EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

CrossMark

Misery loves company: happiness and communication in the city

Aamena Alshamsi[1], Edmond Awad[1], Maryam Almehrezi[1], Vahan Babushkin[1], Pai-Ju Chang[1], Zakariyah Shoroye[1], Attila-Péter Tóth[1] and Iyad Rahwan[1,2]*

*Correspondence:
irahwan@acm.org
[1] Masdar Institute of Science and Technology, Abu Dhabi, UAE
[2] Massachusetts Institute of Technology, Cambridge, USA

Abstract

The high population density in cities confers many advantages, including improved social interaction and information exchange. However, it is often argued that urban living comes at the expense of reducing happiness. The goal of this research is to shed light on the relationship between urban communication and urban happiness. We analyze geo-located social media posts (tweets) within a major urban center (Milan) to produce a detailed spatial map of urban sentiments. We combine this data with high-resolution mobile communication intensity data among different urban areas. Our results reveal that happy (respectively unhappy) areas preferentially communicate with other areas of their type. This observation constitutes evidence of homophilous communities at the scale of an entire city (Milan), and has implications on interventions that aim to improve urban well-being.

Keywords: social media; sentiment; happiness; cities; homophily

1 Introduction

For the first time in history, the majority of humans now lives in cities. A complete theory concerned with the growth and dynamics of cities is still a work-in-progress [1]. However, our phenomenological understanding of cities is growing significantly thanks to progress in our ability to sense the dynamics of human behavior [2], and the urban environment and infrastructure [3, 4].

The high population density in cities is associated with both desirable urban indicators such as innovation, economic growth, and employment opportunities, and with undesirable outcomes such as crimes, diseases and pollution [5, 6]. In particular, cities exhibit consistent sub-linear and super-linear scaling of many of these indicators [7].

These characteristics of cities are attributed to many factors [8, 9]. Among those, special attention is often given to the role of social interaction [10, 11]. Characteristics of human social interaction, such as the role of weak ties [12], structural holes [13] and the diversity in interaction [14] are often seen as important facilitators of success in cities. Recently, it has been suggested that cities are special because the increased urban population density leads to super-linear scaling in social tie density, thus facilitating super-linear scaling of information spreading [15].

On the other hand, the societal success of cities is also a function of urban well-being and happiness. Policy-makers are interested in understanding the drivers of happiness in cities

Springer

in order to sustain or increase it [16]. Thus, it is important to understand the interplay between urban communication, on one hand, and urban happiness on the other. This would help us understand, for example, whether and how urban communication structures facilitate or inhibit the well-being of citizens. In this paper, we take a first step towards investigating this issue.

Traditional studies rely on self-reporting through surveys to measure happiness at the level of entire cities [16–18]. Despite the pervasiveness of surveys in quantifying or indexing happiness measures, they suffer from a number of limitations, such as the unreliability of subjective data [19].

Recently, researchers showed that the abundance of personal data emitted through social media (e.g., short-message broadcast mediums like Twitter) can reliably quantify individual happiness [20–22]. When combined with geographical information, this technique can be used to characterize the geographical distribution of happiness across large areas, such as the continental United States [23].

