San Francisco	Manhattan	Top common local tags
Mission	East Village (0.23)	`graffiti`, `food`, `restaurant`, `mural`, `bar`
Golden Gate Park	Washington Heights (0.26), Upper West Side (0.22)	`park`, `museum`, `nature`, `flower`, `bird`
Financial District	Battery Park (0.29), Midtown Manhattan (0.27)	`downtown`, `building`, `skyscraper`, `city`, `street`
Treasure Island	Roosevelt Island (0.38)	`bridge`, `island`, `water`, `skylines`, `boat`
Chinatown	Chinatown (0.85)	`chinatown`, `chinese`, `downtown`, `dragons`, `lantern`
Castro	West Village (0.06)	`park`, `gay`, `halloween`, `pride`, `bar`

Table 2: Mapping from San Francisco neighborhoods to the most similar ones in Manhattan. For each neighborhood pair n and n', we give the cosine similarity between their local distributions θ_n and $\theta_{n'}$, and list the top "common" local tags ranked according to the product of tag probabilities $\theta_n(t)\,\theta_{n'}(t)$.

4.2 Experimental evaluation

Evaluating models such as the GHM on user-generated content is hard because of the absence of ground truth. There is no dataset that associates, with objectivity, regions to specifically descriptive terms, or assigns terms to levels in a geographic hierarchy. This is due to the intrinsic subjectivity and vagueness of the human conception of regions and their descriptions [15]. In the absence of ground truth, a classic approach is to generate a dataset with known ground truth, and then use it to evaluate the performance of different classifiers. We follow the generative process presented in Algorithm 1 using the geo-tree (3 levels, 68 nodes) built from the dataset presented in Section 4.1. For each node v in our geo-tree, we sample the distribution of tags θ_v from the symmetric Dirichlet distribution $Dir(\alpha)$. We set $\alpha = 0.1$ to favor sparse distributions θ_v. If the node is a leaf, i.e. it represents a neighborhood, we also sample the mixture coefficient $p(z|v)$ from a symmetric Dirichlet distribution $Dir(\beta)$. We set $\beta = 1.0$ to sample the mixture coefficients uniformly over the simplex and have well-balanced distributions. Now that we have the different distributions that describe the geo-tree, we can start generating the tags in each neighborhood. We vary the number of samples per neighborhood in order to reproduce a realistic dataset that might be very unbalanced: data is sparse for some neighborhoods while very dense for some others. For each neighborhood n, we sample uniformly the continuous random variable γ, which represents the order of the number of tags in neighborhood n, from the interval $[3, 6]$. The endpoints of the support of γ are based on the the minimum and maximum values observed in the Flickr dataset presented in Section 4.1, such that an expected number of 18.8×10^6 tags will be generated, similar to the quantity of tags available in the Flickr dataset. Once the number of tags ν to be generated is fixed, we sample, for each iteration, a level of the geo-tree and then a tag from the distribution associated to the corresponding node.

We can now assess the performance of a model by quantifying its ability to correctly predict the level in the geo-tree from which a tag observed in a neighborhood was sampled. In addition to our model, we consider the following methods:

Naive Bayes (NB) is a simple yet core technique in information retrieval [16]. Under this model, we assume that the tags observed in a class (node) are sampled independently from a multinomial distribution. Each class is therefore described by a multinomial distribution learnt from the count of tags that are observed in that class. However, since we do not use the class membership to train our methods, we assign a tag t, observed in neighborhood n, to all the classes (nodes) along the the path R_n, which amounts to having a uniform prior over the classes.

Algorithm 1: Generating tags in neighborhoods

Input: Geo-tree V, neighborhoods $n \in \{1, \ldots, N\}$, tags $t \in \{1, \ldots, T\}$, hyper-parameters α, β.
Output: Tags observed in each neighborhood $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$.

for $v \in V$ **do**
  Sample distribution $\theta_v \sim Dir(\alpha)$;
  if *v is a leaf* **then**
    Sample mixture coefficients $p(z|v) \sim Dir(\beta)$;
for $n \in \{1, \ldots, N\}$ **do**
  Initialize $\boldsymbol{x}_n = \boldsymbol{0}$;
  Sample order $\gamma \sim \mathcal{U}\,[3, 6]$;
  Set $\nu = \lfloor 2 \times 10^{\gamma} \rfloor$;
  while $\sum_t x_{nt} \leq \nu$ **do**
    Sample tree level $z \sim p(v|n)$;
    Sample tag $t \sim p(t|z, n)$;
    Increment tag count $x_{nt} \leftarrow x_{nt} + 1$;

