

Social Media Inequality: Definition, Measurements, and Application

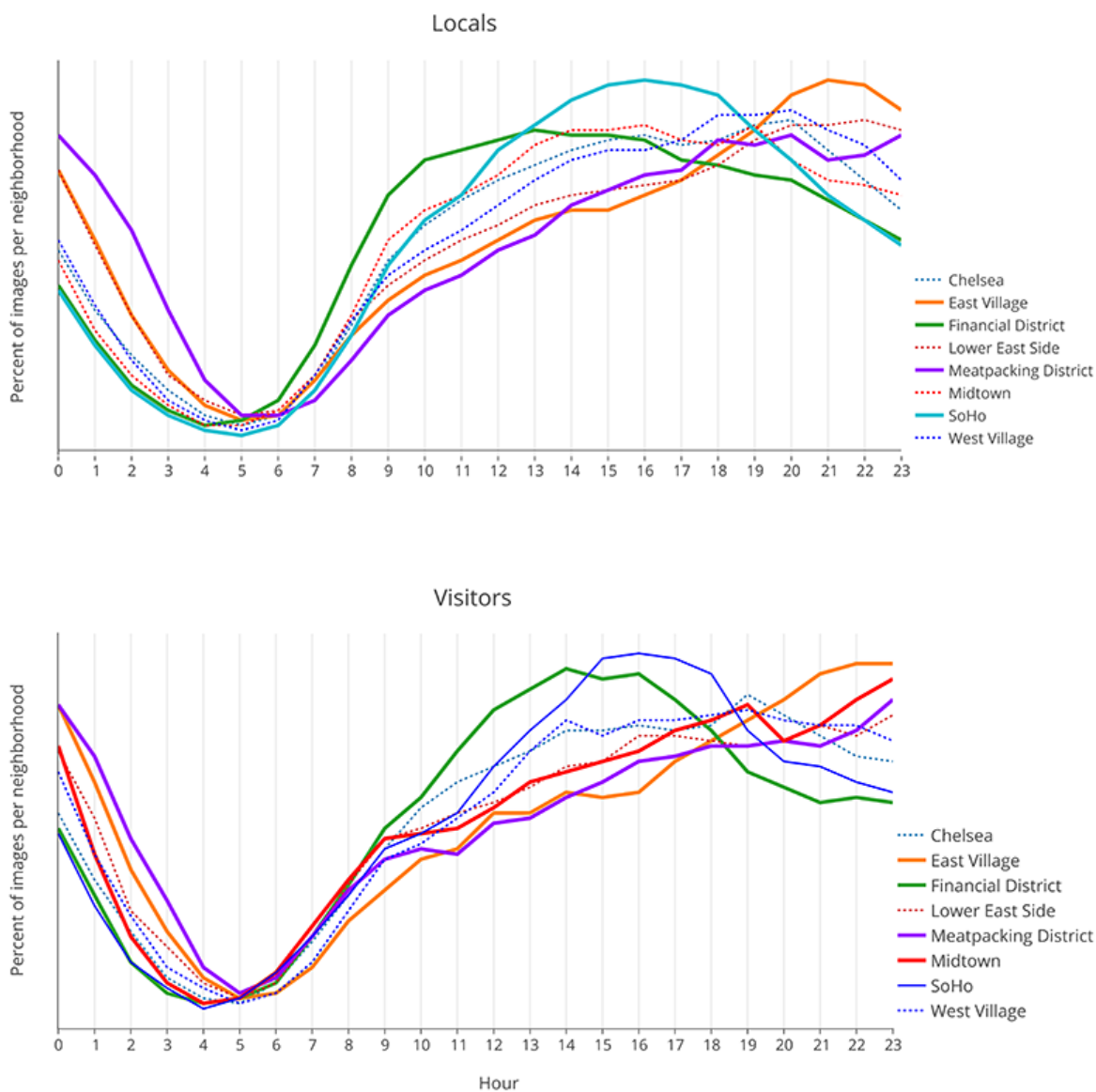
Agustin Indaco, Economics, The Graduate Center, City University of New York

Lev Manovich, Computer Science, The Graduate Center, City University of New York

04/10/2016

Forthcoming in *Urban Studies and Practices* journal, 2016.

The plots below compare patterns in Instagram sharing between local Manhattan residents and visitors. X axis shows hours in a 24-hour cycle (aggregated using our dataset covering



2/2014 - 7/2014). Y axis shows proportions of images shared in each hour.

Keywords:

inequality, social media, Instagram, Gini coefficient, urban analytics, urban science, urban studies, spatial theory

Abstract:

Social media content shared today in cities, such as Instagram images, their tags and descriptions, is the key form of contemporary city life. It tells people where activities and locations that interest them are and it allows them to share their urban experiences and self-representations.

Therefore, any analysis of urban structures and cultures needs to consider social media activity. In our paper, we introduce the novel concept of *social media inequality*. This concept allows us to quantitatively compare pattern in social media activities between parts of a city, a number of cities, or any other spatial areas.

We define this concept using an analogy with the concept of economic inequality. Economic inequality indicates how some economic characteristics or material resources, such as income, wealth or consumption are distributed in a city, country or between countries. Accordingly, we define social media inequality as unequal spatial distribution of social media sharing in a particular geographic area or between areas. To quantify such distributions, we can use many characteristics of social media such as number of people sharing it, the number of photos they have shared, their content, and user assigned tags.

We propose that the standard inequality measures used in other disciplines, such as the Gini coefficient, can also be used to characterize social media inequality. To test our ideas, we use a dataset of 7,442,454 public geo-coded Instagram images shared in Manhattan during five months (March-July) in 2014, and also selected data for 287 Census tracts in Manhattan. We compare patterns in Instagram sharing for locals and for visitors for all tracts, and also for hours in a 24 hour cycle. We also look at relations between social media inequality and socio-economic inequality using selected indicators for Census tracts. The inequality of Instagram sharing in Manhattan turns out to be bigger than inequalities in levels of income, rent, and unemployment.

Introduction

Social media content shared today in cities, such as Instagram images, their tags and descriptions, is the key form of contemporary city life. It tells people where activities and locations that interest them are and it allows them to share their urban experiences and self-representations. Social media also has become one of the most important representations of city life to both its residents and the outside world. One can argue that any city today is as much media content shared in that city on social networks as its infrastructure and economic activities.

For these reasons, any analysis of urban structures and cultures needs to consider social media activity and content. While the industry developed many concepts and measurement tools to analyze social media, these concepts and tools were not developed with the view for the comparative *urban* analysis. Therefore, we need to develop our own concepts that bridge the perspectives of urban studies and design and quantitative analysis of social networks that uses computational methods and “big data.”

In the last few years, one of the most frequently discussed public issues has been the rise in income inequality (Stiglitz 2012, Piketty 2014, Atkinson 2015). But inequality does not only refer to distribution of income. It is a more general concept, and it has been used for decades in a number of academic disciplines besides economics, such as urban planning, sociology, education, engineering, and ecology. The quantitative measurements of inequality allow researchers to characterize a set of numbers or compare multiple sets, regardless of what the data represents. In addition to income inequality, we can measure inequality in wealth, education levels, social well-being, and numerous other social characteristics.

In our paper, we introduce the novel concept of *social media inequality*. We define this concept using an analogy with the concept of economic inequality. Economic inequality refers to how some economic characteristics or material resources, such as income, wealth or consumption are distributed in a city, country or between countries (Ray 1998, Milanovic 2007, OECD 2011). Accordingly, we can define social media inequality as *the measures of distribution of characteristics of social media content shared in a particular geographic area or between areas*.