Motivated by recent results on the geography of happiness and the communication structure in cities, we are interested in understanding the relationship between urban happiness and urban communication. We leveraged datasets provided by the Big Data Challenge that was organized by Telecom Italia [24]. We used the communication intensity data to build a directed network of urban areas whereby the weights of the edges indicate the strength of communication between the areas. To estimate the happiness level of urban areas, we used Dodds et al.'s method [22] to analyze the sentiment of geolocated short-message broadcasts (also known as *tweets*) initiated in these areas, which after aggregation (averaging) gave us a reliable approximation.
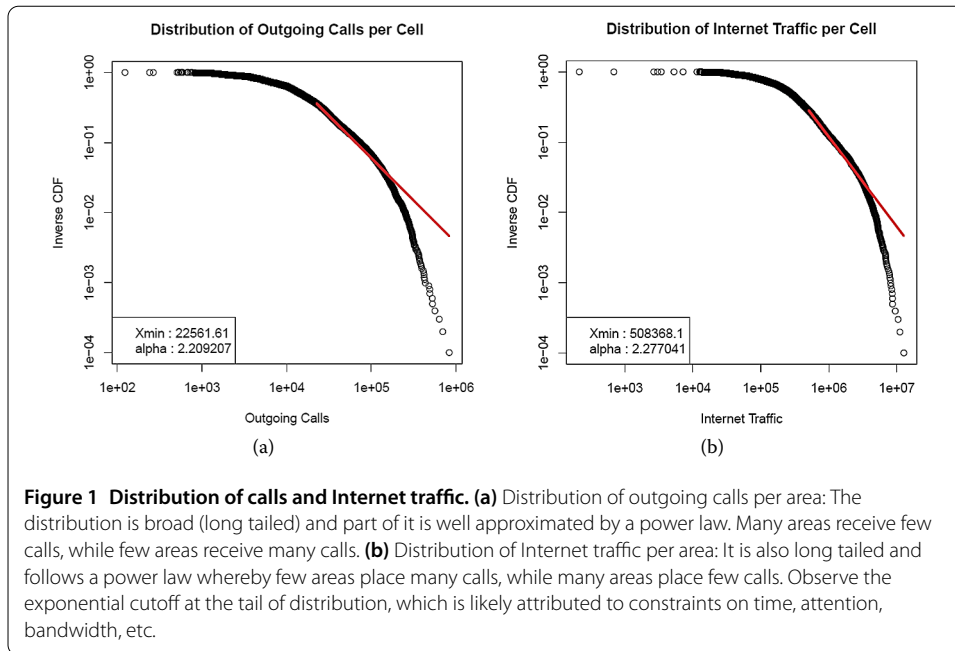
After preprocessing the data, we demonstrated the effectiveness of social media in mapping happiness at a much finer spatial resolution (within an urban area). Then, we investigated the relationship between communication among different geographic areas and their happiness levels. We found that communication patterns of urban areas exhibit homophilous behavior. More precisely, happy urban areas tend to interact with other happy areas more than they interact with unhappy areas. Similarly, unhappy urban areas tend to interact with other unhappy areas more than they interact with happy areas. The urban homophily in happiness that we witnessed in our dataset supports previous findings on homophily in happiness among individual humans [25], and shows that this phenomenon persists at larger scales. Our result is relevant to policy-makers to guide them in setting strategies that increase happiness, which is itself correlated with important outcomes ranging from crime and health, to productivity and innovation.

2 Methods and data
2.1 The dataset
This work uses four datasets released by Telecom Italia for the Big Data Challenge 2014 [24]. The datasets were collected during November and December 2013. Among the released datasets, the following four were used.

- 'Milano Grid'. The city of Milan was divided into a spatial grid of 100×100 cells. This dataset contains the ID of each cell in the grid along with the geometry of the cell. We will use the terms *cell* and *area* interchangeably.
- 'Telecommunications - SMS, Call, Internet - MI'. This dataset contains the activity of incoming and outgoing calls/SMSs and Internet traffic, aggregated over the cells and

Figure 1 Distribution of calls and Internet traffic. (a) Distribution of outgoing calls per area: The distribution is broad (long tailed) and part of it is well approximated by a power law. Many areas receive few calls, while few areas receive many calls. **(b)** Distribution of Internet traffic per area: It is also long tailed and follows a power law whereby few areas place many calls, while many areas place few calls. Observe the exponential cutoff at the tail of distribution, which is likely attributed to constraints on time, attention, bandwidth, etc.

over 10-minute intervals. Figure 1 shows the distribution of outgoing calls and Internet traffic over the cells. One can observe that their distributions are partly characterized by a power law. However, the tail of the distribution exhibits an exponential cutoff, likely caused by cognitive saturation in the communication capacity of people within individual cells [26]. Similar distributions can be found for the incoming calls and the incoming/outgoing SMSs (check Section E of Additional file 1).
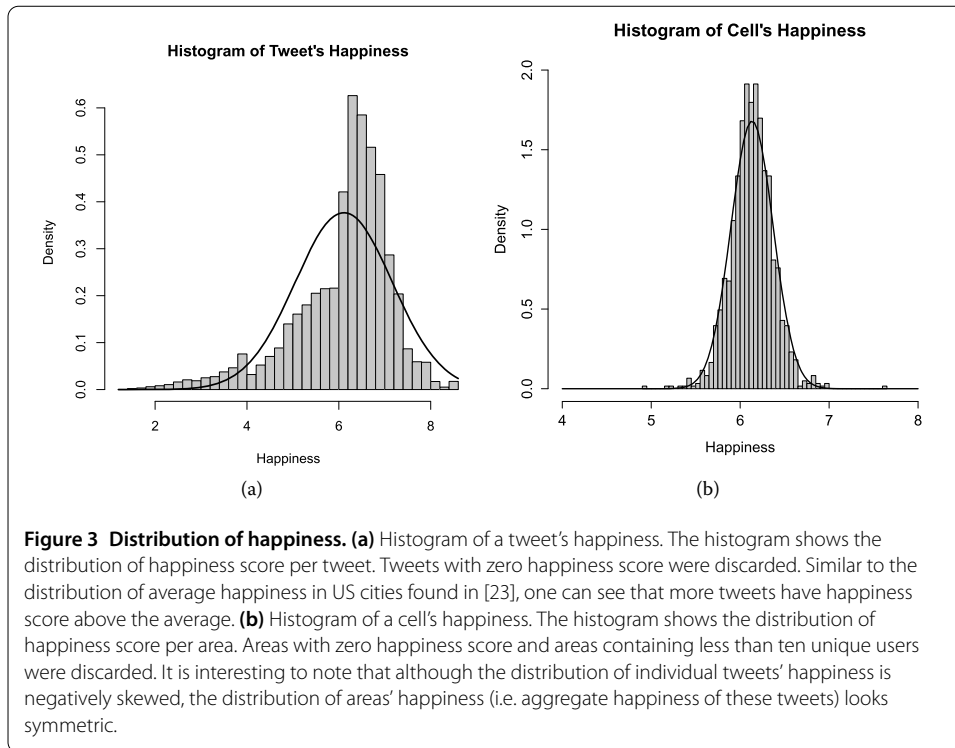
- 'Telecommunications - MI to MI'. This dataset contains the directional interaction strength between the different cells in the grid. This is based on the calls exchanged between mobile phone users in Milan between Nov 1st and Dec 31st, 2013. We used this dataset to construct the weighted directed network of interaction among cells.
- 'Geo Tweets'. This contains about 500,000 tweets that are geo-located in Milan. Figure 2 shows the distribution of the tweets in Milan, highlighting variability among different areas, and peaking at the center.

