

EMPORIO SIRENUSE

By Kimia Mavon, Karan R Motwani, Moreno Vendra and Francesca Morini



Problem

emporiosirenuse ...

Following

Emporio Sirenuse Emporio Sirenuse offers collections of beautiful clothes, accessories & homeware. All designed or handpicked by Le Sirenuse owner, Carla Sersale.
store.emporiosirenuse.com/collections/women

406 posts

3,211 followers

916 following

EMPORIO SIRENUSE
ONLINE STORE

Emporio Sirenuse
@EmporioSirenuse

Tweets 396 Following 447 Followers 210 Likes 171

Tweets Tweets & replies Media

Emporio Sirenuse @EmporioSirenuse · Feb 27

EMPORIO SIRENUSE
ONLINE STORE

Emporio Sirenuse Positano -
Boutique & On-line Store

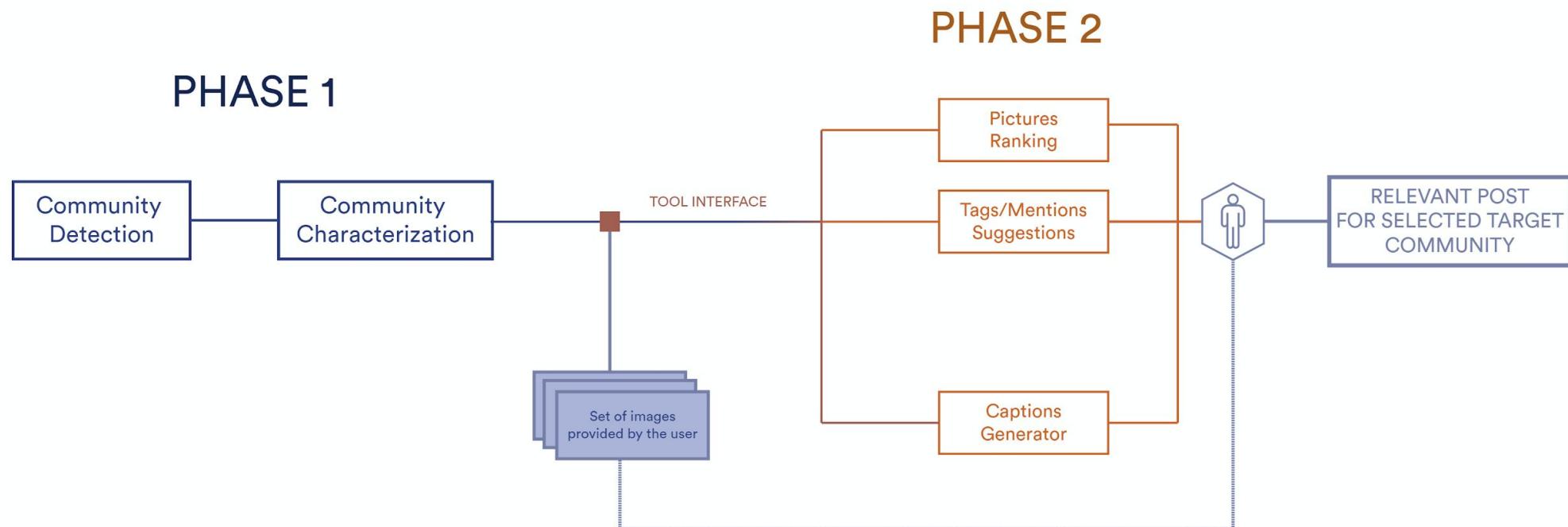
Le Sirenuse Positano began as one Emporio opposite the elegant hotel of Le Sirenuse. With every item designed or handpicked by Carla Sersale

3 followers

FOLLOW

Solution

“What do I post on the brand’s Instagram feed which will make it relevant and popular amongst its audience?”



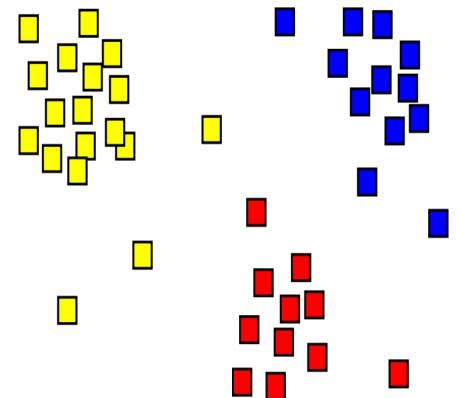
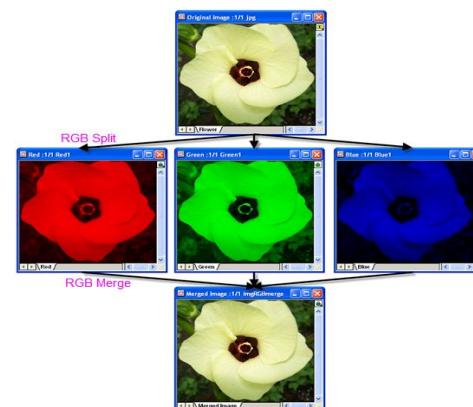
System