Hierarchical TF-IDF (HT) is a variant of TF-IDF that incorporates the knowledge of the geographical hierarchy. This variant was used in the TagMaps method [23] to find tags that are specific to a region at a given geographical level. The method assigns a higher weight to tags that are frequent within a region (node) compared to the other regions at the same level in the hierarchy. We are able to represent each node with a normalized vector in which each tag t has a weight that encodes its descriptiveness.

For the classification, we map a tag t observed in the leaf n to the level $\hat{z}$ that maximizes the probability $p(z|t, n)$ (NB and GHM), or the tag weight (HT). Using our ground truth, we can then approximate the probability of correct classification $p(\hat{z} = z)$ by the proportion of tags that were correctly classified. In our evaluation, we repeat the following process 1000 times: we first generate a dataset, hold out 10% of the data for test purposes and train the model on the remaining 90%. For fair comparison, we initialize and smooth the parameters of each method similarly. Then, we measure the classification performance of each method. The final results, shown in Table 3, are therefore obtained by averaging the performance of each method over 1000 different datasets.

Our GHM model is the most accurate at classifying the tags to the correct level, greatly outperforming NB by 47% and HT by 27%. Even though both GHM and HT take ad-

	Classification Accuracy (std)
Random	0.33 (0.00)
NB	0.51 (0.02)
HT	0.59 (0.02)
GHM	**0.75** (0.01)

Table 3: The average classification accuracy is computed, for each method, over 1000 generated datasets. We also indicate this accuracy if we classify tags uniformly at random.

vantage of the geographical hierarchy in order to classify the tags, the probabilistic nature of GHM enables a more resilient hierarchical clustering of the data, while the heuristic approach of HT suffers from overfitting. For example, if the number of samples available for a neighborhood is low, HT might overfit the training data by declaring a frequent tag as being characteristic, although not enough samples are available to conclude this. This is not case for GHM, because the assumptions of random mixture enable us to obtain a resilient estimate of the distributions, which declare a tag as characteristic of a given level only if it has enough evidence for it. This observation is strengthened if we choose the maximum order γ of the number of tags per node to be 4 instead of 6: the classification accuracy of GHM decreases by 5% only (0.71), whereas the performance of HT decreases by 13% (0.51). Taken together, these results suggest that, if the data observed in a neighborhood is a mixture of data generated from different levels of a hierarchy that encodes the specificity/generality of the data, our method will be successfully able to accurately associate a tag to the level from which it was generated.

5. PERCEPTION FOCUSED USER STUDY

The results produced by our model might not necessarily be intuitive nor expected, especially in the light of people's differing individual geographic perspectives [19, 27, 11]. Trying to objectively evaluate these, without taking into account human subjectivity and prior knowledge, could be misleading. For the Castro neighborhood in San Francisco, for example, the GHM classified the tag `milk` as specifically descriptive. Someone who is not familiar with Harvey Milk, the first openly gay person to be elected to public office in California and who used to live in the Castro, would most probably not relate this tag to the neighborhood. We thus need to understand the correspondences and gaps between our model's results and human reasoning about regions.

We conduct a user-focused study to explore the premise that the posterior probability of a tag being sampled from the distribution associated with a region is indicative of the canonical descriptiveness of this region. We further aim to identify potential challenges in user-facing applications of the model, and to uncover potential extensions to our model. We held ten interviews and conducted a survey with local residents of the San Francisco Bay Area focusing on their reactions to the tags that our model surfaced. To assess the performance of our model while reducing the bias of subjectivity, we used the results of our user survey to obtain the human classification of tags to nodes in the hierarchy, allowing us to approximate the probability that our model classified a tag correctly by the average number of times it corresponded with human classification. We use the interviews to understand the reasons behind matches and mismatches.