2.2 Preprocessing tweets

From 'Geo Tweets' dataset, we extracted tweets written in Italian and English. These tweets constituted about 80% of the overall tweets (check Section H of Additional file 1). Then, using the free Google Translate API (Goslate), we translated Italian tweets to English.

2.3 Measuring happiness of cells

To measure the happiness in tweets and accordingly in cells, we used a total of 10,000 words with their happiness scores on a scale of 1 (unhappy) to 9 (happy). This data was used in various studies [22, 23, 27], and is available online [22]. Following the existing methodology [22, 23], we removed all words with a happiness score between 4 and 6, then the happiness score for each tweet was calculated depending on the words it contains. For a given tweet T, containing N unique words, the average happiness was calculated using

Figure 2 Heat map of number of tweets in Milan. The figure shows the intensity of tweets across Milan. The dark red indicates high number of tweets, while the dark blue indicates low number of tweets. We can see that the bulk of the tweets is located near the center of Milan.

the following formula [22]:

$$h_{\mathrm{avg}}(T) = \frac{\sum_{i=1}^{N} h_{\mathrm{avg}}(w_i) \cdot f_i}{\sum_{i=1}^{N} f_i},$$

where f_i is the frequency of the ith word w_i in T and $h_{\mathrm{avg}}(w_i)$ is the average happiness of the word w_i. Words that do not have happiness scores are given the value of zero happiness. Tweets with zero happiness scores were discarded because they do not provide any information about the sentiment of the area they belong to. Among the considered tweets, 14% of the tweets (55K out of 382K) have zero happiness score. Figure 3(a) shows a histogram of the happiness scores of tweets' in our data.[a]

We used the implementation of Point Inclusion in Polygon Test by W. Randolph Franklin [28] to map tweets to cells. Tweets that do not map into the grid (i.e. are not geo-located in Milan) were discarded. We also discarded cells with zero happiness scores. A cell has a zero happiness score either because it has no tweets (i.e. no tweets are mapped to it), or because all its tweets have zero happiness scores. Then, cells with fewer than ten unique Twitter users were discarded since they provide a very noisy measure of happiness. We are left with 1,213 cells, whose distribution of happiness scores is shown in Figure 3(b). A heat map for cells' happiness score in Milan is visualized in Figure 4(a).

Our investigation of homophily relies on a network of cells with discrete categories (happy and unhappy). So the first step was to classify each cells according to its happiness level according to the following:

Figure 3 Distribution of happiness. (a) Histogram of a tweet's happiness. The histogram shows the distribution of happiness score per tweet. Tweets with zero happiness score were discarded. Similar to the distribution of average happiness in US cities found in [23], one can see that more tweets have happiness score above the average. **(b)** Histogram of a cell's happiness. The histogram shows the distribution of happiness score per area. Areas with zero happiness score and areas containing less than ten unique users were discarded. It is interesting to note that although the distribution of individual tweets' happiness is negatively skewed, the distribution of areas' happiness (i.e. aggregate happiness of these tweets) looks symmetric.

- *Unhappy*: aggregated happiness < 5.91 (lowest 15%)
- *Happy*: aggregated happiness > 6.35 (top 15%)
- *Neutral*: all other cells.

