Probabilistic Modeling of City-scale Image Collections

Diego A. Ballesteros Villamizar

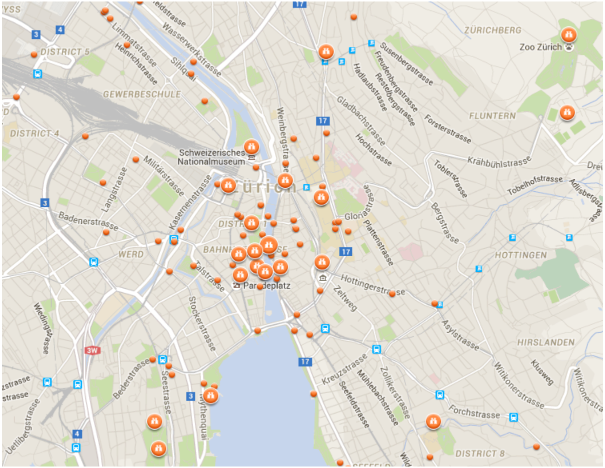
ETH Zürich

April 26th, 2016

Outline

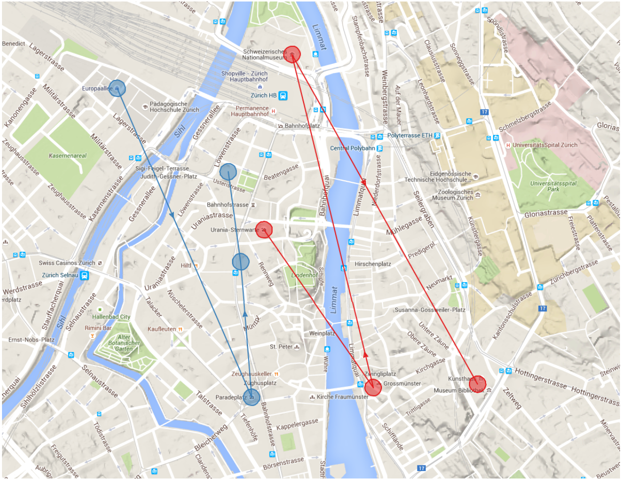
Touristic Route Planning

Let's plan a 2-day trip to Zürich. How can we go from this?



Touristic Route Planning

To this:



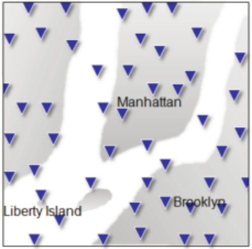
Related Work: Mapping the World's Photos

- "Photo shots are a prime element of the image perceived by visitors and also a path to understanding processes in the symbolic construction of destinations." Donaire et al. (2014)
- Kleinberg et al. (2009) used mean-shift clustering to identify highly photographed locations.

Representative photos for top landmarks in North America identified from Flickr data. Source: Kleinberg et al. (2009)

Related Work: Travel Route Recommendation Using Geotags in Photo Sharing Sites

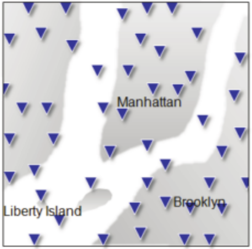
- Kurashima et al. (2010) modeled photographer behavior by combining Markov and topic models in a probabilistic framework.



Highly photographed locations. Source: Kurashima et al. (2010)

Related Work: Travel Route Recommendation Using Geotags in Photo Sharing Sites

- Kurashima et al. (2010) modeled photographer behavior by combining Markov and topic models in a probabilistic framework.



Highly photographed locations. Source: Kurashima et al. (2010)

$$P(l_t \mid l_{t-1}, h^u)$$

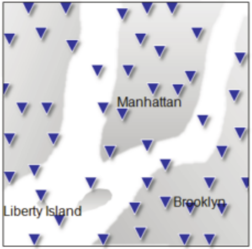
$$P(l_t \mid l_{t-1})$$

$$\sum_{z \in \boldsymbol{Z}} P(z \mid h^u) P(l_t \mid z)$$

Markov and Topic models.

Related Work: Travel Route Recommendation Using Geotags in Photo Sharing Sites

- Kurashima et al. (2010) modeled photographer behavior by combining Markov and topic models in a probabilistic framework.