An example of such characteristics is the number of photos shared by all users of a social network such as Instagram in a given city or city area. Another example is the number of hashtags – how many hashtags users added to the photos, and how many of these hashtags are unique. Other examples include average number of tweets shared by a user in a particular period; numbers of tweets shared per month, per week or per hour of a day; the proportions of tweets that were retweeted, and so on. Of course, we can computer and analyze features of content itself - for example, how many different subjects appear the photos, and what are their proportions. In fact, any metric of social media can be used to compare inequality in social

media activity between areas - for example, number of likes, length of text messages, most frequent and least frequent words, number of unique topics, number of distinct photographic styles, image compositions, styles of video editing, and so on.

We propose that the standard inequality measures used in other disciplines, such as the Gini coefficient, can also be used to characterize social media inequality. We can also compare these measures between content shared on various social networks (Instagram, Twitter, etc.) in the same area or areas. We can do these comparisons for social networks where the main content is text (e.g., Twitter, VK), images (e.g., Instagram, Tumblr), video (e.g., YouTube), or combination of different media (e.g. Facebook, QZone, Sina Weibo, Line, etc.). Finally, we can also compare characteristics of shared content with various social and economic characteristics in the same areas, such as income, rent, the level of education, or ethnic mix.

The paper tests some of these ideas using a large dataset of Instagram images shared in Manhattan borough of New York City. This dataset, which we created for this study, contains 7,442,454 public geo-coded Instagram images shared in Manhattan during five months (February 1, 2014 July 31, 2014). Among these images, 1,524,046 were shared by 515,608 city visitors; the remaining 5,918,408 images were shared by 375,876 city residents. Our analysis of the images shared by two types of users in this paper is inspired by the pioneering project *Locals and Tourists* created by Eric Fischer (Fisher, 2010.)

We used the following method to divide users into “visitors” and “locals.” If a user posted all her photos within a single 12-day period out of the total five months, we consider this person a “visitor”. If a user shared a minimum of two photos within any interval larger than 12 days, we consider this person a local. Although this very simple method is expected to produce some errors, we felt that they are acceptable given the size of our dataset. (Such method is also used a number of published papers that analyze patterns in social media activity.)

Comparing the locations of images shared by visitors (figure 1) and locals (figure 2) gives us an intuition for social media inequality concept. We can immediately notice that in each case these locations are not distributed evenly. Some parts of the city have many more images than other parts. These figures also suggest that the big proportion of images by city visitors are shared only in a few areas, while the locals share images in most areas of the city.

Note that we use the term “shared” rather than “captured” because Instagram allows sharing of any image from user’s phone and not only the ones captured within Instagram app. So users can upload images taken previously in other locations. However, since Instagram captured the geolocation and time when an image was shared (for users who allowed Instagram access to this data), the metadata of images in our dataset tells us about people’s presence at particular place in the city at a particular time.

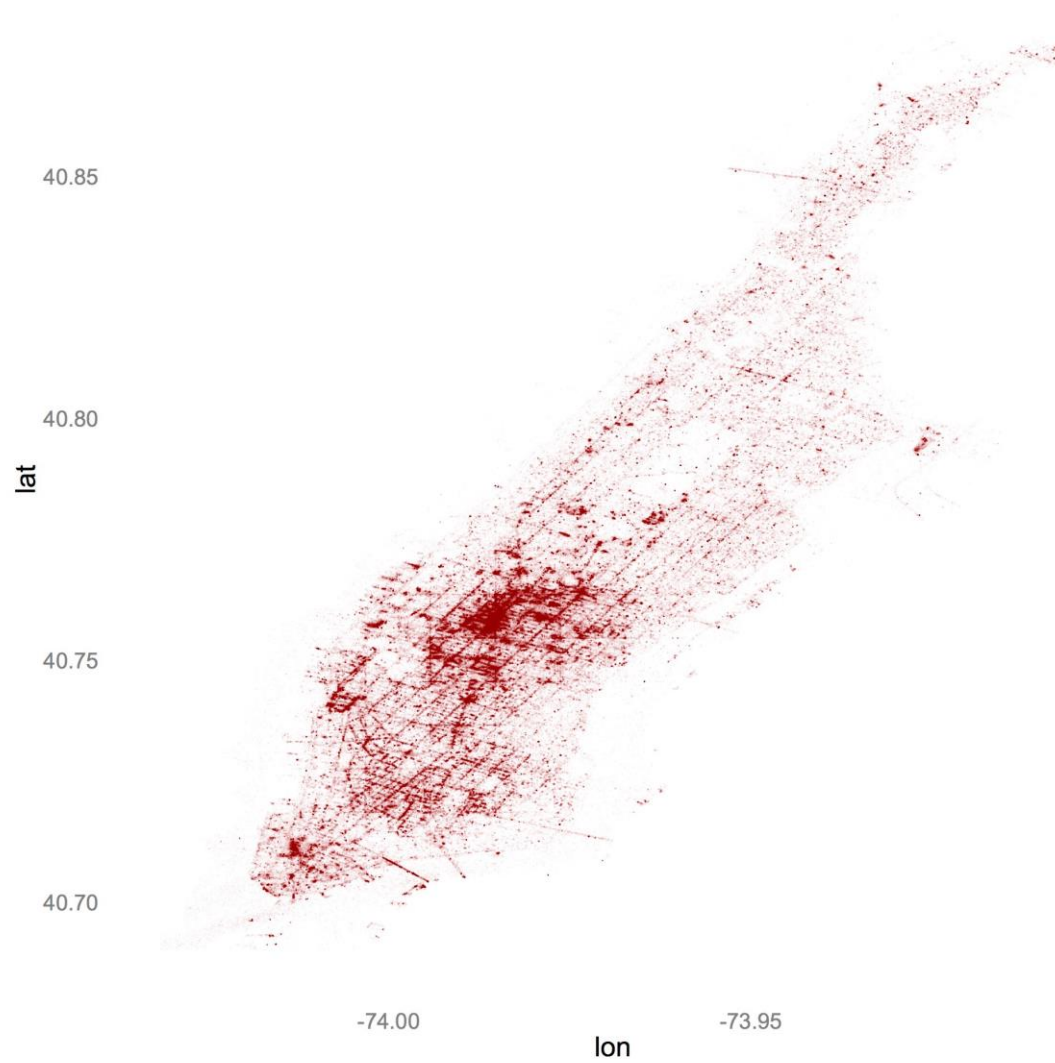


Fig. 1.

