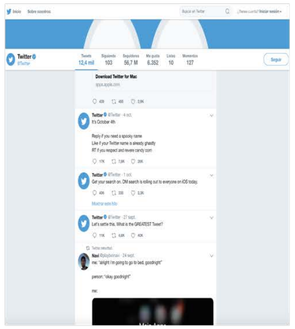
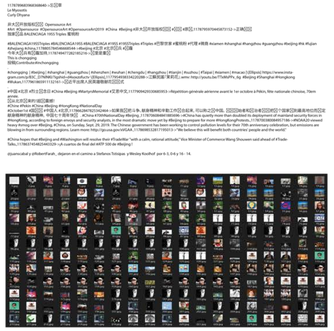

Self Organizing Maps

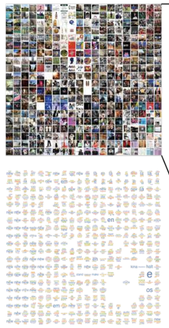


Define your project, and start the crawling process



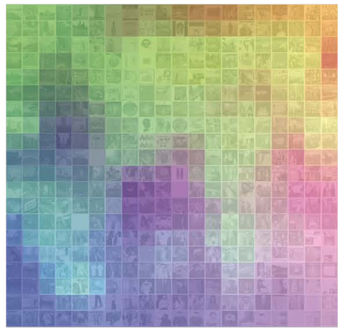
You collected data in two formats

Transform one of the two formats to its numerical representation

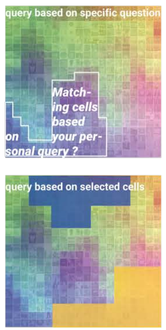


Train and render the SOM

Choose one way of filtering the SOM

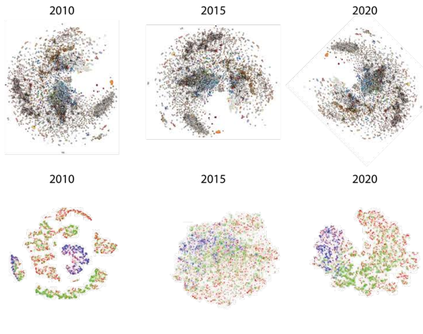


Render the SOM by its weights, assigning colors to each cell



render back to space

or



render by time

FEATURE VECTORS AND EUCLIDEAN DISTANCE

A reminder from the last lecture

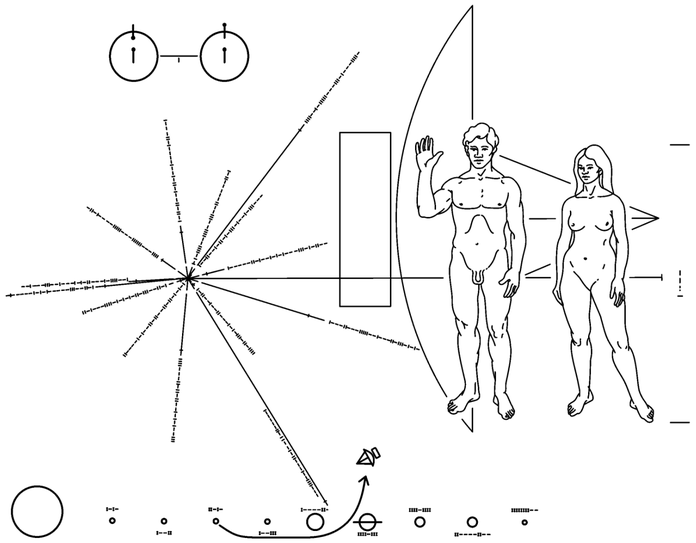
Feature Vector

Numerical features that represent some object

An apple

(color, size, weight, sweetness)

(0 red, 12.3 cm, 180 g, 2 very)



@Pioneer plaque

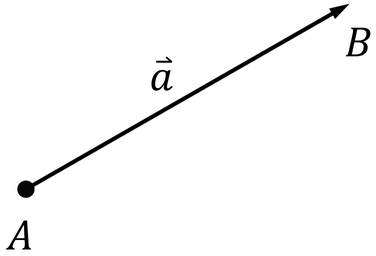
A person

(sex, age, weight, height)

(0 male, 25, 70 kg, 182 cm)

Euclidean distance

distance between two points in Euclidean space



$$\|AB\| = \sqrt{(B_1 - A_1)^2 + (B_2 - A_2)^2 + \cdots + (B_n - A_n)^2} = \|\vec{a}\|$$

Feature Vector + Euclidean Distance

A Common Ground for Comparing Objects

(color, size, weight, sweetness)

(0 red, 12.3 cm, 180 g, 2 very)

A

(color, size, weight, sweetness)

(1 orange, 11.7 cm, 170 g, 1 yes)

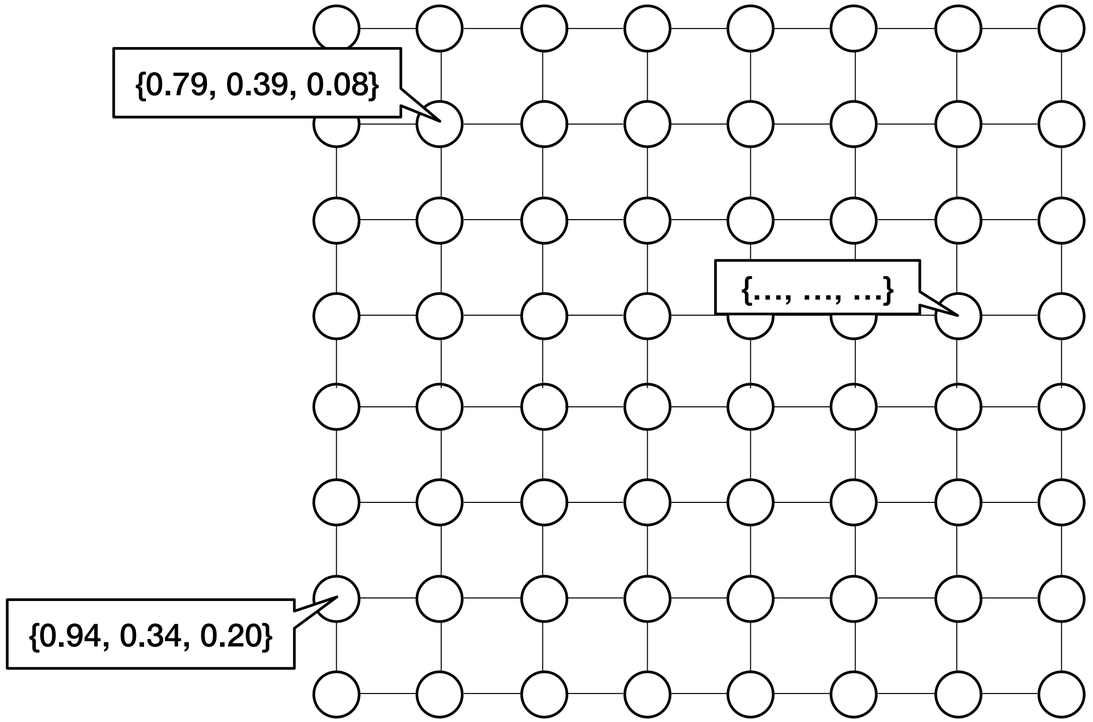
B

$$\|AB\| = \sqrt{(1-0)^2 + (11.7-12.3)^2 + (170-180)^2 + (2-1)^2} = 10.1173..$$

SELF-ORGANIZING MAP (SOM)

Self-organizing map

Kohonen, 1982

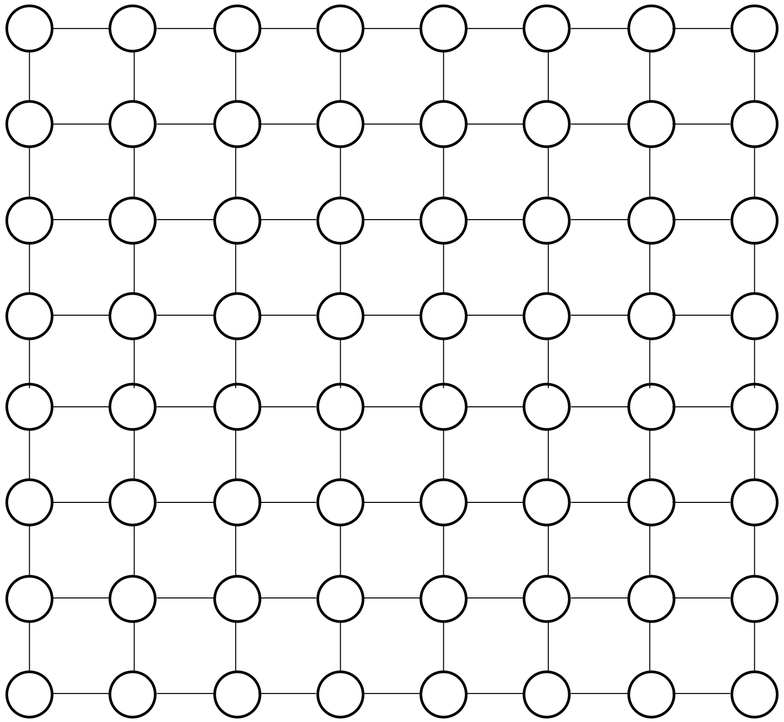


SOM is a grid of feature vectors (cells / units)
above is an example of 8 x 8 x 3 SOM
where 8 x 8 is the size, 3 is the dimensionality of feature vectors

WHAT DOES SOM DO

A Projection from Given Data on Regular Grid

A “compressed” representation of our data



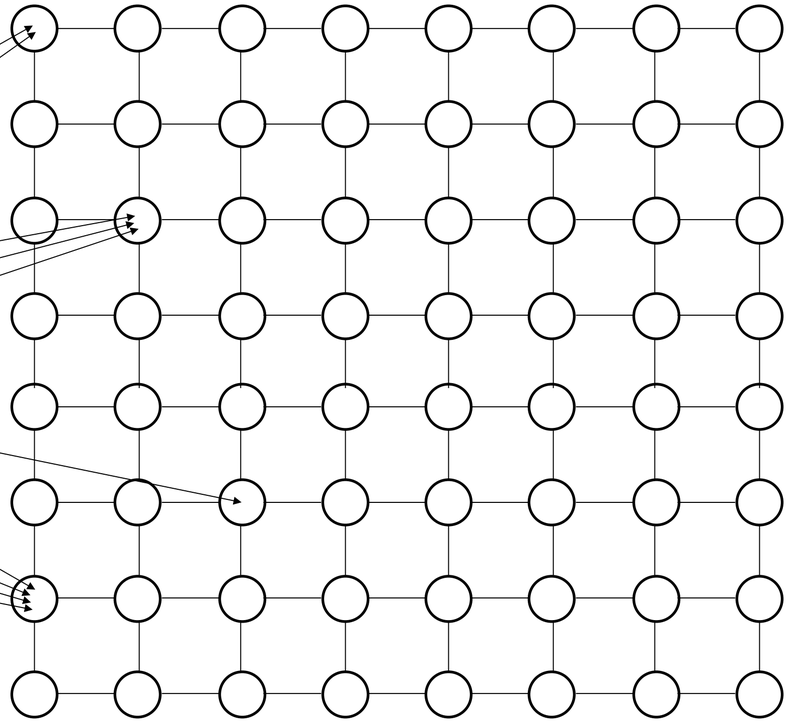
Each SOM cell represents a cluster of similar items from our dataset