Highly photographed locations. Source: Kurashima et al. (2010)

$$P(l_t \mid l_{t-1}, h^u)$$

$$P(l_t \mid l_{t-1})$$

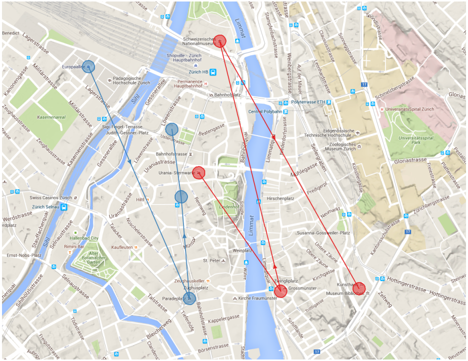
$$\sum_{z \in \boldsymbol{Z}} P(z \mid h^u) P(l_t \mid z)$$

Markov and Topic models.

Routes from the model. Source: Kurashima et al. (2010)

Sets of Locations to Visit

- Let's simplify the problem and focus on which locations to visit.
- The when and how can be determined later.
- Instead of:

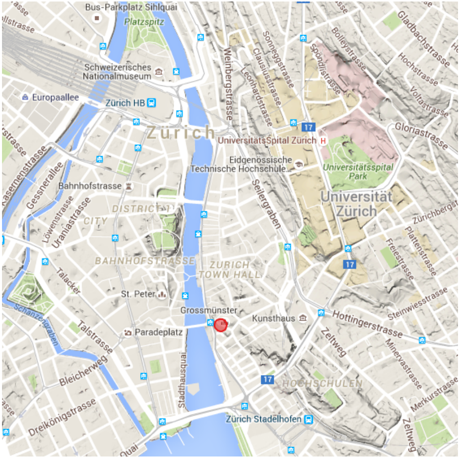


Sets of Locations to Visit

- Let's simplify the problem and focus on which locations to visit.
- The when and how can be determined later.
- We aim to identify:

Good Sets of Locations

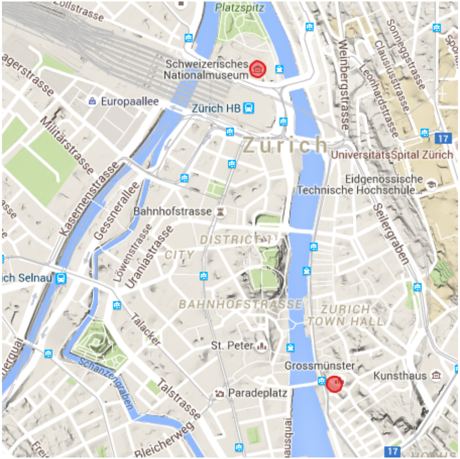
Let's try to build one, first a highly popular location: Grossmünster



Grossmünster. Source: https://commons.wikimedia.org/w/index.php?curid=5082760

Diverse Sets of Locations

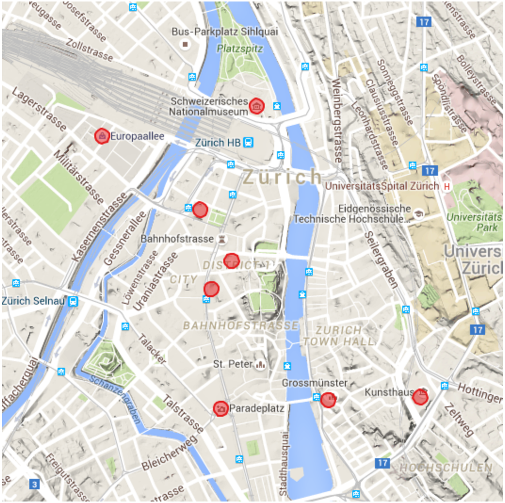
The next on the list of top attractions is Fraumünster, but two churches on the same day? Let's add a museum instead.



Swiss National Museum. Source: `https://commons.wikimedia.org/w/index.php?curid=7974880`

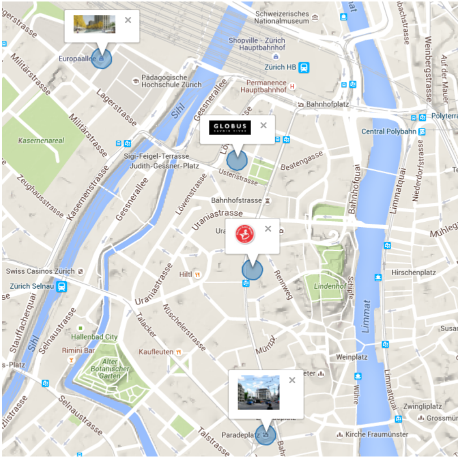
Diverse Sets of Locations

And so on...