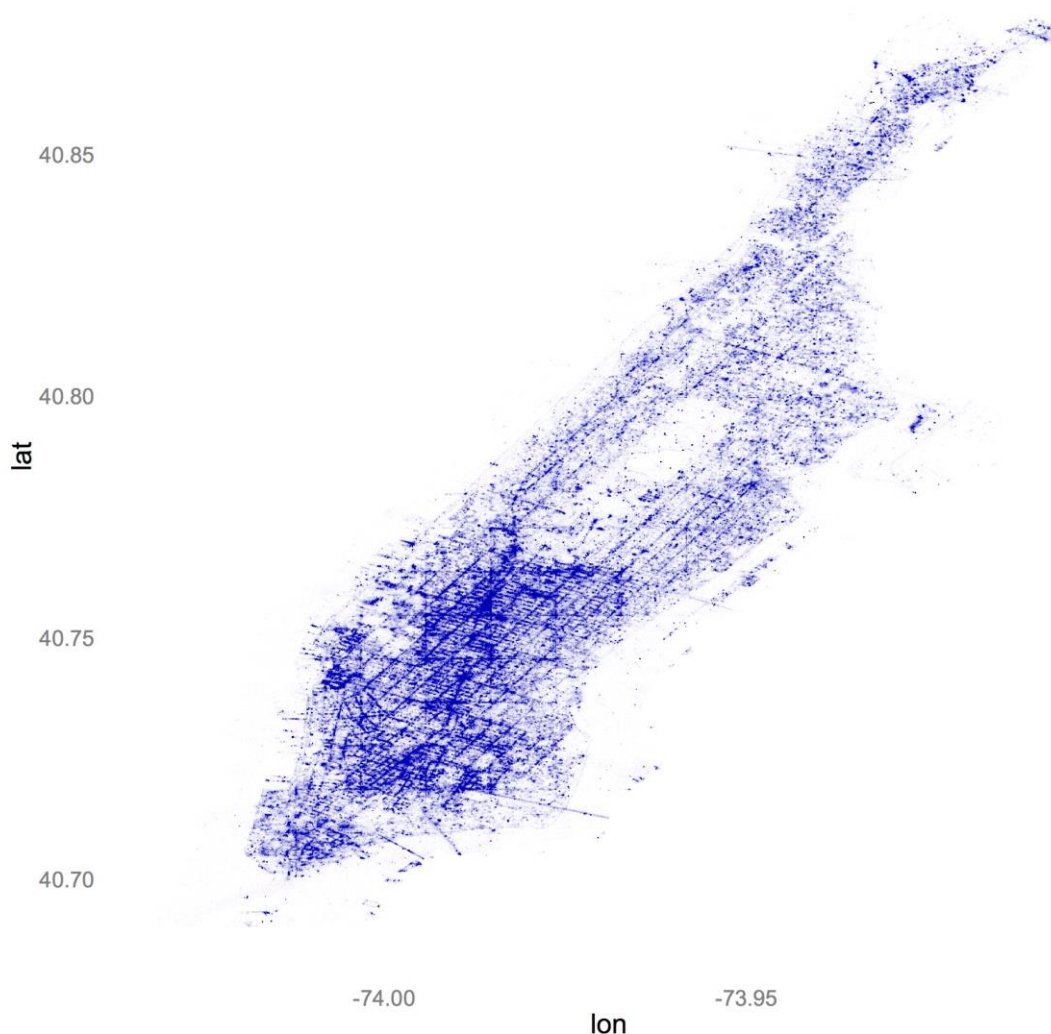


Fig 2.

Visualizing the locations of shared images gives an intuition for spatial social media inequality, but we need some measuring instruments to quantify such inequality. And what if we want to compare inequality not only in the number of images shared, but also in other characteristics we listed above (numbers shared per hour of the day, numbers of unique words in hashtags, etc.)? As characteristics multiply, the need for quantitative measurements becomes stronger. Our paper proposes such measurement instruments and tests them using a few characteristics of images and accompanying metadata in our dataset.

In principle, we could also study social media inequality between individuals living in a city. This will be similar to how economists measure income inequality by comparing people's income, rather than average income by area. However, to do this would require disclosing the

identity of the individual behind a social media account, and thus going against privacy norms accepted in most countries today. At least until now, social networks such as Instagram, Twitter, Facebook and others allowed researchers to download content shared by their users, but they did not disclose any user information beyond what users made visible on their account pages.

While the U.S. Census collects data on individuals, it only reports the data aggregated by geographic areas at different scales. We follow a similar logic in our analysis of spatial social media inequality by dividing a city into hundreds of small areas and aggregating characteristics of social media content shared in each area - as opposed to comparing individuals to each other. The way we measure social media inequality is comparable to how Milanovic defines one of the measures of global economic inequality (Milanovic 2006, Concept 1). This measure uses countries as the units of observation. Milanovic does not directly compare the income of people worldwide. Instead he compares average income across different countries to calculate global inequality. In our case, the Census tracts are our units of observation. We aggregate social media characteristics at the tract level in order to analyze social media inequality across all of Manhattan.

Social media content shared in a given area may combine contributions from different kinds of users: people who reside in this area, people who live in different parts of the city or in suburbs but spend significant time in this area for work during weekdays; international or domestic tourists visiting a city; companies located in this area, and so on. Together, the content shared by all these users create a collective “voice” of a particular area of a city. A city as a whole can be compared to an orchestra of all these voices (although, of course, they are not necessary performing the same composition.) Applying the concept of inequality to a collection of these urban voices can give us new ways of understanding a city, and provide an additional metric for comparing numerous cities around the world.

Social media inequality as we define it refers to the unequal distribution of social media content and its metadata and their characteristics in any type of geographic area – a city, a region, a country, or any other type of area. However, as Fischer’s maps show visually, the density of social media contributions in larger cities is much higher than in non-urban areas, which makes these cities particularly convenient areas of study. We think that our proposed measurements of social media inequality can be useful for urbanism studies, urban planning, urban design, public administration, economics, and other professional and academic fields. While researchers in the fields of social computing, spatial analytics, and “science of cities” have published many quantitative studies analyzing urban data of many kinds (Batty 2013, Goldsmith and Crawford 2014, Townsend 2014, Pucci et al. 2015, Ratti et al. 2006), a significant portion of this analysis cannot be approached without having a degree in computer science. In contrast, social media inequality measurement is a concept that is easy to understand and also easy to calculate.

The locations of social media contributions reflect the presence of people in a particular part of a city at a particular time. However, in comparison to pure location data captured by mobile phones or other body sensors, social media images are much more than simple coordinates and time stamps. The content of these contributions can also tell us what people find interesting and how they are spending their time. Therefore, mapping and measuring inequality in

characteristics of social media can help us understand how social, economic, and urban design characteristics of cities influence life patterns and the overall “dynamism” and “vitality” of a city.

Researchers have never observed perfect equality in any natural, biological or social system or population. In using the term “social media inequality,” we are not suggesting that the goal of urban planners or city administration should be to reduce differences in social media use between various areas to a minimum, or to some optimal level. If people are sharing the same amount of social media in every area of the city, it means that this city does not have any centers or attractions that stand out, or places where many people gather. In terms of modern housing, large American-type suburbs with the same density of houses and same demographics of families and income would probably generate least amount of social media inequality. Today such suburbs are common around the world, from Mexico to China. Given the wide criticism of this classical suburb type, we can assume that *some level of spatial social media inequality is desirable*. In this case, inequality stands for variety and differentiation while complete equality stands for sameness and lack of variety.

But is extreme social media inequality a good thing? For example, do we really want all people living in a city to spend their weekends in a single place? There are certain situations where reducing extreme spatial social media inequality would be very desirable. For example, if city authorities find that most tourists’ social media activity is concentrated in just a few areas surrounding only a few landmarks (like Times Square in New York City), they can change the way the city is promoted to visitors to diversify where tourists go, what they look at, and what they experience. Being able to quantify inequality of social media would allow for better planning and evaluation of such changes.

Formulated as a type of spatial analysis, our study compares the parts of the city that attract more people and generate more content shared on social media networks and thus are “social media rich” with parts of the city that are “social media poor.” What are the relationships between such social media rich and social media poor areas? Is social media inequality larger or smaller than economic or social inequality in the same areas? Does social media inequality increase worldwide, similar to how economic inequality has been growing recently? Which parts of the world have the highest social media inequality and which are the most equal? Although our analysis is focusing on one part of a single megacity (i.e., Manhattan in New York City), it can be expanded to consider hundreds of cities around the world to consider such questions.