A Projection from Given Data on Regular Grid

A “compressed” representation of our data

Our Dataset

{0.831189, 0.433826, 0.114847}
{0.428795, 0.399229, 0.881367}
{0.325843, 0.209812, 0.930411}
{0.246075, 0.468707, 0.972764}
{0.219395, 0.115503, 0.455577}
{0.387547, 0.148913, 0.537185}
{0.0245103, 0.672357, 0.0145895}
{0.233873, 0.419769, 0.494448}
{0.983612, 0.396847, 0.200075}
{0.0644331, 0.453388, 0.899624}
…

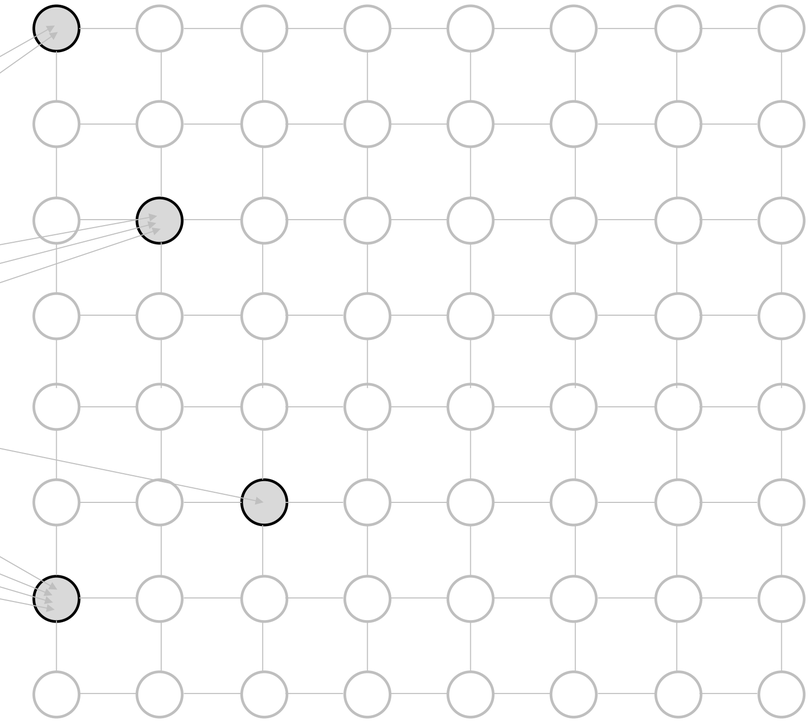


Each SOM cell represents a cluster of similar items from our dataset

The Best Matching Unit (BMU)

Our Dataset

{0.831189, 0.433826, 0.114847}
{0.428795, 0.399229, 0.881367}
{0.325843, 0.209812, 0.930411}
{0.246075, 0.468707, 0.972764}
{0.219395, 0.115503, 0.455577}
{0.387547, 0.148913, 0.537185}
{0.0245103, 0.672357, 0.0145895}
{0.233873, 0.419769, 0.494448}
{0.983612, 0.396847, 0.200075}
{0.0644331, 0.453388, 0.899624}
...



For the items of our dataset, their corresponding SOM cells are called their BEST MATCHING UNIT

MORE IN DETAIL

Training the SOM

A Huge Dataset Impossible to Manage Manually

For example 100k feature vectors of apples

(color, size, weight, sweetness)

(0 red, 12.3 cm, 180 g, 2 very)

(1 orange, 11.7 cm, 170 g, 1 yes)

...

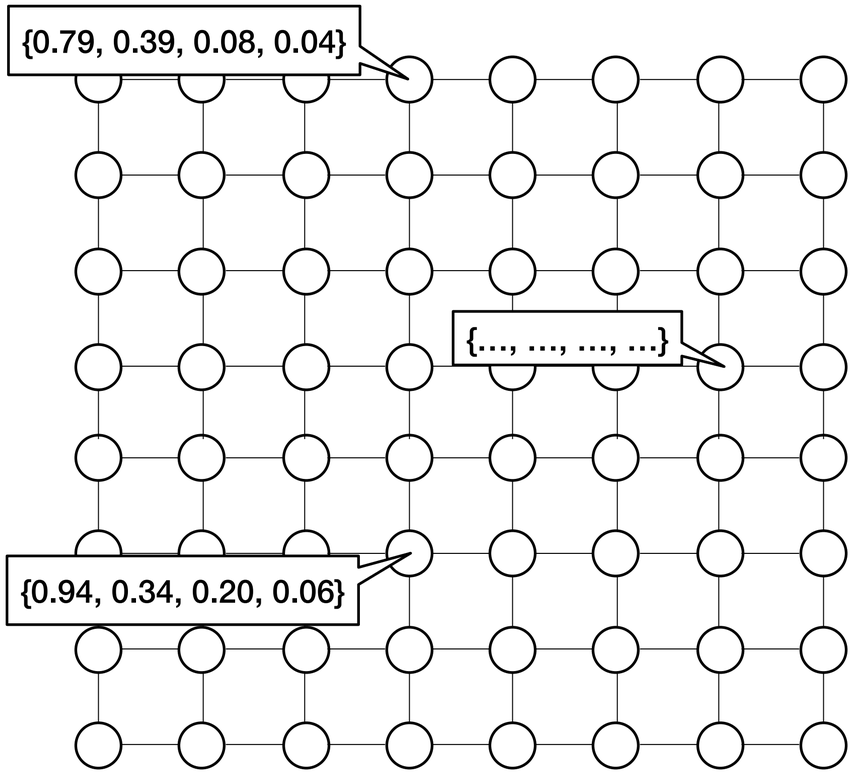
(..., ..., ..., ...)

But we know some of them are similar to each other

A Random SOM

Our dataset

{0.837001,0.727544,0.0394684,0.299028}
{0.684913,0.337634,0.892134,0.794974}
{0.178787,0.837747,0.2533,0.182625}
{0.352355,0.611677,0.643471,0.486624}
{0.0623343,0.330299,0.440126,0.550363}
{0.95928,0.453402,0.250339,0.138137}
{0.906858,0.0194981,0.464387,0.836483}
{0.70246,0.532199,0.463251,0.0710004}
{0.84464,0.606172,0.864091,0.26712}
{0.427036,0.299789,0.128026,0.759577}
…

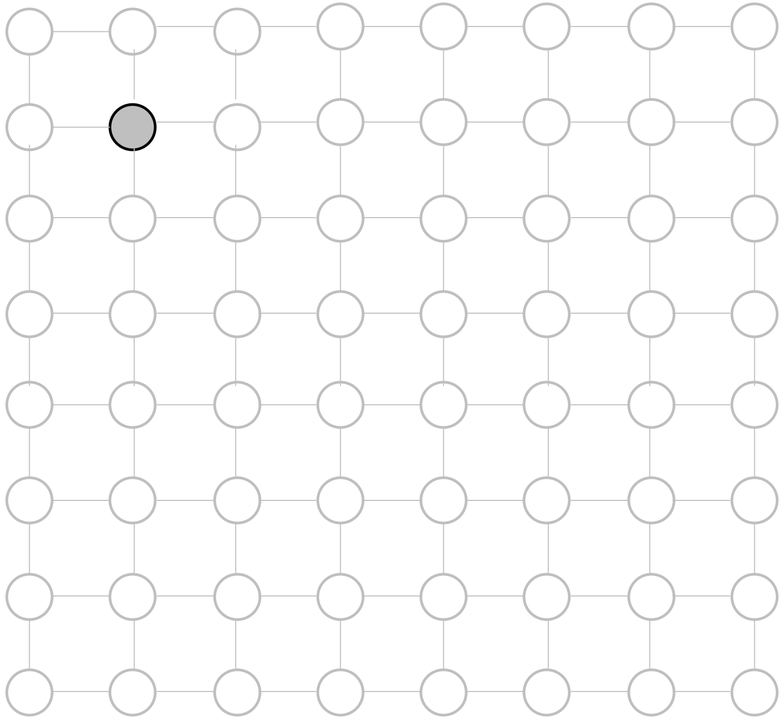


We create a SOM with its feature vectors randomly initialized

Find BMU

Our dataset

{0.837001,0.727544,0.0394684,0.299028}
{0.684913,0.337634,0.892134,0.794974}
{0.178787,0.837747,0.2533,0.182625}
{0.352355,0.611677,0.643471,0.486624}
{0.0623343,0.330299,0.440126,0.550363}
{0.95928,0.453402,0.250339,0.138137}
{0.906858,0.0194981,0.464387,0.836483}
{0.70246,0.532199,0.463251,0.0710004}
{0.84464,0.606172,0.864091,0.26712}
{0.427036,0.299789,0.128026,0.759577}
…



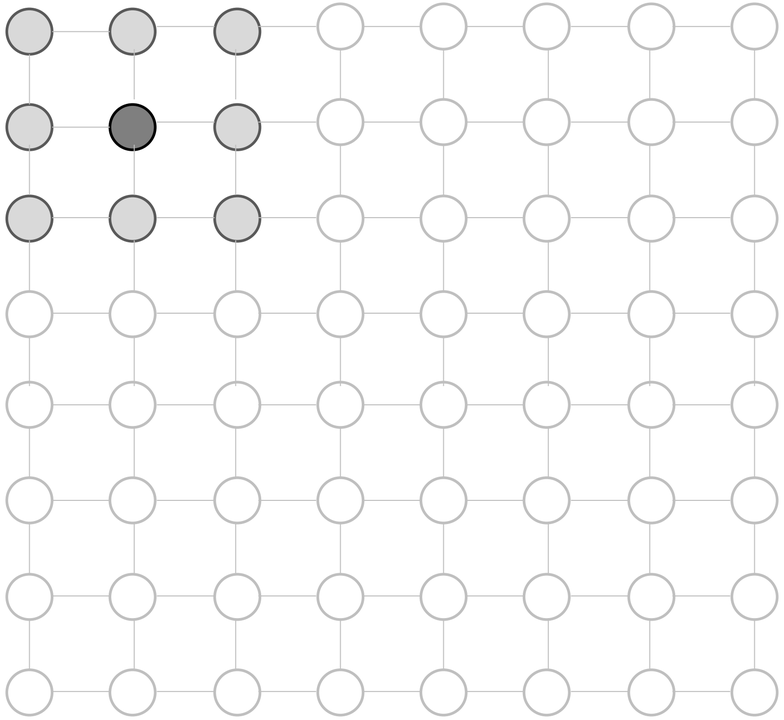
We randomly select one item from our dataset, find its BMU

