Memes

Film
Captain America: The Winter Soldier
Marvel Studios, 2014

Commercial
We Believe: The Best Men Can Be
Gilette, 2019

Videogame
Detroit: Become Human
Sony Interactive Entertainment, 2018

Visualizing and Ranking the Influence of Users who Post *Memes* in Social Networks

Carles Pérez Onielfa

Trabajo Fin de Grado

Doble Grado de Matemáticas e Ingeniería Informática

Supervisors:
Sergio Escalera, Carles Casacuberta

Contributions

1. The definition of a **graph** for visualizing relationships of influence between users who post memes.

2. **Metrics** that rank users concerning meme virality.

3. A **process** for building the graph as mentioned earlier with a new approach to meme detection.

4. A **demonstration** of how this process can be applied to a set of users.

Objectives

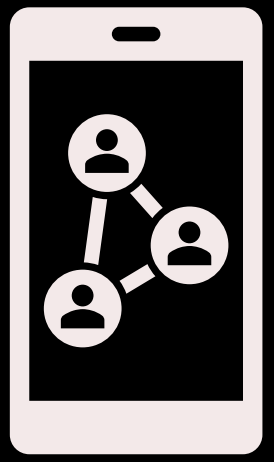
- **Creating** a pipeline for extracting and storing data from a social network.
- **Implementing** algorithms for detecting memes.
- **Extracting** features from images using a Convolutional Neural Network.
- **Clustering** images based on their underlying images.
- **Interpreting** the results of applying our methodology to a set of users.

Hypothesis

*By building the **graph** and computing the defined **metrics**, we will be able to visualize relationships between users and rank them regarding the potential of their memes to become viral.*

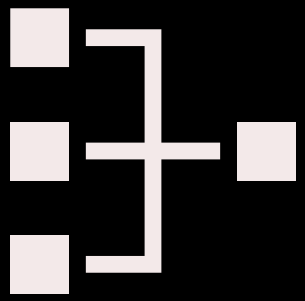
We believe some of these users would be overlooked by using standard metrics for determining influence.