5.1 Interview and survey methodology

Interviewees and survey respondents reacted to a collection of 32 tags per neighborhood. To ensure a certain diversity among the tags presented to the users, we select randomly a subset of tags that are classified by the GHM as being descriptive of (i) neighborhood level (e.g., `graffiti`), (ii) city level (e.g., `nyc`), (iii) country level (e.g., `usa`), or (iv) another neighborhood (e.g., `mission` for the Castro neighborhood). We selected these tags randomly, with the probability of choosing a given tag t proportional to the probability $p(t|n)$. The survey aims to highlight to which extent locals' perspectives match the results produced by our model, while the interviews highlight the reasons why and allow us to understand better the human perception of descriptiveness.

Interview procedure. Our semi-structured interviews focused on how people describe neighborhoods, and investigated their reasoning behind the level of specificity associated to a tag in a given neighborhood. Each one-to-one interview lasted 25–45 minutes. To ensure a wide range of (former) local to newcomer perspectives, we interviewed a total of 10 people (5F, 5M; ages 26–62, $\mu = 37$, $\sigma = 12$) that (had) lived in the San Francisco Bay Area from two months to 62 years. Three of the participants worked in the technology industry, two were students, one was a real estate agent, one a building manager and one a photographer. Each interview covered three different neighborhoods chosen by the participant out of 11 well-known San Francisco neighborhoods. However, one participant described only one neighborhood (due to time constraints), while another participant described four of them. Our interviews addressed:

1. Participants' characterization of the neighborhoods using their own words to get an understanding of the factors that are important to them.
2. Their considerations in whether a tag is perceived as specifically descriptive or not. Interviewees were first asked to classify the 32 tags presented to them as (not) specifically descriptive for the neighborhood, and explain the reasons. Then, they were shown the subset of tags that were classified by our model as specifically descriptive.

We emphasize the fact that the participants were not told about our model, nor that the terms presented to them were actually Flickr tags. This provided broader insight into the factors that led them to classify terms as (not) specifically descriptive of a neighborhood, and allowed for identification of factors not yet addressed by the model, without biasing their judgment towards the assumptions we made (i.e., hierarchical mixture of tags). The interviews were recorded, and transcriptions were iteratively analyzed, focusing on the identification of themes in interviewees' reasoning affecting classifications of tags as (non-)descriptive.

Survey procedure. A total of 22 San Francisco Bay Area residents (5F, 17M; ages 22–39, $\mu = 33$, $\sigma = 4.8$), who had lived there for an average of 5.6 years ($\sigma = 4.2$), participated in our survey about San Francisco and three of its neighborhoods (Mission, Castro and North Beach). Of these 22 respondents, 18 provided tag classifications for the Mission neighborhood, 12 for the Castro and 11 for North Beach. This resulted in 1291 tag classifications, of which 561 for the Mission, 381 for the Castro and 349 for North Beach. The survey asked respondents to describe each neighborhood with their own words using open text fields, and then to classify the 32 tags presented to them as descriptive for a given neighborhood, for a higher level (the city or country), or for another neighborhood. They could also indicate if they did not find the tag descriptive for any level, or did not understand its meaning.

5.2 Results

In this section, we present the results of the survey, and provide examples from the interviews to illuminate the reasoning processes on whether a tag is descriptive for a specific region. We compare the tag classifications provided by the GHM with those supplied by the participants, and we specifically focus on disagreements between neighborhood-level tag classifications in order to identify challenges in human interpretation of modeling results, and potential extensions to our model.

5.2.1 Participant and model congruency

The model's premise that locally frequent content is not necessarily specific to that locale was strongly supported by the interviews and the survey. None of the participants classified all of the tags that occurred *frequently* in a neighborhood as specifically *descriptive* of the given neighborhood. This supports the results of the GHM in classifications of very frequent but wide-spread tags as not being specifically descriptive of the neighborhood. Without prompting, interviewees mentioned terms as being too generic or specific for a given neighborhood. For example, one interviewee (F 28), when describing the Western Addition neighborhood, picked `haight` as a descriptive tag, "because there's Haight Street in this neighborhood", but not the tag `streets`, as "there's streets everywhere". Similar interview examples included: "`california` or `usa` is a generic, or general term" (F 28), "I don't think of ever describing Golden Gate Park as in the USA. Unless I'm somewhere far away, but then I wouldn't even say USA, I would say California or San Francisco" (F 41). This result implies that participants tend to classify tags according to a geographical hierarchy, which supports the validity of the assumptions we make about the hierarchy of tags: tag specificity/generality depends on the hierarchical level from which it was sampled.