Analyzing social media inequality using volumes and locations of shared content

We start the analysis of social media inequality by looking at a single characteristic – number of images shared in different parts of a city during a given time period. To calculate the amount of inequality, we can divide a city into a number of equal size parts using a grid. In the case of complete equality, every part will have exactly the same number of images. In the case of absolute inequality, one part will have all the images, and the rest will have none.

To quantify our perception that the shared images are distributed non-equally (see figures 1 and 2), we use the following procedure. First, we add up the numbers of images shared in every one of 287 Census tracts in Manhattan. (Note that we could have also chosen any other type of areas – for example, we could have divided Manhattan into small parts using a rectangular grid. However, as we will later use selected indicators reported by Census per tract, it is convenient to use tracts to compare images distribution). Figure 3 shows the Census tracts in Manhattan with colors indicating the relative number of images shared in every tract by local users and hashtags they added to these images.

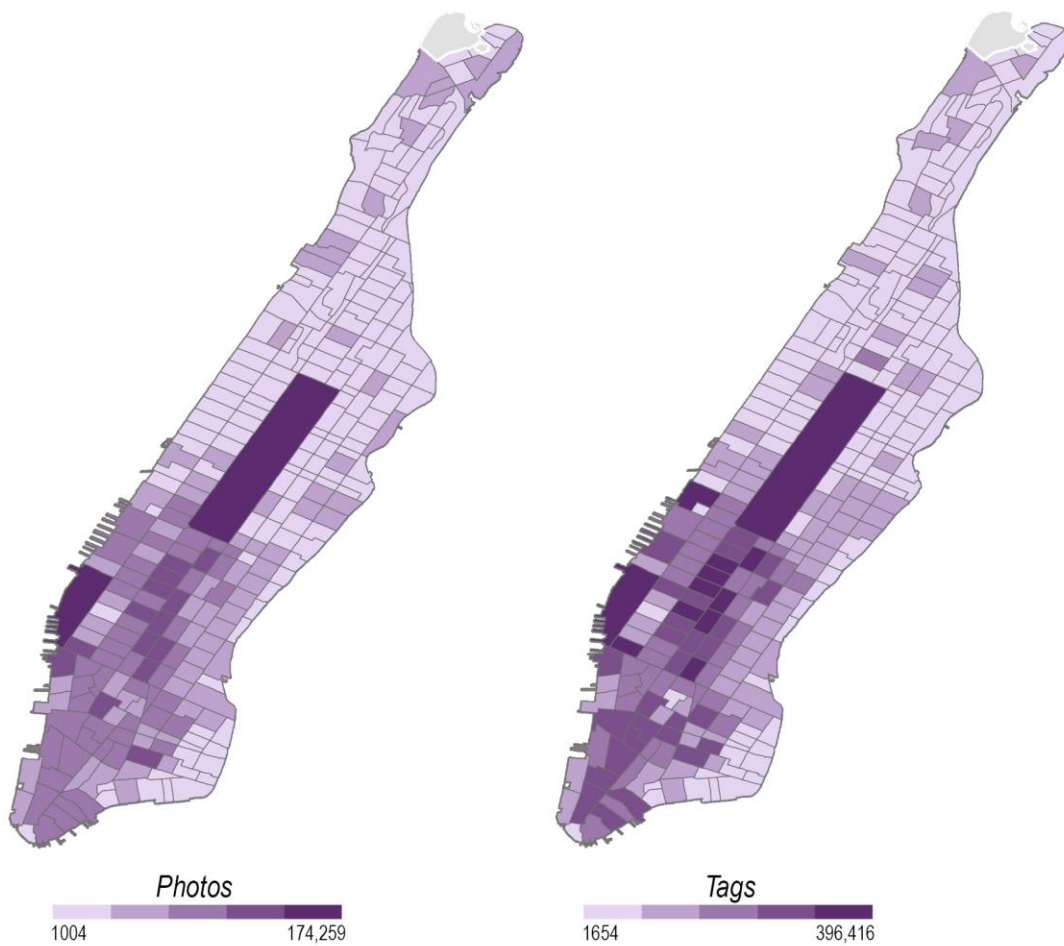


Fig. 3.

Now that we have aggregated number of images per tract, we can use standard measures for measuring inequality. Since the Gini coefficient is the most popular method for measuring inequality used in many fields, we will use it to measure inequality of spatial distribution of

Instagram images. (We use R package *ineq* to calculate all Gini measurements reported in this paper).

Confirming what we already noticed in figures 1 and 2, Gini inequality coefficient turns out to be much larger for visitors than for locals: 0.661 and 0.468, respectively. In other words, visitors' social media inequality is 1.41 times larger than locals' inequality.

The likely explanation is that visitors tend to capture and share images only in particular parts of the city, ignoring many other parts completely. In fact, more than 50% of all images by visitors are shared in only 8.3% of all tracts in Manhattan (24 tracts out of all 287). These tracts cover only 12% of the total area of Manhattan.

While the locations of images shared by locals are also distributed non-equally, the amount of inequality is significantly lower: 50% of their images are shared in 18.4% of all tracts (53 tracts out of 287). These tracts cover approximately 21% of the total Manhattan area.

In general, we may expect that larger areas will have more people living or visiting and therefore these areas will have more shared images. Given that the geographic sizes of Census tracts vary significantly, with largest tracts 10 times bigger than the smallest tracts, we decided to normalize our data by tract size. The rest of this section uses such normalized data.

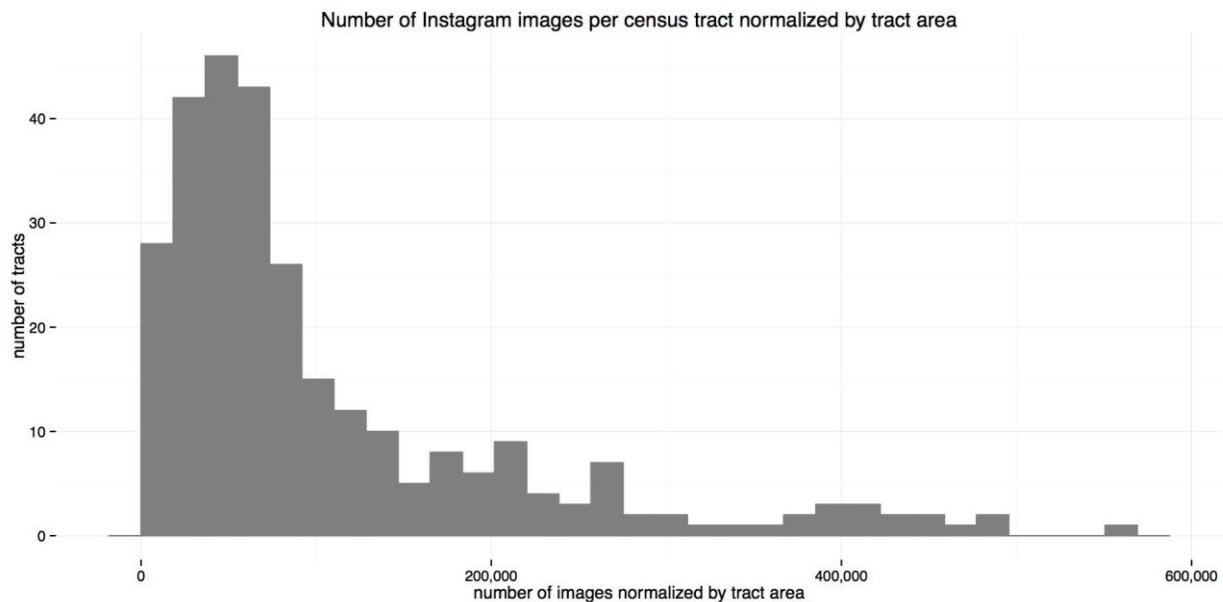


Fig. 4.