Preliminary Concepts



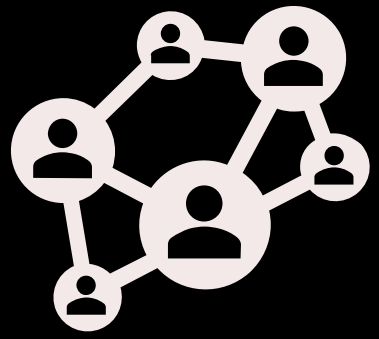
Social Networks

Instagram



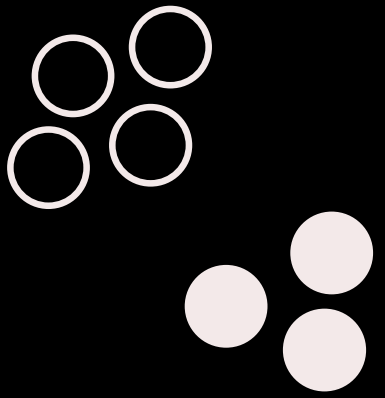
Neural Networks

CNNs, Transfer learning…



Graphs

Directed Graphs, Weighted Graphs, PageRank…



Clustering

DBSCAN

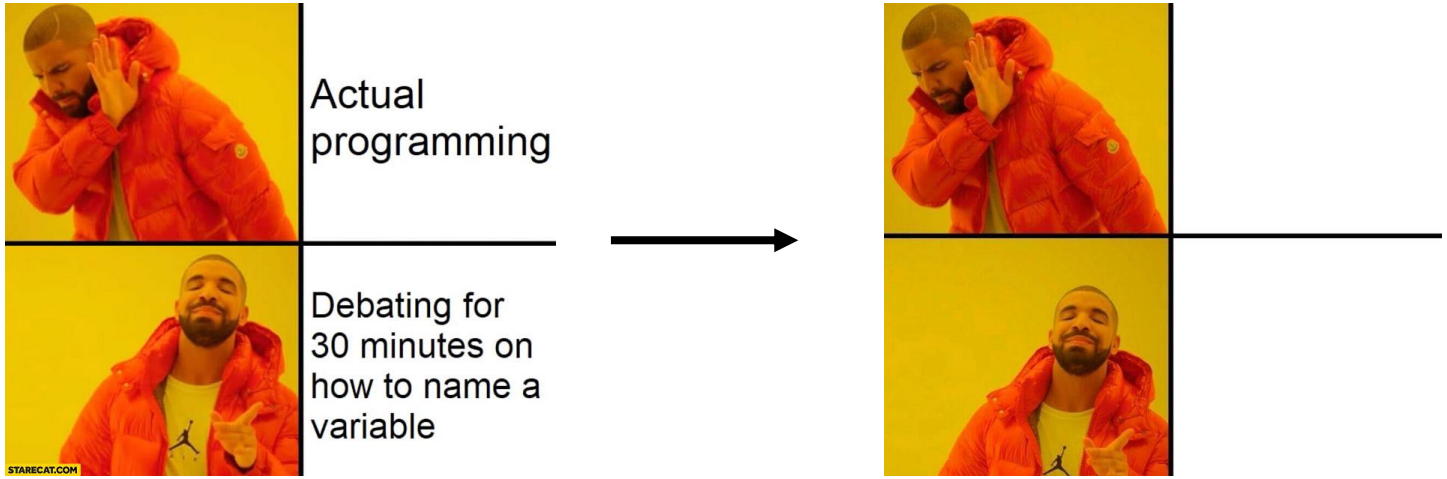
Inpainting

Filling in deteriorated parts of an image

Formalization of the Problem

A *meme* is a virally transmitted image embellished with text, usually sharing pointed commentary on cultural symbols, social ideas, or current events.

Its *meme format* is the underlying image of the composition.



Meme

Meme Format

Formalization of the Problem

The *pioneer* of a meme format F within a set of users U is the user in U who published the oldest meme with the format F.

If the set of users is the set of all users on all social media platforms, we refer to the pioneer as an *absolute pioneer*.

Formalization of the Problem:

The meme influence graph

Given a set of users $U = \{u_0, \dots, u_n\}$, we define the *meme influence graph* of U as a directed weighted graph (U, E, w) with the following properties:

- A pair (u_i, u_j) with $i, j \in \{0, \dots, n\}$ and $i \neq j$ is in the set E of edges if the user u_j has posted a meme whose meme format was pioneered by u_i.
- $w(u_i, u_j)$ is the number of memes posted by u_j whose format was pioneered by u_i.

A meme influence graph is called *maximal* if every user in every social media platform is in the set of users.

Formalization of the Problem:

Metrics

- **Out-degree** for a user *u* on a
  - Directed Graph is the number of outgoing edges of *u*.
  - Meme Influence Graph is the number of users who have used a meme format pioneered by *u*.
  - Maximal Meme Influence Graph is the number of users who have been influenced by *u*.

- **Weighted out-degree** for a user *u* on a
  - Directed Weighted Graph is the sum of the weights of the outgoing edges of *u*.
  - Meme Influence Graph is number of memes published by other users who have used a meme format pioneered by *u*.
  - Maximal Meme Influence Graph is the number of memes that have been influenced by *u*.

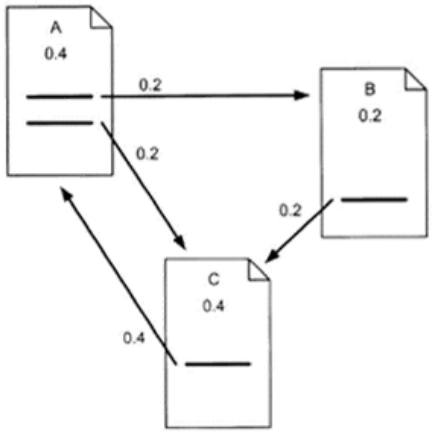
The **in-degree** and **weighted in-degree** are analogous for incoming edges.

Formalization of the Problem:

Metrics

Score: Value that the reverse PageRank algorithm assigns to a user.

PageRank[1] : A vertex is important if other important vertices link to it.



Reverse PageRank: A vertex is important if it links to other important vertices.

[1] Lawrence Page et al. The PageRank citation ranking: Bringing order to the web. Tech. rep. Stanford InfoLab, 1999.

