Strengell, Niklas; Sigg, Stephan

Local emotions - Using social media to understand human-environment interaction in cities

Published in:
2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018

DOI:
10.1109/PERCOMW.2018.8480364

Published: 02/10/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Strengell, N., & Sigg, S. (2018). Local emotions - Using social media to understand human-environment interaction in cities. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018* (pp. 615-620). Article 8480364 IEEE. https://doi.org/10.1109/PERCOMW.2018.8480364

Local emotions - using social media to understand human-environment interaction in cities

Niklas Strengell
School of Electrical Engineering
Aalto University
Otaniemi, Finland
niklas.strengell@aalto.fi

Stephan Sigg
School of Electrical Engineering
Aalto University
Otaniemi, Finland
stephan.sigg@aalto.fi

***Abstract*— Cities have become the most common living environment for humans. With this rising urbanization, urban design has become vital for these growing cities. While measuring objective data like traffic congestion or air quality is important, it does not tell the whole story of how people live in the cities or how cities should be developed to make them more livable. In future for a true smart city a more humane component is needed to understand how the population of cities actually interact with and feel about their surroundings. Surveys are a great and a necessary tool for this and they are already being used in the design process. However, they require effort and and a lot of silent information can be missed. The surveying process also doesn't happen in real time. We suggest that social media data could be used to gather more information about human-environment interaction in cities and compliment the surveys. We show a working prototype of a tool that creates an emotional map of a city by mining social media data for sentiments and heatmapping them. This kind of method could prove to be an useful tool for urban designers, who could take advantage of the visual intuition of humans and see instantly where and how emotional hotspots arise. It could also be of interest for emotion researchers, who could get data on what it really means to *be happy* for a human being - for example eating an ice cream at the beach - instead of only linking conceptual words (such as happy) to external stimuli (such as smiling).**

I. INTRODUCTION

In the 21st century the urban environment is forming to be the most common habitat for human species. With rising urbanization, good urban design is critical. Cities can be conceptualized in many forms: as a social system, as an economic system or even in political terms. Some people see the city as a work of art, some as a historical artifact - some even see it as a living organism. In the same way, the design and planning of cities can be approached from different points of view.

With utopia in mind, the trends of urban design have always shifted: from the green garden city movement to the automobile city where everything is attainable with a car. Fortunately, the current trend seeks to create sustainable urban environments with long-lasting structures and building and overall livability. But as strange as it may sound, the urban life is not necessarily the main viewpoint from which urban design is done. Measurements of traffic congestion or air quality and the usage of resources are already common tools for urban planning. But these measurements are purely objective and lack all humanistic components. They do not reflect how people actually feel towards their environments or experience their surroundings [1]. The social component is much harder to study: surveys can and should of course be done, but the *in situ* observation of people's affections is much harder. But focusing on emotions is important, because emotions are thought to be reactions to the environment and situations that are relevant to the individual's current desires [2]. Emotionally salient data could thus lead urban designers to key-points of human-environment interaction in the urban environment: what and where is important and for whom.

In this research, we study the possibility to use emotional social media data to understand this spontaneous human-environment interaction in cities. We show a "proof-of-concept"-demo of our method by mining Twitter's geolocated tweets for sentiments and visualizing them on map as they arise in time and space. For visualizations of the method see figure 1 and of the results see figures 4, 5 and 6. We also discuss our results and how our methodology could prove useful.

II. RESEARCH AREA

A. AGI from social media

As aforementioned, surveys are a common place in the urban design process. The information is gathered actively from these surveys and is referred to as Volunteered Geographical Information (VGI). But it's counterpart passively gathered Ambient Geographical Information (AGI) is a relatively new phenomenon. This data can be gathered for example from a social media. Recent big turmoils on earth, such as the Arab Spring in 2011 [3] and Japan's earthquake [4] were widely documented on social media. The Arab revolt was even dubbed as "the Twitter revolution" [3]. The studies on these phenomena were among the first to study the expansion and evolution of social interaction in the masses in real time. It is exciting that social media can give us this new possibility to look upon these events that before were visible to us only through ground media - and rarely in real time. AGI available from these events removes the veil and shows us what is happening around us all over the world all the time and in much bigger scale than before.