Fig. 4 shows the distribution of numbers of images shared by locals per square kilometer after the data was normalized. The number of images varies from 2,127 per sq. km to 552,787 per sq. km, and the mean is 106,431. Now that we are comparing the volume of images for equal size

areas after normalization, we see that the differences in “social media coverage” between parts of a city are actually much larger. The ratio between sq. km areas with most (552,787) and least images (2,157) is 256,275, i.e. a quarter of million times!

The Gini coefficients for images calculated using normalized numbers are 0.669 for visitors and 0.494 for locals. To put this in context, we can compare our social media Gini coefficients to the Gini coefficients for countries’ income. While income inequality and social media inequality are defined and calculated differently, these comparisons are relevant. If we find a large inequality in any population on any dimension, this is an important characteristic of this population.

Social media inequality of visitors’ images in Manhattan (Gini = 0.669) is larger than income inequality of the most unequal country in the world (Seychelles where Gini = 0.658). On the other hand, social media shared by locals has a Gini coefficient similar to countries that rank between 25 and 30 in the list of countries by income inequality. These are countries like Costa Rica (0.486), Mexico (0.481), and Ecuador (0.466).

Gini coefficient for income in New York City is 0.594. (It is the most unequal among all American cities according to the U.S. Census Bureau 2014). Interestingly, income inequality in New York City seems to lie approximately in the middle between social media inequality of visitors and locals (0.669 and 0.494, respectively).

Figure 5 visualizes Gini inequality measurements for images shared by visitors and locals using Lorenz curves. Perfect equality (Gini coefficient = 0) corresponds to a straight line at a 45-degree angle. The more curvature a line has for a particular dataset, the more unequal is its distribution. While we already know that distributions for both visitors and locals are highly unequal, and that the former is larger than the latter, the figure also shows that both distributions have similar shapes.

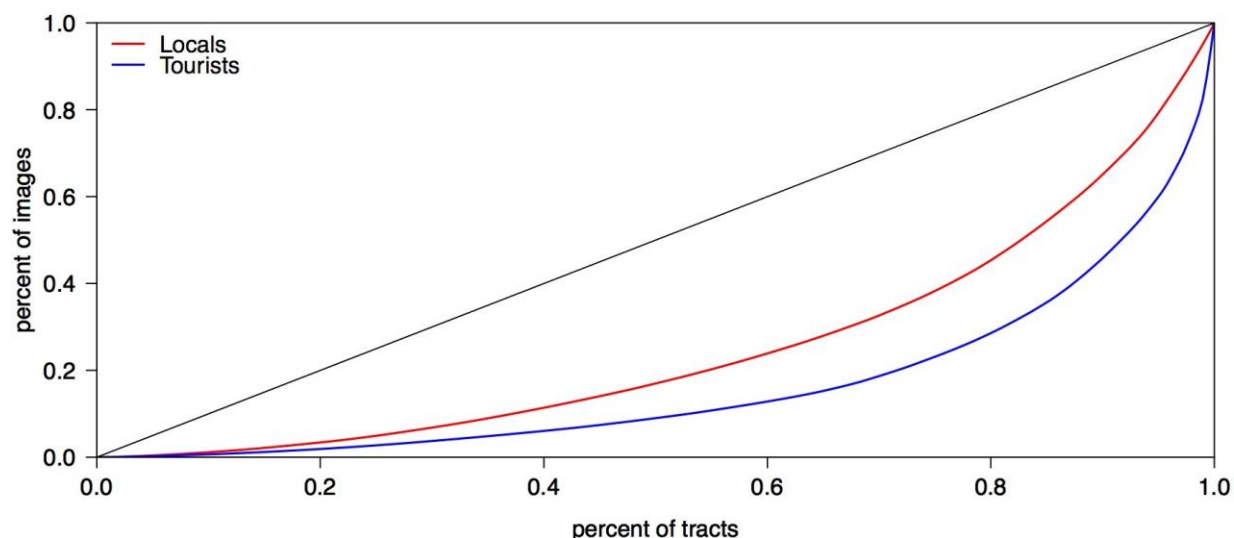


Fig 5.

Adding time dimension to the analysis of social media inequality

In the previous section we analyzed spatial social media inequality for Instagram images shared in Manhattan. To do this, we aggregated locations of millions of images shared over 287 tracts and then compared differences in the volume of images between these tracts. But we have not yet taken advantage of the key difference of social media data from typical 20th century social data – its temporal granularity and density.

Each image shared on Instagram has a time stamp specifying the date, hour, minute, and second when the image was shared. Therefore we can calculate how many images were shared in a given time interval. Therefore, similar to how we did it with space, we can apply inequality measures to compare differences in popularity of time intervals. For example, we can combine data for images shared by locals in Manhattan over five months to calculate average numbers shared per day of a week, and then compare daily volumes. If people share the same number of images every day of the week, the temporal inequality across days for an average week will be 0. If the numbers differ very substantially between days, the “temporal inequality will be close to 1. Using this logic, we can ask all kinds of questions about temporal patterns. Is weekly inequality bigger for visitors or locals? What are the inequality patterns for days of the week, hours of the days, seasons of the year, and so on.

Because the types and the “volume” of human activities change significantly between hours of a day or day of the week – being at work, being at home, sleeping, being active, being with family or friends, commuting, and so on – we need to consider the temporal dimension of social media. But most importantly, the availability of both spatial and temporal metadata for social media content allows us to conceptualize and study cities in new ways. Rather than thinking of social media inequality as a characteristic of a geographic area, as we did in the previous section, we can view it as a dynamic spatiotemporal variable. From this perspective, a city appears not as a static collection of buildings, their residents, firms producing products, and public places but as the aggregations of individuals that follow periodic rhythms in space and in time.

The first temporal analysis of Instagram city patterns was presented in Hochman and Schwartz, 2012. In *Phototrails* project and accompanying paper Hochman and Manovich extended this work by analyzing spatial and temporal patterns for 13 global cities using 2.3 million Instagram images shared over a few months (Hochman and Manovich 2013). Fig. 6 and 7 are two of the visualizations from this paper. Fig. 6 shows temporal patterns in Instagram images shared in parts of Tokyo and New York City over continuous time periods. 50,000 images from each city area are visualized in the order they were shared, top to bottom and left to right. Fig. 6a is New York, and Fig 6b is Tokyo.

We can see repeating day to night patterns in brightness - lighter during the day, and darker at night. But each particular 24 hour interval in every city is also unique. Some days on Instagram are longer (more images are shared), and some are shorter. The colors are also not exactly the same in each period. The uncoordinated images shared by thousands of people at the same time inside city area come together to form a “city symphony,” with each “instrument” adding its own unique signature. As we can see, the temporal image of a city on Instagram alternates between repetition and variation, predictability and unexpected events, following routines and breaking them. (For the purpose of comparison between many city areas, many cities and many periods, we can disregard these variations and create statistical models that account for the regular part. But as representations of complexity of city life, such visualizations consisting from the actual shared images have their advantages, since they show both the regular and the irregular.)



Fig. 6a.

Fig 6b.