We removed all neutral cells since we are only interested in interaction among happy/ unhappy cells. A heat map of happy, unhappy, and neutral cells is shown in Figure 4(b). The results presented in the paper are generated using the 15 percentile. However, we tried other different percentiles (20, 25, 40 and 50) to label cells as happy or unhappy and reported the detailed results in Section A of Additional file 1.

2.4 Building communication network

To build the network, we used communication data for an entire workweek (from Nov 4th to Nov 8th, 2013) to characterize the urban communication network (similar results were obtained using communication data from a single day). We aggregated the calls' weights between cells during this week, then filtered out edges in which the calling/called cells were discarded in a previous stage (either because they had less than ten unique Twitter users or because they had zero happiness score). We also discarded self-edges that capture communication within cells. To remove the effect of transient communications, we used the weight of edges to filter out edges with weak connections (we discarded edges with aggregated weights less than 0.1). Also, as observed in Figure 1, there is a variation in terms of the proportions of communication among areas. Therefore, the communication intensity between two areas can be attributed to their population [29, 30]. Unfortunately, the Big Data Challenge did not provide the population size in each area, and public population data is not available at the same level of granularity. Hence, as a proxy for population, we used the number of Twitter users who initiated tweets in each cell [31]. Our goal is to minimize the effect of population on the intensity of communication. Checking for corre-
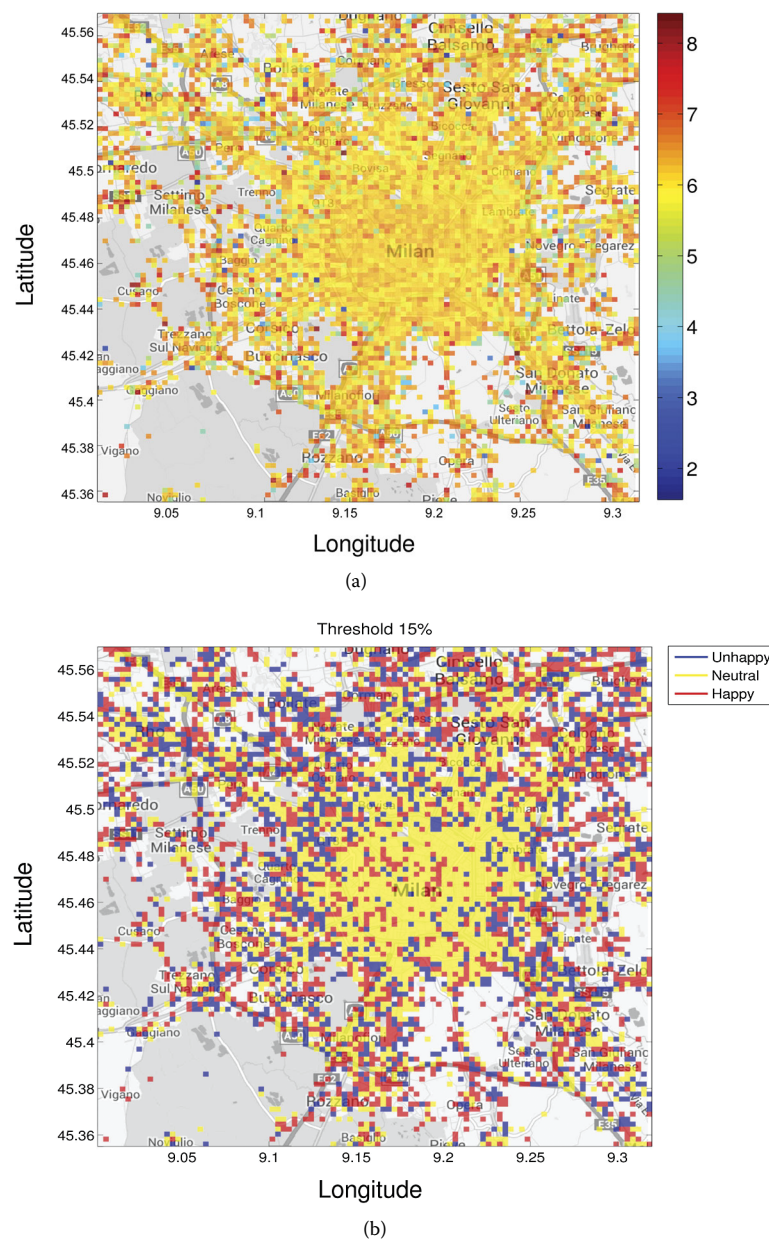
Figure 4 Heat map of happiness in Milan. (a) The map shows the intensity of happiness score of each area. Red areas have high happiness scores, while blue areas have low happiness scores. **(b)** The figure identifies the happiness category (happy, unhappy, or neutral) of each cell. We discretized the happiness score for areas. We considered the areas with the highest 15% happiness scores as happy areas, while the areas with the lowest 15% happiness scores as unhappy areas. Remaining areas are considered as neutral areas.