B. Human-Enviroment Interaction

The mathematical psychologist Anatol Rapoport first documented the importance of human-environment interaction studies for urban planning in his book *Human Aspects of*

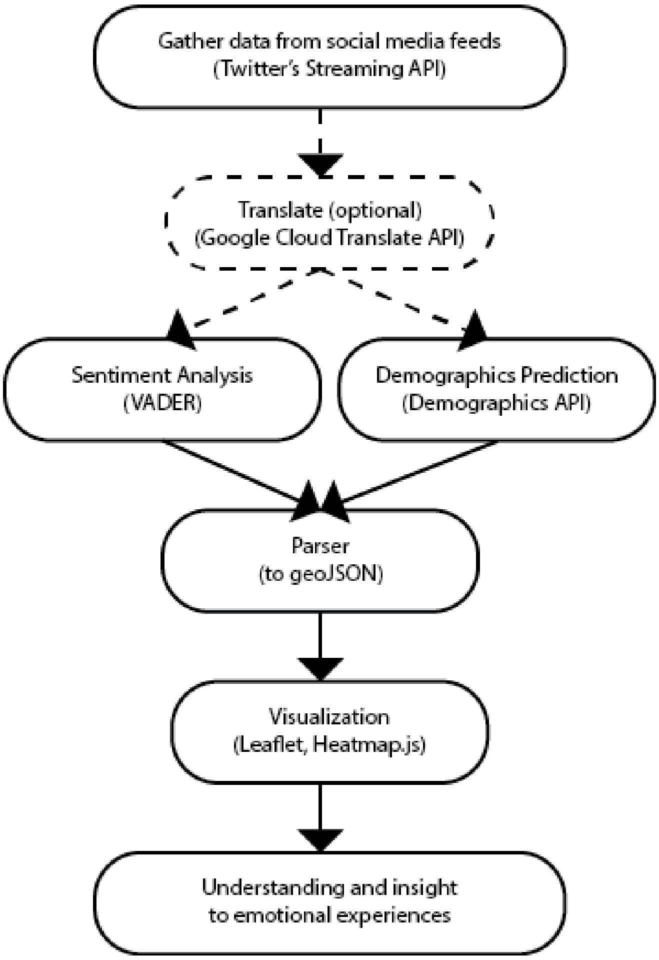

Fig. 1. The workflow in our method.

Urban Form: Towards a Man—Environment Approach to Urban Form and Design [5]. This research revolves around three general questions:

- How do people shape their environment?
- How and to what extent does the physical environment affect people, i.e. how important is designed environment and which contexts?
- What are the mechanisms which link people and environment in this two-way interaction?

By analyzing social media data, with these questions in mind, we could get hints or even answers on how these interactions arise in the environment. We could also study much bigger scales of actions with a lot more people than before. Our method is of course not antagonistic to traditional anthropological research, but rather a complimentary tool, that could point designers and researchers to the right direction. In addition, because the people themselves are actively producing the data in social media, we could get a much more personal and encompassing view. A survey always directs people thoughts and answers.

C. Emotionally aware technologies

We also think that studying emotions is they key point. There are two main motivations.

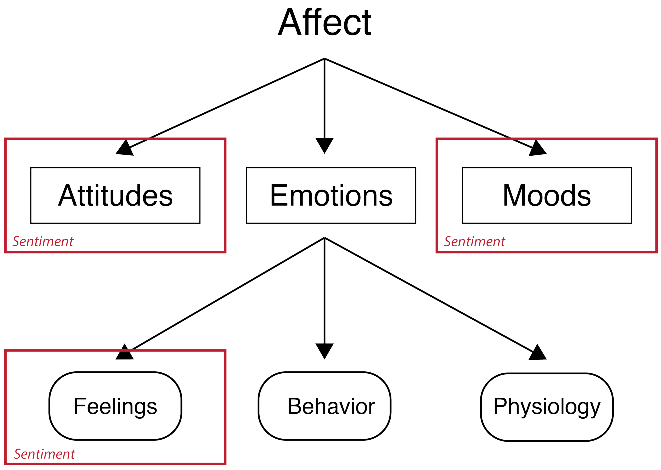

Fig. 2. One way of grouping the affective science. Sentiments are related to attitudes, moods and feelings, but can in theory refer to any of these phenomenon. Figure adopted from [7].