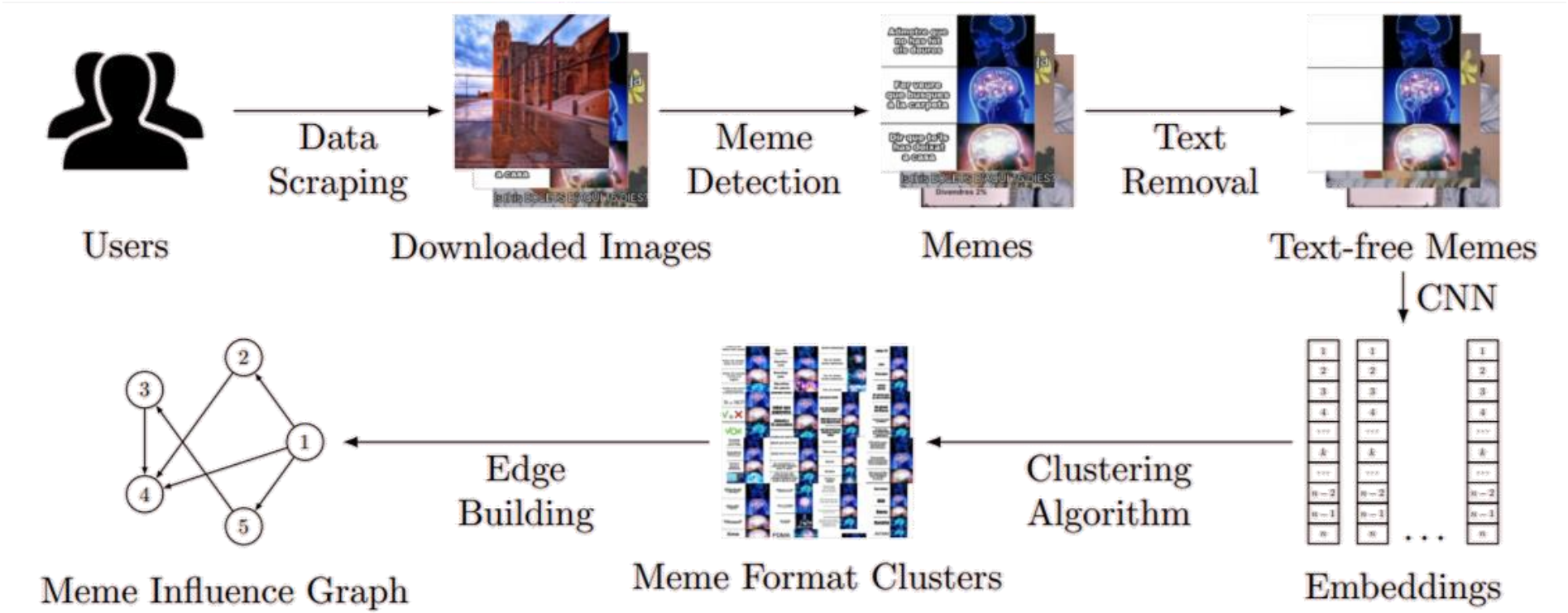
Formalization of the Problem:

Interpretation of the Metrics

Influence from a pioneer to a user is **indirect**

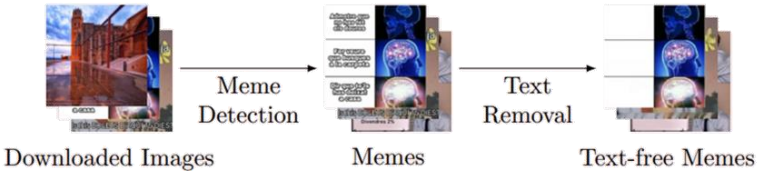
If the influence graph is not maximal → **potential influence**

Implementation



Implementation:

Meme Detection



Original Image

Detect Text using CRAFT[2]

$r \notin (\alpha, \beta)$

Image is not a meme

$r \in (\alpha, \beta)$

Compute r
(text-to-imatge area ratio)

Inpaint text areas with Navier-Stokes[3]