- 01** We analyse the communities (competitors and users) to create content clusters.
- 02** User inputs a set of images and selects a community to target.
- 03** Ranked list of images returned ordered by engagement and similarity to desired community.



Text & Image Clustering

- To identify latent communities, we used clustering approaches such as K-Means, GMM and LDA.
- The results helped us see naive categories of topics that are usually discussed by users.



Results

countryproperties countryestate home
staycation londonproperty
countryandtownhouse listed periodproperty
interiors dreamhome available
highclasshomes townhouse honeymoon
countryside countrylife luxury
countryhome farmhouse london
homesweethome realestate propertyoftheday
property luxuryrealestate houses
familyhomes villa
househunting greatbritishhomes

WordItOut

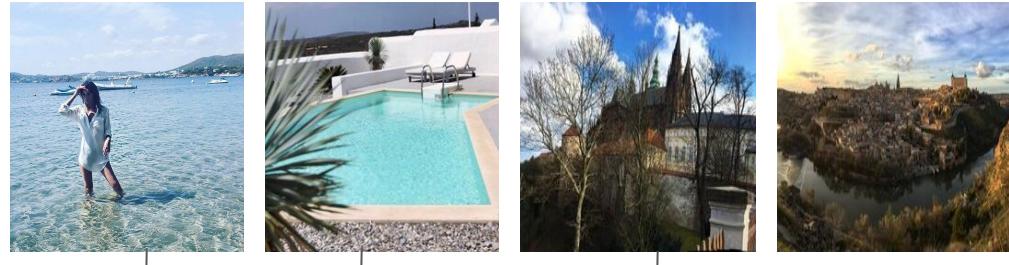
paris shipping weekend luxury whatsapp
link fashion travel morning
repost summer fall shop today
week time wearing good
best new life night like
hair amazing collection
october day beauty make
available world little
just thank love luxurydefined
beautiful happy style die

WordItOut

streetlights makeup nyfw fashion
kendalljenner lanadelrey capsule
harrystyles beauty denim
justinbieber blackfriday gigihadid
style startupbusiness to branding
dance summer kurtcobain
smallbusiness blogger travel
midnight mileycyrus lotd beautyblogger
trilluminati bellahadid socialmedia
cross startup digitalpr

vegan simple hot posting did
styles wanted fishtail
flannels buy super think
went pic updo curls flannel half
new got song fav braid hair
decided jeans recreate weather
recreating pretty comment waterfall
guys ootd farmers make
hairstyles change pictures
money awhile youtube
actually love

WordItOut



What are we missing? Context.

- The naive solutions do not include any information about the features in the image/text.
- We do not account for correlation between text and image.
- The features themselves are too simple to handle the stochasticity of social media data.

Combine Information



Object Detection

Person	Ski Blades	Hill
1	1	1

*Mask RCNN

Image Captioning

"A person skis off a ramp to perform a stunt"

*Flickr Image Captioning

Post

Caption : "She can fly !!!"

Hashtags : None

Likes : 27

*User Input

Combine Information

Object Detection

Person	Ski Blades	Hill
1	1	1

*Mask RCNN

Scene Detection

“A person skis off a ramp to perform a stunt”

*Flickr Image Captioning

Post

Caption : “She can fly !!!”

Hashtags : None

Likes : 27

*User Input

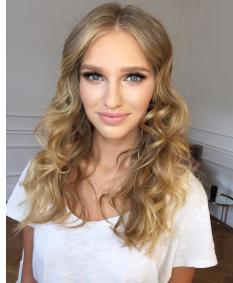


[1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0]

Final Vector Representation

With that, cluster posts with PAM!

Cluster 1



Cluster 2



Cosmetics!

