Studying urban and cultural dynamics using social media

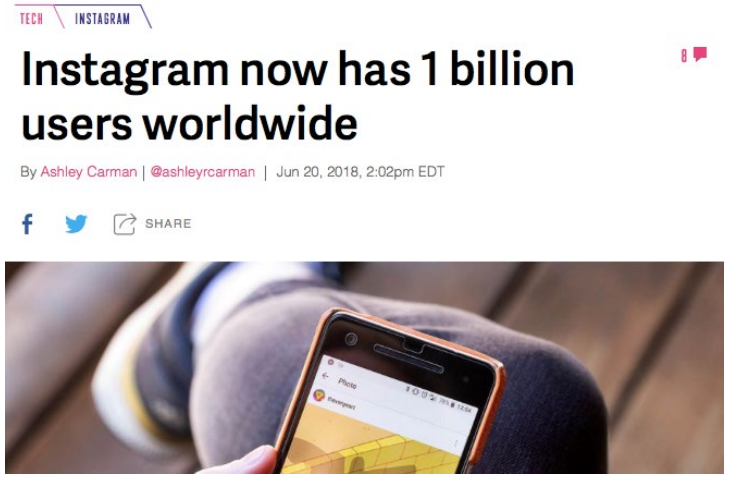
Ksenia Mukhina

Senior Research Fellow at CUDAN Open Lab
School of Digital Technologies, Tallinn University

Instagram Data

Metadata:

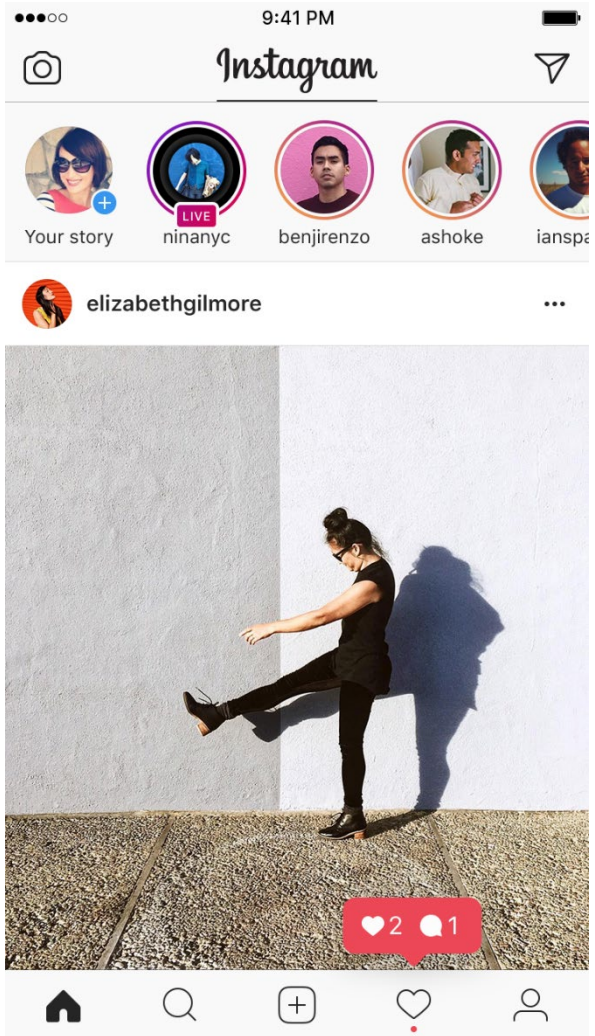
- Photo / video
- Likes & comments
- Caption
- Timestamp
- Location: name, coordinates