Complementary Sets of Locations

- Only diversity? What about tourists interested in a coherent theme?
- Perhaps a trip through stores in Zürich.



Choosing a Model

- Modeling sets of items $\rightarrow$ Functions over sets
- Diversity $\rightarrow$ Submodularity
- Complementarity $\rightarrow$ Supermodularity
- Probabilistic modeling of tourist behavior $\rightarrow$ Probability of visiting a set of locations.

Diversity and Submodularity

- Suppose a function $F(S)$ that sums the number of unique location types in S.
- For example, $F(\{\text{ETH, UZH}\}) = 1$, $F(\{\text{ETH, UZH, Fraumünster}\}) = 2$.

Diversity and Submodularity

- Suppose a function $F(S)$ that sums the number of unique location types in S.
- For example, $F(\{\text{ETH, UZH}\}) = 1$, $F(\{\text{ETH, UZH, Fraumünster}\}) = 2$.
- This is a submodular function.

Diversity and Submodularity

- Suppose a function $F(S)$ that sums the number of unique location types in S.
- For example, $F(\{\text{ETH, UZH}\}) = 1$, $F(\{\text{ETH, UZH, Fraumünster}\}) = 2$.
- This is a submodular function.
- Maximizing $F(S)$ for S of fixed size implies adding locations of different types $\rightarrow$ Diversity.

- Suppose a function $F(S)$ that counts 1 for each pair of repeated location types in S.
- For example, $F(\{\text{ETH, UZH, Fraumünster}\}) = 1$, $F(\{\text{ETH, Fraumünster}\}) = 0$.

Complementarity and Supermodularity

- Suppose a function $F(S)$ that counts 1 for each pair of repeated location types in S.
- For example, $F(\{$ETH, UZH, Fraumünster$\}) = 1$, $F(\{$ETH, Fraumünster$\}) = 0$.
- This is a supermodular function.

Complementarity and Supermodularity

- Suppose a function $F(S)$ that counts 1 for each pair of repeated location types in S.
- For example, $F(\{$ETH, UZH, Fraumünster$\}) = 1$, $F(\{$ETH, Fraumünster$\}) = 0$.
- This is a supermodular function.
- Maximizing $F(S)$ for S of fixed size implies adding locations of the same types $\rightarrow$ Complementarity.

Probabilistic Submodular Models

- Probabilistic Submodular Models (PSMs) are a class of distributions over the powerset of a set V of the form

$$P(S) = \frac{\exp F(S)}{Z}$$

  for all $S \subset V$, where F(S) is submodular or supermodular. These distributions are log-submodular and log-supermodular, respectively.
- Inference in these distributions has been recently studied by Djolonga and Krause (2015) and Gotovos et al. (2015).

Learning PSMs

- We are interested in applying these distributions to a dataset of geotagged photographs. Tschiatschek et al. (2016) studied learning a PSM from data, i.e. estimating the model parameters.
- Maximum Likelihood Estimation is not tractable, because the exact computation of the partition function Z is #P-complete.
- Noise Contrastive Estimation (NCE) was used as an alternative for learning the PSM.
- We'll take a closer look at the model proposed by Tschiatschek et al. (2016).

Noise Contrastive Estimation

- Transform the task of estimating a distribution into a supervised learning task, namely a classification task.
- Noise samples $\mathcal{N}$ are drawn from a known normalized distribution P_n and contrasted with the data samples $\mathcal{D}$ assumed to come from a distribution P_d.
- Partition function Z becomes one of the parameters to estimate.
- Stochastic Gradient Descent can be used to optimize the classification objective.

Outline

Submodular Model for Diversity

- A submodular function for diversity,

$$F(S) = u(S) + D(S)$$

  where $u(S)$ aggregates the utility of the items in S and $D(S)$ quantifies the diversity of the items in S.

- The corresponding PSM is:

$$P(S) = \frac{1}{Z} \exp\left(u(S) + D(S)\right)$$

- This is the general form of the Facility Location Diversity (FLID) model proposed by Tschiatschek et al. (2016).