Update BMU Feature Vector

"*Stretching*" the SOM in feature space

Our dataset

{0.837001,0.727544,0.0394684,0.299028}
{0.684913,0.337634,0.892134,0.794974}
{0.178787,0.837747,0.2533,0.182625}
{0.352355,0.611677,0.643471,0.486624}
{0.0623343,0.330299,0.440126,0.550363}
{0.95928,0.453402,0.250339,0.138137}
{0.906858,0.0194981,0.464387,0.836483}
{0.70246,0.532199,0.463251,0.0710004}
{0.84464,0.606172,0.864091,0.26712}
{0.427036,0.299789,0.128026,0.759577}
…



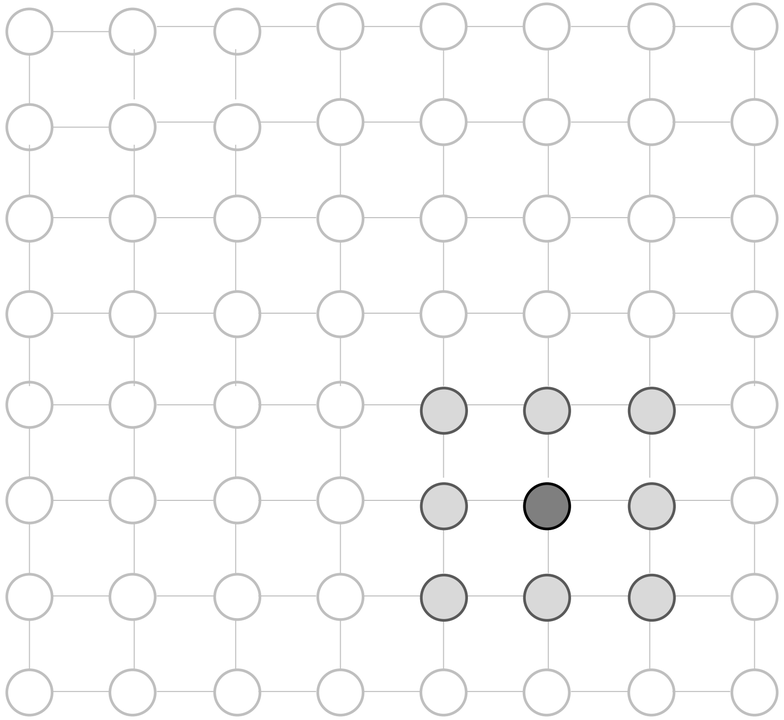
We update the BMU and its neighbors' feature vectors so that they are similar to our selected element
The update is stronger to the BMU than to its neighbors

Update BMU Feature Vector

"*Stretching*" the SOM in feature space

Our dataset

{0.837001,0.727544,0.0394684,0.299028}
{0.684913,0.337634,0.892134,0.794974}
{0.178787,0.837747,0.2533,0.182625}
{0.352355,0.611677,0.643471,0.486624}
{0.0623343,0.330299,0.440126,0.550363}
{0.95928,0.453402,0.250339,0.138137}
{0.906858,0.0194981,0.464387,0.836483}
{0.70246,0.532199,0.463251,0.0710004}
{0.84464,0.606172,0.864091,0.26712}
{0.427036,0.299789,0.128026,0.759577}
…



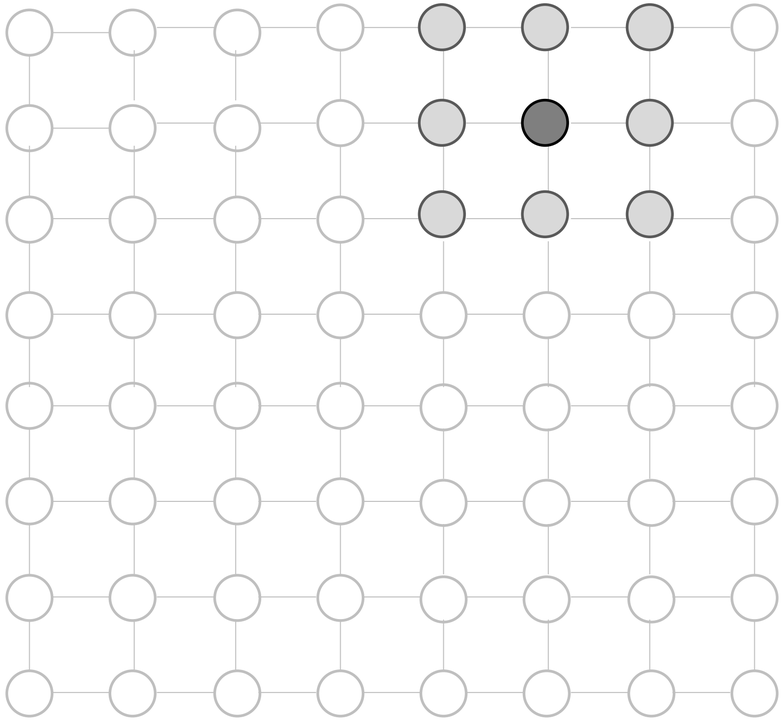
We move to another randomly selected item, and repeat the process

Update BMU Feature Vector

"*Stretching*" the SOM in feature space

Our dataset

{0.837001,0.727544,0.0394684,0.299028}
{0.684913,0.337634,0.892134,0.794974}
{0.178787,0.837747,0.2533,0.182625}
{0.352355,0.611677,0.643471,0.486624}
{0.0623343,0.330299,0.440126,0.550363}
{0.95928,0.453402,0.250339,0.138137}
{0.906858,0.0194981,0.464387,0.836483}
{0.70246,0.532199,0.463251,0.0710004}
{0.84464,0.606172,0.864091,0.26712}
{0.427036,0.299789,0.128026,0.759577}
…

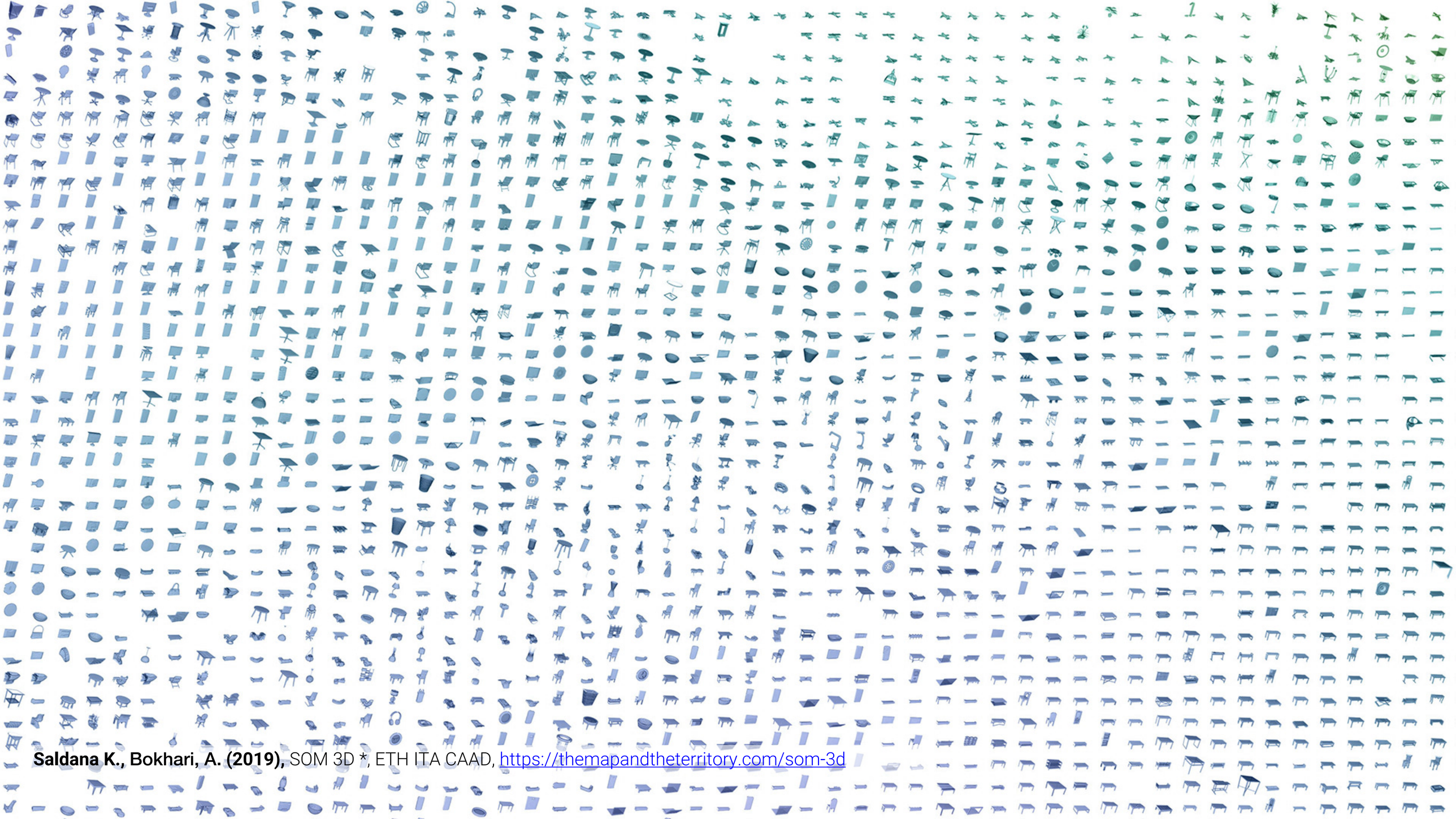


Repeating the process over the entire dataset called an Epoch
We do many Epochs with the radius of neighbor cells decreasing