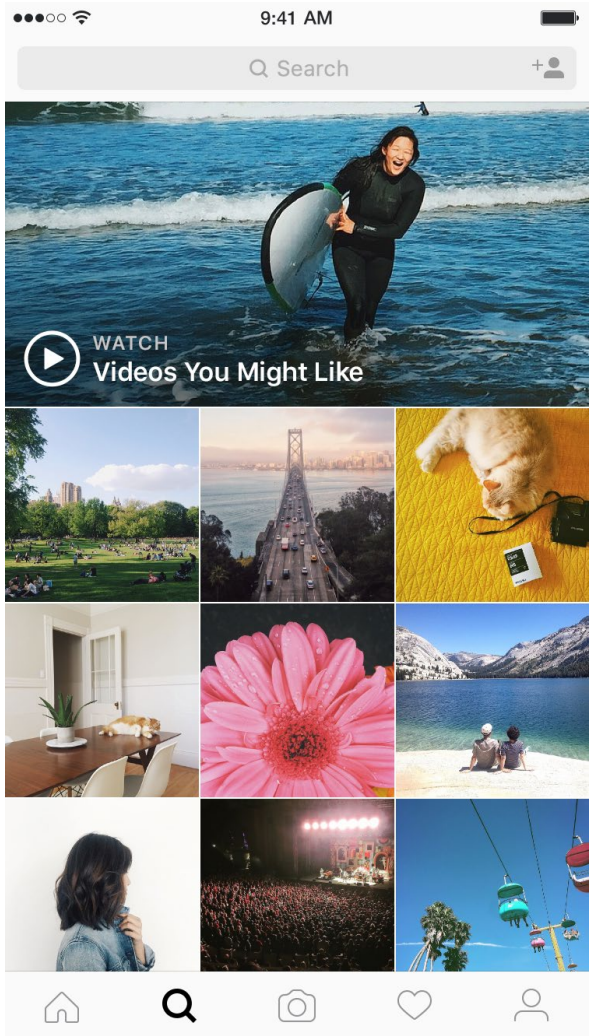

TECH | INSTAGRAM

Instagram now has 1 billion users worldwide

By Ashley Carman | @ashleyrcarman | Jun 20, 2018, 2:02pm EDT

SHARE

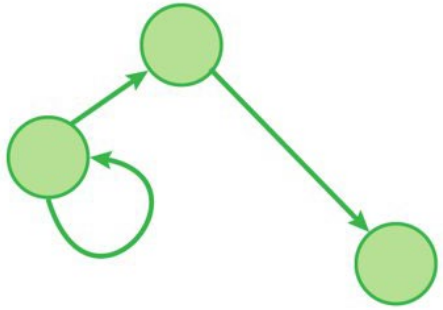




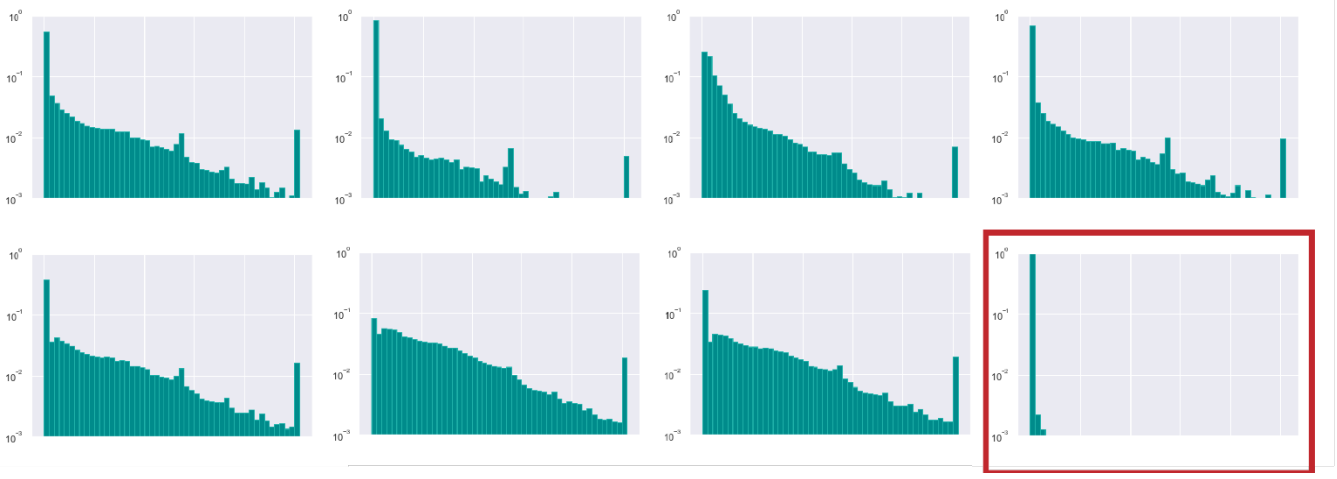
User profile analysis

Business profile

User profile

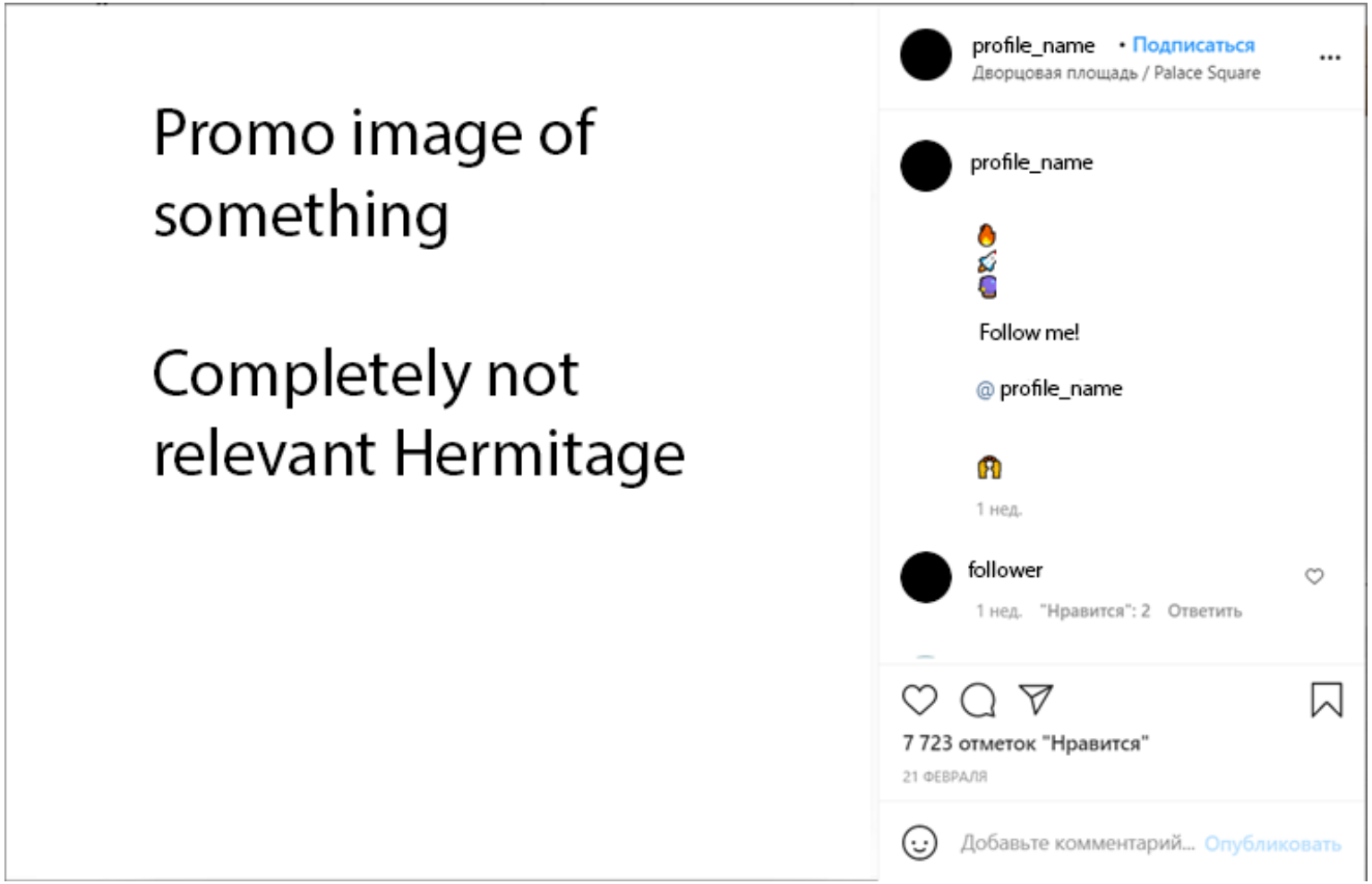


Shape of user profiles' clusters



Example of business profile

Spam detection



Train set:
1065 – usual posts
529 – spam, ads, etc.

Test set:
463 – normal posts
84 – spam

Metrics

Accuracy is fraction of correct predictions.

$$A = \frac{TP + TN}{TP + FP + TN + FN}$$

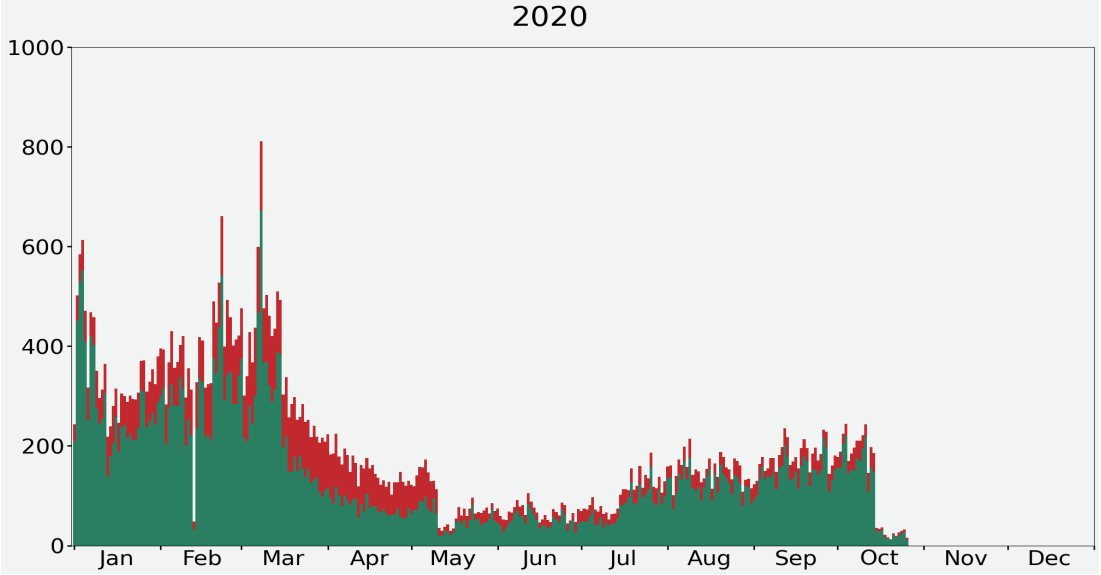
Precision is the fraction of relevant instances among the retrieved instances.

$$P = \frac{TP}{TP + FP}$$

Recall is the fraction of relevant instances that were retrieved.

$$R = \frac{TP}{TP + FN}$$

$$F_1 = 2\frac{P \cdot R}{P + R}$$



Confusion matrix

Actual/Prediction	True (Normal)	False (Spam)
True (Normal)	424	39
False (Spam)	17	67

Selected features & results

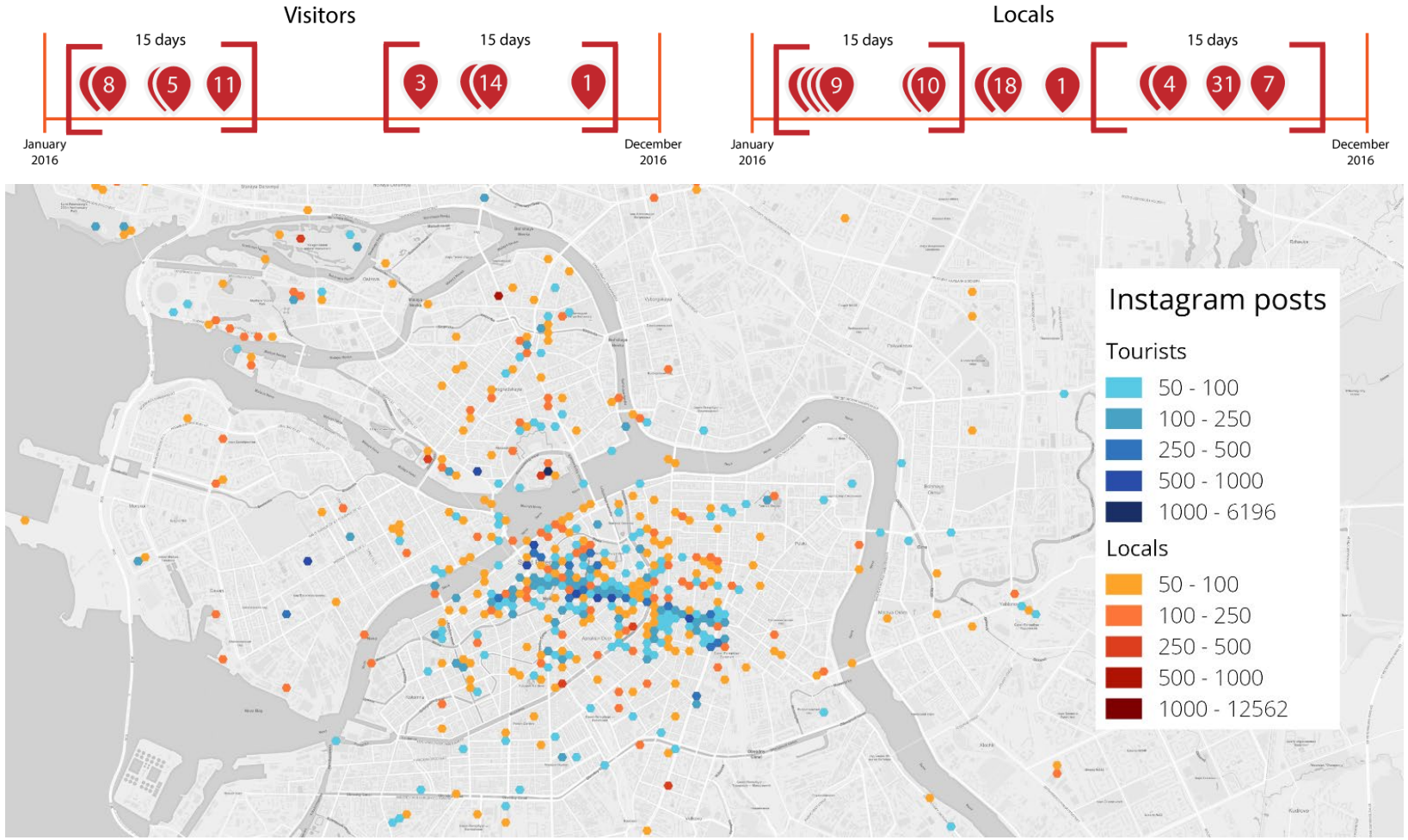
- Likes;
- Comments;
- Numbers in text;
- Number of words;
- Number of URLs;
- Number of email addresses;
- Number of mentions;
- Number of unique accounts mentioned;
- Number of emoji;
- Number of unique emojis;
- Number of places connected to place;
- Similarity of tags;
- Number of spam words;
- Number of lines.