We are aware that people will not agree with every classification made by our model. Tags classified by the GHM as specifically descriptive of a neighborhood, were not necessarily perceived as such by all respondents; variations occurred between neighborhoods and between participants. As a consequence, evaluating the results of an aggregate model 'objectively', as if there were only a single correct representation of a neighborhood, is difficult. However, to place the the GHM classification into context with human classification, we use the results of our survey to reduce subjective biases: we obtain a majority vote human classification by assigning each tag to the class that users have chosen most often. We then approximate the probability that the GHM classifies a tag 'correctly' by the average number of times its classification matches this human majority classification. For most tags the majority assignment is aligned with the assessment of the model (Table 5). We obtained an average classification correspondence of 0.77, with a highest classification accuracy of 0.84 for the Mission neighborhood. The alignment between the model and human classification for the Castro neighborhood was 0.81. Alignment was lowest for the North Beach neighborhood with a correspondence of 0.66, mainly caused by tag classifications as 'another neighborhood' (Table 4). Such mismatches occur for a multitude of reasons. First of all, as a very basic requirement, participants have to understand what a tag refers to, before they can assign it to a specific level. For example, in the survey, 8% of the answers given for the Mission neighborhood were "I don't know what this is" (Table 4). This issue occurred less for the selection of tags for the Castro (1%) and North Beach (0%). The terms that users understood were necessarily perceived as either descriptive (i.e. assignable to a level in the geographical hierarchy) or non-descriptive (i.e. not belonging to any level in the hierarchy). The proportion of individual answers that are "non-descriptive" is around 34%, which included answers to tags such as `cat`, `sticker` and `wall` for the Mission. The fact that these tags indeed describe content that occurs frequently in the Mission doesn't imply that they are perceived as descriptive per se by *all* users.

People's local experiences shape and differentiate their perceptions. The interviews illustrated how tags were interpreted in multiple ways; `church` was taken to refer to a church building, or to the streetcars servicing the J-Church light-rail line, and was not seen as descriptive by the majority of participants (see Castro in Table 5). Local terms surfaced by our model, such as `walls` in the Mission, represented the neighborhood's characteristic murals to a long-time local interviewee (M 49), whereas this term was a meaningless for others. While the majority of survey respondents classified `night` and `coffee` as unspecific for the Mission, the same interviewee (M 49) for example saw `night` as descriptive for its bars, restaurants and clubs and thought `coffee` referred to the copious amounts of coffee shops. Similarly, while one interviewee described North Beach as "party land" (M 46), another claimed "I don't believe there's much of a nightlife there" (F 26). While these results highlight an opportunity for discovery and recommendation of local content that users may not be aware of, it also means that a careful explanation might be necessary.

5.2.2 Model extensions

Beyond misunderstanding tags or finding them not specifically descriptive of a certain neighborhood, the interviews provided additional clues about the mismatches identified in our survey, as well as individual differences. They are organized below in potential opportunities for extensions of the GHM model.

Sub-region detection. According to the majority of our survey participants, 11 tags classified by the GHM as specifically descriptive of North Beach were actually descriptive of another neighborhood. From our interviews, we learned that the terms surfaced by our model for the North Beach neighborhood included references to the bay's waterside and to the tourist attraction Fisherman's Wharf (see for example `wharf`, `seal`, `crab`, `bay`, `pier` in Table 5). The locals

	Mission		Castro		North Beach	
Neighborhood-level tag count	17		17		17	
Neighborhood-level tag classifications	287	(100%)	201	(100%)	185	(100%)
Tags not understood	22	(8%)	3	(1%)	0	(0%)
Tags seen as non-descriptive for any level	116	(40%)	50	(25%)	62	(33%)
(Part of) neighborhood	109	(39%)	67	(33%)	37	(20%)
Higher-level node (CA/USA)	24	(8%)	38	(19%)	7	(4%)
Other neighborhood	16	(5%)	43	(22%)	79	(43%)

Table 4: The distribution of answers given by survey participants. The (rounded) percentages are computed with respect to the total number of answers given.

