3 Results

#### 3.1 Global Entropy and City-Level Analyses

We begin by applying global entropy measures to the commuting networks of the 25 different cities. We reiterate that because entropy values are normalised with respect to network size, we can make comparisons not only between cities, but also between commuting networks of different socioeconomic groups. Figure 2 elucidates how, for every city, the global entropy of commuting out-flows is consistently larger than the global entropy for commuting in-flows. This pattern is to be expected, as residential locations are known to be more evenly distributed than employment hubs [11]. Notably, the lower values of global in-flow entropy in cities such as New Orleans, San Francisco, and Boston indicate the presence of larger employment hubs.

By disaggregating the commuting networks based on workers' economic profiles, we can analyse the networks of low-income and high-income workers separately. Interestingly, within the cities we analyse, the global in-flow entropy of high-income commuters is persistently lower than that of the low-income group. These lower values imply less structural diversity in high-income commuting origins, which translates to a higher level of monocentricity for high-income jobs. What is clear is that there is a distinct difference in global entropy values for commuting in-flows when considering networks of different socioeconomic groups. Whether these differences are indicative of socioeconomic inequality is explored in the following sections.



**Fig. 2.** Global entropy values for 25 US cities. The grey and white bar capture global entropy of the entire commuting network for in-flow and out-flow commutes, respectively. The orange point measures  $H_{GN}^{in}$  for the low-income network, while the purple point does the same for the high income network.

### 3.2 Local Entropy as a Measure of Segregation

In this section, we aim to disentangle whether overall structural diversity entails other forms of diversity, such as lower levels of segregation. In doing so, our goal is to clarify if the monocentricity of job opportunities, which is more present in high-income networks, is a privilege or a burden.

**Residential Segregation.** To better understand the differences in how housing and employment landscapes intersect for socioeconomic groups, we evaluate whether a relationship exists between residential segregation and the diversity of employment destinations for residents of a particular tract. The fourth column of Table 1 lists the Pearson correlation coefficients when comparing the local out-flow entropy ( $H_L^{out}$ ) with the residential ICE value (Eq. 3) of census tracts in a city. We recall that the higher the  $ICE_{res}$  value, the larger the proportion of high-income residents in that area. All but three of the 25 cities have a significant, negative correlation. This reveals that census tracts characterised by more affluent residents tend to have lower local diversity values (i.e., concentrate commutes to fewer tracts), which indicate their dependence on particular tracts for supplying labour to their residents. On one hand, one can argue that these higher diversity values for less affluent tracts makes them less vulnerable to any shortages in labour supply, such that if an employment location stops providing opportunities, they have other options of commuting destinations. However, this negative correlation could also indicate the presence of inequalities in the housing landscape. Thus, we proceed to analyse the opposite dynamic, comparing the diversity of commuting origins ( $H_L^{in}$ ) to the degree of employment

segregation ( $\text{ICE}_{\text{emp}}$ ), to explore whether the identified negative correlations imply socioeconomic disparity on a residential or employment level.

**Employment Segregation.** For completeness, we have included the results for the correlation patterns between segregation of a tract’s workforce ( $\text{ICE}_{\text{emp}}$ , Eq. 3) and the diversity of commuting origins for that workforce ( $H_L^{\text{in}}$ ) in the fifth column of Table 1, which shows significant positive correlations for 20 of the cities. When considering the high-income group, we see that, from an employment perspective, tracts in which more affluent employees work are more diverse, expressing heterogeneity in commuting origins. However, from a residential perspective, tracts with more affluent residents are less diverse, expressing homogeneity in commuting destinations. Thus, we show that higher values of structural diversity do not necessarily imply an advantage.

This section elucidated how the structural diversity of census tracts often corresponds with their urban characteristics. We measure the local in and out-flow entropy of entire commuting networks to highlight how diversity of the employment landscape can be reflective of employment and residential segregation. The next section wraps up this analysis by disaggregating the entire commuting network of each city into low and high-income networks. This allows us to examine how unequal labour distribution may be exacerbating existing inequalities.

### 3.3 Socioeconomic Disparities in Diversity of Commuting Origins

In this section, we illustrate how local in-flow entropy measures of disaggregated networks can capture experienced segregation. For improved readability and due to space constraints, we show the results for four representative cities of different sizes and socioeconomic profiles (Milwaukee, San Francisco, New York City and Detroit). Nevertheless, our findings are based on the analyses of the 25 cities.