	DecisionTree	KNN	SVC	RandomForest	CatBoost	Ensemble
Precision	0.4954	0.4809	0.4947	0.6436	0.3195	0.6121
Recall	0.6750	0.7875	0.5875	0.8125	0.9625	0.8875
Accuracy	0.8519	0.8446	0.8519	0.9068	0.6947	0.9013
F-score	0.5714	0.5971	0.5371	0.7182	0.4797	**0.7245**

Tourist detection

The time-window method is used for identification of tourists and locals.

Person is identified as a tourist if all his or her posts by two 15-days long time windows with a gap between the windows at least 30 days.

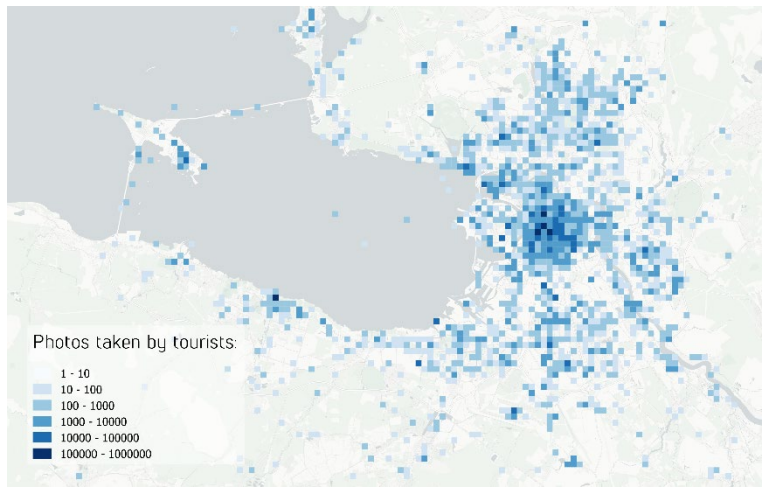


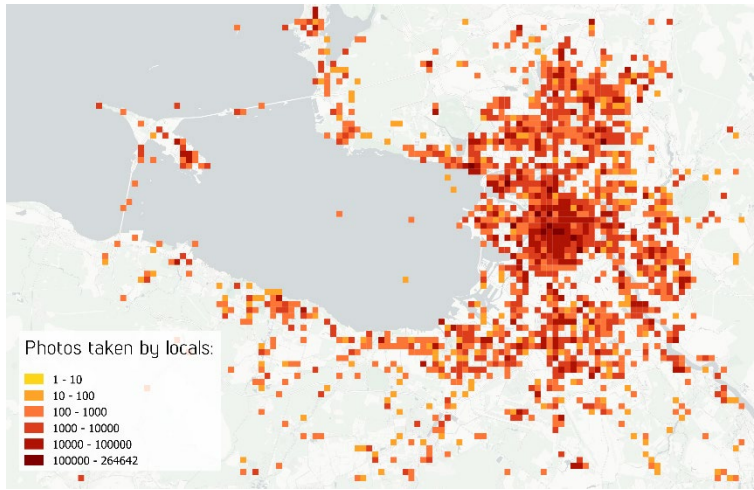
Tourist detection

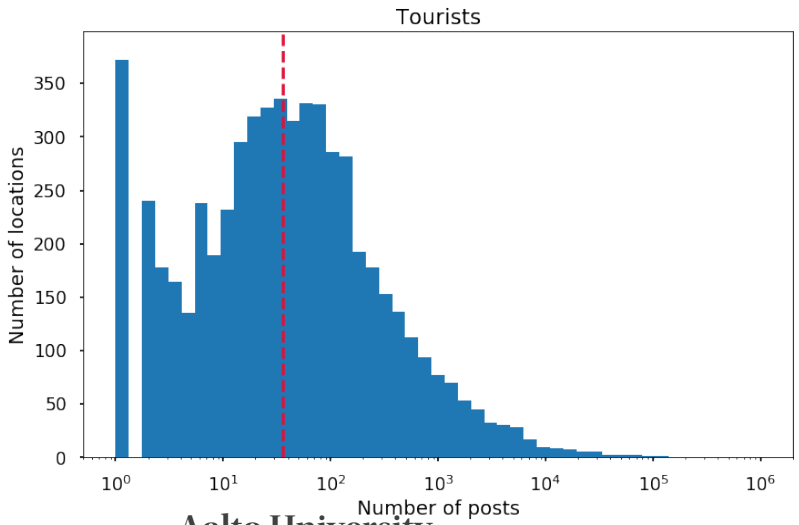
Findings:

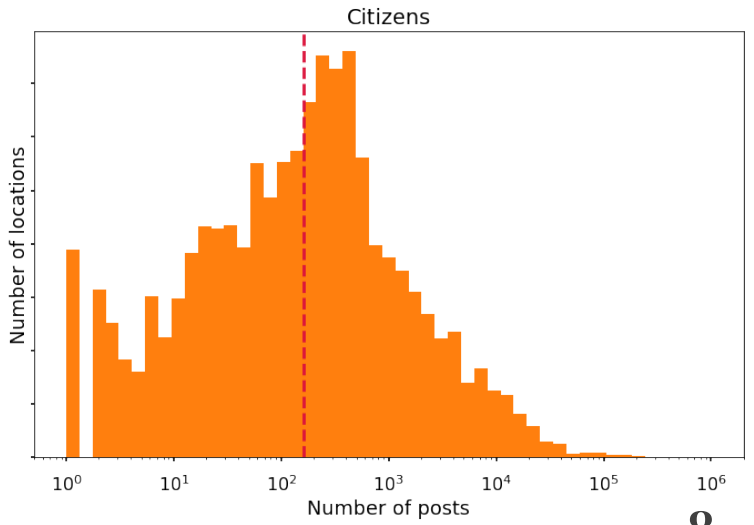
- Larger part of tourists' posts is located within the city center;
- Only 4,322 locations out of 6,385 contain 10 or more posts;
- Half of the locations were mentioned in less than 50 posts (median is 36);
- Number of PoIs mentioned once is about two times more for tourists (372 places compared to 195 for citizens).

Tourists do not have sufficient knowledge about city.