Fig. 7 shows patterns in time and space together on the level of individuals. It compares 289 Instagram users in Tel Aviv that uploaded most Instagram images with geo-locations during three months in spring 2012. Each plot shows locations of photos shared by a particular person during this period. The green to red color gradient indicates the time when an image was shared (green — morning, yellow — afternoon, red — evening). A line is drawn between two dots if

corresponding photos were shared within the same hour. If we consider total social media content as a type of resource produced by people in the city, quantifying inequality allows us to understand how this resource is distributed spatially and temporally. We may expect that every city will have its own distinct signature of spatial-temporal social media inequality. These signatures reflect where people who share content on a particular social media service or services spend their time, including the waves of commuters traveling daily for work, locals going to other areas for leisure activities, visitors shopping and sightseeing, and so on. Many areas get “activated” during different days of the week and hours of the day. Each can also have different types of users being more or less active at different times.

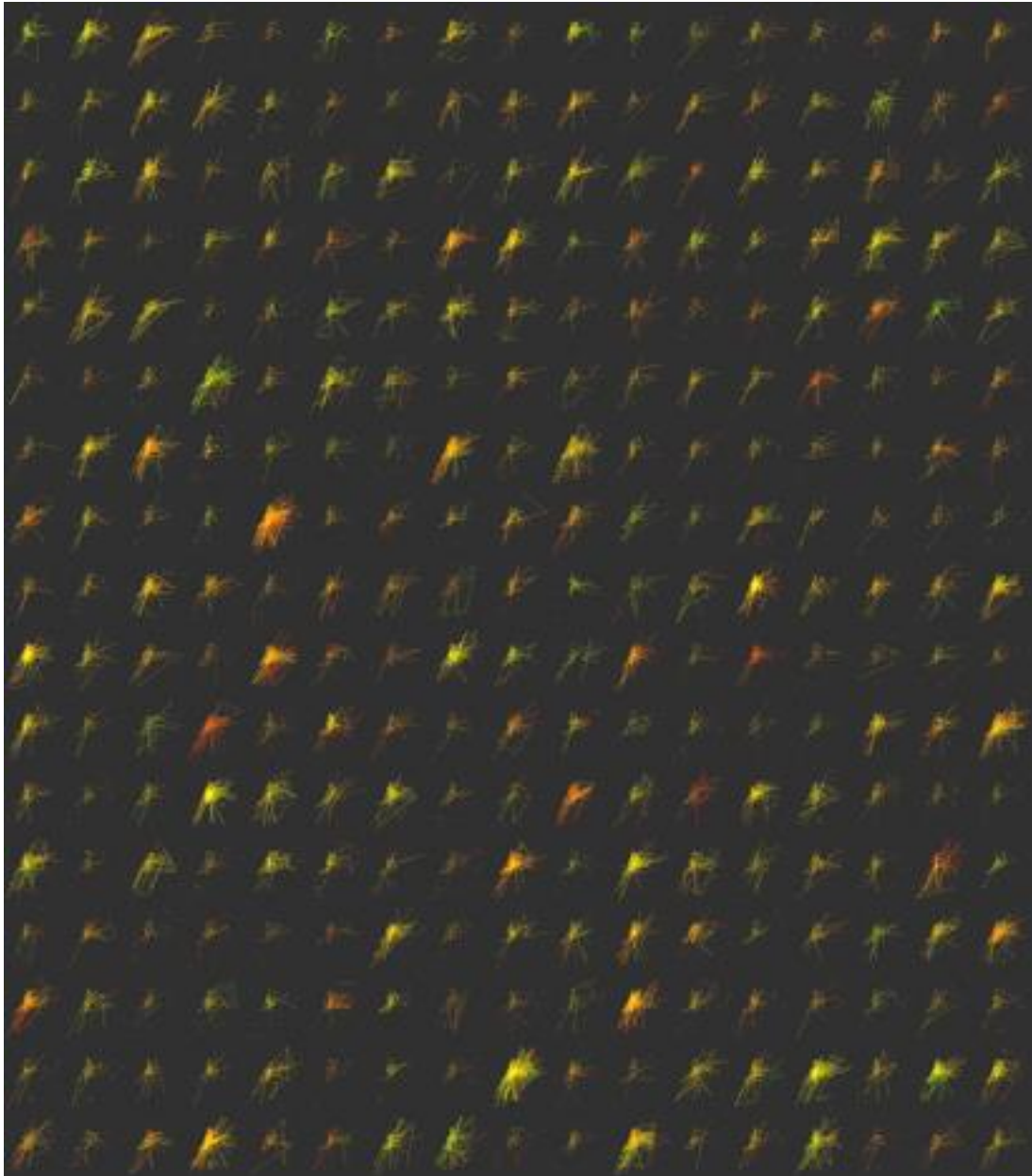


Fig. 7.

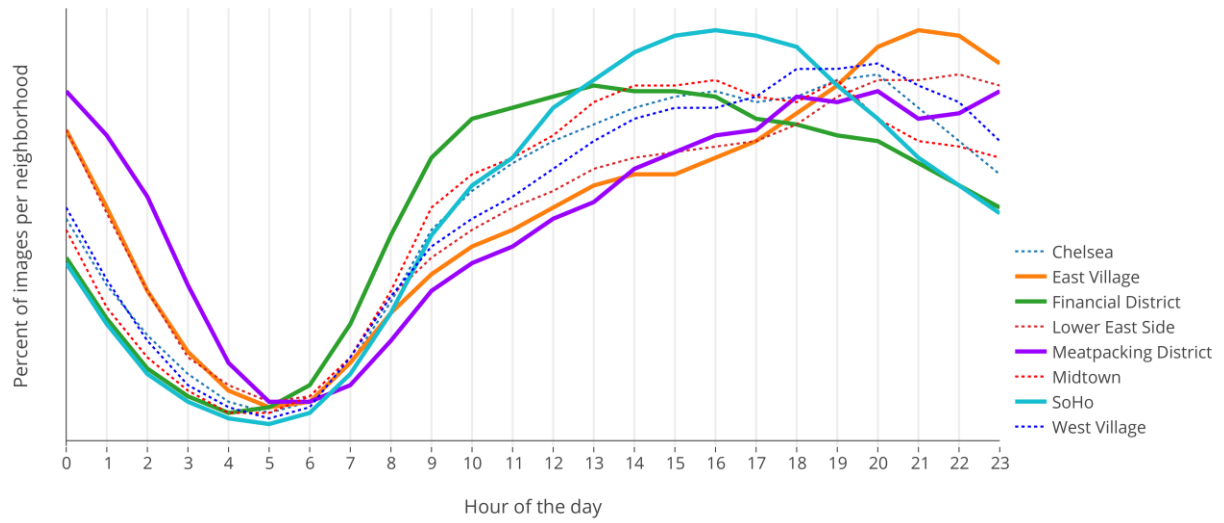
As we already noted, analyzing and visualizing these patterns moves us away from the image of a city as a static map of physical structures that change very infrequently. Instead, we get a multi-dimensional “volume” that reflects where people are and what they do every hour. Three dimensions of such volume would correspond to space; another dimension would correspond to time; others can indicate types of users; still others would code different kinds of social media characteristics such as volumes of messages, their content uniqueness, etc.

Figure 8 shows three such slices of the “volume” of Manhattan - hourly proportions of images shared by locals and by visitors in selected neighborhoods. To show patterns in a 24 hour cycle, we aggregated all data we have for five months. The graphs reveal important differences in Instagram sharing between neighborhoods and between locals and visitors. (Because the volume of images shared by visitors in the neighborhoods above 110 street is quite low, these patterns are not as reliable, so we are not including this graph).