lation between intensity of communication among two areas and the population in these areas would help to determine a good way to minimize this effect. We tested the correlation between the intensity of communication among a pair of areas with the minimum of the two areas' population, the product of the two areas' population, the average of the two areas' population, the calling area's population, and the receiving area's population. They all provided significant positive correlation values (Spearman's rank correlation: 0.3, 0.29, 0.25, 0.22, and 0.2 respectively). Thus, we divided the communication between each pair

of areas by the lower population of the two:

$$\frac{Communication(A, B)}{\min\{population(A), population(B)\}}.$$

The resulting network consists of 363 nodes which represent cells/areas, each of which is labeled as happy or unhappy. These nodes are connected using 52,417 weighted, directed edges that represent the intensity of calls between areas.

3 Results

We start with a visual exploration of whether cells communicate preferentially with cells of their own type, we ran a community detection algorithm (the *multi-level modularity optimization* algorithm [32]) on the communication network among the urban areas that we classified as 'happy' and 'unhappy.' The output is shown in Figure 5. Nodes labeled with + represent happy areas and nodes labeled with − represent unhappy areas. The different colors represent different communities identified by the algorithm. Most communities are dominated ($\geq \frac{2}{3}$) by a particular class (happy or unhappy).

To further quantify this effect, we statistically investigated the variation in communication between areas of different happiness levels. We conducted a *two-way Analysis of Variance (ANOVA)* to compare the effect of levels of happiness (i.e. happy versus unhappy) of the *source* and *target* areas on the strength of communication, a relationship that can

Figure 5 Network of calls between happy/unhappy areas in Milan. Nodes represent areas, and edges represent calls between these areas. Nodes labeled with + represent happy areas, while nodes labeled with − represent unhappy areas. Colors represent communities generated by the *multi-level modularity optimization* algorithm [32]. One can see that most communities are dominated ($\geq \frac{2}{3}$) by a particular class (happy or unhappy).

be described by the following linear model:

$$strength_{\text{communication}} = B_1 \cdot Source + B_2 \cdot Target + B_3 \cdot Source \times Target.$$

The dependent variable is the aggregate directional communication between the two areas (continuous number). The independent variables are: (1) *Source*, a factor with two levels of happiness representing the area that initiates the communication, (2) *Target*, a factor with two levels of happiness representing the area that receives the communication (happy or unhappy), (3) the interaction between these two factors (*Source* × *Target*).

The results of the *ANOVA* show that the interaction between *Source* and *Target* is significant ($F(1, 52,418) = 72.733$, $p \ll 0.00001$). The interaction effect might indicate the existence of homophily or heterophily in the communication patterns of urban areas in Milan. Homophily means that areas of the same level of happiness tend to interact more with each other than they interact with areas of the other level of happiness, whereas heterophily means that areas tend to interact with areas of the other level of happiness more than they do with areas of the same level.

We produced an interaction plot to visualize the interaction effect. Figure 6 shows a tendency for homophily in communication. Taking into account the whole volume of communication, happy areas tend to call happy areas more than they call unhappy areas. Similarly, unhappy areas tend to call unhappy areas more than they call happy areas. The same behavior can be also noticed regarding receiving the calls. Happy areas receive more calls from happy areas than from unhappy areas, and unhappy areas receive more calls from unhappy areas than from happy areas.

In order to study the significance of the previous observations, we conducted post hoc comparisons using the Tukey HSD test [33] to compare all the six possible combinations of weights from the interactions. Tukey HSD is a statistical test that is used with an *ANOVA* (a two-way ANOVA in our case) to do pairwise comparisons between the means of the
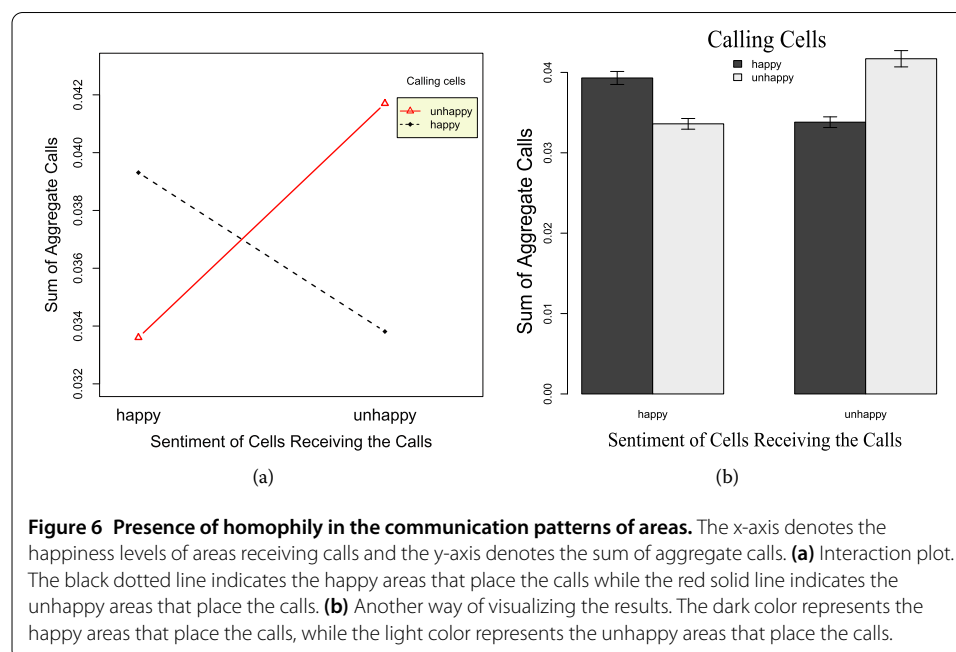