We proceed, analysing how differences in heterogeneity of commuting origins coincide with levels of employment segregation. We begin by splitting the commuting network of the city into separate networks that measure the residential-work patterns of socioeconomic groups separately. In doing so, we can understand the extent to which the diversity of commuting origins differs between the low-income and high-income workers in a census tract. Panels A-D in Fig. 3 plot the local in-flow entropies of every tract in the low-income network ( $H_{L,lo}^{\text{in}}$ ) against their respective entropy values in the high-income network ( $H_{L,hi}^{\text{in}}$ ), for four of the analysed cities. The black line expresses the case in which a census tract has equal diversity of commuting origins in both socioeconomic networks, which we see a few cases of in New York City and Detroit, indicated by the white points on the diagonal. Orange points reflect census tracts in which the low-income individuals have a more even distribution of commuting origins than the high-income workforce. The purple points capture census tracts with the opposite characteristics: greater diversity in residential locations for the affluent workforce. We observe that most census tracts in San Francisco and New York

**Table 1.** General network properties and Pearson correlation coefficients for different forms of local entropy and socioeconomic segregation. The last column refers to local entropy differences between the disaggregated networks, discussed in Sect. 3.3. Boldface is used to indicate significant correlations, with asterisks reflecting p-value.

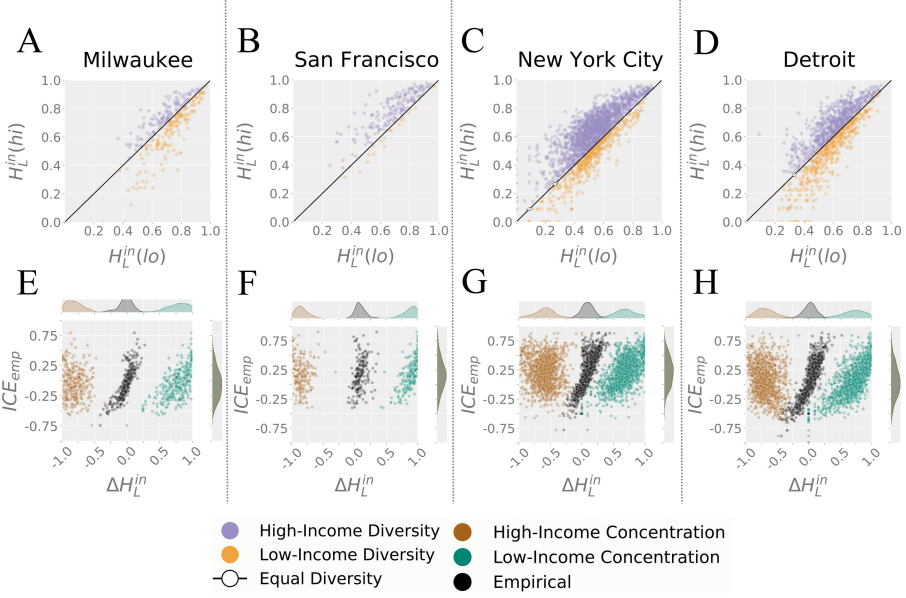
City	Network properties		$H_L^{out}$ vs. $ICE_{res}$	$H_L^{in}$ vs. $ICE_{emp}$	$\Delta H_L^{in}$ vs. $ICE_{emp}$
	<i>Nodes</i>	<i>Edges</i>	$r^a$	$r^a$	$r^a$
Charleston	85	6, 162	<b>-0.313**</b>	0.116	<b>0.453***</b>
San Francisco	196	25, 886	<b>-0.373***</b>	<b>0.614***</b>	<b>0.341***</b>
Gainesville	56	2, 803	-0.243	<b>0.313*</b>	<b>0.496***</b>
Greenville	111	10, 431	<b>-0.386***</b>	0.033	<b>0.596***</b>
Albuquerque	153	18, 871	<b>-0.385***</b>	0.116	<b>0.632***</b>
New Orleans	176	13, 680	<b>-0.309***</b>	0.140	<b>0.843***</b>
Houston	921	398, 876	<b>-0.287***</b>	-0.018	<b>0.716***</b>
Boston	204	20, 608	<b>-0.401***</b>	<b>0.652***</b>	<b>0.646***</b>
Indianapolis	224	33, 147	<b>-0.494***</b>	<b>0.231***</b>	<b>0.754***</b>
Las Vegas	487	124, 083	0.027	<b>0.093*</b>	<b>0.818***</b>
Philadelphia	384	69, 364	<b>-0.547***</b>	<b>0.289***</b>	<b>0.754***</b>
Columbus	347	76, 995	<b>-0.486***</b>	<b>0.272***</b>	<b>0.735***</b>
Hartford	224	33, 107	<b>-0.507***</b>	<b>0.430***</b>	<b>0.710***</b>
Jacksonville	173	23, 610	<b>-0.441***</b>	<b>0.473***</b>	<b>0.722***</b>
Cincinnati	222	32, 801	<b>-0.185**</b>	<b>0.306***</b>	<b>0.751***</b>
Milwaukee	297	50, 210	<b>-0.517***</b>	<b>0.377***</b>	<b>0.829***</b>
Cleveland	446	85, 609	<b>-0.182***</b>	<b>0.330***</b>	<b>0.834***</b>
Bridgeport	210	28, 259	<b>-0.425***</b>	<b>0.278***</b>	<b>0.665***</b>
Fort Worth	357	76, 883	<b>-0.483***</b>	<b>0.280***</b>	<b>0.798***</b>
Memphis	221	31, 507	<b>-0.135*</b>	<b>0.212**</b>	<b>0.816***</b>
Chicago	1,318	441, 406	<b>-0.287***</b>	<b>0.358***</b>	<b>0.855***</b>
New York City	2,164	990, 302	<b>-0.478***</b>	<b>0.455***</b>	<b>0.837***</b>
Detroit	1,163	385, 185	<b>0.139***</b>	<b>0.527***</b>	<b>0.865***</b>
Dallas	529	129, 486	<b>-0.569***</b>	<b>0.350***</b>	<b>0.830***</b>
Kansas City	283	49, 377	<b>-0.165**</b>	<b>0.328***</b>	<b>0.781***</b>