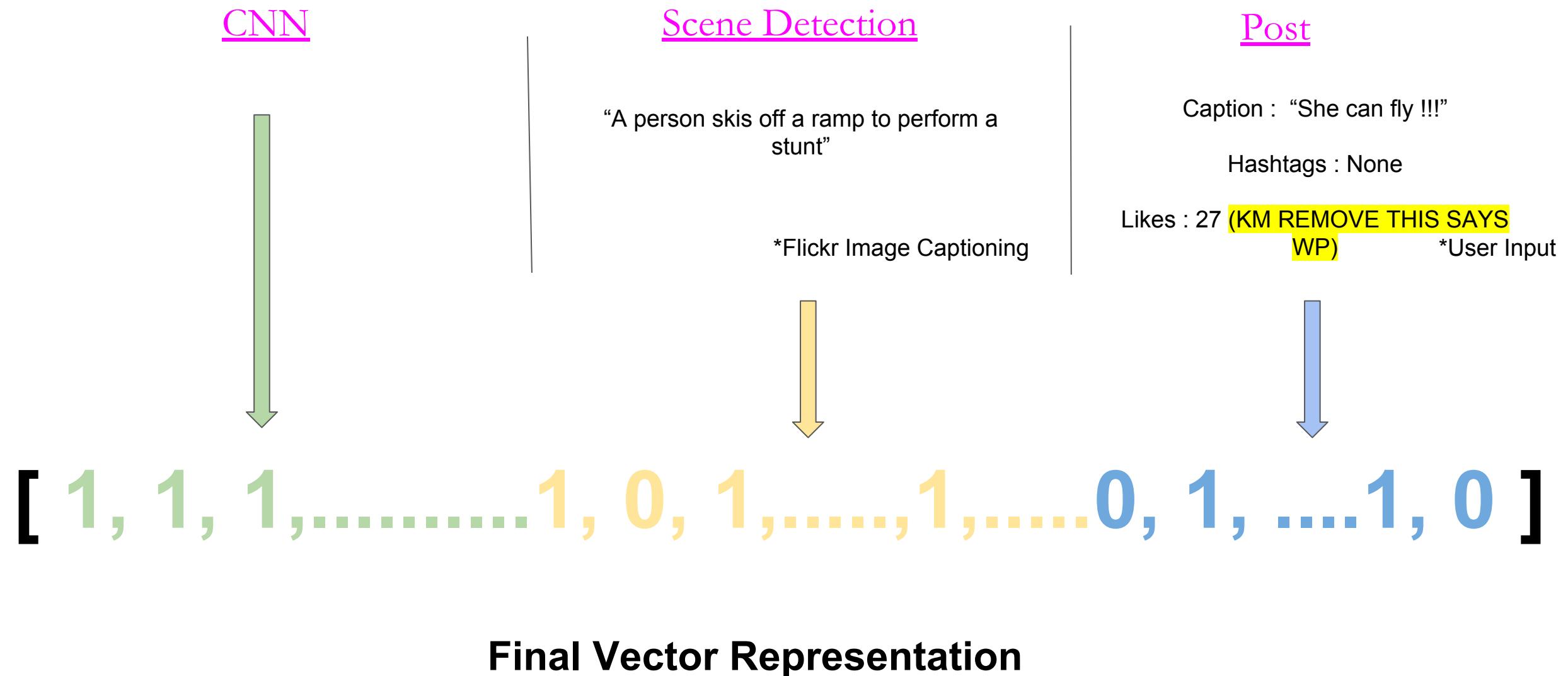
Modelling!

Lifestyle!

Moving forward...

- Inferences:
 - While intuitive, the validation of the techniques is hard without ground truth.
 - Moving forward, a curated dataset will be used for validation.
- Techniques to try:
 - Using Convolutional Neural Networks for recognizing Latent Image features.
 - Try more advanced clustering algorithms.
 - Try embedding Word+Image, together or separately, into a latent space for a more effective representation.

Combine Information



Progress?

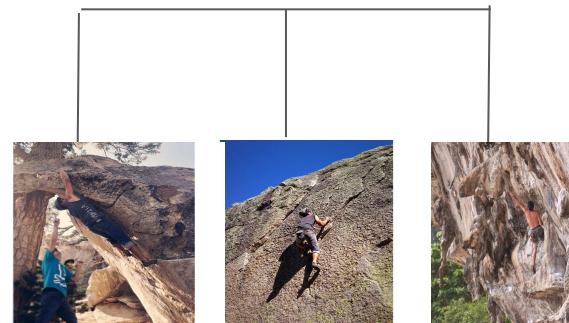
- Initial hiccups:
 - While complex and accurate, the network needs to be trained on a GPU cluster since over 50x speedup was observed.
 - We can't directly store large data on the JupyterHub, script to link S3 AWS and Jupyter needed to be written.
 - To validate future performance, curated dataset generated of different image groups scraped from instagram. 10K images of 4 categories.

