Title
Crowd-Sourced Neighborhoods - User-Contextualized Neighborhood Ranking

Authors
Sandoval Olascoaga, Carlos S
Xu, Wenfei
Flores, Hector

Publication Date
2016-08-01

Peer reviewed

Crowd-Sourced Neighborhoods

User-Contextualized Neighborhood Ranking

Carlos Sandoval Olascoaga[1], Wenfei Xu[2], Hector Flores[3]
[1,2,3]Massachusetts Institute of Technology
[1]csandova@mit.edu [2,3]{xu.wenfei|hectorcflores}@gmail.com

Finding an attractive or best-fit neighborhood for a new resident of any city is not only important from the perspective of the resident him or herself, but has larger implications for developers and city planners. The environment or mood of the right neighborhood is not simply created through traditional characteristics such as income, crime, or zoning regulations - more ephemeral traits related to user-perception also have significant weight. Using datasets and tools previously unassociated with real-estate decision-making and neighborhood planning, such as social media and machine learning, we create a non-deterministic and customized way of discovering and understanding neighborhoods. Our project creates a customizable ranking system for the 195 neighborhoods in New York City that helps users find the one that best matches their preferences. Our team has developed a composite weighted score with urban spatial data and social media data to rank all NYC neighborhoods based on a series of questions asked to the user. The project's contribution is to provide a scientific and calibrated understanding of the impact that socially oriented activities and preferences have towards the uses of space.

Keywords: *Textual Semantic analysis, machine learning, participatory planning, community detection, neighborhood definition*

INTRODUCTION

With contemporary computational advances urban models have become richer, and represent more aspects of the city, they have not necessarily become more accurate in their predictions (Batty 2016). We believe that the kind of problem the city is cannot be addressed through a single comprehensive model, but through contextual individual representations of the world -our platform presents a first step towards that by harnessing personal participation in the computation process, and going beyond traditional analytical models.

A Contextual Urban Computing

The city is constructed iteratively and in an interconnected manner by its different inhabitants; the most quantitatively optimal representation may not, in fact, be the most socially accurate one. As Herbert Simon suggested in "The Sciences of the Artificial", urban growth is comparable to the process of painting (Simon 1981), where the construction of the overall result does not occur in a single homogenous layer. Rather, local social interactions and their accretions constantly reframe, modify, and construct the larger whole. Cities are continuously reframed and reinterpreted by their inhabitants.

In order to construct a holistic representation that allows the contextualization of the vast amount of urban datasets and layers there is a need to incorporate the use of granular, socially produced datasets and a direct interaction with the user. Cities and neighborhoods ought to be understood beyond data; by conjoining traditional data with user produced data, we hypothesize that it is possible to embed interrelated spatial and non-spatial conditions and construct personal representations of place. Open-ended online-systems can turn urban models into instruments that facilitate interaction between designers and stakeholders. Current analysis tools and urban models are largely based on digital mapping; however, an attempt of urban legibility through digital modeling can only present a single, incomplete static representation of the world.

As an incomplete description of the city, urban models aiming to predict urban processes such as neighborhood ranking ought to be complemented by alternative approaches. By incorporating crowdsourced datasets and user participation in the construction of heterogeneous urban models, it is possible to contextualize urban information: exploring and reframing urban information with different levels of aggregation highlights outliers and serendipity. In this project we explore bottom-up user driven computation to give planners and users the capacity to continuously reframe urban datasets and conditions, discovering and creating new relationships within the urban system, and allowing engagement with different actors of city-making. To understand the complex dynamics of cities, models should be transparent and open to transformation, contextualization and collaboration; rather than simulating an average, or optimal urban condition, their assumptions should be tuned to the contextual and heterogeneous nuances of the city.

BACKGROUND

Our project is part of a collaborative effort to understand how open data and social media can be layered above traditional municipal data sources in order to provide a more fine-grained and individualized understanding of neighborhood characteristics. In the following sections, we provide an overview of the qualitative and methodological research and projects on top of which our research is built. The relevant qualitative context are those projects that address the topic of personalized neighborhood recommendations based on an extensive set of features. Our methodological context consists primarily of research that fine-tunes clustering algorithms from both a spatial and lexical perspective.