For locals, we see a clear separation between work areas (Financial District), shopping areas (SoHo) and nightlife areas and nightlife areas (East Village, Lower East Side, Meatpacking District). For visitors, these patterns are similar but even more clearly pronounced (for example, volume in East Village keeps increasing after 9 pm). One exception is Midtown: here visitors’ activity keep increasing late in the evening. The possible reason is that this areas has biggest concentration of hotels in the city and also key attractions such as Time Square.

The patterns in the neighborhoods located above 59th street are different. Overall, the activity keeps increasing and peaks latter in the evening. And in the areas with lower average income (East Harlem, Morningside Heights, and Washington Heights), the activity keeps increasing until midnight. We will come back to these differences in the next section where we will offer a possible explanation.

Locals in selected Manhattan neighborhoods - 1



Locals in selected Manhattan neighborhoods - 2

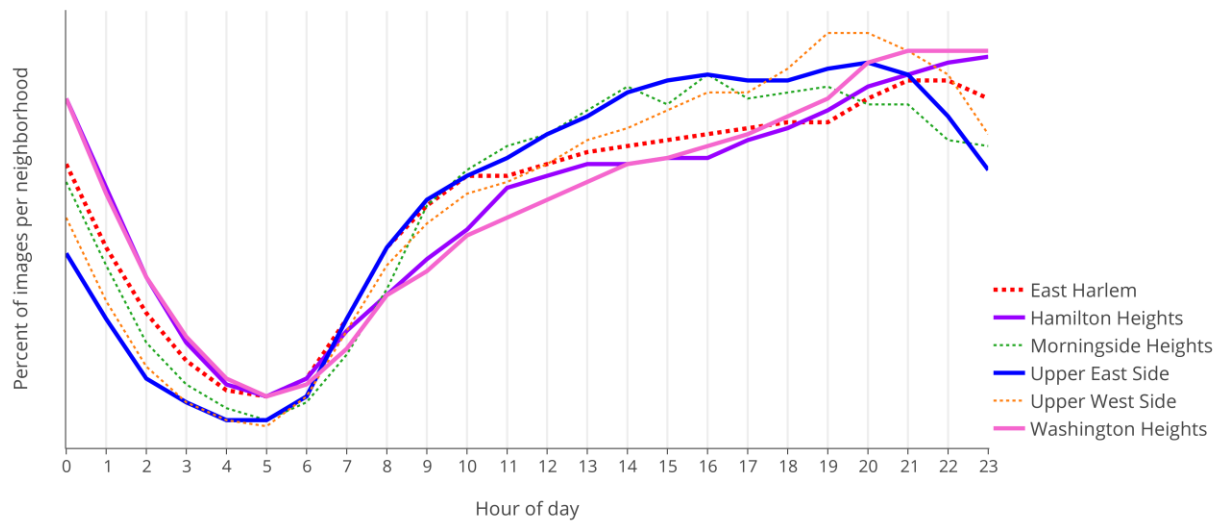


Fig. 8.

Many other slices can be equally revealing. Analyzing data in every slice will produce a different social media inequality measurement. This suggests that to better characterize social media activity in a city, we need to measure the “inequality of inequalities” – for example, the distribution of inequality indexes for each day in a year, or the distribution of spatial inequalities at different spatial scales, and so on. Construction of such “super-index” can be the interesting subject for future research.

Finally, we should also mention another important consequence of considering both time and space together in social media analysis. So far we looked at both time units such as hours and spatial “semantic” units such as neighborhoods as fixed entities. However, the continuously changing patterns of sharing create clusters in time and in space. In one time intervals some neighborhoods may have similar patterns, thus forming a single larger cluster. In another time interval we may see smaller parts of different neighborhoods having the same pattern, but not a neighborhood as a whole. This does not mean that the boundaries of “neighborhoods” (or another type of spatial unit) are completely irrelevant for a “social media city.” Rather, social media shares are likely to construct their own map of divisions that change periodically over time. Sometimes they may overlap with neighborhood divisions, and other times they may have little in common with them. The same holds for time. A 24-hour cycle may get divided into a few periods depending on volumes of shared images, gradual or rapid increase or decrease, or other patterns. We can see such patterns for some of Manhattan neighborhoods in Fig. 8. (Manovich 2014 presents the analysis of the temporal patterns in central areas of six global cities.)

Comparing Social Media Inequality and Socio-Economic Inequality

How is social media inequality related to socio-economic inequality? For example, in a place like Manhattan, is social media inequality smaller, bigger or similar to the inequality of various socio-economic characteristics? In this section we will do these comparisons using selected economic and social indicators for Census tracts on the city and volume of Instagram shares in the same tracts. These indicators come from American Community Survey 2012 estimates (American Community Survey, 2012). This is the yearly estimate published by U.S. Census Bureau based on the responses of a sample of U.S. residents. We downloaded ACS data using R *acs* package.

In preparation for the analysis, we have considered a number of socio-economic indicators available in the Census publications per tract: number of households surveyed, median age, median household income, median rent, total population, race, employment status, time of commute to work, educational attainment, health insurance coverage, and Gini coefficient for income. Because most of these indicators are correlated, we decided to use a smaller subset: median household income, median rent, and unemployment rate.

First we consider relations between numbers of Instagram images shared by Manhattan residents and median family income in the same tracts. (See Goddemeyer, Stefaner, Baur, Manovich, 2014 for the initial analysis of the part of the data corresponding to 68 tracts crossed by Broadway.) More affluent tracts have more images and less affluent tracts have less images. According to Pew Research Center 2015 report, people in the U.S. who have higher education levels and household income are more likely to use social media than people who have less education and income, but the differences they report are all less than 12%. Therefore, they cannot account for massive differences in numbers of images shared in parts of the city - ranging from 2,157 to 552,787 per sq. km area, i.e. over 256,000%. Note also that images shared in a given tract are not necessary only from residents of that tract. They can also come from people who live outside that tract but are spending some time in this tract.

Therefore, simply correlating income level and images volume is not meaningful. We need a more nuanced analysis. We consider three variables: average income per tract, the numbers of images shared during day time (7am - 7pm) and the numbers of images shared during night time (7pm - 7am). We also divide the tracts into two kinds: the ones where median family income is less than the average for Manhattan (\$74,693), and the ones where it is bigger.

The analysis reveals that in most tracts with income higher than Manhattan average more images are shared during day time. Conversely, in most tracts with income lower than Manhattan average more images are shared during night time. We propose the following explanation for this pattern. During weekdays the residents of less prosperous areas (such as most parts of Manhattan above 100th Street) work in more prosperous parts of the city below 100th Street where more big businesses in the city are located (County Business Patterns 2012). This is where they share images on Instagram during the day, so their shares get added to these areas.

Since these people are absent from their home areas during these working hours, the volumes of images in these areas during day time is relatively small. In the evening, they return to their areas of residence, and this is why these less prosperous areas have higher volume of Instagram shares at evening and night hours.

Note also that the areas of Manhattan below 100th street with most businesses are also the ones that are the most popular among visitors. Thus, we have the effect of *double amplification* – social media contributions by affluent residents of these areas get amplified with contributions of people who travel there for work and also by contributions from city visitors. This amplification may be the key reason why spatial social media inequality we calculated for Manhattan using Gini coefficient is so high. One part of the city gets images from three groups (residents, commuters, and visitors), while the other part gets the potential images “subtracted” (these are images that would be shared if the residents in this part did not commute).

We may expect that in other geographical areas around the world different relations between places of residence and work, income distributions, and tourist areas lead to different spatial and temporal patterns of sharing. For example, in many European cities, small historical centers

are popular with visitors but most businesses employing lots of people are located outside these centers.