Figure 6 Presence of homophily in the communication patterns of areas. The x-axis denotes the happiness levels of areas receiving calls and the y-axis denotes the sum of aggregate calls. **(a)** Interaction plot. The black dotted line indicates the happy areas that place the calls while the red solid line indicates the unhappy areas that place the calls. **(b)** Another way of visualizing the results. The dark color represents the happy areas that place the calls, while the light color represents the unhappy areas that place the calls.

Table 1 Weight of communications from happy/unhappy cells (*Source*) to happy/unhappy cells (*Target*)

From	To	Mean	Standard deviation
Happy	Happy	0.039	0.1
Unhappy	Unhappy	0.041	0.1
Unhappy	Happy	0.033	0.075
Happy	Unhappy	0.033	0.075

different treatments (in our case, we have four treatments as shown in Table 1, and thus we have six pairwise comparisons).
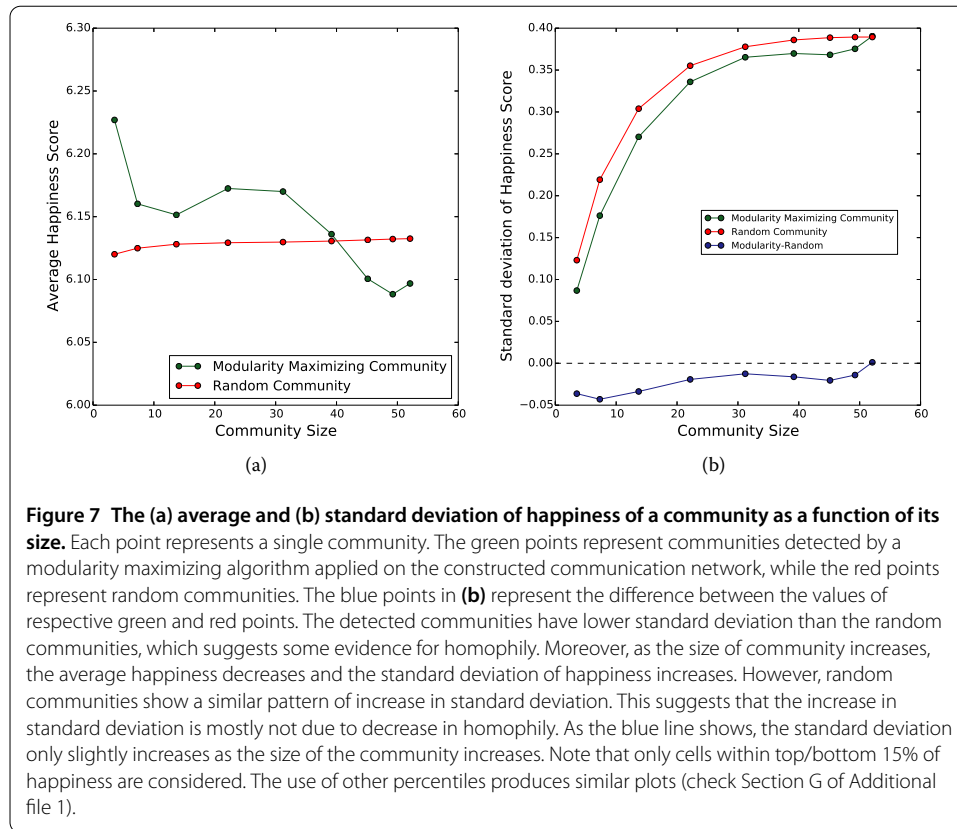
Table 1 shows the mean and the standard deviation values for the interactions between happy/unhappy cells. The mean of the communication weights of the unhappy areas calling and receiving calls to/from other unhappy areas is statistically significantly different from the mean communication weights of the unhappy areas communicating with happy areas ($p \ll 0.00001$). Similarly, the mean of the communication weights of the happy areas calling and receiving calls to/from other happy areas is statistically significantly different from the mean communication weights of the happy areas communicating with unhappy areas ($p \ll 0.00001$).