Qualitative Context

To help us frame our research question and to better understand the landscape of neighborhood characterization, we first consulted popular tools and interfaces that help users personalize their real estate decisions. Through these resources, we gain a better understanding what is currently available to users and what is understood to be important for those people such as home buyers and renters, who seek to relocate.

Livability Calculator. The Livability Calculator developed by Nate Silver (Silver n.d.) is a web application that ranks New York City neighborhoods according to urban dimensions like affordability, school accessibility, housing quality and green space. It also provides a number of pre-defined profiles, like "Young, Single, and Cash-Strapped", or "Double Income, No Kids", that sets the score given to each of the livability dimensions to specific values and provides a custom ranking of the neighborhoods.

Walk Score. Walk Score is a static neighbor 'walkability' assessment tool. It calculates walking distances to nearby amenities, which indirectly assesses the proximity to amenities, and decays these distances by time to create a composite score. Additionally, Walk Score aggregates scores within an entire neighborhood in order to rank the 'walkability' of that neighborhood. Walk Score is an earlier commercialized method for assessing livability, but aided in generating interest in this field.

Trulia and Zillow. Increasingly, real estate database companies such as Trulia and Zillow are offering services which give evaluative powers to the user in the decision-making process. These tools, such as Zillow's Zestimate or Trulia Estimates, are primarily used for assessing the value of a building, but include such features for surrounding schools, nearby home values, historical housing prices in the neighborhood, and Walk Scores.

Methodological Context

A line of more academic research has focused on refining methods to categorize neighborhoods and extrapolating groups from those data that characterize a neighborhood. The primary strategy is clustering through different methods of unsupervised learning of spatial and semantic data, often with a temporal element.

Spatial Methods. The baseline spatial clustering technique used in this field, on top of which most other methodologies is created, is density-based spatial clustering of activity with noise (Ester et al. 1996). This algorithm groups points that are close together, while marking as outliers those points that are in a low point-density region. More recent neighborhood description techniques build upon the idea of clustering. We look at two main methods of spatial differentiation: One is premised on the primacy of activity mobility, and the other is premised on homogeneity of points-of-interest.

CitySense (Loecher and Jebara 2008) is a discovery tool for temporal and spatial hot spots of activity in the city that has been implemented as a mobile application. The CitySense algorithm, in its first iteration, was a clustering algorithm based on GPS data from different sources such as taxis and people. The model was intended to not only create hot spots, but also to match users who have similar activities during similar points in the day in order to create predictions on future behavior or preferences.

The Hoodsquare project (Zhang et al. 2013) developed an algorithm for extracting neighborhood boundaries in cities, including New York City based on social media data. The algorithm works by using data related to Foursquare venue types, spatial distribution of local and tourists in the city. Their neighborhood definitions goes beyond the neighborhood as administrative or politically defined units in order to unearth geographies that are much more in accord with temporal activities in the city. The Livehoods project (Cranshaw et al. 2012) combines both points-of-interest clustering with the clustering of similar activity 'schedules'. It is based on the development of an algorithm that uses Foursquare data in order to produce neighborhoods classifications based on spatial and social proximity of venues in New York City. The Livehood project has developed a compelling method to unearth a classification of New York City neighborhoods by taking advantage of massive data sets and unsupervised learning approaches. A significant result of this approach is the Livelihood's definition of a neighborhood as an "an urban area [...] defined not just by the type of places found there, but also by the people that choose to make that area part of their daily life"(Cranshaw et al. 2012). Vaca, Quercia, Bonchi, and Fraternali (2015) recently looked at a neighborhood description technique that involves both clustering and labeling simultaneously. Whereas previous techniques first cluster nearby points on a map and then assign labels to the points, Vaca et al. use a technique for hierarchical clustering that merges branches only when it increases an objective function.

Semantic Methods. Unique in our methodology for neighborhood definition is the usage both traditional geo-spatial data and user-generated social media data. Because much of the social media data is both textual and spatial, we explored methods of parsing tweets and tags that would allow us to specify neighborhood qualities. Urban dynamics as explained by geolocated social media data has been explored by the likes of Noulas et al. 2011, and Chang et al. 2011 using geolocated data on categories or tags. Other papers such as Chang et al. 2012 and Frias-Martinez and Frias-Martinez 2014, have looked more closely at Twitter tweets texts and patterns.