Aggregating Utility

- A (sub)modular function

$$u(S) = \sum_{i \in S} u_i$$

where u_i represents the utility of each item i in the ground set V.

- For example, the utility of a location could be the number of users that have taken photos of it.

Quantifying Diversity (Simplified)

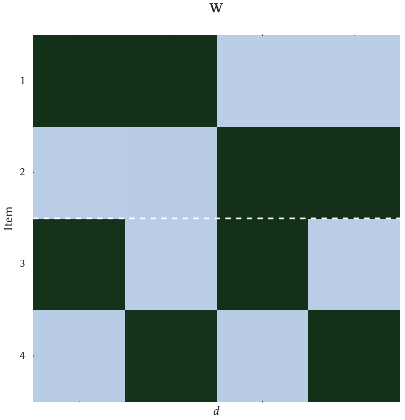
- A submodular function

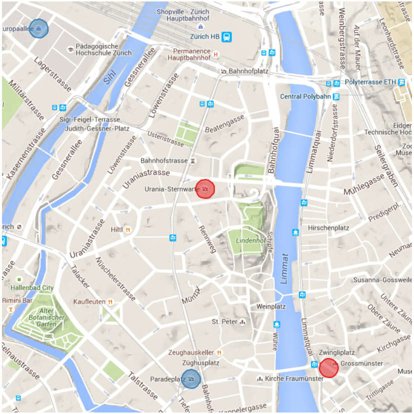
$$D_d(S) = \max_{i \in S} w_{i,d} - \sum_{i \in S} w_{i,d}$$

  where w_i quantifies the contribution of item i to some concept d related to the diversity of the set.
- For example, for the concept "is a museum" locations may have binary weights $w_d \in \{0, 1\}$. Then a set with many museums is penalized with a negative value of $D_d(S)$

Quantifying Diversity

- FLID uses L dimensions capturing some diversity concepts. Each item can be represented with a vector $\mathbf{w}_i \in \mathbb{R}^L_{\geq 0}$.





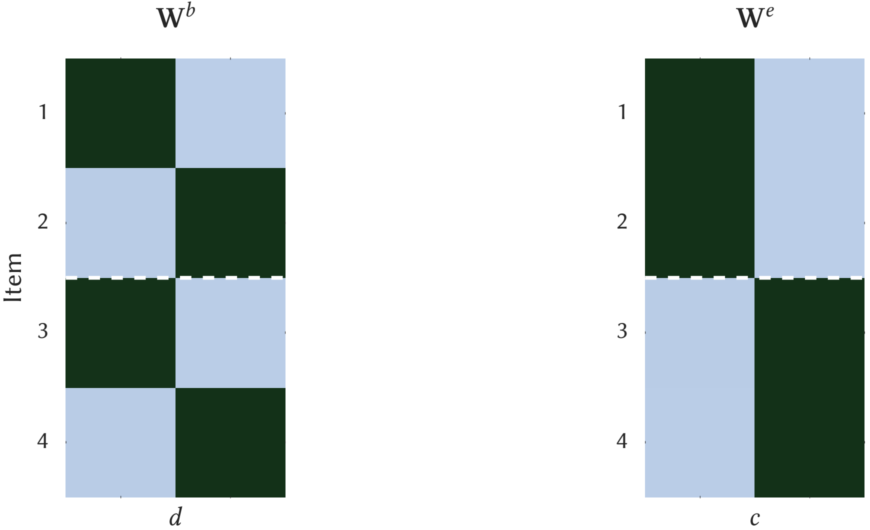
FLID

$$P(S) = \frac{1}{Z} \exp \left(\sum_{i \in s} u_i + \sum_{d=1}^{L} \max_{i \in S} w_{i,d} - \sum_{i \in S} w_{i,d} \right)$$

- Learning can be performed efficiently, in $\mathcal{O}(|\mathcal{D} \cup \mathcal{N}| L\kappa)$ where $\kappa = \operatorname{argmax}_{S \in \mathcal{D} \cup \mathcal{N}} |S|$.
- For small L, exact computation of Z is feasible.
- Applications shown by Tschiatschek et al. (2016):
  - Product recommendation with Amazon baby registries.
  - Summarization of an image collection.