EXAMPLES OF SOMS ON DIFFERENT TYPES OF DATA

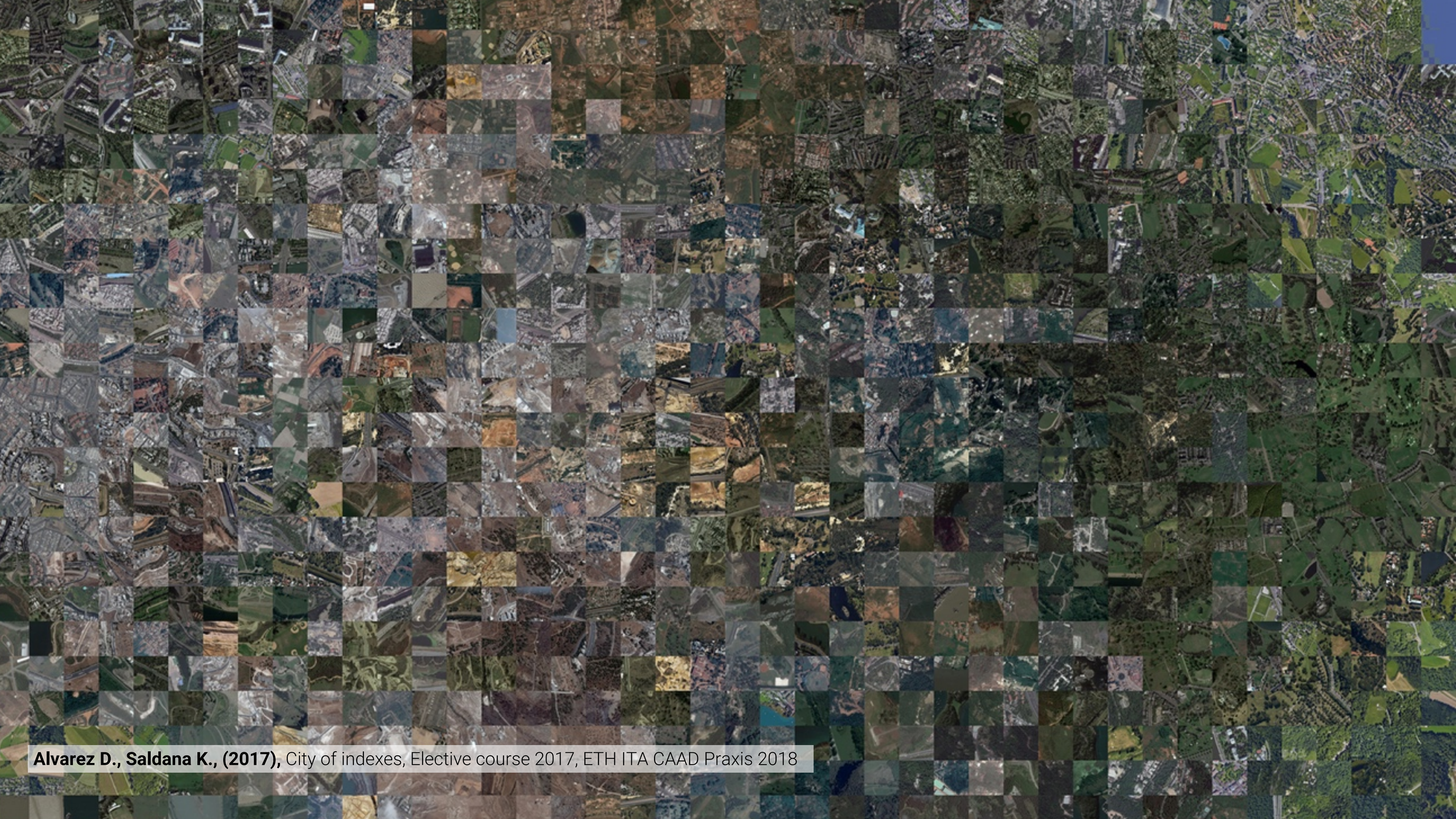
We can render SOM cells by the items they contain

Saldana K., Bokhari, A. (2019), SOM 3D *, ETH ITA CAAD, https://themapandtheterritory.com/som-3d

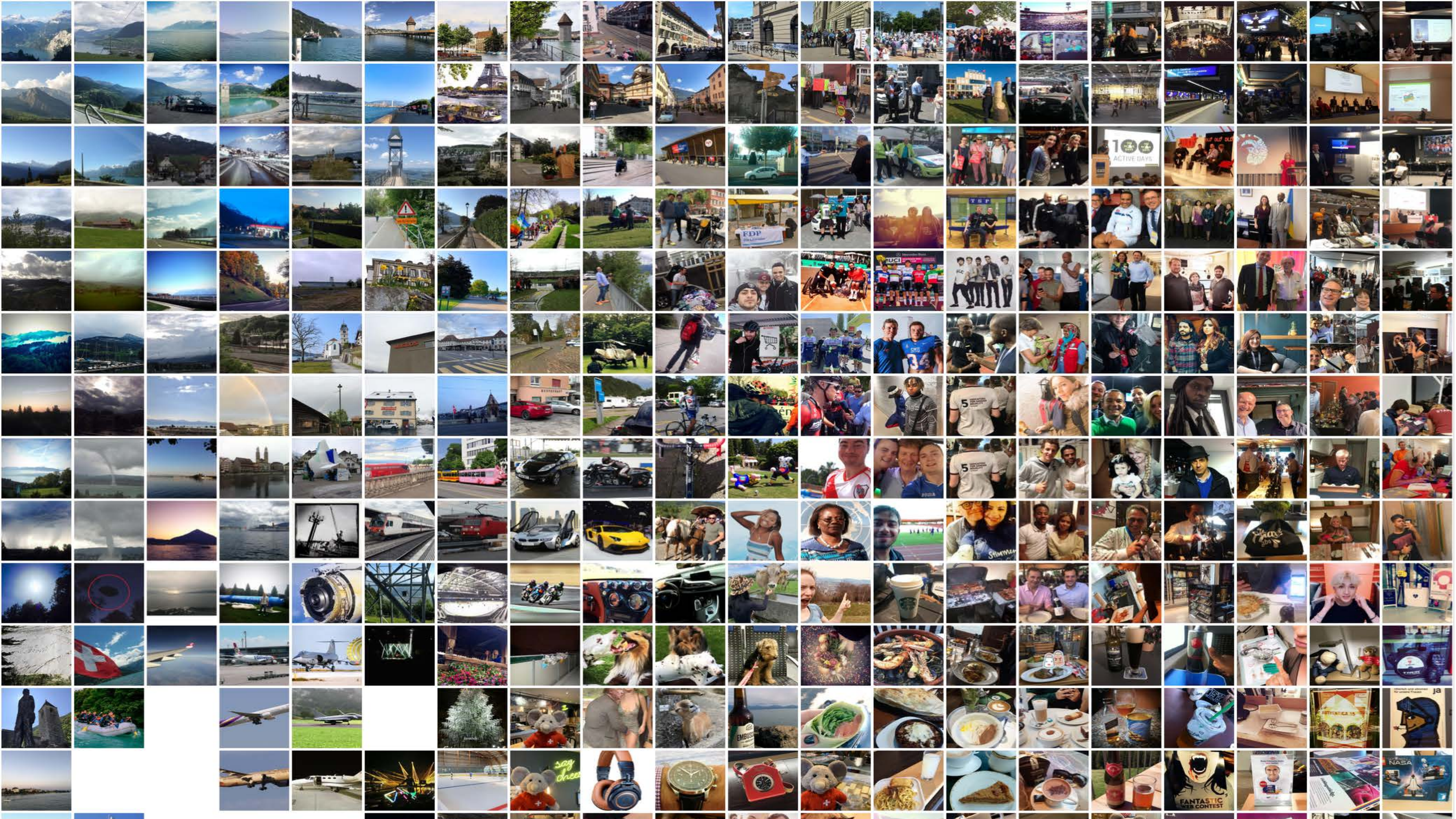


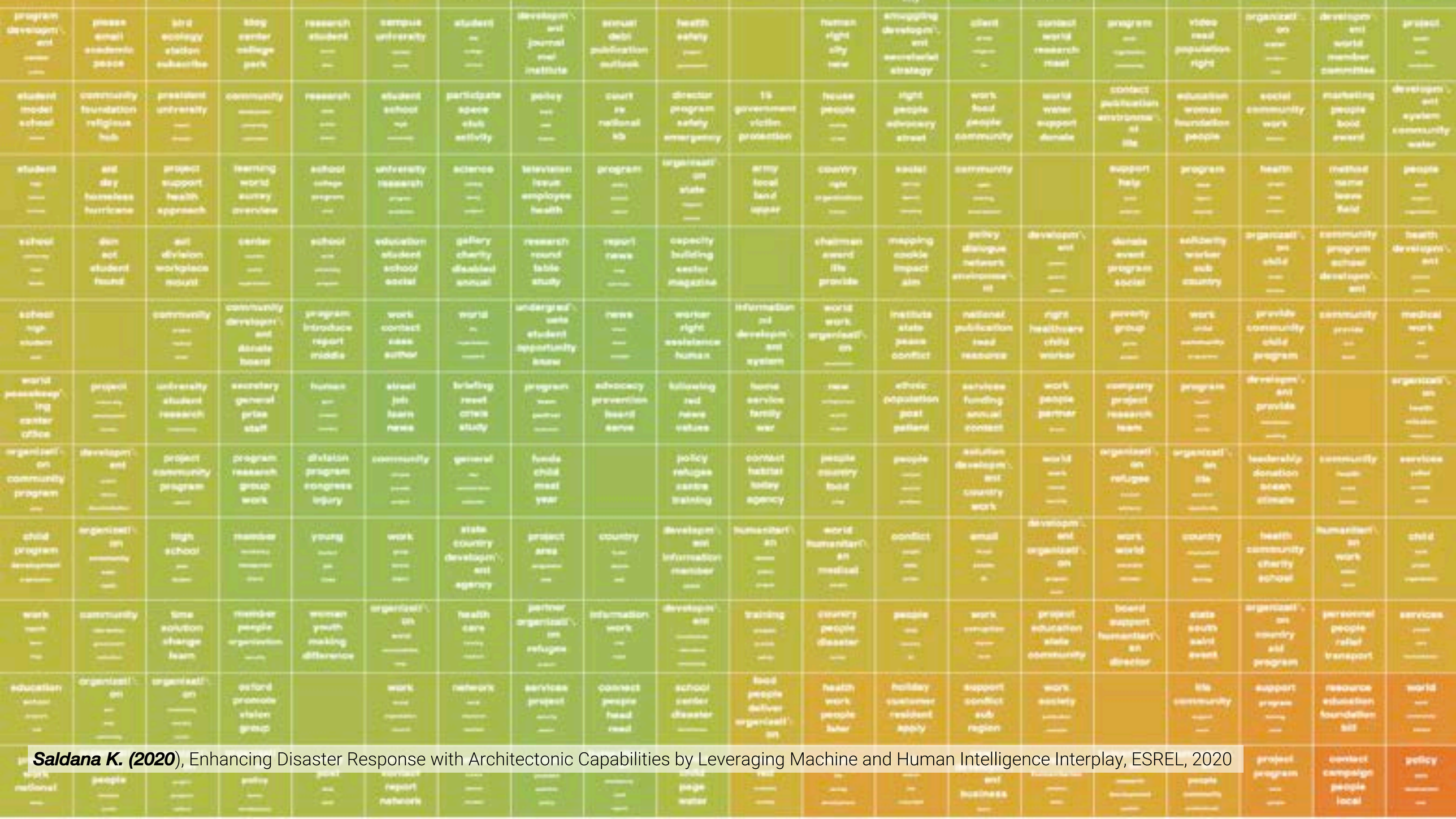
Saldana K., Bokhari, A. (2019), SOM 3D *, ETH ITA CAAD, https://themapandtheterritory.com/som-3d