	Mission	Castro	North Beach
Tag count	32	32	32
Tag classications	561	381	349
GHM alignment	0.84	0.81	0.66
Misaligned tag count	5	6	11
Misaligned tags	`night`, `coffee`, `sidewalk` `cat`, `brannan`	`mission`, `streetcar`, `dolorespark` `church`, `sign`, `night`	`embarcadero`, `bridge`, `water` `alcatraz`, `seal`, `sea`, `crab`, `wharf`, `boat`, `bay`, `pier`

Table 5: Alignment of GHM tag assignments with the majority of survey respondents' assignment for each survey neighborhood. Misalignment would for example be a tag classified as neighborhood level by the GHM while the majority of survey respondents had assigned it to another neighborhood or another level.

that responded to our survey (see Table 4) and also the interviewees clearly made a distinction between Fisherman's Wharf and North Beach, whereas the set of administrative regions we used for our model did not: according to the data from the city planing department of San Francisco, Fisherman's Wharf is not a neighborhood in itself and is simply a sub-region of North Beach. Clarifications of region borders and detecting emerging sub-neighborhoods (e.g. "as you go further down south it's a totally different neighborhood"), would improve the quality of the neighborhood tags presented to the user. From the modeling perspective, the GHM already allows for considering predefined sub-regions in certain neighborhood, since the geo-tree can be unbalanced and have a different depth for certain sub-trees.

Permeable adjacency and topography. Using *permeable adjacency* by considering adjacent neighborhoods appears promising. For example, the mismatching tag `dolorespark` (see Table 5) refers to the park right on the edge of the Mission and Castro, which however was placed outside the Castro neighborhood by the majority of our survey respondents. Interviewees defined neighborhood boundaries differently, and sometimes indicated they didn't know neighborhoods well "It's a little difficult because it's right next to the Presidio. I kind of, maybe confuse them from each other" (F 26). Interviewees extended their reasoning about activities or points-of-interest that could spill over into adjacent neighborhoods. For example, the Golden Gate Bridge, officially part of the Presidio neighborhood, but photographed from a wide range of other neighborhoods was mentioned as a distant-but-characteristic feature: "you can see the Bay Bridge from there" (M 46). Extensions that consider wider topography and fuzzy boundaries could improve the quality of the results.

Representative lower levels. Interviewees at times cited neighborhoods' unique character as representative of wider regional developments, e.g. gentrification of neighborhoods, or "California is known for being... very liberal... and it's almost as if a lot it comes from Castro" (M 35). Hierarchical modeling offers opportunities beyond the identification of locally descriptive content; it could also help find neighborhoods that are representative of a higher level in the geographical hierarchy.

Temporality. Time of day, shifting character of a neighborhoods, events, long-term history, and even change itself were referenced by interviewees: "I think of night, ... there's a lot of activity during night time with the bars and the clubs... but before... you wouldn't be caught there at nighttime... 20 years ago it was a different neighborhood." (M 49). Since the Flickr tag collection we used spanned multiple years, some aspects of neighborhoods, represented in tags surfaced by our model, that were characteristic at one time but no longer as prominent at present (such as the Halloween celebrations in the Castro neighborhood), were considered out of place (M 32). Yet other interviewees still found such tags characteristic, and freely associated: "I think of outrageous costumes, and also the costumes that people see when it's not Halloween" (F 26). Combining hierarchical modeling with features detecting diurnal rhythms [13] or events [25] would provide additional insight in such content changes over time. However, we must be aware that this necessitates a dataset that is much more prolific than the one at hand: we need to have a sufficient number of geotagged samples per time period.

The non-tagged. More recent socio-economic developments were mentioned by the interviewees, but not represented in the dataset of tags. For example, one interviewee argued that, while the terms for SoMa did capture the neighborhood ("I think it's really descriptive and pretty accurate"), he would himself add "something about startups and expensive apartments that really aren't all that great. $3500 a month for one bedroom. That would be a good descriptor." (M 32). Note that even the absence of a feature could be characteristic: "There's not a lot to do around there." (M 35). Distinguishing however between what is absent in a neighborhood, and what is not represented in a dataset is a challenge [24]. Findings proxies for such features, such as proposed by Quercia et al. [22], requires rigor because it might introduce error and biases. Combining different data sources and model features can however provide additional opportunities.