METHODS

Discussed here are the structural components of our ranking system for the 195 neighborhoods in New York City. We present the practicalities of constructing the ranking system, producing personalized user results through machine learning techniques, and representing the data for user interaction.

The project proposes the use of machine learning techniques to analyze aggregate datasets, and predict regularities across regions, such as mood, theme, or related trends. In particular, we make extensive use of natural language processing (NLP) (Blei 2012, Blei et al. 2003) to characterize and contextualize geo-located social media datasets. The diverse set of analysis techniques implemented and developed for the project result in a non-spatial representation of the city, where, in contrast to geospatial analysis, which prioritizes Euclidian distances, relationships are discovered across spatially diverse locations through a sameness of social and environmental characteristics within the city. The development of a publicly available web interface provides agency to the user, allowing the analysis and construction of the city to be made according to their individual preferences, encouraging a continuous rediscovery of the city.

EXPLORATORY ANALYSIS

In order to determine the differences of neighborhood classifications when using conventional datasets or the non-conventional datasets and methodologies proposed by the project, we performed an initial exploratory data analysis with conventional GIS datasets. The initial exploratory data analysis also served to uncover basic relationships in the data that were later used for the development of the composite neighborhood ranking. Among others methods, we used unsupervised algorithms such as Principal Component Analysis, K-Nearest Neighbors, K-Means and hierarchical clustering algorithms, to investigate possible classifications of the urban and social data we had, as well as the neighborhoods themselves. The initial study allowed us to identify the dimensions or urban and social media data that would be the most helpful for users to decide which neighborhoods to live in.

GIS. Over more than month in November 2014, the authors gathered social media and live datasets from publicly available web APIs; datasets like tabular and GIS files were acquired from state and municipal open-source repositories (see Table 1).

We used New York City's Socrata Open Data portal to understand the static, underlying characteristics of the city. The portal contains a wide variety of municipal data including housing, transportation, business, and environmental data. Additionally, much of the administrative and neighborhood boundary was also from Socrata.

The 311 information includes all the complaints that were reported to the city. The relevant fields that were used for the final analysis are complaints regarding construction noise, dirty conditions, graffiti, and derelict vehicles, as determined that these factors combined would result in a generalized indication of an undesirable place to live. From these data, we were able to achieve a better understanding of how these neighborhoods compare in terms of the more traditional desirable housing qualities.

Hierarchical Clustering. We utilized K-Nearest Neighbors and K-Means, unsupervised learning algorithms that iteratively learn and classify datasets based on a number of predefined clusters. Validating and calibrating the number of predefined clusters proved to be "sticky", as the traditional datasets could not capture the diversity of the actual neighborhoods. We additionally implemented a hierarchical clustering algorithm, a bottom-up greedy search algorithm that pairs similar clusters together, moving up on the cluster hierarchy (see Figure 2). Beyond the broad clustering performed through the PCA analysis, the hierarchical clustering allowed us to provide an initial classification of the neighborhoods. The

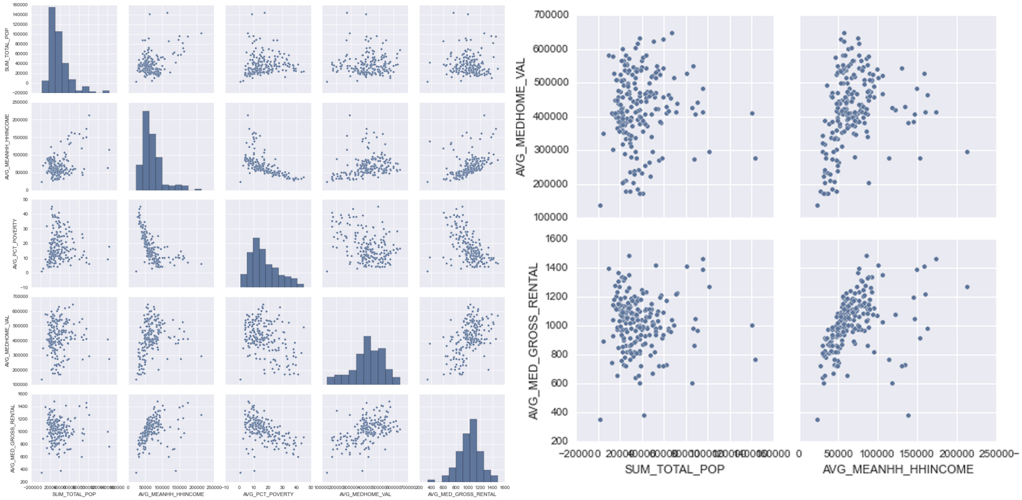

Figure 1
Exploratory data analysis of urban data of NYC neighborhoods. Zoom (left).