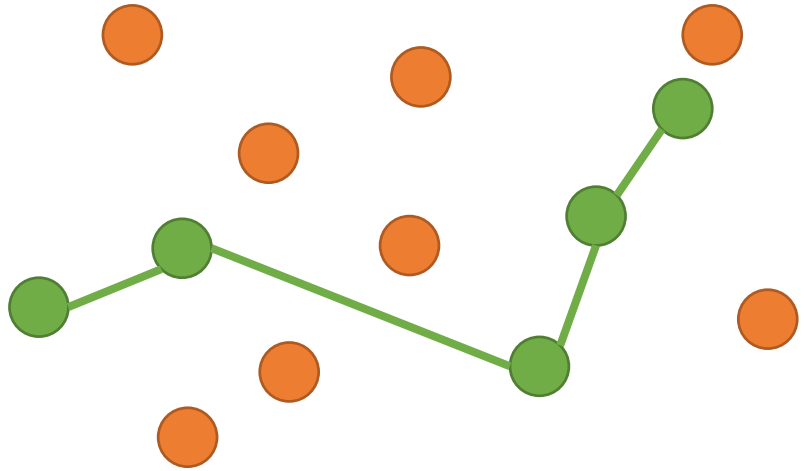


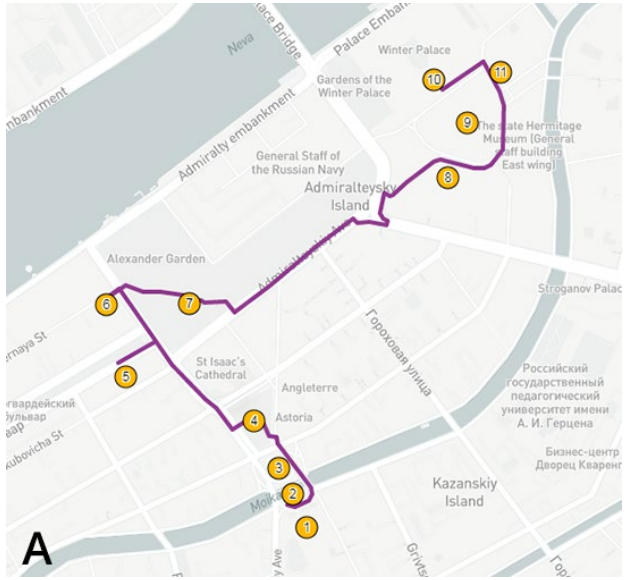


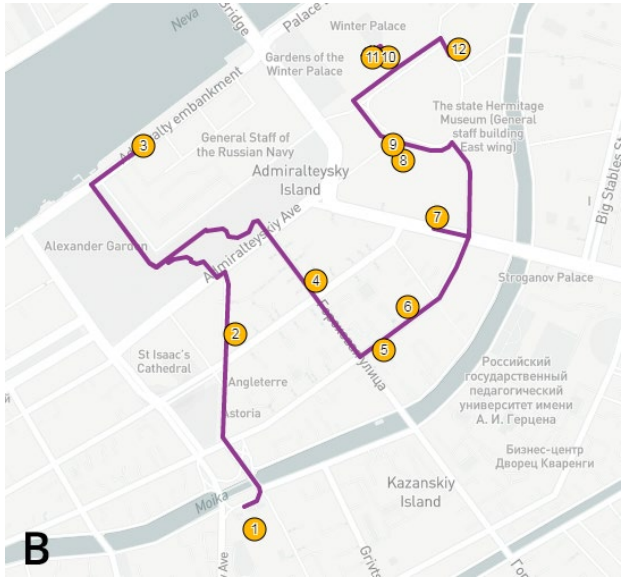


Tourist walking path

The aim is to find the most profitable way to reach terminate node from the start node whose total cost does not exceed a budget.





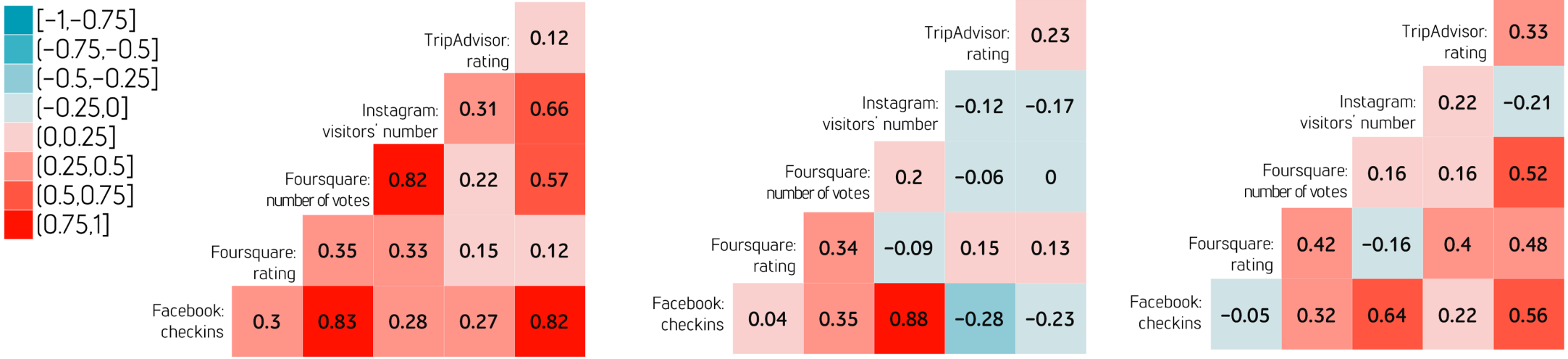


Data

Popularity does not always lead to enjoyment and satisfaction:

- there is no correlation between the amount of people who visited a place and their evaluation of it even for the same sources;
- correlation between ratings on different sources is even smaller. This happens because some services are used by tourists and another are used mostly by local residents.

To construct high quality tourist itineraries the combination of different sources is required.

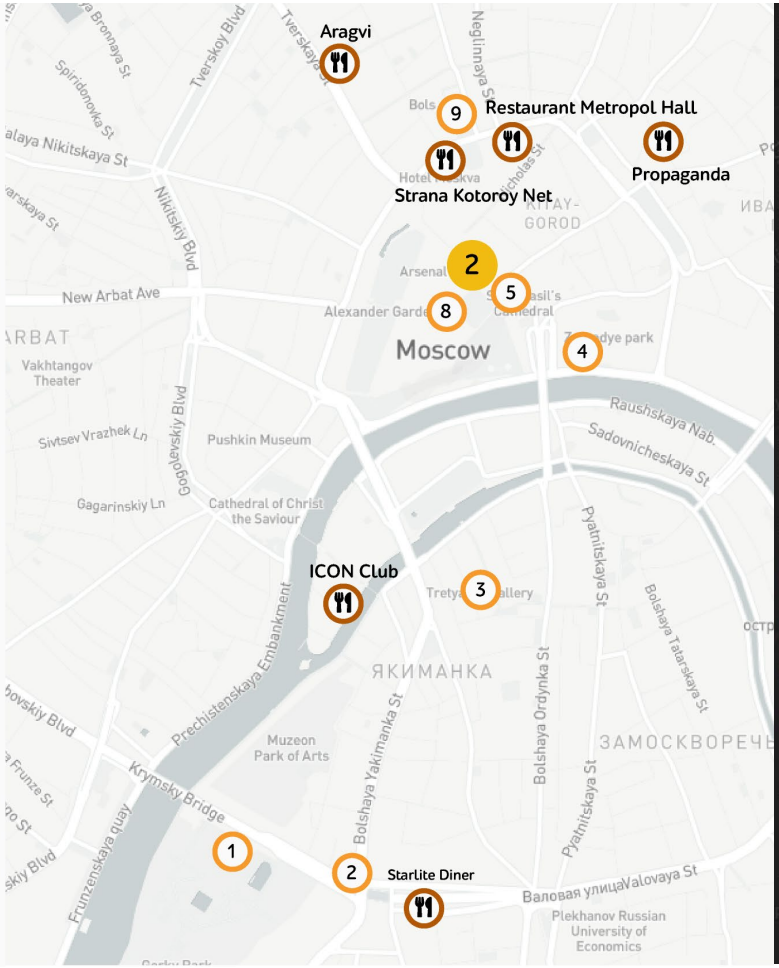


Tourist walking path

Legend:

- - USSR

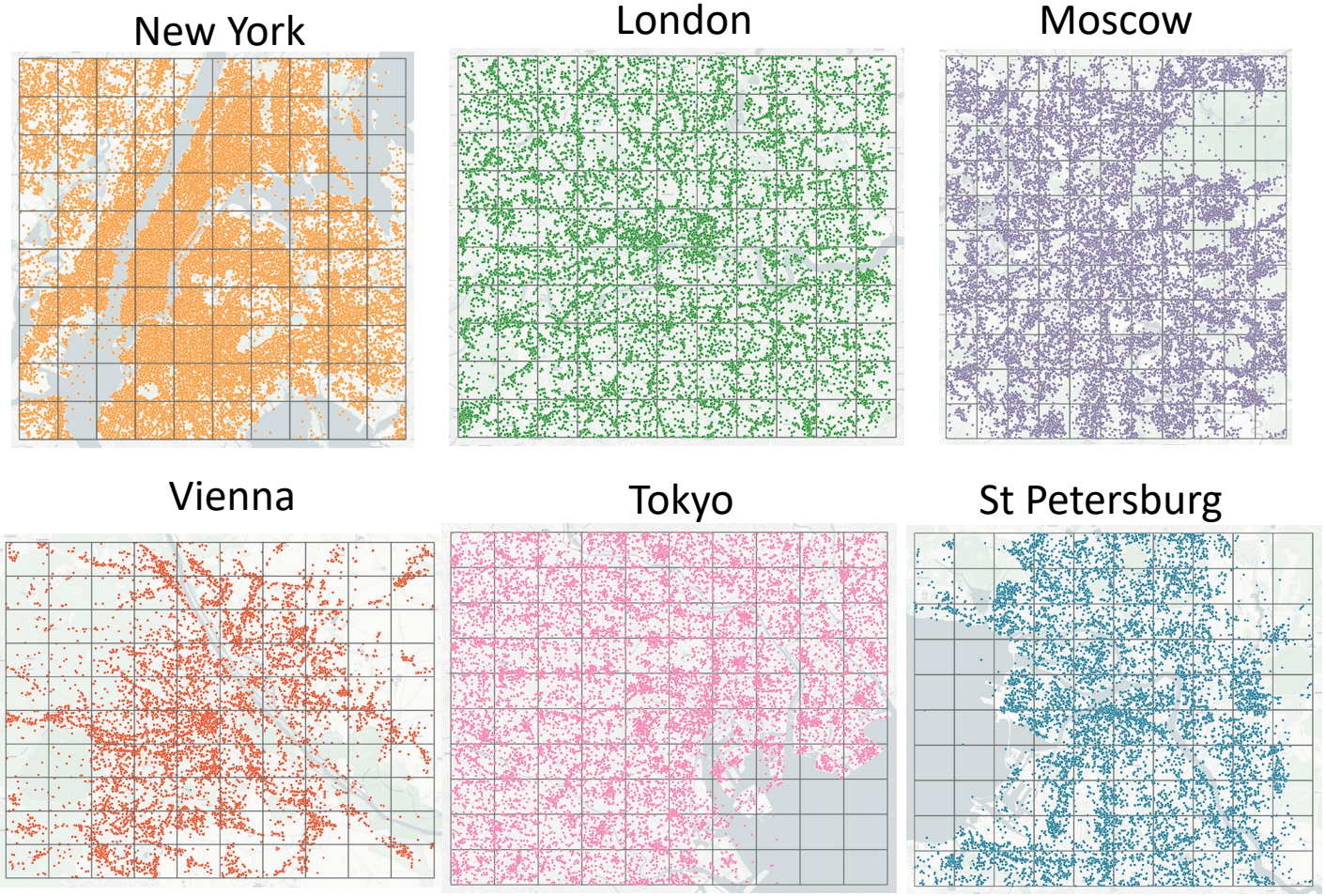
1. Gorkiy Central Park of Culture and Recreation
2. Muzeon Art Park
3. The State Tretyakov Gallery
4. Park Zaryadye
5. St. Basil's Cathedral
6. Red Square
7. Lenin's Mausoleum
8. Moscow Kremlin
9. Bolshoi Theatre



Spatiotemporal analysis

City coverage by LBSN depends on Internet access, popularity of particular service, landscape, etc.

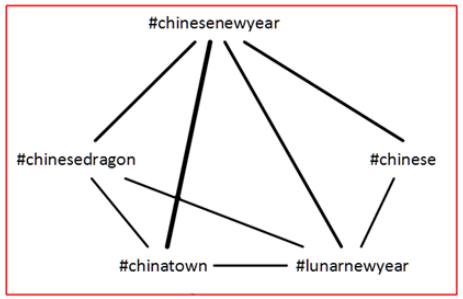
- Social networks stimulate people to react **immediately** as they observe something
- Modern decision support systems use data from social networks