<sup>a</sup>\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

City tend to have higher homogeneity in commuting origins for low-income workers than compared to high-income workers. This comparison opens the gateway into using local entropy values to extend our understanding of segregation from a residential dimension to an employment one as well.

In order to accomplish this, we evaluate how socioeconomic disparities in the homogeneity of residential locations for a region relate to the demographic balance of that region's workforce. For each tract in a city, we compare a census





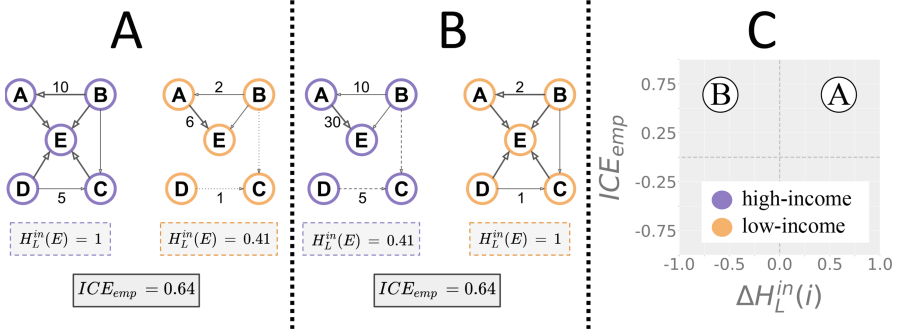
**Fig. 3.** Local in-flow entropies for high and low-income networks in 4 US cities. Panels A-D compare entropy values for tracts in the low and high-income network, with points below the diagonal reflecting tracts in which low income workers have more diverse commuting origins. Panels E-H show how the differences in these values (black points) compare to null models derived from Fig. 4

tract’s in-flow entropy value in the high and low-income networks:

$$\Delta H_L^{in}(i) = H_{L,hi}^{in}(i) - H_{L,lo}^{in}(i), \quad (4)$$

where  $H_{L,hi}^{in}(i)$  captures the local in-flow entropy of census tract  $i$  in the high-income commuting network, whereas  $H_{L,lo}^{in}(i)$  describes the in-flow entropy for  $i$  in the low-income commuting network. Thus,  $\Delta H_L^{in}(i)$  can range from  $-1$  to  $1$ , where negative values represent more heterogeneity of commuting origins for the low-income population. Positive values capture scenarios in which higher-income workers have more heterogeneous commuting origins than lower-income workers.