First of all, emotionally aware technologies offer better usability and can even help people improve their own emotional intelligence (EI). It has been given a lot of attention in consumer goods, such as the smart phone. This same design paradigm could and should be applied further - in our case to cities. Cities should not be thought as mechanical clockworks, but as living organisms: a separate and social units made up of highly interconnected people and places. For a true smart city an understanding of the layer of human behavior, emotions and experiences is needed [6].

Second of all, if emotions are thought to be reactions to situations that are relevant to the individual's current desires [2], studying their location can reveal us much about how human emotion ties into places and thus experiences.

D. Sentiment analysis and opinion mining

Strictly speaking, sentiments are not emotions but related to emotions. Both are a type of affective phenomena. Sentiments are similar to opinions or attitudes, which are relatively stable beliefs about the goodness or badness of something or someone [7]. They bias how a person will think about, feel towards or behave regarding a person or a thing [8]. In some sense, *sentiment* can be used as synonym to *feeling*. However, there is a slight difference: sentiment refers to a specific feeling or attitude behind something, while feeling is a more general term. What we mean by sentiment could best be understood as the valence of an emotion in the dimensional emotion theory (see figures 2 and 3). We are searching how *negative or positive sentiments people have towards something*.

Sentiment analysis is a viable tool for social media data for three primary reasons. First of all, much of the data on social media is opinionated [9]. Second of all, it is very humane and thus interesting to want to know how people feel towards us (or in this case a city or a certain location). In fact it is an important part of our information gathering habits [10]. And third of all, it can be very simple to do because it only

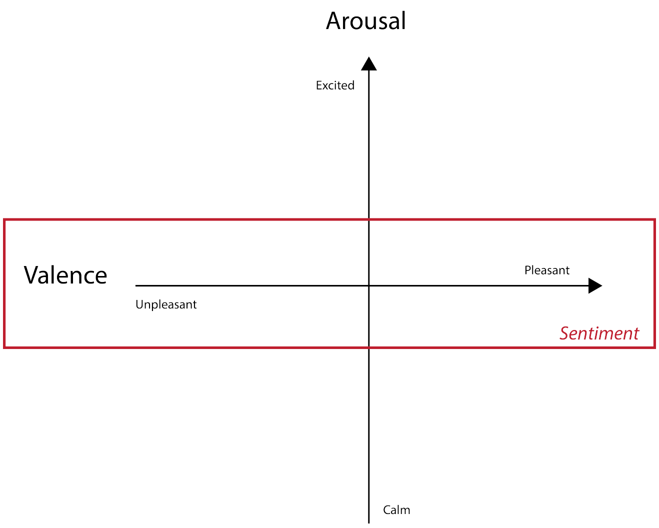

Fig. 3. One way of visualizing the dimensional emotion space. Our definition of sentiment refers to the valence of the affect, i.e. its pleasantness-unpleasantness.

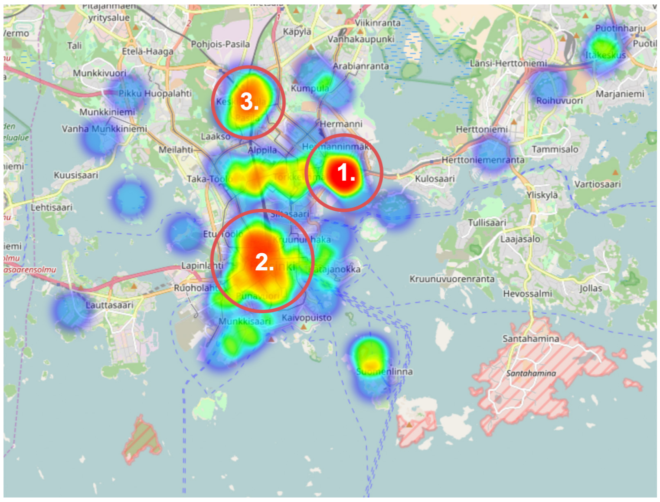

Fig. 4. Positive sentiment hotspots from the weekend of week 32.

Hotspot 1	Flow Festival (urban music festival)
Hotspot 2	City Center (no one specific event, but many smaller ones. Also the highest absolute amount of tweets.)
Hotspot 3	WorldCon (a sci-fi convention) and Tube-Con (a youtube convection)

TABLE I
POSITIVE HOTSPOTS EXPLAINED.

has three possible prediction outputs: positive, negative or neutral.