Images with Captions!

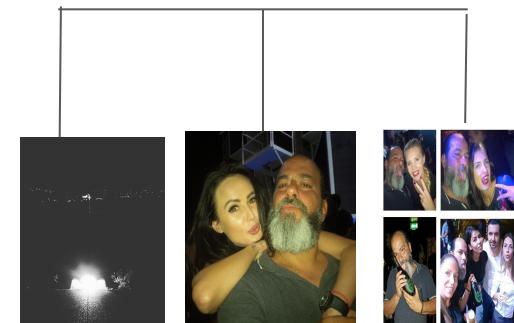
Category 1



Category 2



Category 3



Category 4



#Cosmetics

#RockClimbing

#Nightlife

#FoodPorn

Short Term Goals:

- Regularizing the network to prevent overfitting. Currently Training Accuracy : 100%, Testing Accuracy : 60%
- Taking Latent Representation from intermediate layers for Image Representation to yield better clustering.

Long Term Goals:

- Try Deep Embedding of Image and Text & Autoencoders for above Image Representation.

Ranking

PHASE 1



Community Detection

Community Characterization



TOOL INTERFACE

Set of images provided by the user

Kullback–Leibler divergence



PHASE 2

Pictures Ranking

Tags/Mentions Suggestions

Captions Generator



RELEVANT POST FOR SELECTED TARGET COMMUNITY

Pictures uploaded by user



$p = [0.6 \ 0.4, \ 0.0]$



$p = [0.3, \ 0.7, \ 0.0]$

Target Community

Community 1

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.99	0.0	0.01

Step 1:

The user (the social media manager, or Carla) uploads images taken from their photoshoot

Step 2:

The user identifies the target community he/she would like to target

Each of these (the uploaded images, and the clusters) has a distribution, as it is made up of images that have been passed through the clustering algorithm. Therefore, you can use KL Divergence to compare the distributions

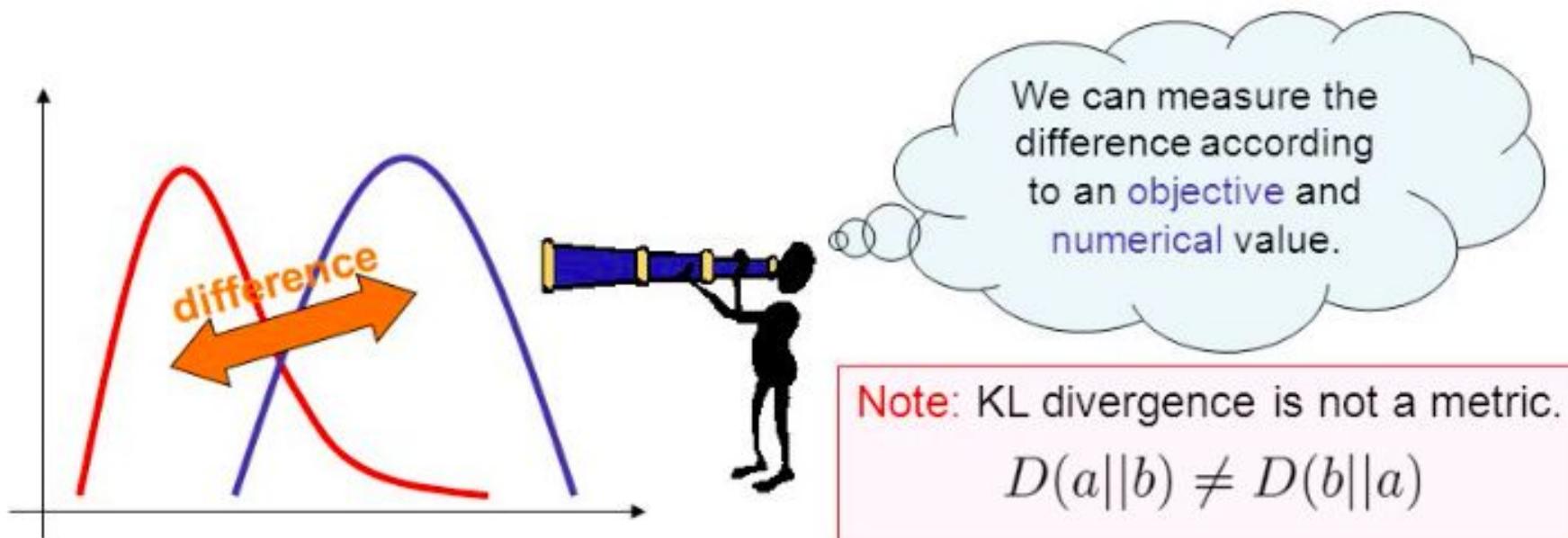
Community 2

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.4	0.0	0.6

KL Divergence

- A **measure** of the difference between two probability distributions: $a(x)$ and $b(x)$

$$D(a||b) \equiv \int dx a(x) \log \left(\frac{a(x)}{b(x)} \right)$$



KL Divergence (objective)