Outline

Quantifiying Complementarity

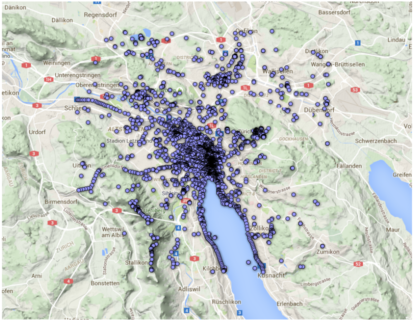


Facility Location Diversity and Coherence

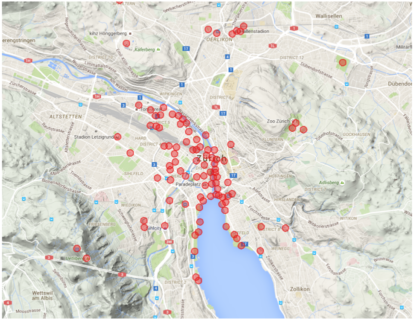
- We propose adding a similar supermodular term to FLID to model complementarity.

$$P(S) = \frac{1}{Z} \exp\left(u(S) + D(S) + \sum_{c=1}^{K} \sum_{i \in S} w^e_{i,c} - \max_{i \in S} w^e_{i,c} \right)$$

- Complexity of learning is only increased by a factor of K.



Geotagged photos from Flickr.



Top photographed locations.

Experimental Setup: Completing Sets

- Geotagged photos were grouped by user and day they were taken.
- Each group represents a set. Test sets are constructed by taking out one element at a time.
- For example, if the original set is *{Fraumünster, Grossmünster, Hauptbahnhof}*, the test sets are:
  - *{Fraumünster, Hauptbahnhof}*
  - *{Grossmünster, Hauptbahnhof}*
  - *{Fraumünster, Grossmünster}*

Experimental Setup: Baselines

- The test set is $S = \{l_1, \ldots, l_n\} \setminus \{l_t\}$.
- Modular

$$\operatorname*{argmax}_{i \notin S} u_i$$

- Proximity

$$\operatorname*{argmin}_{i \notin S} \sum_{j \in S} d(i, j)$$

- FLID

$$\operatorname*{argmax}_{i \notin S} P_{FLID}(S)$$

Experimental Setup: Ordered Baselines

- The ordered test set is $S_n = [l_1, \ldots, l_{t-1}, l_{t+1}, \ldots, l_n]$.
- Markov model

$$\operatorname*{argmax}_{i \in V} P(i \mid l_{t-1}) P(l_{t+1} \mid i)$$

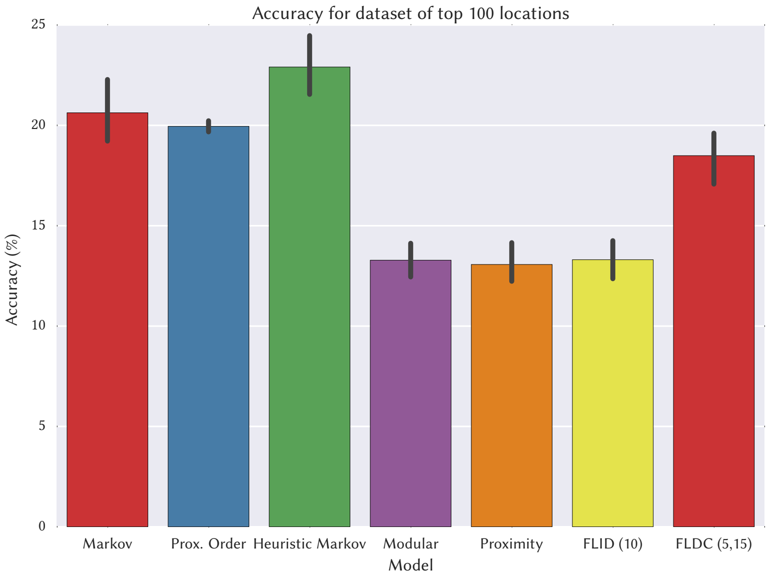
- Heuristic Markov model

$$\operatorname*{argmax}_{i \notin S_n} P(i \mid l_{t-1}) P(l_{t+1} \mid i)$$

- Proximity ordered model

$$\operatorname*{argmin}_{i \notin S_n} d(i, l_{t-1}) + d(i, l_{t+1}) \tag{1}$$

Results - Accuracy



Outline

How to Generalize?