Alvarez D., Saldana K., (2017), City of indexes, Elective course 2017, ETH ITA CAAD Praxis 2018

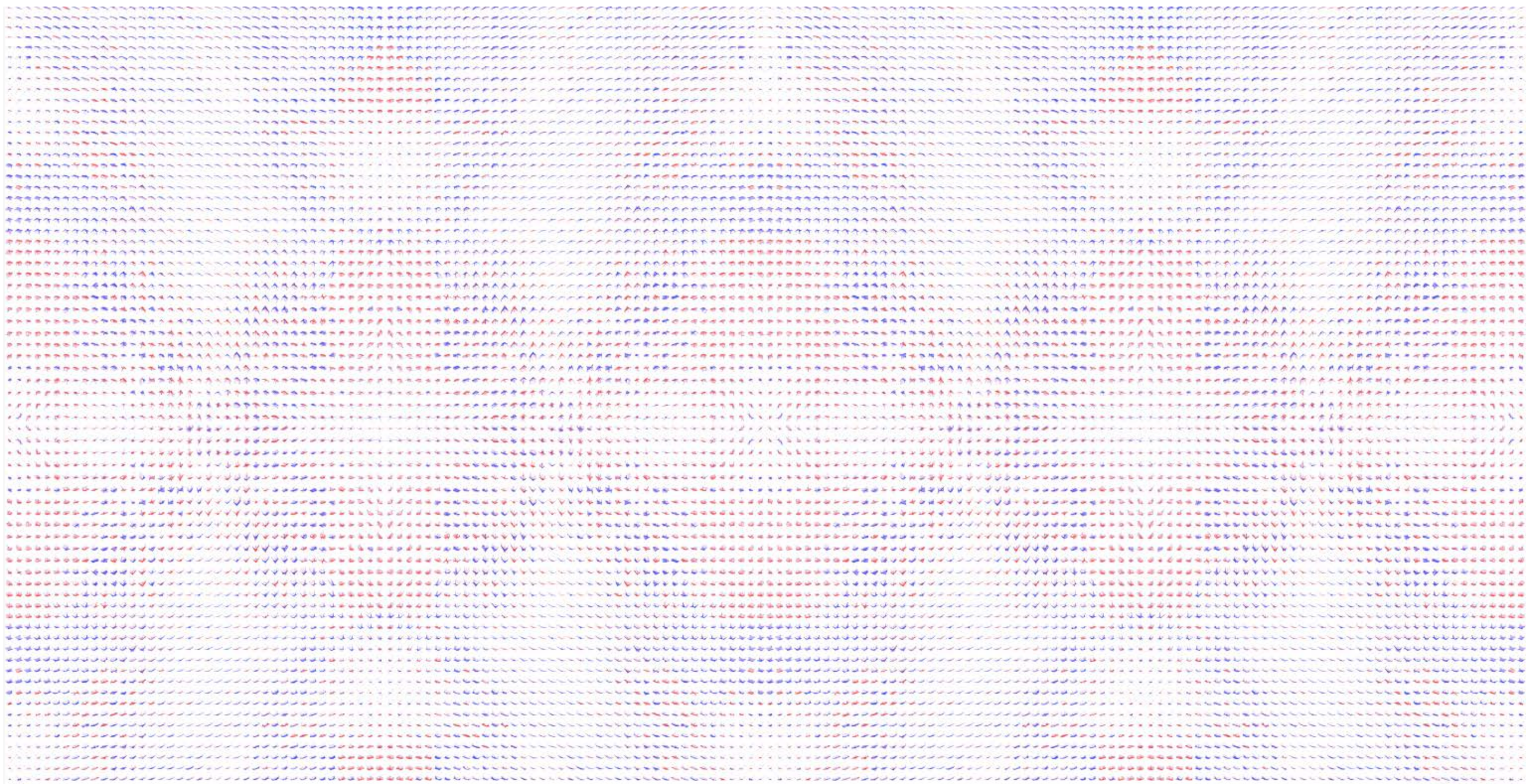


Alvarez D., Saldana K., (2017), City of indexes, Elective course 2017, ETH ITA CAAD Praxis 2018


FDP
100
ACTIVE DAYS
say
cheese
NASA



Saldana K. (2020), Enhancing Disaster Response with Architectonic Capabilities by Leveraging Machine and Human Intelligence Interplay, ESREL, 2020



Saldaña K., Ohlbrock O., D'Acunto .P, Moosavi V., (2020), Beyond Typology Beyond Optimization, International Journal of Architectural Computing