E. *Visualization*

To analyze complex data like sentimental social media data with location information, visualization is the key. Of all the five senses, the human brain relies most on vision. With visualization we can quickly show thousands of data points, from which we can quickly understand difficult concepts or identify new patterns [11].

Using geographical maps also ties our data to real world locations. With this emotional mapping we can visually spot emotional hotspots and clusters that emerge in the city. These visualizations can also tell us stories that we would otherwise miss and help us to understand bigger scales of action [12].

III. RELATED WORK

Pioneering research in emotional mapping has been done in the 21st century. Some notable examples are Christian Nold and his Emotional Cartography that began in 2004 [13]: a combination of technological vision and performing arts, where people were fitted with bio-monitoring devices and asked to walk around the city and go on with their hobbys and chores. The data was then visualized on a map and also inspired a collection of essays, on how our world looks like when emotional monitoring becomes a standard.

More recently emotional mapping and it's participatory potential has been studied and applied by Jirka Panek. His findings conclude that emotions and perceptions are a crucial part of modern cartography and that emotional maps can be valuable tools for bridging the gap between citizens and urban planners [14][15]. Studies done by Marketta Kyttä and Maarit Kahila in Aalto University also emphasize the importance of understanding human behavior and experiences in the context of city design [6].

IV. METHOD

The method consists of three major parts: gather data, analyze it for emotions and then visualize it on the map. The workflow can be seen in the figure 1. In this paper we show the visualizations from two time spans: of tweets gathered from Helsinki Area for the weekend from 11th of August to the 13th of August and for the weekdays from 14th of August to 18th August. This is enough to show us meaningful clusters and differences in time (weekend versus workweek), but isn't it too much so that individual stories would drown in noise.

We start by gathering data from Twitter's Streaming API with the help of Python programming languages tweepy library [16]. Twitter was chosen as the social media to study for several reasons. First of all, the data available is easy to analyze - each tweet consists of only 140 symbols. Second of all, there is plenty of data available as the principle ethic of Twitter is to share everything publicly. This is very different from for example Facebook, where everything is available only to your friends. Third of all, Twitter's API is open compared to all other the major social media sites. Twitter also provides lots of secondary metadata such as user information and location.

Because for this demonstration we are interested in Helsinki, and it's immediate surroundings, most of the tweets were not in English, but Finnish (and some other languages,

Hotspot 1	City center (no one specific event, but many smaller ones. Also the highest absolute amount of tweets.)
Hotspot 2	Flow Festival, TubeCon and WorldCon
Hotspot 3	Traffic accidents and congestion

TABLE II
NEGATIVE HOTSPOTS EXPLAINED.

such as Russian). Natural Language Processing tools for Finnish at the time are next to inexistent or not readily available. Thus before we can apply sentiment analysis, we need to translate the tweets to English. For this we used Google's Cloud Translation API [17].

A. SENTIMENT ANALYSIS

For sentiment analysis we used the VADER library that is built for Python [18]. VADER utilizes a rule based lexicon model, where each word is given a sentiment value as a number. The sentiment of a complete sentence is acquired as a combination of these words and a sentiment score is given: -1 is the most negative and 1 the most positive. 0 is neutral.

How VADER differs from other state-of-the-art libraries, is that it's rules are fine-tuned for social media text. It gives much better accuracy on predicting the sentiment on social media text. It's performance, 96% correct classification, is much higher than other benchmark applications such as the psycholinguistic analysis software called Linguistic Inquiry and Word Count (LIWC) or custom-built and trained neural networks [18].

B. VISUALIZATION

No complicated soft- or hardware is needed to use our visualizations as all of it happens in the common web browser. For mapping we utilize the Leaflet.js library [19] with OpenStreetMaps [20]. Both are community-driven and open source technologies.

To visualize the amount of tweets in an area, we use Leaflet.js-libary's plugin heatmap.js [21]. The more tweets an area has, the more red it is displayed. The visualization does not utilize the relative valence of the data but only it's category. In other words, each positive or negative tweet is given the same score of -1 or 1 (neutral tweets with score 0 are discarded). We could utilize the relative score for each tweet by using the continuous scale from -1 to 1 given to us by VADER, but we thought that the score given this way would be too arbitrary.

C. DEMOGRAPHICS

If our methodology would be used in citizen science or urban planning, it is important not only to gather data, but to know from whom it is gathered. Thus we also show that it is possible to analyze demographics for the tweets. For this we use two different means.