**Comparing Local Entropies in Disaggregated Networks.** For each of the 25 cities, we find significant positive correlations between the difference in local entropy values ( $\Delta H_L^{in}(i)$ , Eq. 4) to the level of segregation in employment areas ( $ICE_{emp}$ , Eq. 3). These correlations are outlined in the last column of Table 1. If we consider the local in-flow entropy of tract  $i$  in the high-income network, we can measure how evenly the total number of affluent individuals commuting to  $i$  is distributed across all other census tracts in the city. We use Fig. 4 to explain how entropy values are not strictly correlated with in-degrees.



**Fig. 4.** Toy example highlighting the distinction between  $\Delta H_L^{in}(i)$  and  $ICE_{emp}$ . Panel A shows concentration of high-incoming commuting origins, while B captures homogeneous origins for the low-income group. Panel C conveys how the two scenarios of commuting diversity relate to employment segregation.

The purple network captures the high-income commuting network and the orange network reflects that of the low-income workers. We set the total number of individuals working in node  $E$  to be 50. For Scenarios A and B, we can observe that the socioeconomic composition of individuals working in node  $E$  remains the same (40 high-income workers, 8 low-income workers). Thus, values of employment segregation ( $ICE_{emp}$ ) are consistent throughout the examples. What changes across the scenarios are the values of  $H_L^{in}$ , which describe how evenly commuting origins are distributed across other nodes in the network. Scenario A depicts a case where low-income commutes are more concentrated and high-income commuting origins are more diverse. Meanwhile, Scenario B captures homogeneity in high-income commuting origins and heterogeneous commuting origins for the low-income workers in the node  $E$ . We can, then, understand that  $\Delta H_L^{in}(i)$  serves to measure differences in heterogeneity of commuting origins between the high and low income networks. The right most panel in Fig. 4 uses Scenarios A and B to elucidate how positive correlations between  $ICE_{emp}$  and  $\Delta H_L^{in}(i)$  are non-trivial.

**Null Model Comparisons.** The black scatter plots in Fig. 3E-H illustrate the aforementioned positive correlations between socioeconomic differences in the heterogeneity of residences ( $\Delta H_L^{in}(i)$ ) and employment segregation  $ICE_{emp}$ , for four cities. We use null models, inspired from the examples in Scenario A and B of Fig. 4, to emphasise how the observed positive correlations are not a result of measuring related network attributes. The null models elucidate how the correlations between employment segregation and local entropy of networks are not always significant or positive. Both null models retain the in-degree of the disaggregated, empirical commuting networks, only changing how evenly a node's incoming edges are distributed across other nodes in the network.

The Low-Income Concentration null model, represented by the blue scatter plot, hypothesises that the observed positive correlation between commuting origin diversity and employment segregation is a consequence of greater diversity of origins for the affluent population than the low-income group, captured by positive values of  $\Delta H_L^{in}(i)$ . We construct this null model, such that the residential locations for a tract’s high-income workforce is uniformly sampled across all other tracts. The origins of the low-income workforce are defined by one, randomly sampled tract, producing smaller values of  $H_{L,lo}^{in}(i)$ . All the while, we maintain the empirical workplace composition for each census tract, thus retaining empirical values of  $ICE_{emp}$  and aligning the Low-Income Concentration model with Scenario A in the toy example. Meanwhile, the High-Income Concentration model, inspired by Scenario B in the toy example, and the brown scatter plot in Fig. 3, capture negative values of  $\Delta H_L^{in}(i)$  by simulating high diversity of low-income commuting origins and homogeneity for high-income origins.

We emphasise that, in both null models, the only difference from the empirical commuting network is from the distribution of incoming edges to a tract. The equivalent distributions for  $ICE_{emp}$ , shown along the y-axis, demonstrate how measures of employment segregation remain consistent across the empirical and null model scenarios, despite having disparities in residential-work patterns. On the other hand, the distributions along the x-axis, illustrate how  $\Delta H_L^{in}(i)$  can capture these structural inequalities by identifying tracts in which socioeconomic groups express stark differences in their dependence on commuting origins. While the empirical scatter plot does not show signs of extreme structural disparities, as expressed by the null models’ scatter plots, the positive correlations indicate that areas which have more heterogeneity of residential locations for a particular socioeconomic group tend to be areas in which that socioeconomic group is more concentrated. The sign of  $\Delta H_L^{in}(i)$  specifies which demographic is segregated, with values less than zero implying low-income employment segregation. Thus, this section highlights how measuring disparities in structural diversity of disaggregated networks can expose dimensions of segregation in residential-workplace dynamics, that conventional metrics of segregation may overlook.