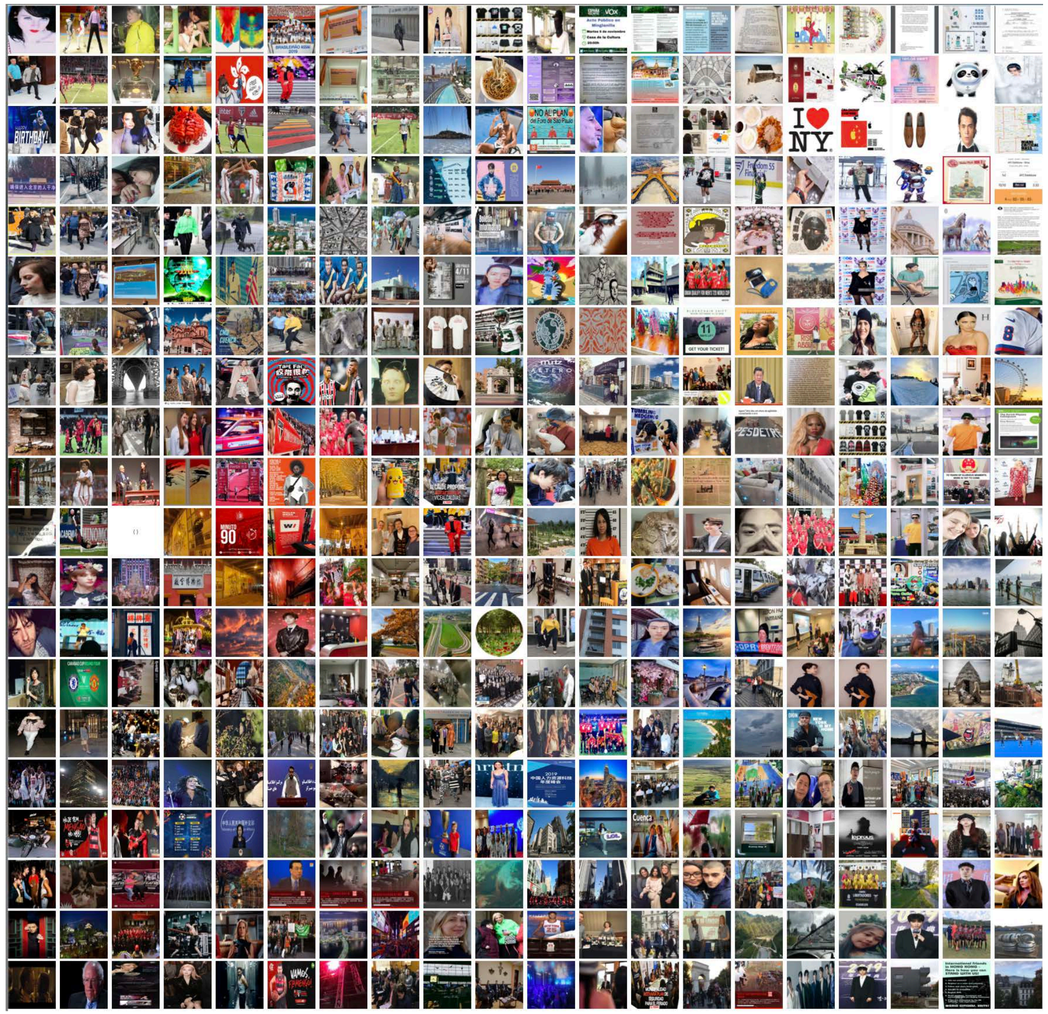
SOM ON TWEETS

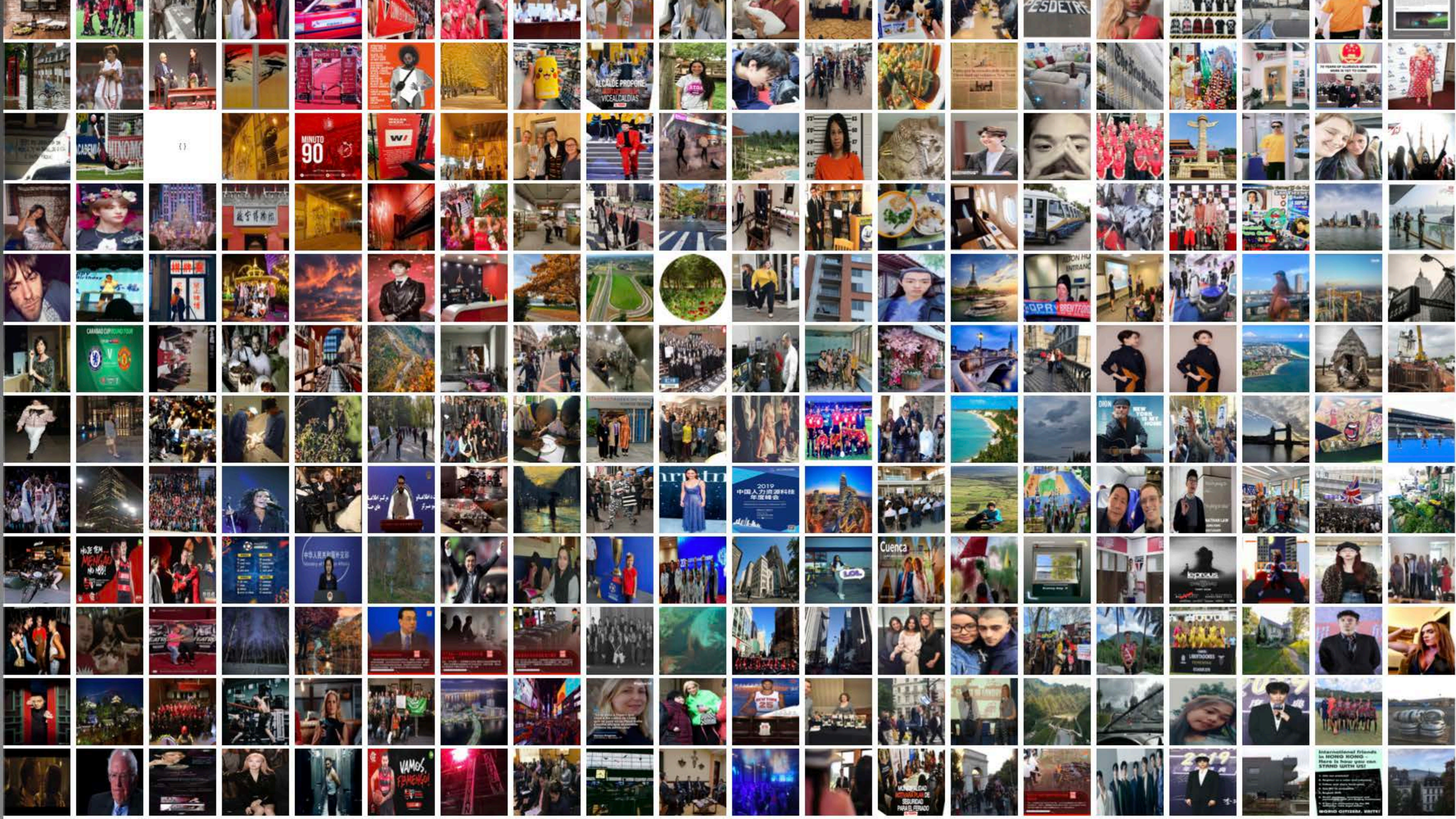
Demonstration with the previous collected tweets

image

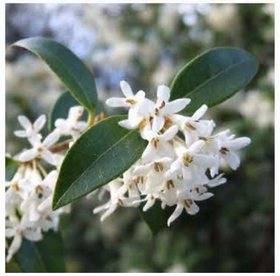
-> "colors"

```
1600   0  1600   0  1600
   0 1600   0  1600   0
1600   0  1516  84  1536
  64 1600   0  1562  38
1477 123  1494 106  1558
  42 1600   0  1600   0
1600   0  1600   0  1512
```

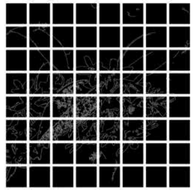



MINUTO 90
VAMOS, FLAMENGO!
Cuenca
International friends in HONG KONG - Here is how you can STAND WITH US!

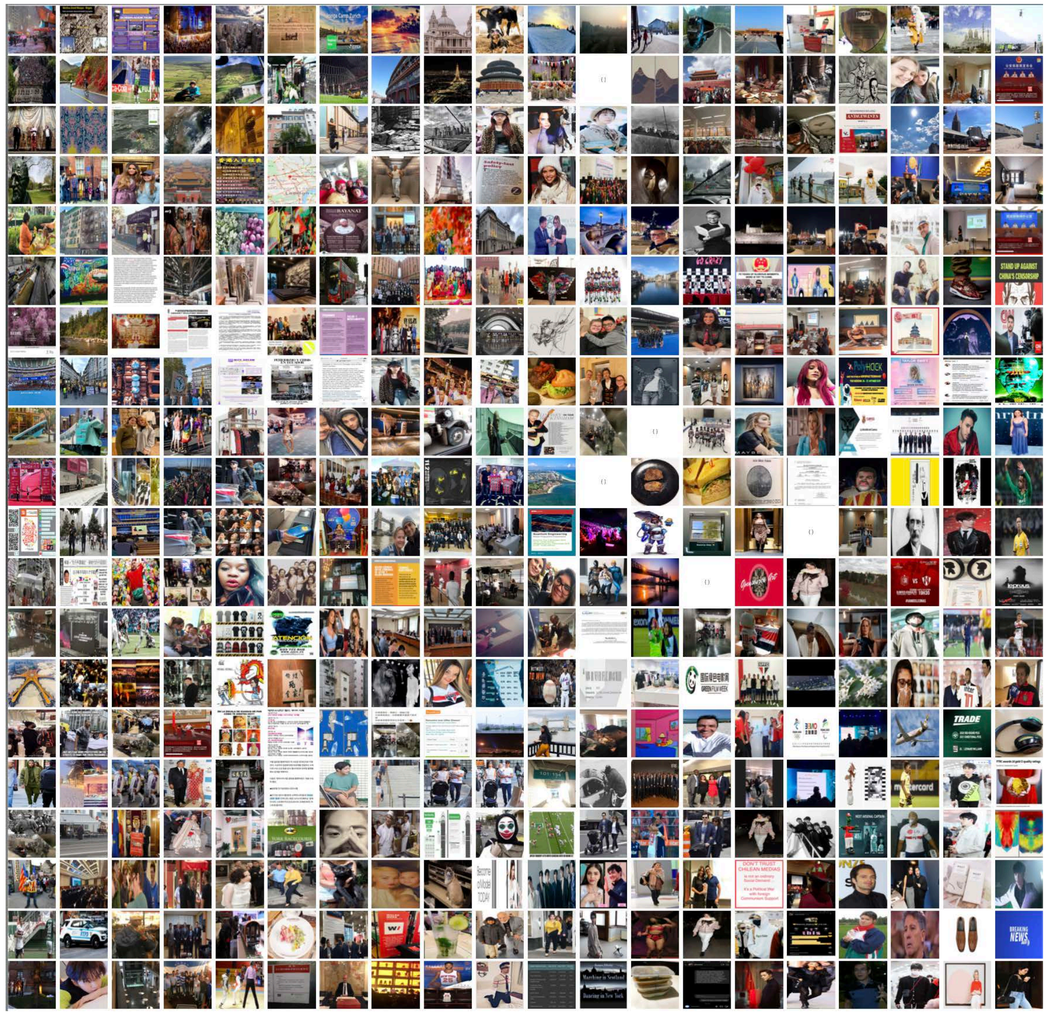
image

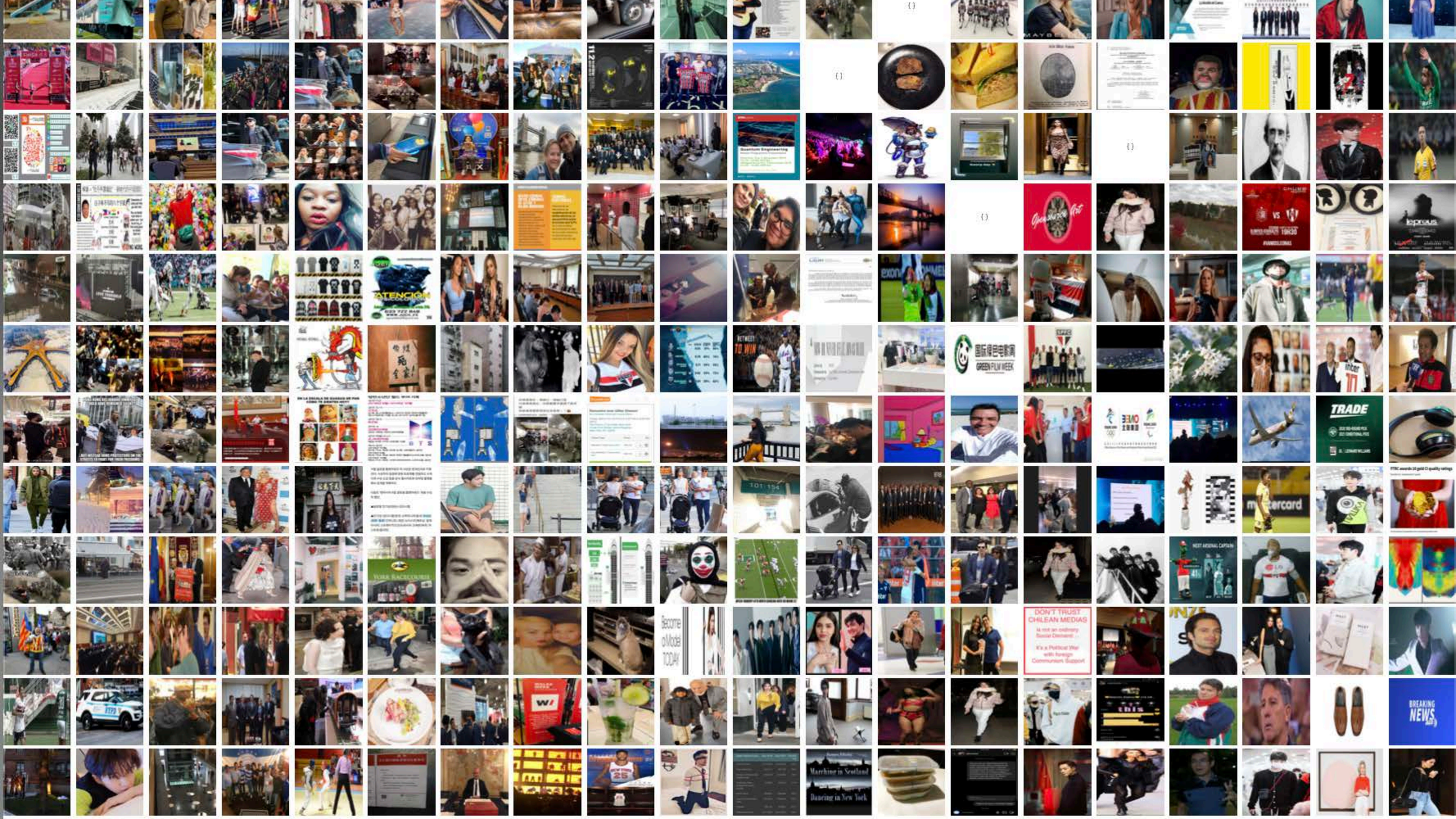


-> "edges"