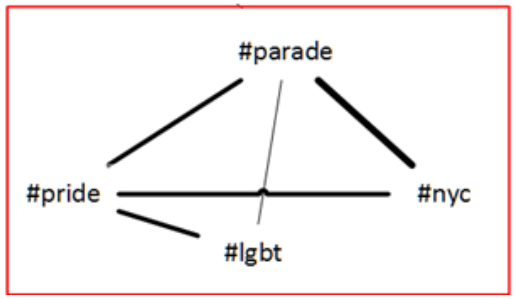


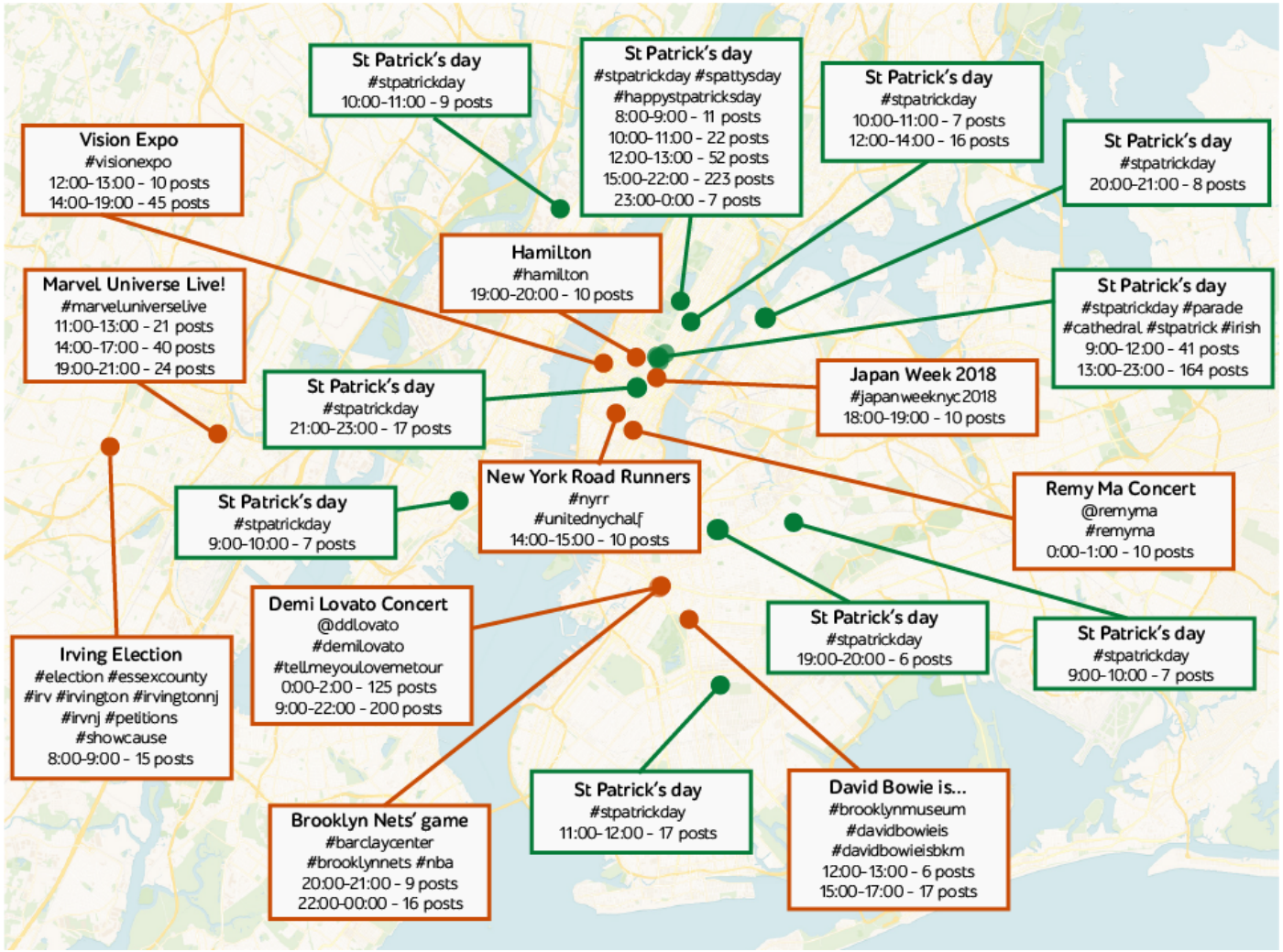
Event detection

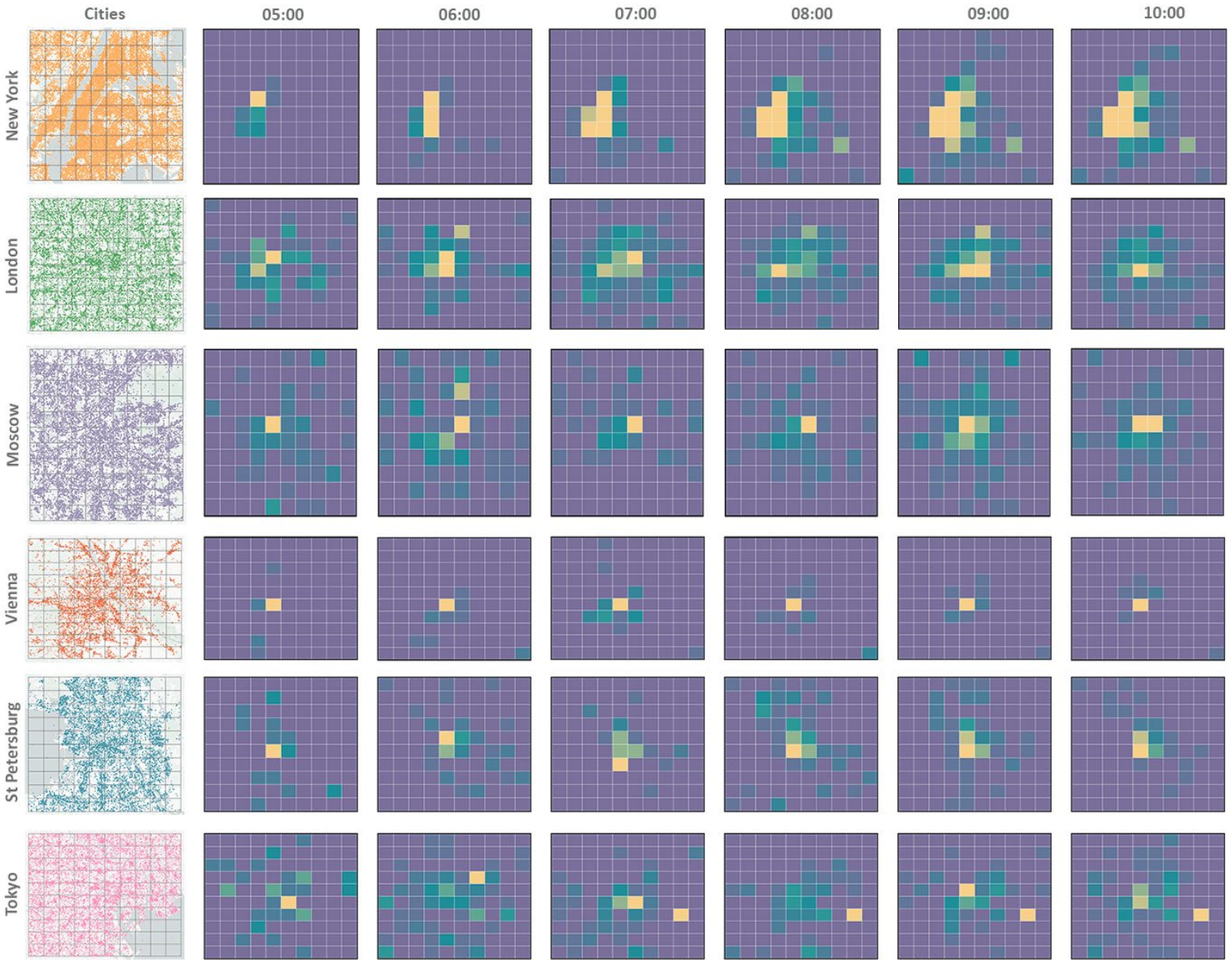
Stages:

1. load data into historical grids and search for cells that significantly differ from their standard states;
2. analyze posts in the candidate cells through co-occurrence graphs.