$\sigma < \gamma$

Image is not a meme

$\sigma \geq \gamma$

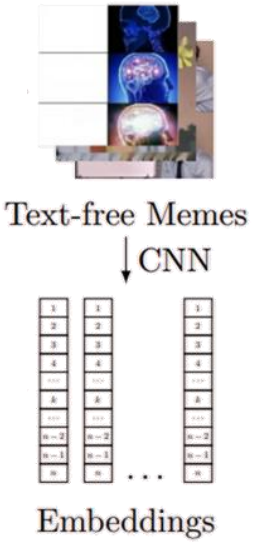
Image is a meme

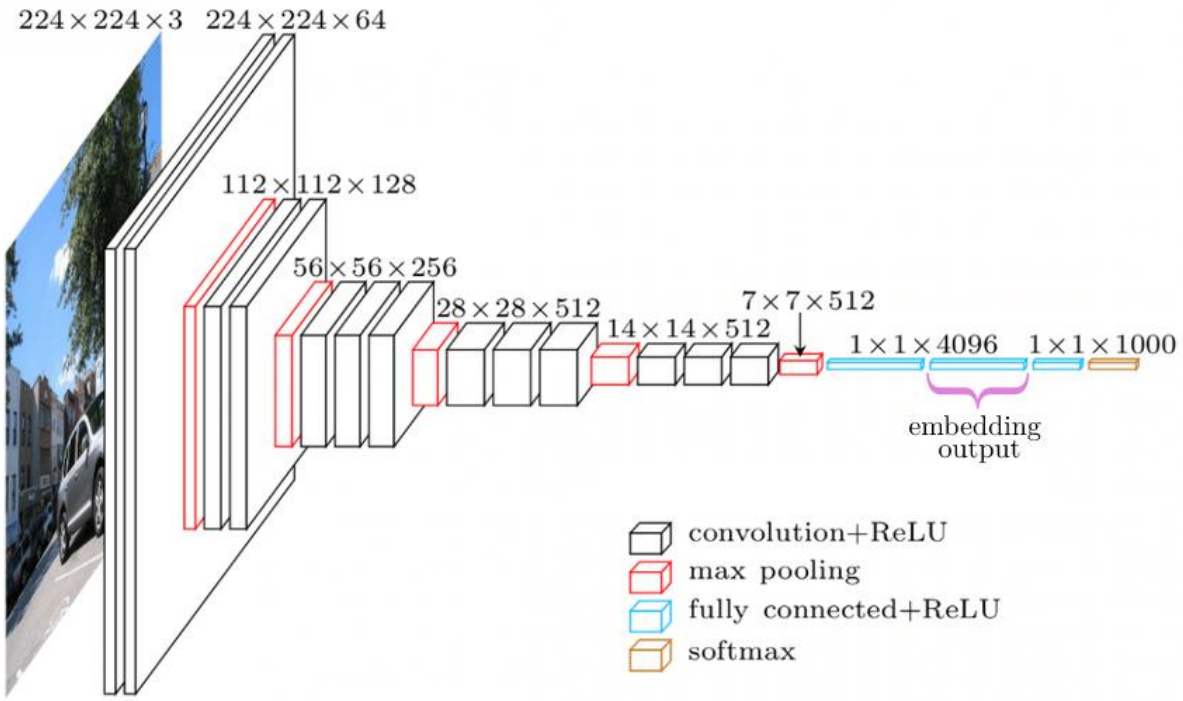
Compute σ
(standard deviation of gray values of the inpainted image)

[2] Youngmin Baek et al. “Character region awareness for text detection”. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 9365–9374.

[3] Marcelo Bertalmio, Andrea L Bertozzi, and Guillermo Sapiro. “Navier-Stokes, fluid dynamics, and image and video inpainting”. In: 2001 IEEE Computer Society Conference on CVPR. Vol. 1. IEEE. 2001, pp. I–I.