Another interesting issue is the effects of changing patterns of work - especially in creative and software industries. Many cities now act as distributed workspaces where designers, programmers, bloggers, and other culture industry professionals work from cafes close to where they live. Note that this demographic is also likely to be most active on social networks such as Instagram in many parts of the world.

Finally, consider another issue which is changing many cities worldwide today – gentrification. An area which previously only had less affluent residents, who may be commuting to work during daytime in other parts of the city, may now have a growing proportion of creative class workers and other freelancers who also stay there during the day to work from homes or cafes. Some parts of Manhattan above 100th Street have been undergoing gentrification for a while now. Although our dataset only covers five months and therefore does not allow us to qualitatively analyze the effects of this gentrification on Instagram sharing, the elevated volumes of images in certain areas described as being gentrified suggests that the two are related.

For our final analysis, we compare inequality for volume of Instagram images shared by locals and three socio-economic indicators for Manhattan: median household income, median rent, and unemployment rate. The Gini coefficients for these indicators are 0.32 (median income), 0.22 (median rent), 0.35 (unemployment rate), and 0.49 (numbers of Instagram images shared by local residents). Figure 9 shows Gini measures for these variables using Lorenz curves.

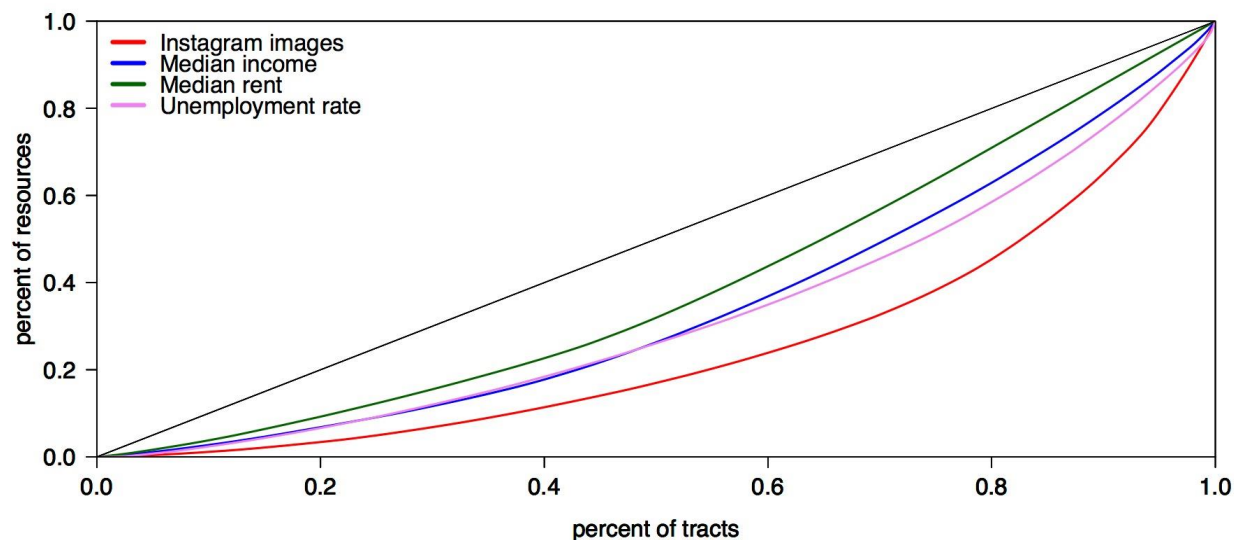


Fig. 9.

The inequality of Instagram images shared in Manhattan turns out to be bigger than inequalities in levels of income, rent, and unemployment. This is a very interesting and original

result. Note that we are only considering images shared by local residents, which is what makes the comparison between distributions of social media and distributions of socio-economic indicators meaningful. We could have expected to see this result for visitors, given the concentration of most tourist landmarks and shopping areas in particular parts of the city. Finding that the inequality in Instagram shares is also larger than socio-economic inequality for local residents was really unexpected.

It is too early to draw big conclusions from this finding since we only looked at a single urban area (i.e., Manhattan). Nevertheless, recall that Manhattan has the highest income inequality among all urban areas in the U.S. (U.S. Census Bureau 2014). Does this mean that in many other cities social media inequality will be even higher than the socio-economic indicators? Or does it mean that social media signal amplifies already present social and economic inequalities in our societies? What are the relations between social media “portraits” of the cities created by postings of its residents and visitors, spatial patterns of socio-economic inequality, and locations of places of residence, work, and tourist attractions? Looking at data from many cities should help us answer these interesting questions.

REFERENCES

- American Community Survey. 2012 Estimates. (2012). Retrieved from <https://www.census.gov/programs-surveys/acs/news/data-releases.2012.html>.
- Atkinson, A. B. (2015). *Inequality: What Can Be Done?* (1st edition). Cambridge, Massachusetts: Harvard University Press.
- Batty, M. (2013). *The New Science of Cities*. Cambridge, Massachusetts: The MIT Press.
- Fischer, E. (2010). *The Geotagger's World Atlas*. Retrieved from <https://www.flickr.com/photos/walkingsf/sets/72157623971287575/>
- Goddemeyer, D., Stefaner, M., Baur, D., & Manovich, L. (2014) *On Broadway*. Retrieved from <http://on-broadway.nyc>
- Goldsmith, S., & Crawford, S. (2014). *The Responsive City: Engaging Communities Through Data-Smart Governance* (1st edition). San Francisco, CA: Jossey-Bass.
- Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday*, 18(7). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/4711>
- Hochman, N., & Schwartz. (2012). Visualizing Instagram: Tracing Cultural Visual Rhythms. In *The Workshop on Social Media Visualization (SocMedVis) in conjunction with The Sixth International AAAI Conference on Weblogs and Social Media (ICWSM-12)*. Dublin, Ireland. Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4782/5091>

- Manovich, L. (2014). When do people share? Comparing Instagram activity in six global cities. Retrieved from <http://lab.softwarestudies.com/2014/11/when-do-people-share-comparing.html>
- Milanovic, B. (2006). Global Income Inequality: What It Is And Why It Matters? DESA Working Paper No. 26.
- Milanovic, B. (2007). Worlds Apart: Measuring International and Global Inequality. Princeton, N.J.: Princeton University Press.
- OECD. (2011). Society at a Glance 2011. Paris: Organization for Economic Co-operation and Development. Retrieved from http://www.oecd-ilibrary.org/content/book/soc_glance-2011-en
- Pew Research Center. (2015). Social Media User Demographics. Retrieved from <http://www.pewinternet.org/data-trend/social-media/social-media-user-demographics/>
- Piketty, T. (2014). Capital in the Twenty-First Century. (A. Goldhammer, Trans.). Cambridge Massachusetts: Belknap Press.
- Pucci, P., Manfredini, F., & Tagliolato, P. (2015). Mapping Urban Practices Through Mobile Phone Data. Springer.
- Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. Environment and Planning: Planning and Design, 33(5), 727–748. <http://doi.org/10.1068/b32047>
- Ray, D. (1998). Development Economics. Princeton, N.J: Princeton University Press.
- Stiglitz, J. E. (2012). The Price of Inequality: How Today's Divided Society Endangers Our Future (1st edition). New York: W. W. Norton & Company.
- Townsend, A. M. (2014). Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia (1st edition). New York: W. W. Norton & Company.
- U.S. Census Bureau. Country Business Patterns. (2013). Retrieved from <http://www.census.gov/econ/cbp/>
- U.S. Census Bureau. State and County QuickFacts. (2014). Retrieved from <http://quickfacts.census.gov/qfd/states/36/36061.html>