- KL divergence between the two distributions

$$\begin{aligned} D(r||P) &= \int dx r(x) \log \left(\frac{r(x)}{P(x|\theta)} \right) \\ &= \underbrace{\int dx r(x) \log r(x)}_{\text{Constant: independent of parameter } \theta} - \underbrace{\int dx r(x) \log P(x|\theta)}_{\text{}} \end{aligned}$$



Constant: independent
of parameter θ

To minimize KL divergence, we have only
to maximize the second term with respect
to the parameter θ .

Pictures uploaded by user



$p = [0.6 \ 0.4, \ 0.0]$



$p = [0.3, \ 0.7, \ 0.0]$

Target Community

Community 1

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.99	0.0	0.01

Community 2

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.4	0.0	0.6

Pictures uploaded by user



$p = [0.6 \ 0.4, \ 0.0]$



$p = [0.3, \ 0.7, \ 0.0]$

Target Community

Community 1

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.99	0.0	0.01
Community	Rockclimbing	Nightlife	Foodporn

KL Scores

	Community 1	Community 2
2.1		15.9
7.8		4.8

Calculate the KL divergence.
Rank from lowest to highest

Community 2

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.4	0.0	0.6
Community	Rockclimbing	Nightlife	Foodporn

Pictures uploaded by user



$p = [0.6 \ 0.4, \ 0.0]$



$p = [0.3, \ 0.7, \ 0.0]$

Target Community

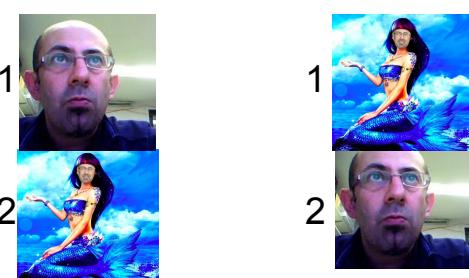
Community 1

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.99	0.0	0.01

KL Scores

	Community 1	Community 2
	2.1	15.9
	7.8	4.8

R
a
n
k



Community 2

Community	Rockclimbing	Nightlife	Foodporn
Probability distribution	0.4	0.0	0.6

Inferences, Improvements

1. Cluster Evaluation -

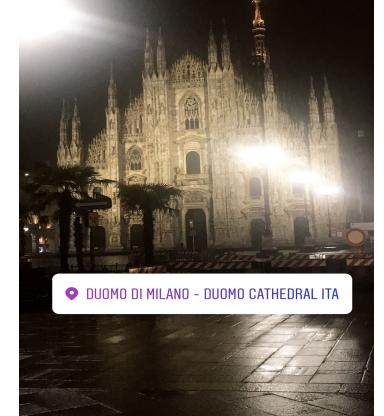
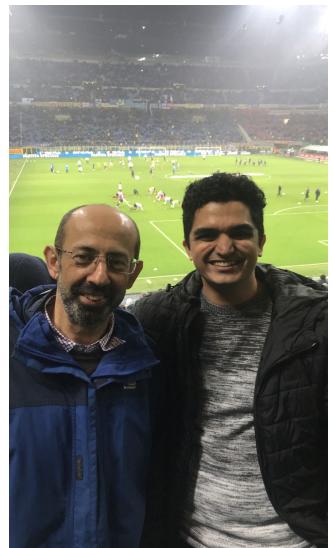
- a. Are the clusters meaningful? Clearly defined? How do we qualitatively assess the clusters (as we are already quantitatively doing so)

2. Ranking for the Community *and* for Engagement

Kimia's to-dos for the next week

1. Wrap up Recommendation System
 - a. How to balance KL divergence (low number) with Engagement (high number) for a final score?
2. Modularize
3. Add tests
4. Gather evaluation data with Moreno, to make more robust clusters

Also, we went to Italy.



DUOMO DI MILANO - DUOMO CATHEDRAL ITA

Thank you 😊