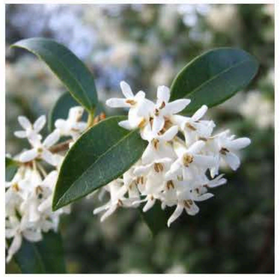


```
0.439216 0.470588 0.541176
0.694118 0.733333 0.862745
0.780392 0.811765 0.898039
0.733333 0.741176  0.74902
0.462745 0.490196 0.458824
0.505882 0.533333 0.458824
0.662745 0.694118 0.701961
```

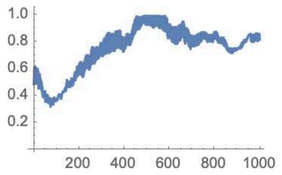



DON'T TRUST CHILEAN MEDIAS
TRADE
BREAKING NEWS
Marching in Scotland
Dancing in New York

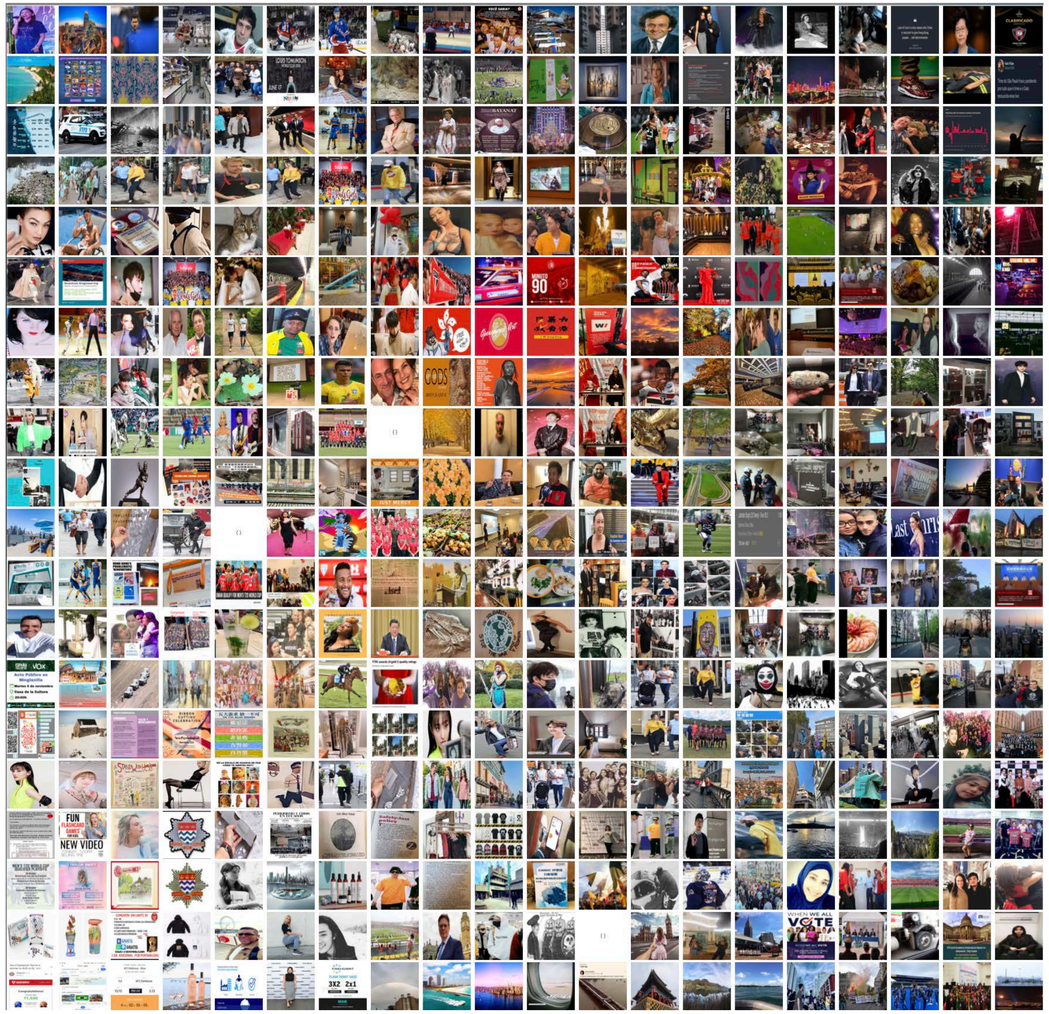
image

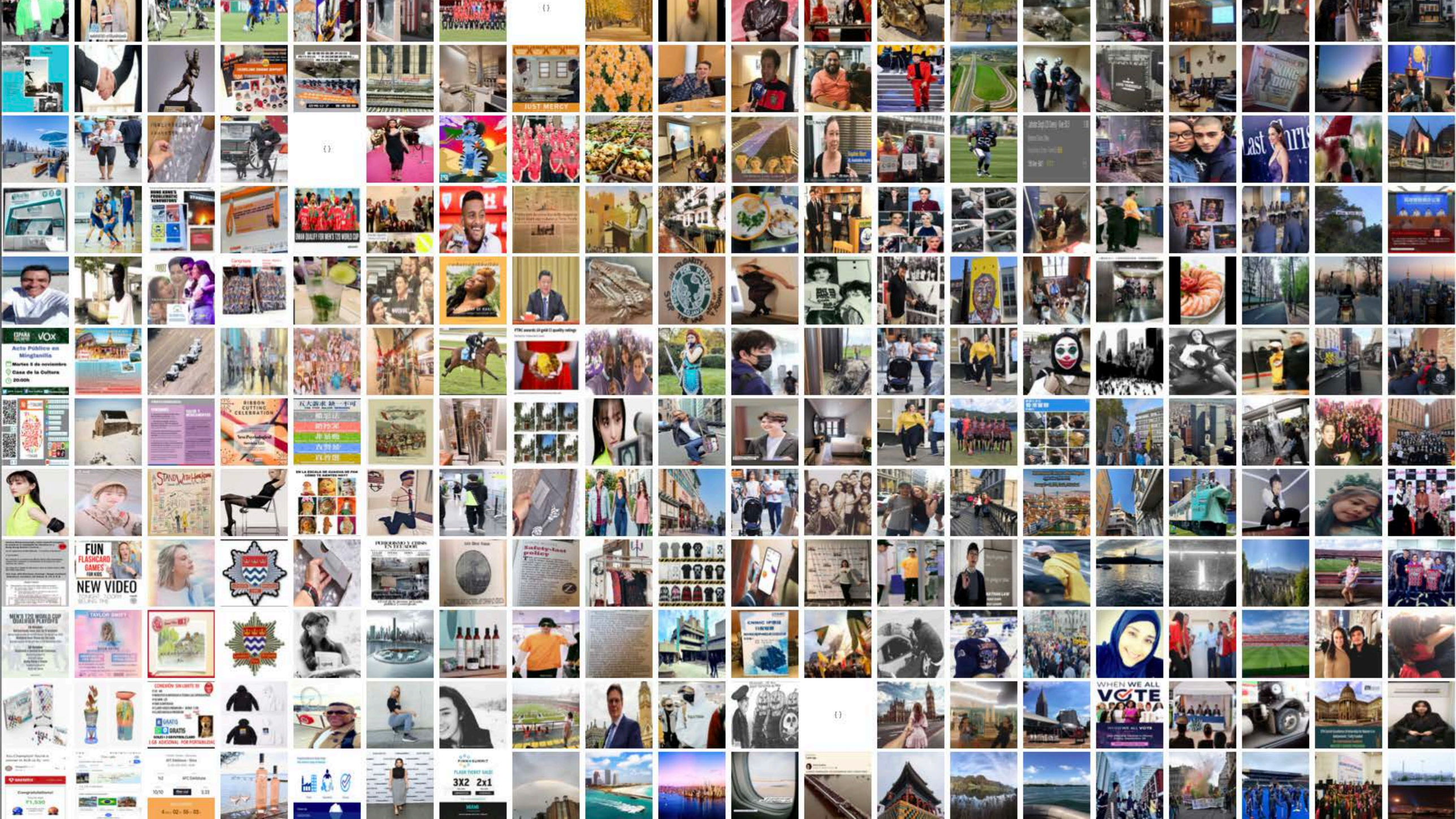


-> "fourier"



311.196 + 0. i	-2.844
-2.8447 - 8.53134 i	-18.802
0.0507741 - 1.67722 i	-0.85581
11.2859 + 32.356 i	2.2017
1.98793 + 0.293197 i	-2.4941
1.08693 - 0.0374585 i	-0.72863
-7.28062 + 7.17637 i	0.61670



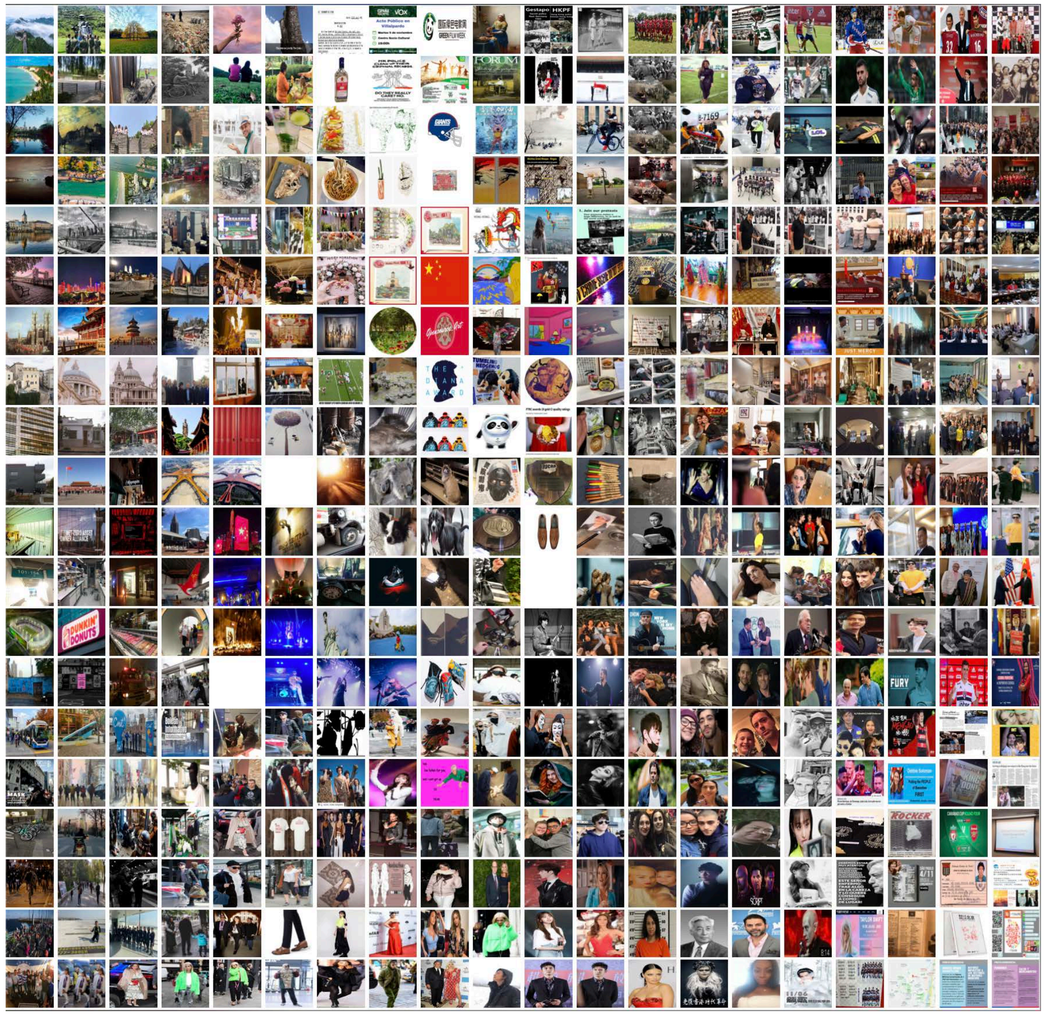

Acto Público en Minglanilla
Martes 8 de noviembre
Casa de la Cultura
20:00h
FUN FLASHCARD GAMES FOR KIDS
NEW VIDEO
JUST MERCY
3X2 2x1
WHEN WE ALL VOTE