## 4 Conclusion

This work highlights how network entropy can be used to capture sociodemographic inequalities in commuting patterns that classical segregation metrics fail to detect. We compare global in-flow entropies of 25 commuting networks across the U.S. to identify cities, such as San Francisco and Boston, that have higher degrees of monocentricity.

Then, by incorporating local entropy measures for the entire commuting network, we uncover that census tracts with a higher concentration of affluence have residents that travel to more homogenous workplaces. Meanwhile, tracts that attract a higher-income workforce express trends of heterogeneity in commuting origins. Finally, by splitting cities’ commuting networks into high and low-income networks, we demonstrate the strength of network entropy in

identifying disparities in resident-workplace trends across socioeconomic groups. We find that larger differences in node-level entropies, which measures socioeconomic disparities of a work force in terms of their residential distribution, correspond with higher levels of employment segregation.

This brings into question what has long been deliberated in sociological fields: how the consequences of segregation may change depending on which demographic is segregated. Future work can examine how these disparities in commuting flows reflect in other aspects of urban life. Specifically, other null models can be explored to understand what mechanisms may be fuelling the strong correlation between entropy and segregation. Moreover, this framework can be applied to other mobility networks in the context of various demographic dimensions, such as gender or age.

## References

1. Bokányi, E., Juhász, S., Karsai, M., Lengyel, B.: Universal patterns of long-distance commuting and social assortativity in cities. *Sci. Rep.* **11**(1), 1–10 (2021)
2. Booth, A., Crouter, A.C.: The prodigal paradigm returns: ecology comes back to sociology. In: Does it take a village?, pp. 53–60. Psychology Press (2001)
3. Bureau, U.C.: 2019 American community survey 5-year estimates, table b19001 (2022). Accessed 15 Dec 2022
4. Bureau, U.C.: LEHD origin-destination employment statistics data (2002–2019) (2022). Accessed 15 Dec 2022. Longitudinal-Employer Household Dynamics Program, LODES 7.5
5. Diaz, R., Garrido, N., Vargas, M.: Segregation of high-skilled workers and the productivity of cities. *Regional Sci. Policy Pract.* **13**(5), 1460–1478 (2021)
6. Goetz, S.J., Han, Y., Findeis, J.L., Brasier, K.J.: Us commuting networks and economic growth: measurement and implications for spatial policy. *Growth Chang.* **41**(2), 276–302 (2010)
7. Hackl, A.: Mobility equity in a globalized world: reducing inequalities in the sustainable development agenda. *World Dev.* **112**, 150–162 (2018)
8. Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., González, M.C.: The TimeGeo modeling framework for urban mobility without travel surveys. *Proc. Natl. Acad. Sci.* **113**(37), E5370–E5378 (2016)
9. Lenormand, M., Samaniego, H., Chaves, J.C., da Fonseca Vieira, V., da Silva, M.A.H.B., Evsukoff, A.G.: Entropy as a measure of attractiveness and socioeconomic complexity in Rio de Janeiro metropolitan area. *Entropy* **22**(3), 368 (2020)
10. Louail, T., et al.: Uncovering the spatial structure of mobility networks. *Nat. Commun.* **6**(1), 1–8 (2015)
11. Marin, V., Molinero, C., Arcaute, E.: Uncovering structural diversity in commuting networks: global and local entropy. *Sci. Rep.* **12**(1), 1–13 (2022)
12. OECD: Compendium of productivity indicators 2013 (2013)
13. Qian, H.: Diversity versus tolerance: the social drivers of innovation and entrepreneurship in us cities. *Urban Stud.* **50**(13), 2718–2735 (2013)
14. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Techn. J.* **27**(3), 379–423 (1948)
15. Veneri, P., Comandon, A., Garcia-López, M.À., Daams, M.N.: What do divided cities have in common? an international comparison of income segregation. *J. Reg. Sci.* **61**(1), 162–188 (2021)