Implementation:
Embeddings



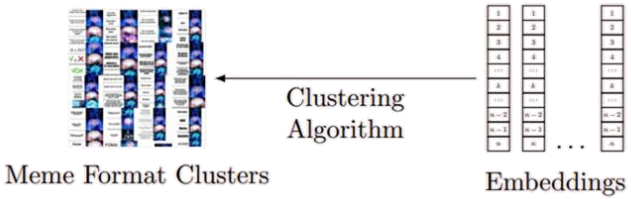


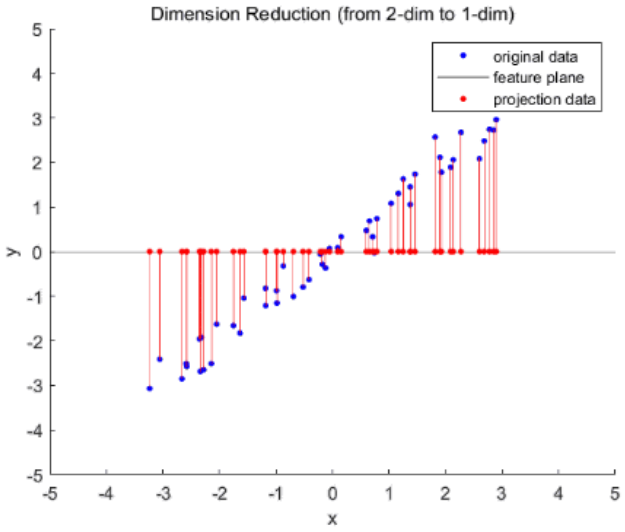
VGG16 [4] pretrained with ImageNet weights

Our output is highlighted in pink

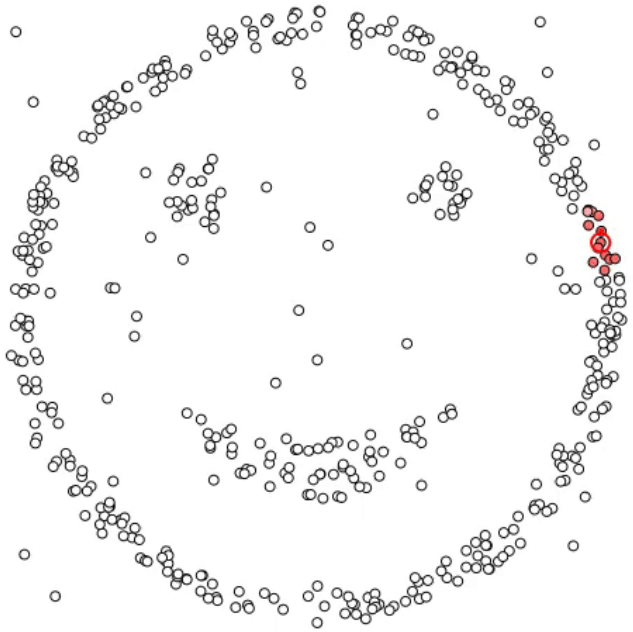
[4] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: arXiv preprint 1409.155 (2014).

Implementation:
Clustering





Principal Component Analysis
(PCA)
4096 dimensions to 1024 dimensions



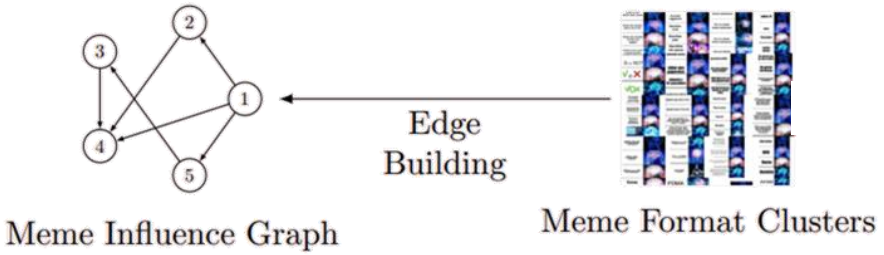
Density-Based Spatial Clustering of Applications with Noise [5]
(DBSCAN)

1. Does not require previous knowledge of the **number of clusters**.
2. Detects **outliers**.
3. Two parameters: **threshold distance** (ε), **minimum samples per cluster**.

[5] Martin Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: kdd. Vol. 96. 34. 1996, pp. 226–231

Implementation:

Meme Influence Graph



For each cluster, we create edges from the pioneer of a cluster to the authors of the rest of the memes of that cluster.