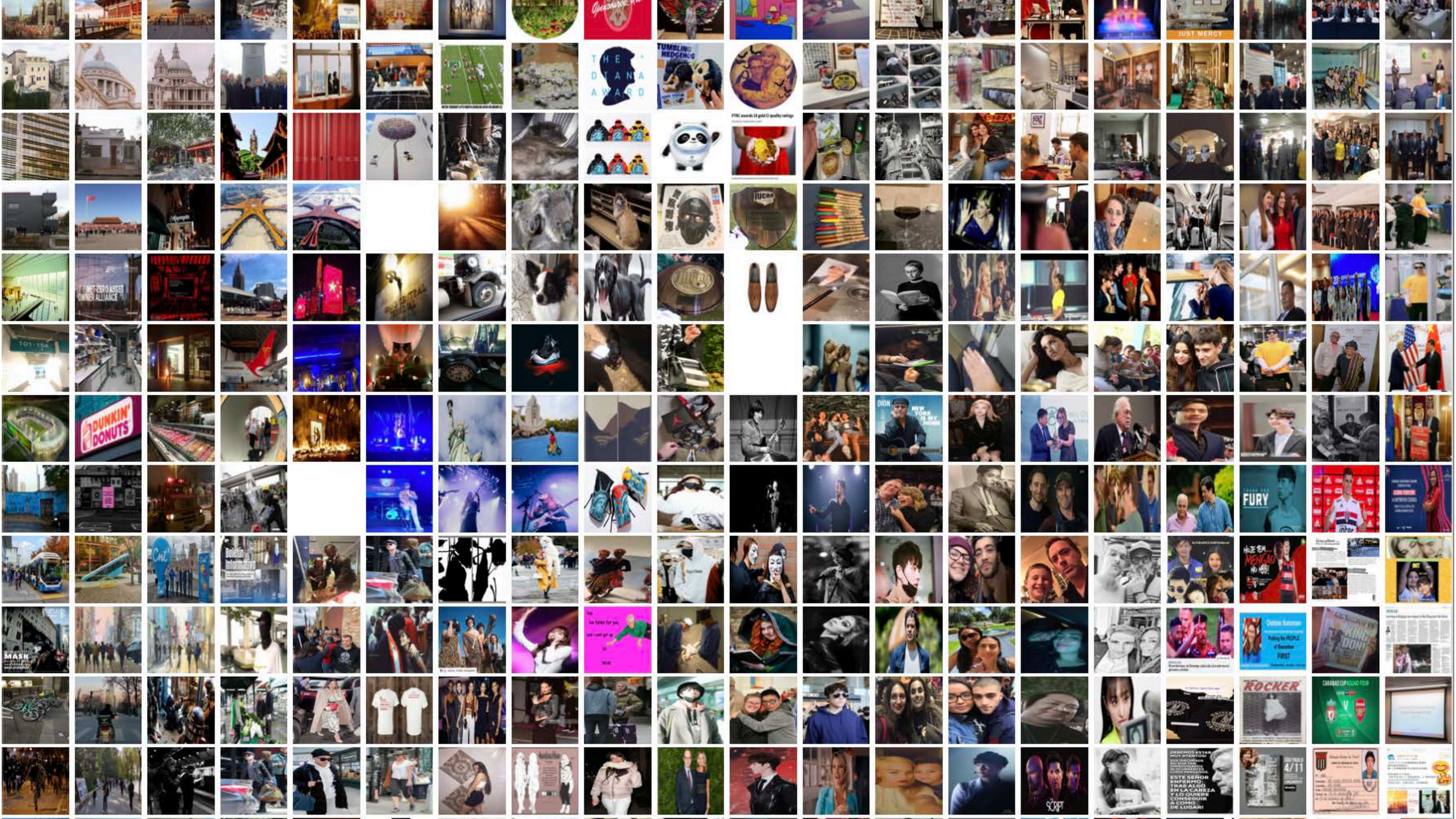
image

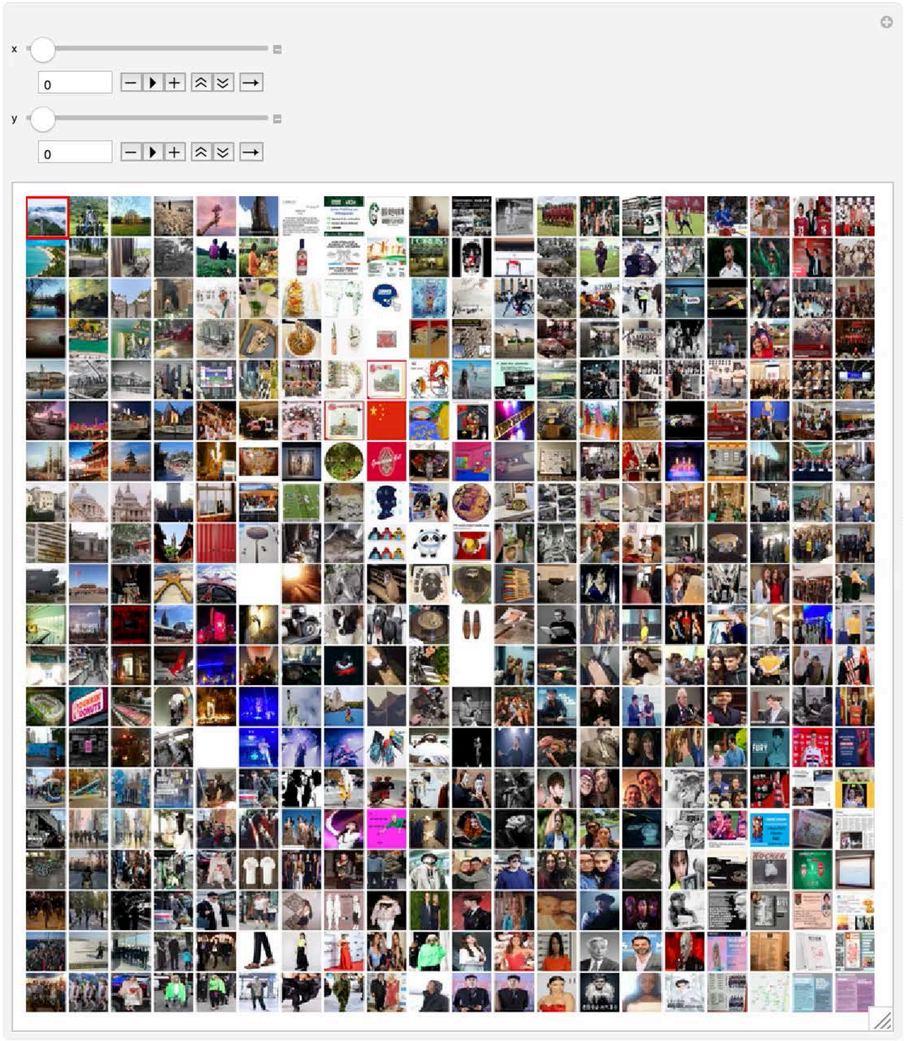
-> "feature extraction"

- common privet
- common jasmine
- laurel
- California laurel
- fruit tree
- pride-of-rochester

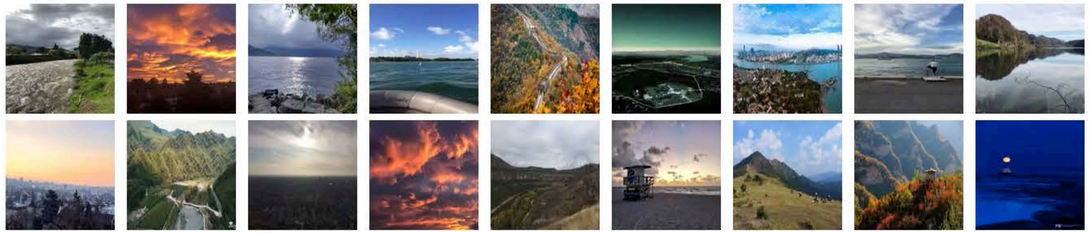
```
 8.22561  -8.62608  28.0713
-3.54004  12.2321   3.18221
-6.41089  1.24627  -2.62163
```




JUST MERCY
THE DIANA AWARD
TUMBLING HEDGEHOG
DUNKIN' DONUTS
FURY
MASK
ROCKER
SCRAP



* cell N = 1 * events= 34



WORKING WITH SOM IN MATHEMATICA

Fitting your questions into the techniques

Functions »

Creating SOM

Out[•]= some sample data (490MB, 3K images, crawled by keywords, non-geo-taged)
https://drive.google.com/open?id=1t0F1t4Qygi-HtbRZoVyrD3Fmrnc8psna

Out[•]= swiss main cities sample data (1GB, 26 cities, 8k images, geo-taged)
https://drive.google.com/open?id=1F9i0mSq-BOMFodDKVT5Rvcn8vYAzaUlA

Versions »

Build Self Organizing Maps

Specifying Working Folder »

Import Data + Train a SOM »

OR Import a Saved SOM »

Visualize SOM »

Save SOM as File »

Working With SOM

Overlay two Datasets (for comparison) »

Highlight SOM by Location/Time/Keywords »

Select Cells on the SOM »

Query to SOM »