6. DISCUSSION AND CONCLUSIONS

In this work we proposed a probabilistic model that allows for uncovering terms that are *specifically descriptive* of a region within a given geographical hierarchy. By applying our model to a large-scale dataset of 20 million tags associated with approximately 8 million geotagged Flickr photos taken in San Francisco and Manhattan, we were able to associate each node in the hierarchy to the tags that specifically describe it. Moreover, we used these descriptions to quantify the uniqueness of neighborhoods, and find a mapping between similar but geographically distant neighborhoods. We further conducted interviews and a survey with local residents in order to evaluate the quality of the results given by the GHM and its ability to surface neighborhood-characteristic terms, which span both terms known and unknown to locals. The classification accuracy of GHM, measured with respect to the classification of tags made by locals, provides strong support for the validity of our approach. However, the results of the interviews also highlighted the difference between the performance of a model at classifying community-generated data, and its performance as judged subjectively by an individual. This highlights the importance of taking user subjectivity into account, and the need for explanation and framing. As a consequence, the traditional evaluation of modeling approaches that are fed with user-generated data faces the challenge of both the representation and the subjectivity inherent to vernacular geography. It is important to note that the dataset itself was not formally labeled and rather only contained free-form, often noisy user-generated tags. While unstructured user-generated tags can be cleaned, for instance through canonicalization [28], one can utilize other structures inherent in the data to find meaning. Our work takes advantage of the latent geographical patterns exhibited by geotagged data, and shows how we can relate tags to regions within a geographical hierarchy.

Representation. We used a specific type of dataset (tags) in this paper to describe geographical regions. We are aware that the tags generated by other online communities that use different social photo sharing services, such as Instagram, may yield different descriptions of regions. For example, if tagging in such a service is more emotive rather than descriptive of the content, our method would surface different local qualities. User-generated data is only representative of the activity of a community within it, and not necessarily of all people in a certain region. Large-scale social media datasets, whether they consist of check-ins, status updates, or photos uploaded to a community-based service, are the result of communicative acts influenced by service design features and evolving community norms [24, 26, 8]. In our case, not all local features are captured and shared on Flickr. The content that is present can however be used as relative hints at local trends, and provide comparative insights. Consequently, when we here state locally descriptive, we therefore mean descriptive for the content generated in that locale, not necessarily for all human activity.

Vernacular geography. Spatial knowledge that is used to communicate about space and regions has been referred to as vernacular geography [15] or naive geography [11] that carries with it a certain intrinsic vagueness in its nature [10]. Individuals' descriptions of places are inherently subjective, as are interpretations of what is and what is not descriptive for a locale [4]. Towards any community, online or offline, the perception of place is considered to be of a shared frame of reference [19]. Photo tags generated by the Flickr community are no exception. However, our qualitative exploration is aimed at understanding *what factors come into play* when describing space, and how might they manifest into a probabilistic model. Fuzzy regional boundaries, temporal events, hidden landmarks all underscore types of regions and terms we surfaced through the interviews, as well as the kinds of tag descriptors discovered quantitatively. We find a mixed method approach brings a clearer understanding of online communities and naive geography alike.

Future work. As illustrated by our generative interviews, the presentation of our model's results ought to take into account subjective interpretations, support the discovery of new conceptual clusters and their explanation. The interviews also pointed to the possibility of using certain neighborhoods as embodying the spirit of its encompassing region, e.g., the Castro described as an influential San Francisco neighborhood, or technology companies moving into SoMa reflecting changes in the city as a whole. The relationship between the leaves of a geographical tree, temporal change and events, sub-regions, and fluid local boundaries that consider wider topographical features represent potential viable extensions to our model, conditioned on the fact that the dataset at hand includes enough data and spans a long period of time. For example, we could extend our model to account for the porosity of frontiers between neighborhoods and the fact that there is no consensus about the exact boundaries of a neighborhood: we define, for each region, a set that is composed of the regions that are adjacent to it. The probability of observing tag in a certain region then includes a new term that accounts for the possibility of sampling tags from adjacent regions. The learning procedure for this extension would be very similar to the GHM's learning procedure.

7. REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. Yang. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In *Proceedings of Digital Libraries*, pages 1–10. ACM Press, 2007.

[2] A. Ahmed, L. Hong, and A. Smola. Hierarchical geographical modeling of user locations from social media posts. In *Proceedings of WWW'13*, 2013.