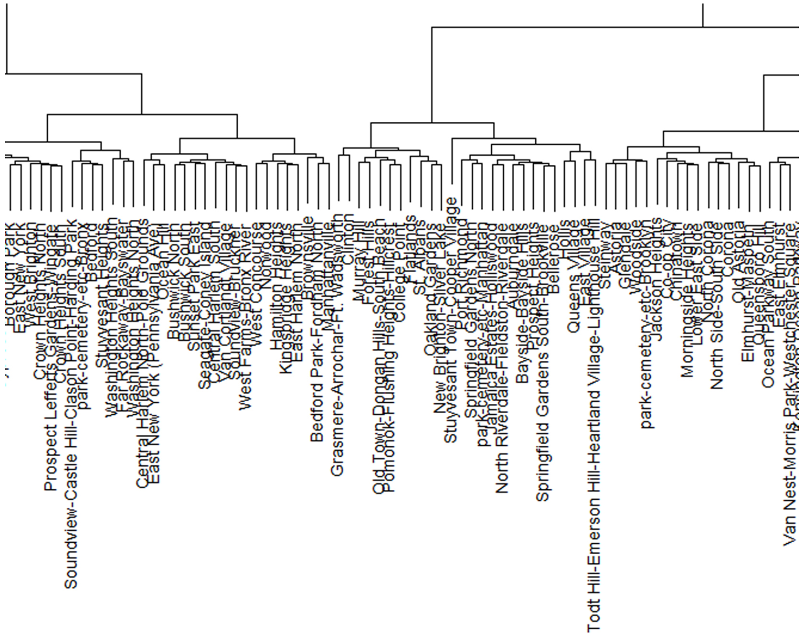

Figure 2
Unsupervised Clustering of Neighborhoods (Zoom).

hierarchical clustering algorithm suggested that the optimal number of clusters for the 195 New York City neighborhoods was between 3 and 7 clusters.

The results of these classification methods were used to develop the overall weights for the composite ranking that will be described in the next sections, suggesting which data dimensions were the most meaningful to urban dynamics. We identified which of the datasets (i.e., safety, urban density, social) provided useful dimensions for users to rank (in terms of how much relevance they placed in them) and define distinguishable groups of neighborhoods.

Semantic Analysis

The project's methodology proposes the development of a personalized ranking through the use of non-traditional and dynamic datasets to understand different perspectives of urban livability. Our methodology uses machine-learning techniques combining supervised and unsupervised classifiers and clustering algorithms to make sense out of such disaggregated datasets. At the same time, such machine learning techniques are used to construct spatial measures of wellness that are hard to capture with standardized social measurements.

Semantic analysis is used to generalize the syntactic structures of text enabling the creation of relationships among large amount of social data obtained through web APIs. Specifically, we implemented a Naïve Bayes algorithm for sentiment analysis classification and a Latent Dirichlet Allocation (LDA) clustering algorithm for parent-child classification of the social media datasets. The results of the machine learning algorithms were then combined with the rest of the datasets utilized in the project to inquire about the implications of social dynamics in the interpretation and construction of urban form and city dynamics.

Furthermore, the classification algorithms allow users to develop subjective and personal choices about the way they differentiate neighborhood conditions. By interacting with the web interface they can refine their queries and neighborhood classifications based on the real-time feedback obtained from the system.