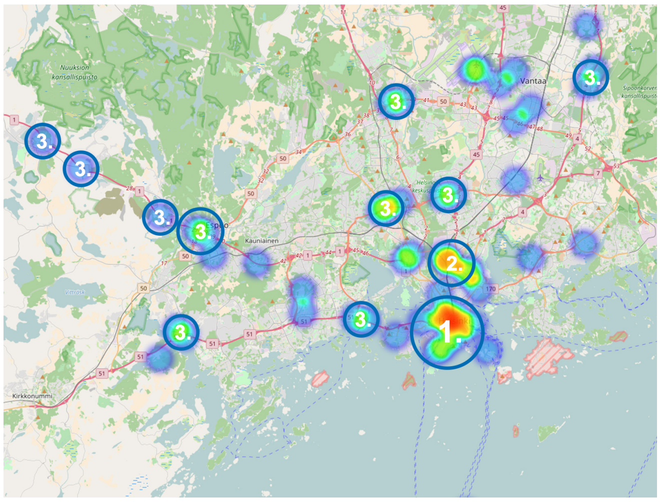

Fig. 5. Negative sentiment hotspots from the weekend of week 32. Many positivity clusters are also negativity clusters.

Demographics API [22], a text analysis software, by a Dutch company called Applied.ai, predicted the age and gender. According to their documentation, their API can achieve 65% accuracy on gender prediction and 58% accuracy on age category. It is strictly text-based. The operating principle is that certain use of phrases and certain naming conventions are neccesarily generation-bound and gender-bound. The gender prediction was further improved by comparing the twitter user's given names to a list of Finnish first names by Väestörekisterikeskus [23].

D. THE CODE

The code can be downloaded from GitHub (urlhttps://github.com/solarii/Local-Emotions) and run locally. The end product with test data visualization can also be seen at the author's website: `http://niklasstrengell.fi/dev/localemotions`. We can not publish the original dataset used, because it would violate The Terms of Service of Twitter's Streaming API.

E. RESULTS

We are able to gather data and classify it as positive or negative and show it on map with the demographic groups added as demonstrated here.

How much data we are able to gather varies. Twitter's Streaming API itself gives only access to around 1% of all the tweets. For Helsinki area during a week or a busy weekend this amounts to around 10 000. Of these 10 000 we discarded almost 9000 for this visualization, because they do not have exact location coordinates, but a vague bounding box which can range in size anywhere from a city to a country. For 1152 tweets with exact coordinates from the weekend, VADER gives us 419 tweets classified as positive and 121 as negative. For the weekdays we have 915 tweets with exact coordinates, of which 319 are classified as positive and 131 as negative.

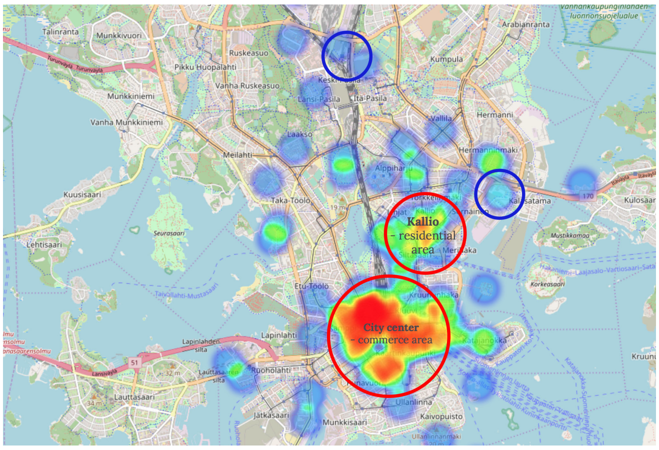

Fig. 6. Positive sentiments during working days of the week 33. The hotspots have changed from weekend oriented culture locations (marked in blue) to business and residential areas.