We have also quantified the level of assortativity mixing in the network of areas by using a weighted version of the assortativity coefficient defined by [34]. For more information about how we implemented it, please refer to Section D of Additional file 1. We found that the assortativity coefficient is 0.21 which could be considered relatively high. Hence, this is another evidence of the assortative behavior in the communication patterns of urban areas in Milan.

4 Homophily on community level

Given these findings, it would be interesting to know whether homophily exists on the community level. To investigate this, we used the output of the community detection algorithm (namely *multi-level modularity optimization* algorithm [32]). This algorithm uses the notion of modularity, which is a quality measure for graph clustering proposed by Newman [35]. After we found communities, we studied the effect of community size on the average and the standard deviation of a community's happiness score. In general, if the standard deviation is small, then one might conclude the existence of homophily on community level. Additionally, we are interested in finding whether the average and the standard deviation of a community's happiness score will change as the size of the community (measured as the number of cells) changes. For comparison, we generated random communities of similar sizes of the communities we have. A random community of a size h is formed by randomly assigning h cells into it. The process is repeated 200 times and the average value (of averages or of standard deviations) is calculated for the community. For more details, check Section G of Additional file 1.

Figure 7(a) shows that small-size communities have slightly higher average happiness than a random community of the same size. As community size increases, its happiness score decreases to become less than that of a same-size random community. This suggests that smaller communities enjoy higher level of happiness than larger ones. Figure 7(b) shows that detected communities have lower standard deviation than the random communities, which suggests some evidence for homophily within communities. Additionally, it shows that the standard deviation of happiness score increases as the community size

Figure 7 The (a) average and (b) standard deviation of happiness of a community as a function of its size. Each point represents a single community. The green points represent communities detected by a modularity maximizing algorithm applied on the constructed communication network, while the red points represent random communities. The blue points in **(b)** represent the difference between the values of respective green and red points. The detected communities have lower standard deviation than the random communities, which suggests some evidence for homophily. Moreover, as the size of community increases, the average happiness decreases and the standard deviation of happiness increases. However, random communities show a similar pattern of increase in standard deviation. This suggests that the increase in standard deviation is mostly not due to decrease in homophily. As the blue line shows, the standard deviation only slightly increases as the size of the community increases. Note that only cells within top/bottom 15% of happiness are considered. The use of other percentiles produces similar plots (check Section G of Additional file 1).

increases. However, random communities show a similar pattern of increase in standard deviation. This suggests that the increase in standard deviation is only slightly due to decrease in homophily. That is, homophily within a community is slightly influenced by the size of the community.

5 Discussion

We have taken a first step towards understanding the interplay between communication and happiness in urban areas at a high resolution. We found evidence of assortative mixing (homophily) in communication between different urban areas based on their happiness level. We also found that the mean of happiness seems to vary with community size, where community is defined in terms of communication structure.

Obtaining our main result required developing a data science pipeline that combines data from a variety of sources and conducts social media data scraping, translation, sentiment scoring, aggregation, geo-location, and statistical hypothesis testing. We believe this type of pipeline can be used beyond the scope of this particular paper. For example, our claims about homophilous communication are limited to a particular indicator, namely happiness measured through public social media production. It may be possible to apply the same technique to measure homophily based on other sentiment indicators that can be extracted from social media, such as consumer confidence, or political opinions.

Certainly, Twitter is not the only way to measure happiness in cities, and it may be possible to establish assortativity using other measures of happiness such as collecting self-reported happiness of a sample of people through surveys [36]. However, these measures are expensive, particularly at the high spatial resolution obtained by this study. Moreover,

using social media posts like Tweets provides a real-time indicator of happiness, and is therefore better suited for applications that require this information at higher temporal resolution.

The main limitation of the present study is that it involves a single city. This is caused by the limited availability of data. In the future, it would be necessary to conduct similar investigations for other cities to see if homophily holds consistently across a variety of urban centers. Even if the pattern does hold, it would be interesting to investigate whether different cities exhibit homophilous communication to different degrees.

Another opportunity for further work is to explore the role (if any) played by other urban indicators, such as income, in mediating our observations. It may be possible, for example, that our observed effect is more (or less) pronounced for areas with similar per-capita income.