Supervised Learning. The platform uses a sentiment analysis algorithm, to determine the overall contextual polarity of a textual dataset. We implemented a Naive Bayes Classifier (Norvig 2003), a probabilistic classifier that was previously trained with a corpus of over 2,000 document rating classifications. While Naïve Bayes classifiers have been employed since the 1960s, their use on social media data is inherently problematic -the inconsistencies and highly contextual and specific lexicon difficulties the use of popular out-of-the-box classifiers. To deal with such textual inconsistencies, the media posts were cleaned and parsed -we eliminated non-English words, since our classifier was trained with an English word corpus words not in the English dictionary, and we eliminated excessive punctuation. Although other tests involved the analysis of hashtags, ultimately, only English words in the tweet were analyzed in order to be consistent with the training set used for the algorithm.

After training the classifier, we evaluate every data point with the classification algorithm, which returns a normalized polarity value for each social media data point that is then aggregated spatially by neighborhood. The aggregated values are then used as part of the composite neighborhood ranking that will be introduced in the next section. Additionally, through the web interface, users are able to relate different neighborhood sentiment thresholds to a number of social activities and physical characteristics of the space.

Unsupervised Learning. We used topic modeling to discover the abstract "topics" that occur in a collection of social media posts -essentially a collection of textual data points are classified into k-number of topics that share similar topics in their content. Through topic modeling a set of text data points are evaluated; based on the statistical similarities of the words in each text data point their topics are evaluated and classified. We used Latent Dirichlet Allocation (LDA), an unsupervised learning algorithm -

Table 1
GIS Sources.

Source	*Size (Row count)*	*Year*
311	11,398,056	2010 - present
Legally Operating Businesses	74,590	2014
SAT Results	498	2014
Department of Finance Condominium Comparable Rental Income	2,627	FY 2011/2012
Land use	2,000	2014
Subway stop information	1,905	2014
Neighborhood boundaries	195	2014

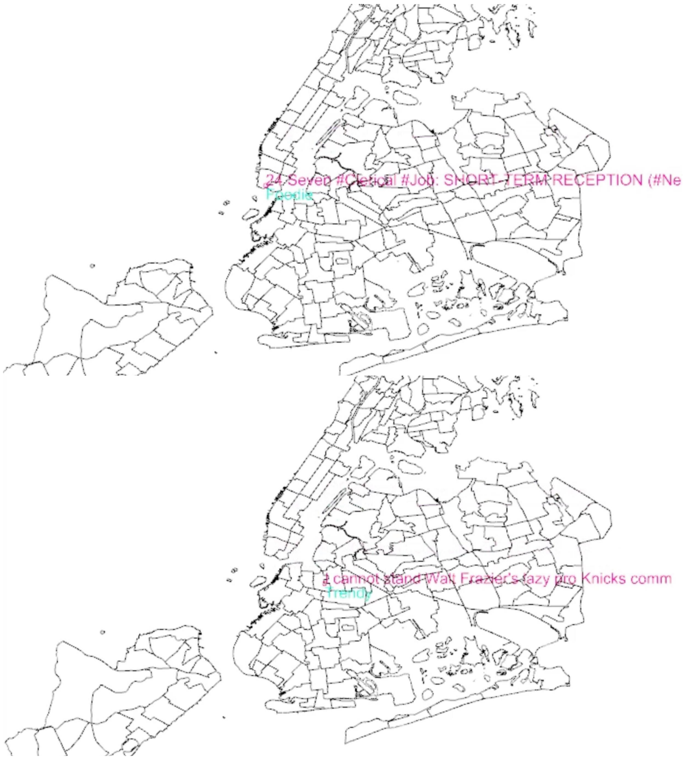

Figure 3
Tweets
Classification.

the most common topic modeling method (Blei et al. 2003). The algorithm iteratively divides a dataset of text data points into a number of k-subgroups with common topics.

Once the entire dataset of social media data points was classified, the most common words of each subgroup and their overall topics were assigned to each individual data point within the group. Every data point in the dataset was based on hashtags of tweets; this attempted to reduce the noise in the messages due to the wide use of emoticons, and non-English words in longer tweet texts. The advantage of such classification algorithms over a Naïve Bayes classifies is that since it does not require a pre-trained dataset, non-traditional textual elements such as social media hashtags can be iteratively classified according to their statistical similarities to the rest of the set.