Results

Data Scraping

91 Users **447,101** Images

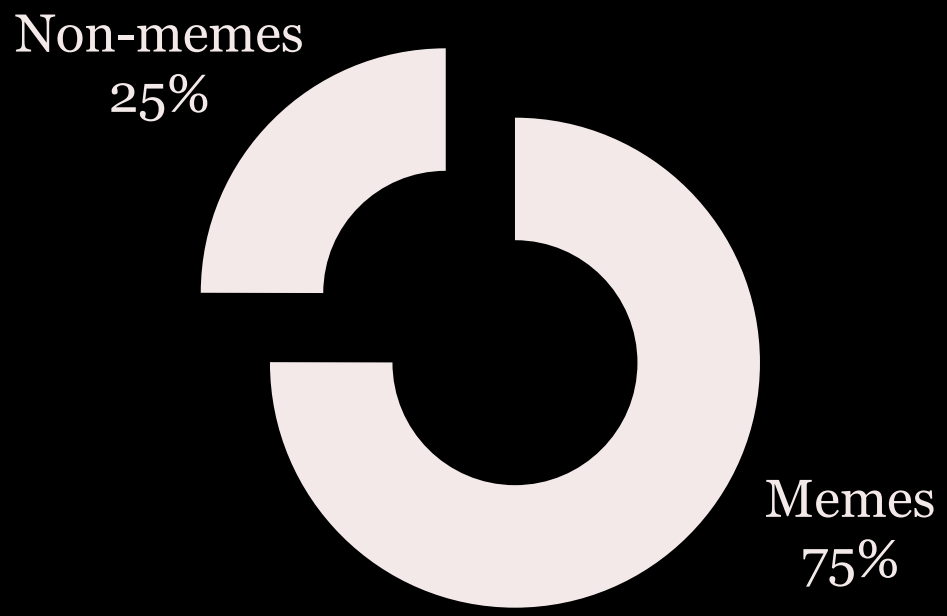
January 2017 – April 2022

Results:

Meme Detection

Parameters found for the meme detection algorithm:

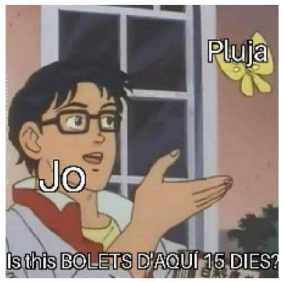
$$\alpha = 0.018,\ \beta = 0.4,\ \gamma = 26$$



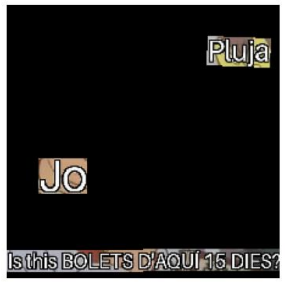
342,984 Memes

Results:
Meme Detection

Text-to-image ratio $r = 12.07\%$ | Standard deviation $\sigma = 44.50$
Is meme? True



Original



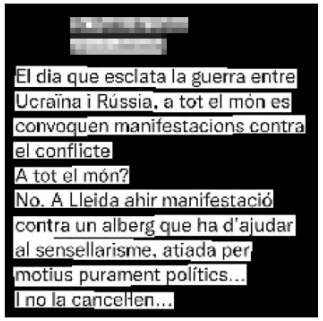
Text detection

Expanded mask

Result

Text-to-image ratio $r = 43.00\%$
Is meme? False

Original



Text detection

Text-to-image ratio $r = 35.43\%$ | Standard deviation $\sigma = 18.51$
Is meme? False