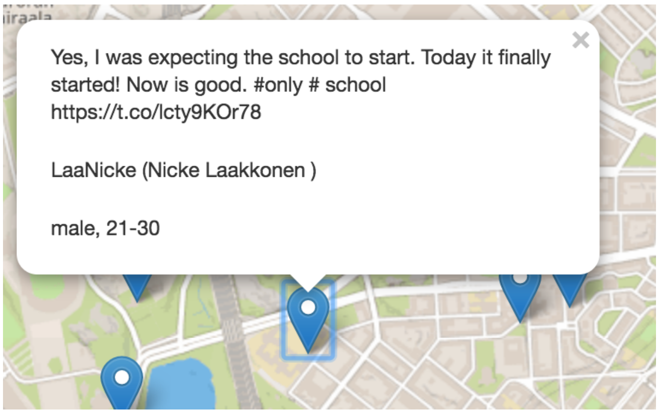

Fig. 7. A true positive. VADER can correctly distinguish that the sentiment is positive even with the weird translation. The demographics are also in the right category.

On the sentiment classification VADER toolkit gives a 96% classification accuracy, which does not significantly drop even with the added translation. Most of the tweets visualized in our demo are true positives.

There are some mistakes, but some of these are partly due to sarcasm or irony. I.e. the sentence "I thought you were a better person than that" is falsely classified as a positive tweet. To compensate, VADER seems to be able correctly classify very hard sentences with a lot of negative words as correctly positive, such as the one where a passing thunder storm destroyed electrical devices but the tweeter's streaming devices survived. There are also some ambivalent tweets classified as positives: is it a positive or a negative sentiment if a road is closed due to a marathon? It depends whether you are a partaker in the marathon or someone rushing to work.

Whatismore, our application shows some distinctive hotspots of emotions arising in Helsinki during the weekend. We can see the positive clusters in the figure 4 and negative in the figure 5. Further details about the hotspots can be seen in tables I and II. The hotspots also differ during the week

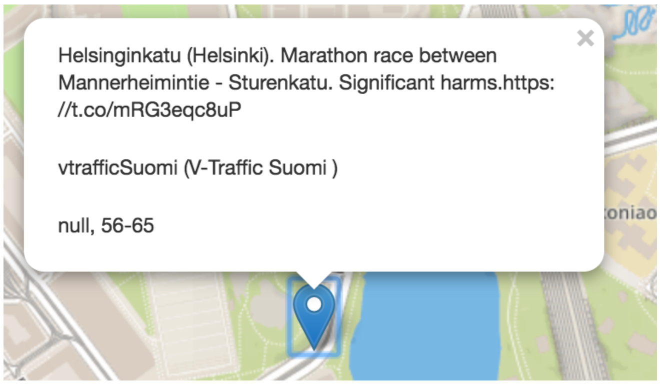

Fig. 8. An ambivalent positive. Is a road closure due to marathon a positive or negative thing?

as shown in figure 6. The cultural areas, such as Suvilahti and Messukeskus, where people gather during the weekend are not at all active during the week. During weekdays, much more activation is seen in the city center, which is the retail and business center, and Kallio area, which is a trendy neighbourhood of Helsinki with a lot of bars and cafeterias.

With animation we can also visualize how the hotspots arise in time domain, giving us an interesting view how location tied social media usage evolves during the day and night. The animation is available in the demo at the author's website: `http://niklasstrengell.fi/dev/localemotions/`.

F. SHORTCOMINGS

Social media data is not all encompassing. Most of it is positive and the content is shared mostly by males and younger people. We are thus not getting a view of the society as a whole. However, this might even be advantage if we specifically wanted to research younger people.

We also only focus on geolocated tweets here which narrows down our sampling. Not all content is shared with exact location coordinates and some content creators are not willing to share their location at all. Even with location tagging enabled, almost 90% of the tweets must be discarded because they don't have exact coordinates. Of course, if we would study with bigger granularity such as a city or a county we would get much more data.

V. CONCLUSIONS AND FUTURE WORK

As argumented in the introduction, a true smart city should understand it's citizens feelings and behaviour. It should study and observe where and how value and quality is created and where it is destroyed. For this social media can be a very valuable tool - to study human behavior in natural context and *in situ*. This could be combined with surveys to cross-validate and complete each other. Surveys are comprehensive and thorough, but are tedious and results might take time. Social media data can be shallow, but its reach is much wider and it can be gathered and analyzed in almost real time.

We've demonstrated in this proof-of-concept model, that we can pick up and visualize emotional patterns arising in locations and evolving through time. Urban planners could use this data and visualization to see how their choices affect the environment or to predict where intervention or redesign of city infrastructure is required. Even though our model might not be 100% accurate, it is valuable in it's simplicity and ability to show a lot of data and at once. Furthermore, it shows data that would otherwise be lost in the city design process.