Once the LDA model was trained, the tweets were divided in k-number of groups. Essentially, each group is classified based on how related the hashtags are to each other. Some of the topics that were extracted from the subgroups of the dataset were: trendy, foodie, nightlife, public space, music, social justice, and fitness. The individual topics of each data point were aggregated within the NYC neighborhoods, and then used to quantify the occurrences of different topics within neighborhoods -every neighborhood got an aggregated count of different emergent topics extracted from the conversations happening in the area through social media. By interacting with the user through the web interface, the users can discover the relationships among specific topics within a given neighborhood.

Composite Neighborhood Ranking

In our ranking, we used conventional urban data, like median home values or access to transit and safety measures, and also semantic and spatial social media data to gain insight on the mood and social life within New York City neighborhoods. In order to do so, our team developed a composite weighted score with urban and social media data to rank neighborhoods.

Given the diversity of users and preferences, the built-in request in our web app for users input is a way to give them control when defining what matters most to them, as opposed to providing a unique ranking system for all of them. The end product for our user is a ranking of all the 195 neighborhoods built according to the categories the user cares the most, and that will help make a better decision on which New York City neighborhood to live in.

Components of the Ranking. The five dimensions of the composite score are: social capital, affordability, urban density, safety, and the mood or sentiment within a neighborhood. Each of these dimensions is a simple average of its components variables, except for the safety score in which a much higher weight was assigned to the count of shootings (weighted by population) in order to reflect the severity of shooting events in comparison to other crimes that might be less relevant to deciding where to live (i.e. occurrence of graffiti in streets). We realize that our dimensions may have some challenges.

We used average SAT score in each neighborhood as a proxy for education quality, average mean household income as a proxy of resourcefulness, and average percentage poverty (in such a way that a higher average percentage of poverty per neighborhood will decrease the social capital dimension, as explained in the previous section) These levels are then averaged to create the social capital component.

In the case of the affordability dimension, we incorporated data related to the average median home value, average median gross rental prices, and the rent price percentage increase, and we include them as indicators more expensive places to live in. This does not reflect, however, that higher prices might reflect simply better value for each dollar spent in the case of neighborhood with a high level of amenities and relatively low prices.

For the urban density dimension, we used the average total number of units, the average floor per area ratio (FAR), as well as other FARs (for example, commercial, residential, retail) to reflect neighborhoods with more density. We also included the average percentage of occupied (as opposed to vacant) house units and the weighted number of subway stations. In this dimensions, we have assigned most of the weight (65%) to the average FAR as we take it to be the single most important indicator of how dense a neighborhood is.

For the safety score we used counts of shootings, graffiti, dirty conditions events, and derelict vehicles reports, all of which came from the city's 311 data. Our safety measure was an average of these categories.

Finally, for the mood dimension, we use the sentiment analysis data extracted from our tweets geocoded by neighborhood (the proportion of positive tweets out of the total tweets in a given location).

Front-end Web Interface
In this section, we present the technical specifications and properties of the front-end web interface developed to enable the interaction with the computed neighborhood values, the construction of neighborhood choices, and the visualization of the values related to the specific combination of weighted values. The front-end is a D3 app that leverages D3's data binding to synchronize and compute the data displayed and edited by the UI with specific data on the computed matrices. The graphical component of the web interface is a choropleth map of NYC that is constructed from a geoJSON of NYC.

The UI is built with a series of sliders that allow the modification and construction of weighted queries that modify the neighborhood rankings on the fly according to the user choices; the sliders allow weight the different traditional and nonconventional neighborhood values that have been pre-computed through the composite neighborhood ranking. Additionally, a number of text boxes allow the user to input alternative characteristics of a neighborhood that can be queried from the social media datasets on the fly, adding additional weights to the composite neighborhood ranking. The color values of the map are generated dynamically by parametrically blending the personalized neighborhood rankings between two colors.

DISCUSSION
Our research straddles the traditional realm of real estate research, which is rooted in investigating the relationship between housing values and income, race, age, proximities to amenities, and transportation, and the more dynamic, interactive realm of social media. An example of this type of research is Nate Silver's Livability Calculator; it uses these more traditional characteristics to determine livability, but allows users to prioritize their own needs and desirable living characteristics. This calculator, however, misses the voice of the neighborhood's residents. One of the main promises of social media research is that it gives us a sense of vivid cultural pulse. In this way, our project aims to differ from the traditional real estate tools in that it incorporates the more ephemeral social elements that constitute a neighborhood.