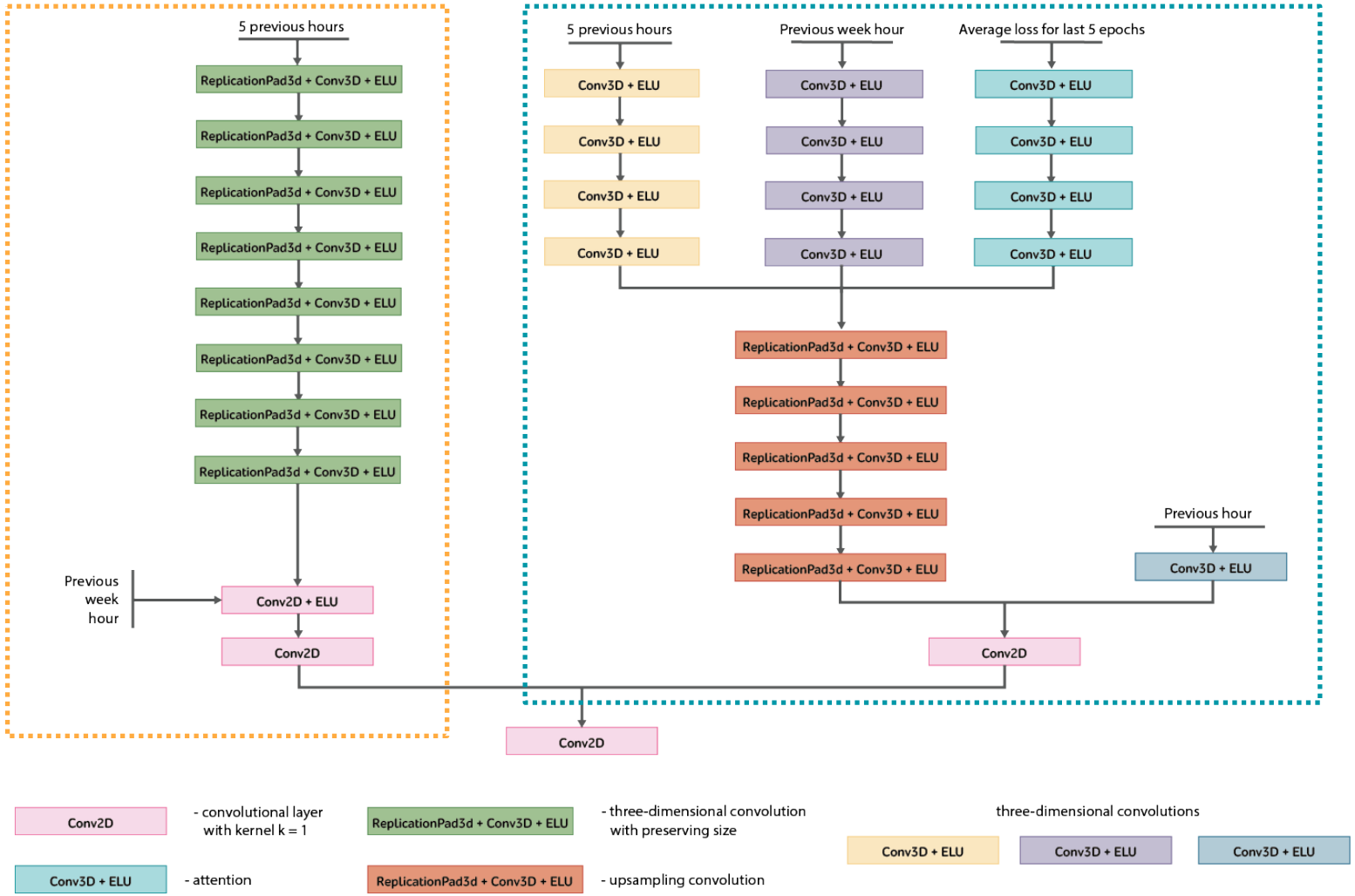


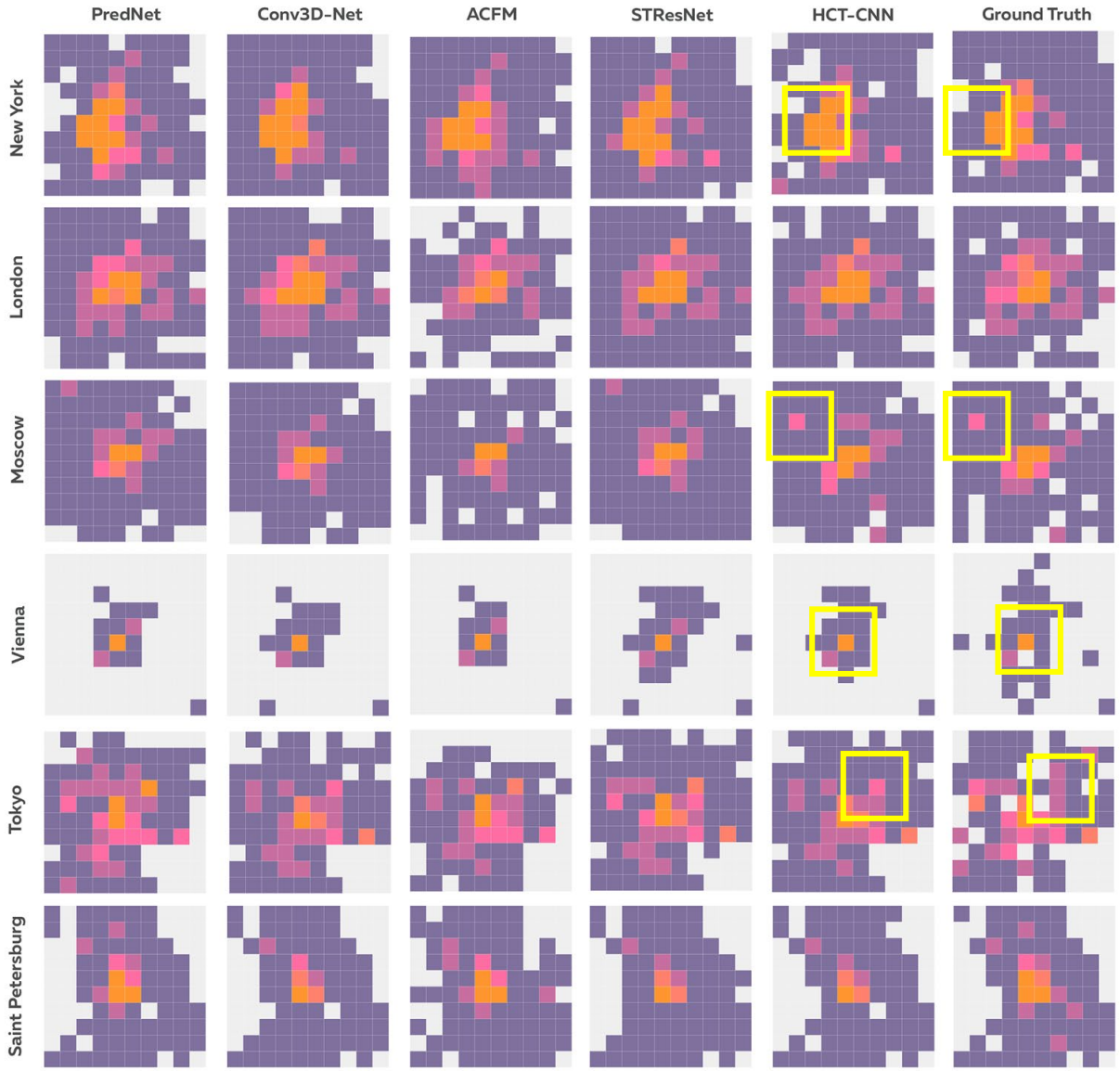


The main challenges of prediction task:

- the absence of a strong influence on sequential hours;
- a lot of zeroes in monitoring area;
- important to predict each cells as precise as possible.

Model architecture





HCT-CNN is able to correctly capture the high activity areas as well as majority of low activity areas.

Orange color represents the maximum level of activity, grey cells indicate zero posts.

Yellow squares indicate correct predictions of HCT-CNN where other models struggled.

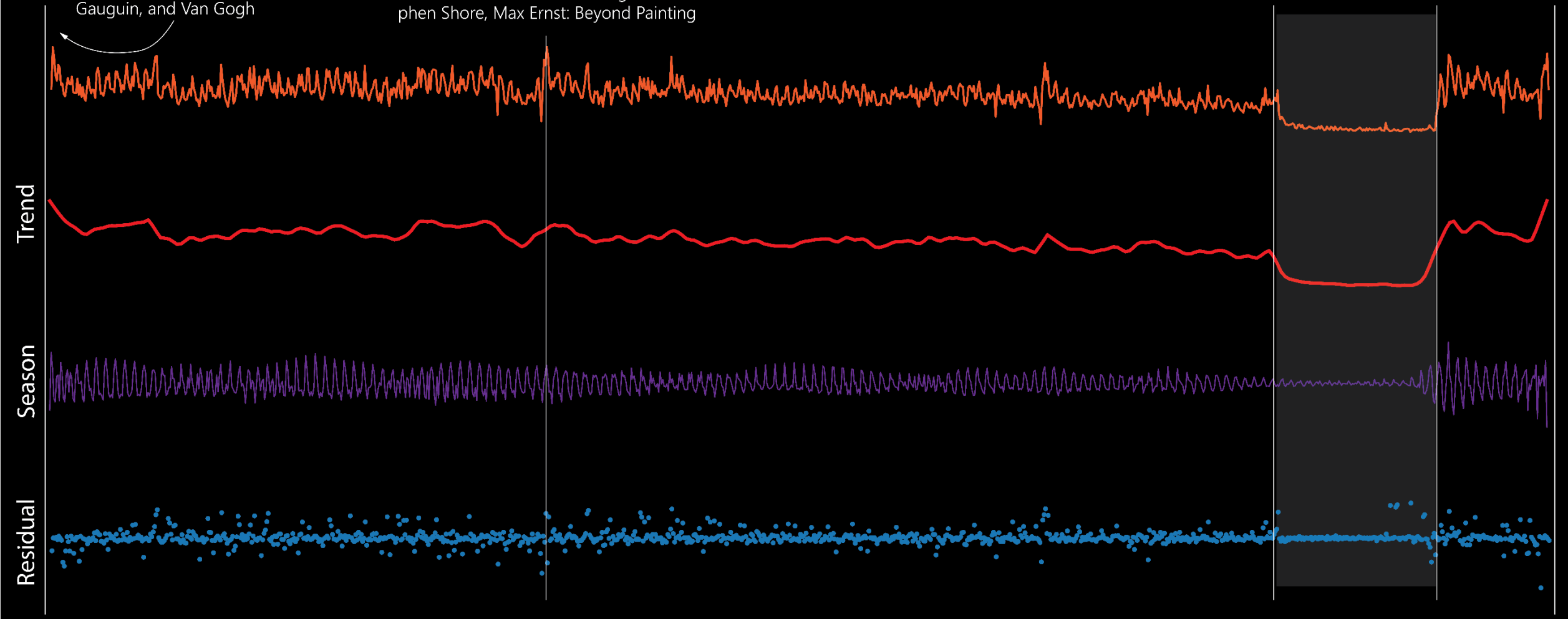
MoMA

The Museum of Modern Art

1st January 2017
Evening Tour: Cézanne, Gauguin, and Van Gogh

27th December 2017
Member After Hours: The Long Run, Stephen Shore, Max Ernst: Beyond Painting

15th June - 21st October 2019
MoMA Expansion
Museum closed

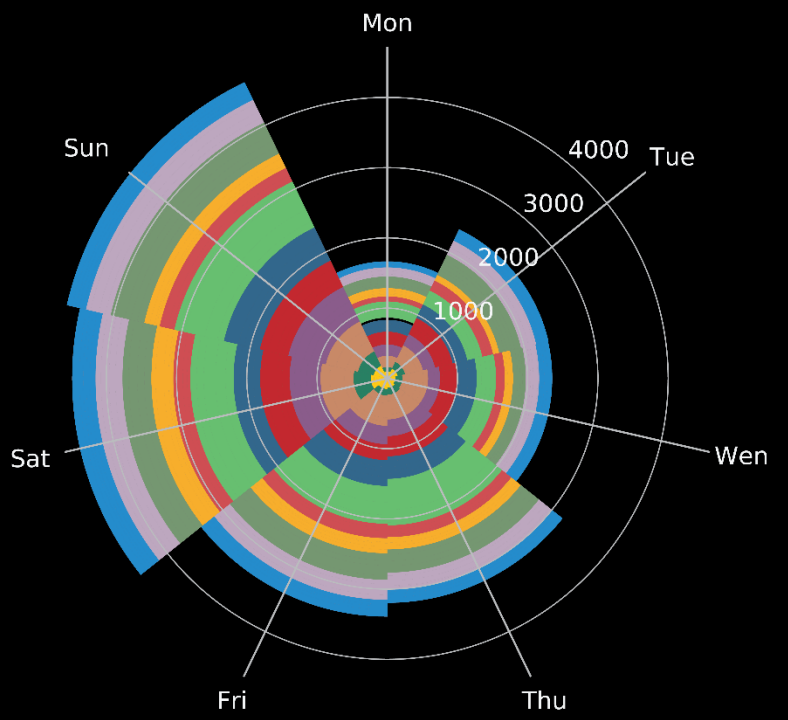


DMA
DALLAS MUSEUM OF ART

6th July- FRIDA FEST
Frida Kahlo's 110th birthday Anniversary

13th March - Exhibition Opening México 1900–1950: Diego Rivera, Frida Kahlo, José Clemente Orozco, and the Avant-Garde

15-16th July- Exhibition Closing México 1900–1950: Diego Rivera, Frida Kahlo, José Clemente Orozco, and the Avant-Garde