Neither does our method give straight answers to the questions posed by Rapoport to study human-environment-interaction. But it can give us valuable hints where and how people gather and experience and how those experiences tie into emotions. The data from emotional experiences could be gathered and analyzed to build a real emotional database: not just mapping facial expressions to words, but really creating semantic features on what it means to experience a certain emotion. To be happy is not just to smile: to be happy can be many things, such as going back to school or eating an ice cream or finding a lost phone.

Modern data-analytics could take this workflow even further with bigger amounts of data and more accurate emotion classification and demographics prediction. Building language specific sentiment analyzers or demographics prediction toolkits would improve the prediction models. Location prediction without any coordinates could also improve the method. However, surveying people's feelings also raises ethical questions: is it viable for our well-being and development as individuals and as a species? Could monitoring emotions give us a better understanding of our own-selves and enable us to evolve as a species? Or does it push us closer to an utilitarian dystopia of mistrust where even our own inner feelings are constantly being watched upon?

References

[1] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions", in *Proceedings of the 12th annual international digital government research conference: Digital government innovation in challenging times*, ACM, 2011, pp. 282–291.

[2] R. W. Levenson, "The intrapersonal functions of emotion", *Cognition & Emotion*, vol. 13, no. 5, pp. 481–504, 1999.

[3] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, *et al.*, "The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions", *International journal of communication*, vol. 5, p. 31, 2011.

[4] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors", in *Proceedings of the 19th international conference on World wide web*, ACM, 2010, pp. 851–860.

[5] A. Rapoport, *Human aspects of urban form: Towards a man—environment approach to urban form and design*. Elsevier, 2016.

[6] M. Kahila and M. Kyttä, "Softgis as a bridge-builder in collaborative urban planning", *Planning support systems best practice and new methods*, pp. 389–411, 2009.

[7] J. J. Gross, "The future′s so bright, i gotta wear shades", *Emotion Review*, vol. 2, no. 3, pp. 212–216, 2010.

[8] N. H. Frijda *et al.*, "Varieties of affect: Emotions and episodes, moods, and sentiments.", 1994.

[9] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining.", in *LREc*, vol. 10, 2010.

[10] B. Pang, L. Lee, *et al.*, "Opinion mining and sentiment analysis", *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[11] J. Koponen, J. Hildén, T. Vapaasalo, *et al.*, "Tieto näkyväksi: Informaatiomuotoilun perusteet", 2016.

[12] E. R. Tufte, S. R. McKay, W. Christian, and J. R. Matey, "Visual explanations: Images and quantities, evidence and narrative", *Computers in Physics*, vol. 12, no. 2, pp. 146–148, 1998.

[13] C. Nold *et al.*, "Emotional cartography: Technologies of the self", *Online book). Last accessed in December 5th*, 2009.

[14] J. Pánek, "Emotional maps in the context of participatory planning-examples from the czech republic, spaces semiotic-ethnographies",

[15] J. Pánek and K. Benediktsson, "Emotional mapping and its participatory potential: Opinions about cycling conditions in reykjavík, iceland", *Cities*, vol. 61, pp. 65–73, 2017.

[16] J. Roesslein, "Tweepy documentation", *Http://tweepy.readthedocs. io/en/v3*, vol. 5, 2009.

[17] I. Google, *Cloud translation api*, 2017. [Online]. Available: `https://cloud.google.com/translate/` (visited on 08/28/2017).

[18] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text", in *Eighth international AAAI conference on weblogs and social media*, 2014.

[19] V. Agafonkin, *Leaflet - a javascript library for interactive maps*. [Online]. Available: `http://leafletjs.com/` (visited on 10/27/2017).

[20] O. Foundation, *Openstrreetmap*. [Online]. Available: `http://www.openstreetmap.org/` (visited on 10/27/2017).

[21] P. Wied, *Heatmap.js : Dynamic heatmaps for the web*. [Online]. Available: `https://www.patrick-wied.at/static/heatmapjs/` (visited on 10/27/2017).

[22] Ai-Applied.nl, *Demographics api*, 2017. [Online]. Available: `http://ai-applied.nl/demographics-api/` (visited on 08/28/2017).

[23] Avoindata.fi, *Etunimitilasto 2017-04-06 vrk*, 2017. [Online]. Available: `https://www.avoindata.fi/` (visited on 08/28/2017).