Final Composite Score Results
In order to generate the final score for each of the 195 New York City neighborhoods and retrieve an ordered list of them for the user, we calculated a weighted composite score of the five dimensions as described above (social capital, affordability, urban density, safety and mood or sentiment). The way the application builds these scores is by asking the user to rank all dimensions from 0 to 5. These weights are calculated by adding the 5 ranking values the user has given for each dimension and dividing each one of them by the total. For example, if a user has given a score of 3, 5, 5, 4 and 4 to the social capital, affordability, urban density, safety and sentiment score, their weights would be 15% (3/20), 25% (5/20), and 20% (4/20), respectively. Finally, we multiply each dimension by its score and calculate and overall score that will rank all 195 neighborhoods in New York City.

We found that, when more traditional measures such as social capital, safety, affordability, and den-

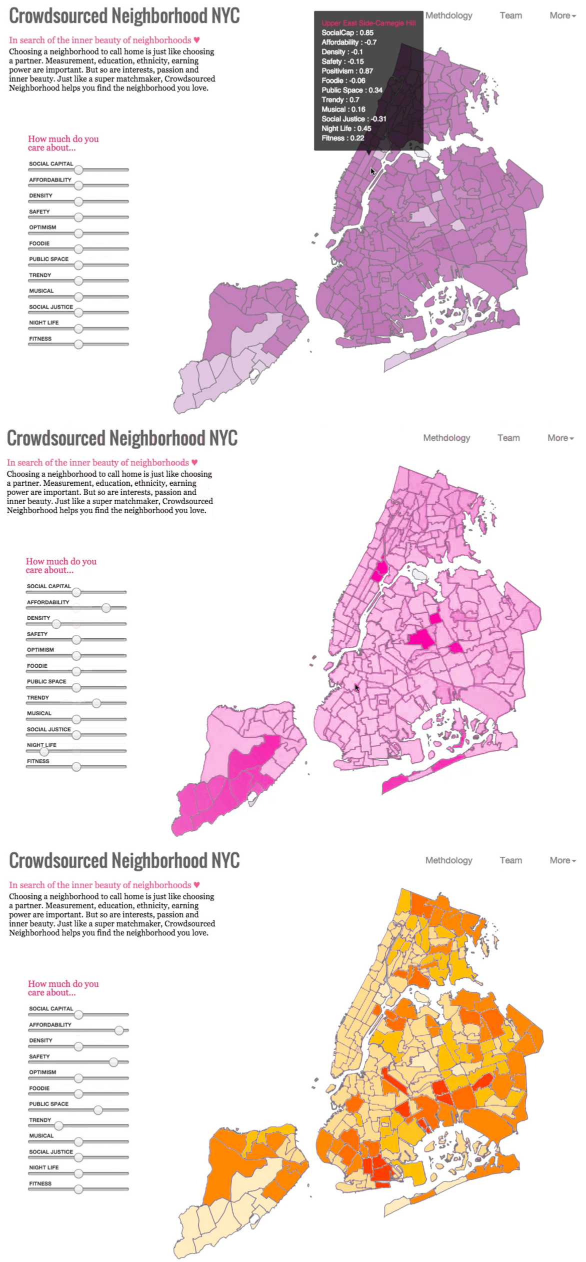

Figure 4
Changes in neighborhood rankings based on user-adjusted preferences.

sity are given preference, our results align fairly closely to those that Zillow or Trulia would produce. For instance, preference for high density and affordability above the other measures creates suggestions for neighborhoods in upper Manhattan, while preference for social capital produces neighborhoods on Upper East and West sides on Manhattan. When we prioritize mood, however, our results become more specific to particular neighborhood eccentricities that help shape the quality of those places. For instance, our 'night life' topic brought neighborhoods such as the Lower East Side higher in rankings, while the 'public space' highlighted those neighborhoods with easy access to Prospect Park and Central Park.