- Is it possible to transfer knowledge about Zürich to Geneva?
- If new locations are identified, how can they be included in the model without learning it again?
- Our proposed solution: **Featurized representations**.

Featurized Facility Location Diversity and Coherence

- Define a feature matrix for all items in V, i.e. $\mathbf{X} \in \mathbb{R}^{|V| \times M}$.
- Factorize the model parameters $\mathbf{u}, \mathbf{W}^b, \mathbf{W}^e$

$$\mathbf{u} = \mathbf{Xa}$$
$$\mathbf{W^b} = \mathbf{XB}$$
$$\mathbf{W^e} = \mathbf{XE}$$

- Model is then defined in the space of features, not items.

Experimental Setup: Locations

A smaller dataset with 10 locations.

- Hauptbahnhof
- Fraumünster
- Grossmünster
- Hallenstadion
- Prime tower
- Bürkliplatz
- Paradeplatz
- Bellevueplatz
- Rathaus
- Zoo

Experimental Setup: Features

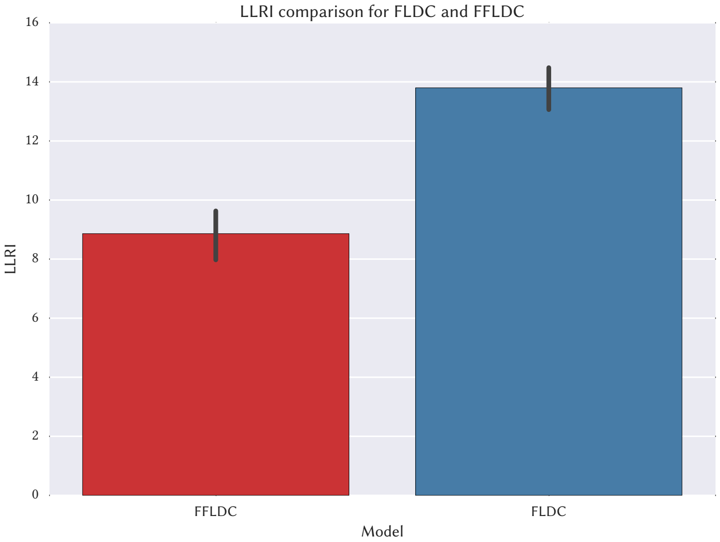
- Is it a transit station? E.g. Hauptbahnhof, Bürkliplatz.
- Is it a church? E.g. Fraumünster.
- Is it a historic building? E.g. Grossmünster, Rathaus.
- Is it an indoors location? E.g. Prime tower, Hallenstadion.
- Normalized number of photographs n_p, $\sqrt{n_p}$ and $\sqrt[4]{n_p}$.
- Number of users per photograph.

Experimental Setup: Evaluation

$$\mathrm{LLRI} = 100 \frac{\mathcal{L}_{model} - \mathcal{L}_{log-modular}}{|\mathcal{L}_{log-modular}|}$$

- It measures the model fit on test data.
- For the small dataset, the accuracy is already at 40% with the modular model and does not improve.

Results - LLRI



Outline

Conclusions

- We have proposed an extension of FLID to model complementarity or attractiveness in sets.
- Experiments in a real world application show how this can improve the quality of the model.
- Recommendation of tourist locations can be modeled using PSMs with positive results. However, modeling of ordered sets is an important direction to explore.
- We have proposed an extension of FLID to generalize the model to unseen items through features.

References I

Djolonga, J. and Krause, A. (2015). Scalable variational inference in log-supermodular models. In *International Conference on Machine Learning (ICML)*.

Donaire, J. A., Camprubí, R., and Galí, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Management Perspectives*, 11:26–33.

Gotovos, A., Hassani, S. H., and Krause, A. (2015). Sampling from probabilistic submodular models. In *Neural Information Processing Systems (NIPS)*.

Kleinberg, J., Huttenlocher, D., Backstrom, L., and Crandall, D. (2009). Mapping the World's Photos. In *Proceedings of the 18th International Conference on World Wide Web*, pages 761–770.

Kurashima, T., Iwata, T., Irie, G., and Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 579–588.

References II

Tschiatschek, S., Djolonga, J., and Krause, A. (2016). Learning probabilistic submodular diversity models via noise contrastive estimation. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*.