Jan 1, 2017

Dec 31, 2017

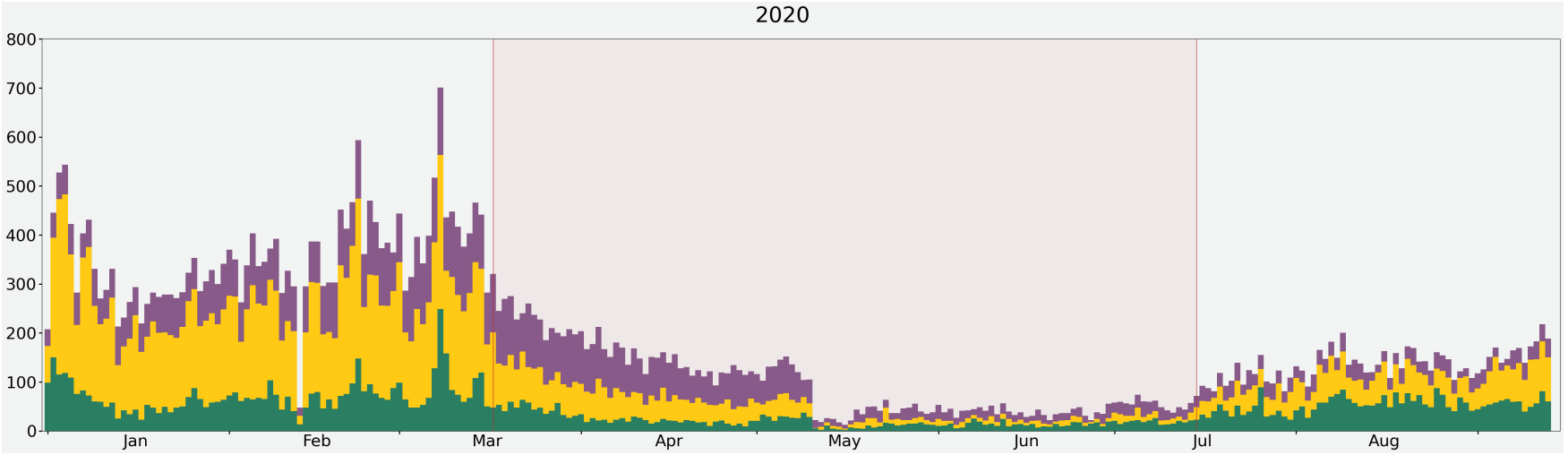
Number of Instagram's posts

Indoor vs. Outdoor

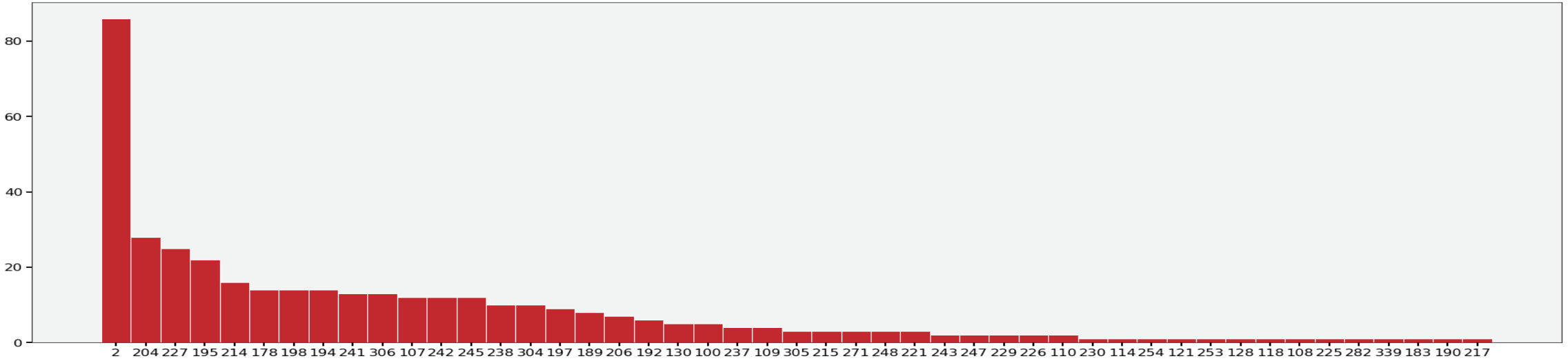
State-of-the art convolutional neural networks trained on Places Database, a repository of 10 million scene photographs.

<u>Results:</u> Accuracy: 0.9607

<u>Read more and check online demo:</u> **Places: A 10 million Image Database for Scene Recognition.** Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba.
http://places2.csail.mit.edu/demo.html



What to see in the Hermitage?



Random sample: 460 images

- The most popular place is the main staircase – 86 photos
- Da Vinci's paintings – 16 photos (5th place)
- Jupiter Hall – 12 photos (11th place)

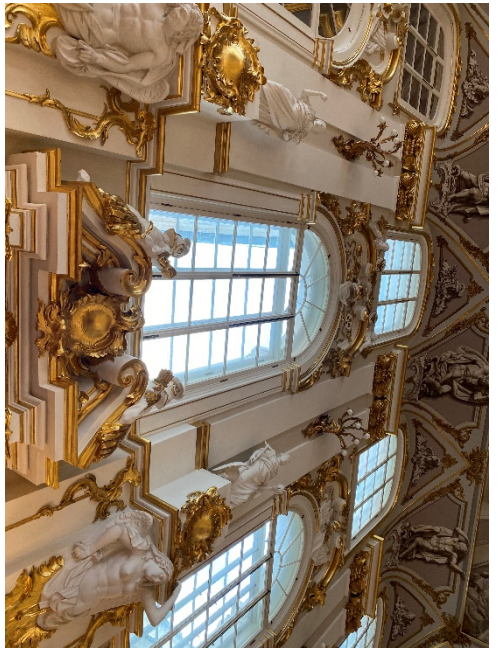
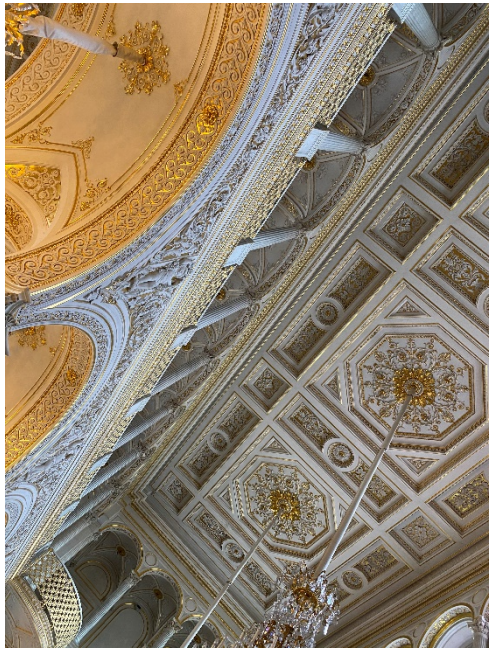
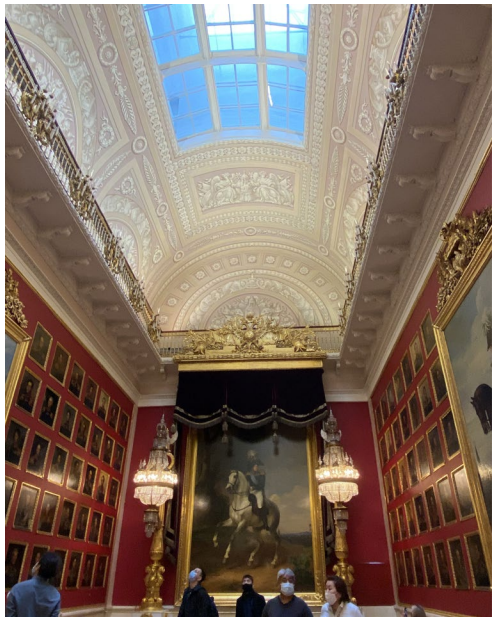
Room detection

Disadvantages:

- Human behavior on video and photo is different;
- Filters and color adjustments;
- Less variety.

Accuracy: 0.717949

Title	Duration	Short description
State Hermitage Museum Tour - St Petersburg Russia	637	Walking tour
Saint Petersburg - Walking Hermitage - Russia / Санкт-Петербург 4K	4725	Walking tour
Эрмитаж. Экскурсия по музею.	4098	Walking tour
Hermitage Museum walking tour 4k 60fps - Saint Petersburg, Russia	2907	Walking tour
12 HOURS IN THE WINTER PALACE + HERMITAGE MUSEUM | Saint Petersburg, Russia	693	Walking tour

Ksenia D. Mukhina
mukhinaks@gmail.com
mukhina@tlu.ee
@mukhinaks at Twitter | Telegram | Instagram

Open Lab

TALLINN UNIVERSITY

Relevant publications:

1. Spatiotemporal Filtering Pipeline for Efficient Social Networks Data Processing Algorithms
2. Detection of tourists attraction points using Instagram profiles
3. Orienteering Problem with Functional Profits for multi-source dynamic path construction
4. Multiscale event detection using convolutional quadtrees and adaptive geogrids
5. Forecasting of the Urban Area State Using Convolutional Neural Networks

ORCID: 0000-0002-2926-3037
Research Gate | Google Scholar | Scopus