Contributions

In this project, we take the notion of crowd-sourcing qualitative information, following the model of services like Yelp or Foursquare, and apply them to decision-making about environmental preferences regarding urban spaces. The result is a tool that improves upon current ones like Trulia or Airbnb's recommendation engine in two ways: the first is that it gives agency and primacy to the decision making-abilities of the public which reflects the more subjective aspects of urban planning; the second is that gives the user a wider set of tools in their decision making process and incorporates more aspects of neighborhood description, creating a scientific metric for previously unquantifiable qualities.

Another important concept in our project, which should be highlighted, is the capacity for discovery: due to the non-spatial nature of certain types of activities, sub-communities form around certain lifestyle and recreational interests, which can exist across different neighborhoods. Our tool follows in the lineage of Tony Jebara's work on activity-based tribes - we create a real-world implementation of the concept that our choice of location can be determined by the social activities we choose to be involved with. For instance, a devoted yogi who wants to live close to an advanced yoga center in addition to being located near public transit and a holistic foods store, can choose between Tribeca, Greenpoint, Sunset Park, amongst other neighborhoods across the city. In this way, the yogi using the tool can discover new areas of the city that satisfy his social or recreational needs.

More importantly, the project is a first step towards the creation of urban models that through a transparent construction and exploration of urban meaning can minimize the power structures between stakeholders and citizens. Urban tools that promote civic engagement contribute to a collaborative construction of cities.

Future Work

Future work involves testing and refining the tool through feedback data. Currently, we have a 'test set' in the sense that our tool is not refined through calibration with user feedback. Tools in the real estate realm are beginning to incorporate a more detailed description of neighborhoods to give a unique feel. For instance, Airbnb has begun to create neighborhood profiles for some of their host cities which include labels that summarize the character of the area as well as qualitative images and descriptions. We envision our tool could directly connect to sites such as Airbnb, Craigslist, or Zillow and provide a more scientific and qualified version of Airbnb's neighborhood recommendation tool. This would give users who start on our site the ability to find places to stay in the neighborhoods recommended from our tool and provide us data on the successfulness of engine.

REFERENCES

Blei, D. 2012, 'Probabilistic Topic Models', *ACM*, 55, pp. 77-84

Blei, D., Ng, A. and Jordan, M. 2003, 'Latent dirichlet allocation', *J. Mach. Learn*, 3, pp. 993-1022

Chang, J and Sun, E 2011 'Location3: How Users Share and Respond to Location-Based Data on Social', *IC-SMW*

Chrisman, N 2006, *Charting the Unknown: How computer mapping at Harvard became GIS*, ESRI Press

Cranshaw, J, Schwartz, R, Hong, J and Sadeh, N 2012 'The Livelihoods Project: Utilizing Social Media to Understand the Dynamics of the City', *Assoc. Adv. Artif. Intell*

Frias-Martinez, V and Frias-Martinez, E 2014, 'Spectral Clustering for Sensing Urban Land Use Using Twitter Activity', *Engineering Applications of Artificial Intelligence*, 35, pp. 237-245
Jacobs, J 1961, *The Death and Life of Great American Cities*, Vintage Books
Light, J 2003, *From Warfare to Welfare: Defense Intellectuals and Urban Problems in Cold War America*, John Hopkins University Press
Loecher, M and Jebara, T 2009 'CitySense: Multiscale space time clustering of gps points and trajectories', *Proceedings of the Joint Statistical Meeting*
Martin, E, Krieger, HP, Sander, J and Xu, X ND, 'A Density-Based Algorithm for Discovering Clusters in LargeSpatial Databases with Noise', *Kdd*, 96, pp. 226-231
McHarg, I 1971, *Design with Nature*, Natural History Press
Norvig, R 2003, *Artificial Intelligence: A Modern Approach*, Prentice Hall
Noulas, A, Scellato, S, Mascolo, C and Pontil, M 2011 'Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-Based Social Networks', *SMW Workshop at ICWSM*
Townsend, A 2013, *Smart cities: Big data, civic hackers, and the quest for a new utopia*, WW Norton & Company
Vaca, C, Quercia, D, Bonchi, F and Fraternali, P 2015 'Taxonomy-Based Discovery and Annotation of Functional Areas in the City', *AAAI Conference on Web and Social Media*
Zhang, A, Noulas, A, Scellato, S and Mascolo, C 2013, 'Modeling and Recomending Neighborhoods in Location-based Social Networks', *Hum J*, 2, pp. 1-15