[3] L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak. Spatial variation in search engine queries. In *Proceedings of WWW'08*, pages 357–366. ACM, 2008.

[4] F. Bentley, H. Cramer, S. Basapur, and W. Hamilton. Drawing the city. differing perceptions of the urban

environment. In *Proceedings of CHI 2012*. ACM Press, 2012.

[5] C. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, New York, 2006.

[6] D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. In *NIPS*, 2003.

[7] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

[8] M. Cameron, M. Naaman, d. boyd, and M. Davis. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of HYPERTEXT'06*, pages 31–40. ACM Press, 2006.

[9] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh. The livehoods project: utilizing social media to understand the dynamics of a city. In *Proceedings of ICWSM 2012*. AAAI, 2012.

[10] C. Davies, I. Holt, J. Green, J. Harding, and L. Diamond. User needs and implications for modelling vague named places. *Spatial Cognition & Computation*, 9(3):174–194, 2009.

[11] M. Egenhofer and D. Mark. *Naive geography.* Springer, 1995.

[12] S. Elwood, M. Goodchild, and D. Sui. Prospects for VGI research and the emerging fourth paradigm. In *Crowdsourcing Geographic Knowledge*, pages 361–375. Springer, 2013.

[13] N. Grinberg, M. Naaman, B. Shaw, and G. Lotan. Extracting diurnal patterns of real world activity from social media. In *ICWSM*, 2013.

[14] B. Hillier and L. Vaughan. The city as one thing. *Progress in Planning*, 67(3), 2007.

[15] L. Hollenstein and R. Purves. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, (1):21–48, 2010.

[16] D. D. Lewis. Naive (Bayes) at forty: The independence assumption in information retrieval. In *Proceedings of the 10th European Conference on Machine Learning*, ECML '98, pages 4–15, London, UK, UK, 1998. Springer-Verlag.

[17] K. Lynch. *The Image of the City.* MIT Press, 1960.

[18] A. McCallum, R. Rosenfeld, T. M. Mitchell, and A. Y. Ng. Improving text classification by shrinkage in a hierarchy of classes. In *Proceedings of the Fifteenth International Conference on Machine Learning*, ICML '98, pages 359–367, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[19] D. Montello, M. Goodchild, J. Gottsegen, and P. Fohl. Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation*, 3(2-3):185–204, 2003.

[20] M. Naaman and R. Nair. Zonetag's collaborative tag suggestions: What is this person doing in my phone? *IEEE MultiMedia*, 15(3), 2008.

[21] Y. Pang, Q. Hao, Y. Yuan, T. Hu, R. Cai, and L. Zhang. Summarizing tourist destinations by mining user-generated travelogues and photos. *Comput. Vis. Image Underst.*, 115(3):352–363, March 2011.

[22] D. Quercia, H. Cramer, and N. O'Hare. Aesthetic capital: What makes London look beautiful, quiet, and happy. In *Proceedings of CSCW'14*, 2014.

[23] T. Rattenbury and M. Naaman. Methods for extracting place semantics from flickr tags. *ACM Transactions on the Web*, 3(1), 2009.

[24] M. Rost, L. Barkhuus, H. Cramer, and B. Brown. Representation and communication: Challenges in interpreting large social media datasets. In *Proceedings of CSCW'13*. ACM Press, 2011.

[25] D. A. Shamma, L. Kennedy, and E Churchill. Peaks and persistence: Modeling the shape of microblog conversations. In *Proceedings of CSCW'11*, 2011.

[26] D. A. Shamma, R. Shaw, P. Shafton, and Y. Liu. Watch what I watch: using community activity to understand content. In *Proceedings of MIR'07*, pages 275–284. ACM, 2007.

[27] R. Stedman. Toward a social psychology of place: Predicting behavior from place-based cognitions, attitude, and identity. *Environment & Behaviors*, 34(5), 2002.

[28] B. Thomee and G. De Francisci Morales. Automatic discovery of global and local equivalence relationships in labeled geo-spatial data. In *Proceedings of the ACM Conference on Hypertext and Social Media*, pages 158–168, 2014.

[29] B. Thomee and A. Rae. Uncovering locally characterizing regions within geotagged data. In *Proceedings of WWW'13*, pages 1285–1296, 2013.