We believe there are many opportunities for further exploration of the role of urban communication in urban well-being. An interesting, though challenging, experiment can involve running interventions aimed at manipulating the urban communication structure to see if a causal link can be made to urban well-being.

Additional material

Additional file 1: Supporting information.

Endnote
   a   Note that the happiness score of a tweet might not be meaningful per se, but rather contributes to the calculation
      of happiness score of an area. With that being said, the distribution of happiness of individual tweets are shown
      only for reference, while the (aggregated) happiness scores for areas are the ones being used in the study.

References
 1. Bettencourt LM (2013) The origins of scaling in cities. Science 340(6139):1438-1441
 2. Eagle N, Pentland A (2006) Reality mining: sensing complex social systems. Pers Ubiquitous Comput 10(4):255-268
 3. Crane P, Kinzig A (2005) Nature in the metropolis. Science 308(5726):1225
 4. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 38(4):393-422
 5. Heilig GK (2012) World urbanization prospects: the 2011 revision. United Nations, Department of Economic and Social Affairs (DESA), Population Division, Population Estimates and Projections Section, New York
 6. Glaeser EL, Sacerdote B (1996) Why is there more crime in cities? Technical report, National Bureau of Economic Research
 7. Bettencourt LM, Lobo J, Helbing D, Kühnert C, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. Proc Natl Acad Sci USA 104(17):7301-7306
 8. Bettencourt LM, Lobo J, Strumsky D, West GB (2010) Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. PLoS ONE 5(11):e13541
 9. Glaeser E (2011) Triumph of the city: how our greatest invention makes US richer, smarter, greener, healthier and happier. Pan Macmillan, London
10. Audretsch DB, Feldman MP (1996) R&D spillovers and the geography of innovation and production. Am Econ Rev 86(3):630-640
11. Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108(3):577-598
12. Granovetter M (1973) The strength of weak ties. Am J Sociol 78(6):1360-1380

13. Burt RS (2009) Structural holes: the social structure of competition
14. Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. Science 328(5981):1029-1031
15. Pan W, Ghoshal G, Krumme C, Cebrian M, Pentland A (2013) Urban characteristics attributable to density-driven tie formation. Nat Commun 4:1961
16. Leyden KM, Goldberg A, Michelbach P (2011) Understanding the pursuit of happiness in ten major cities. Urban Aff Rev 47(6):861-888
17. Helliwell JF, Layard R, Sachs J (2013) World happiness report 2013
18. Cutler DM (2009) Measuring national well-being. In: Measuring the subjective well-being of nations: national accounts of time use and well-being, pp 107-112
19. Bertrand M, Mullainathan S (2001) Do people mean what they say? Implications for subjective survey data. Am Econ Rev 91(2):67-72
20. Davidov D, Tsur O, Rappoport A (2010) Enhanced sentiment learning using Twitter hashtags and smileys. In: Proceedings of the 23rd international conference on computational linguistics: posters, pp 241-249
21. Thelwall M, Buckley K, Paltoglou G (2011) Sentiment in Twitter events. J Am Soc Inf Sci Technol 62(2):406-418
22. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS ONE 6(12):e26752
23. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM (2013) The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. PLoS ONE 8(5):e64417
24. Telecom Italia (2014) Big Data Challenge. http://www.telecomitalia.com/bigdatachallenge
25. Bollen J, Gonçalves B, Ruan G, Mao H (2011) Happiness is assortative in online social networks. Artif Life 17(3):237-251
26. Miritello G, Lara R, Cebrian M, Moro E (2013) Limited communication capacity unveils strategies for human interaction. Sci Rep 3:1950
27. Kloumann IM, Danforth CM, Harris KD, Bliss CA, Dodds PS (2012) Positivity of the English language. PLoS ONE 7(1):e29484
28. Haines E (1994) Point in polygon strategies. In: Graphics gems IV, pp 24-46
29. Expert P, Evans TS, Blondel VD, Lambiotte R (2011) Uncovering space-independent communities in spatial networks. Proc Natl Acad Sci USA 108(19):7663-7668
30. Krings G, Calabrese F, Ratti C, Blondel VD (2009) Urban gravity: a model for inter-city telecommunication flows. J Stat Mech Theory Exp 2009(07):L07003
31. Botta F, Moat HS, Preis T (2015) Quantifying crowd size with mobile phone and Twitter data. R Soc Open Sci 2(5):150162
32. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):P10008
33. Yandell BS (1997) Practical data analysis for designed experiments
34. Newman ME (2003) Mixing patterns in networks. Phys Rev E 67(2):026126
35. Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci USA 103(23):8577-8582
36. Leedle BR, Clifton J (2014) State of well-being 2013 state, community and congressional district analysis. Technical report, Gallup Inc. and Healthways Inc.