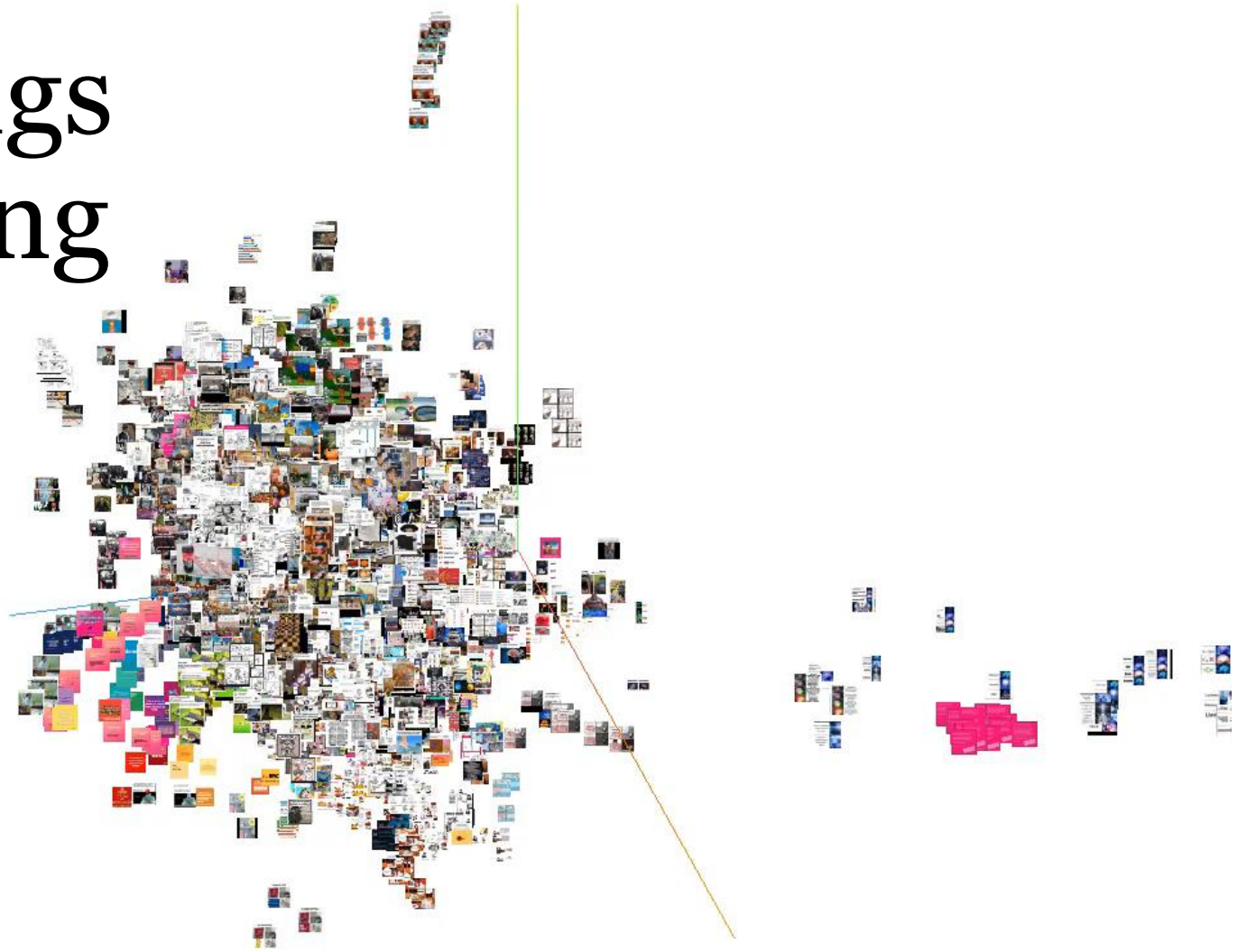
Original

Text detection

Expanded mask

Result

Embeddings & Clustering

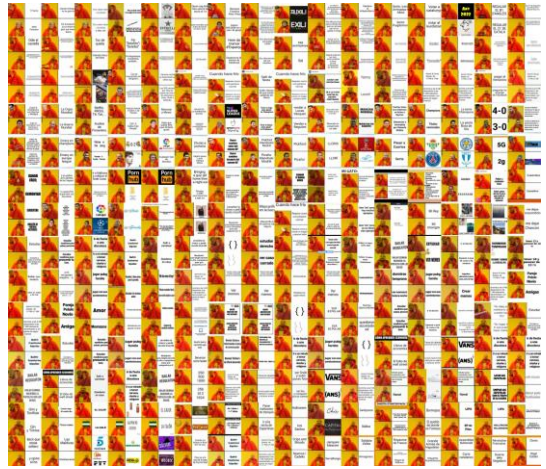


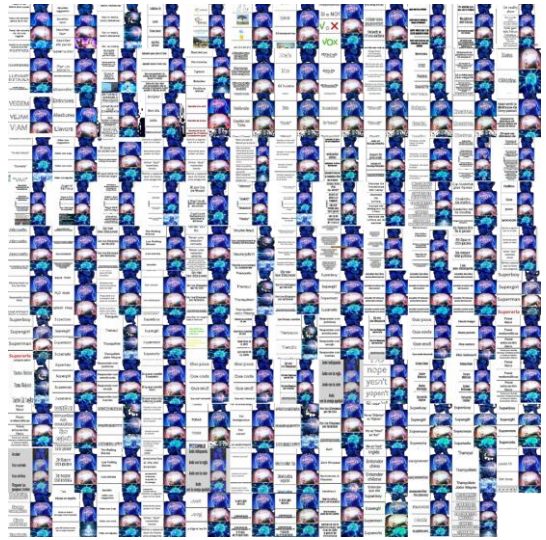
Clustering

113,663 Clusters

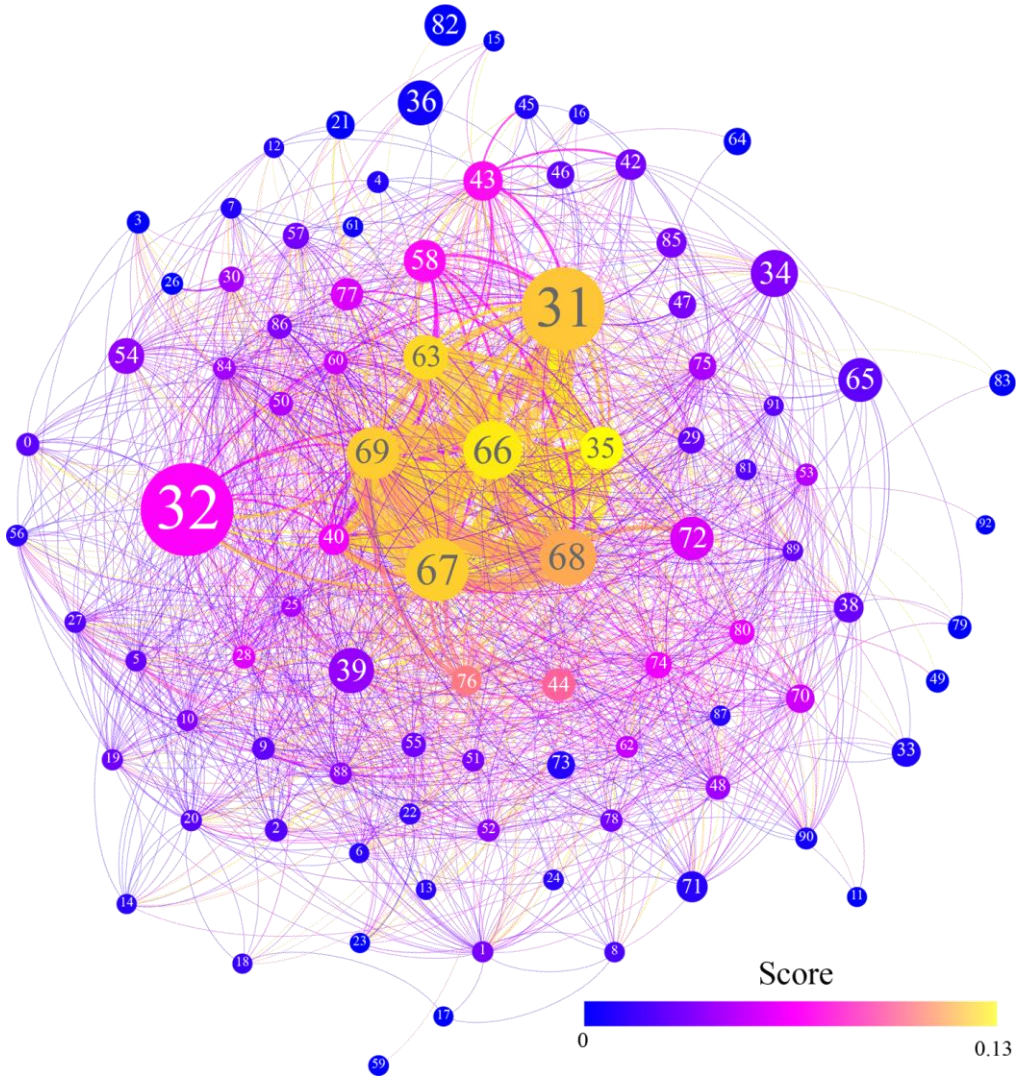
82,801 Memes (24.14%)

260,183 (75.86%) Memes labeled as noise





Influence Graph and Metrics



Correlation coefficient between metrics and variables in the dataset

	follower count	average likes	average comments	media count	meme count	meme ratio
out degree	0.30	0.05	0.19	0.58	0.66	-0.19
in degree	0.28	-0.01	0.12	0.59	0.63	-0.18
weighted out degree	0.30	0.10	0.22	0.73	0.80	-0.10
weighted in degree	0.27	0.08	0.16	0.85	0.91	0.01
score	0.29	0.08	0.19	0.69	0.79	-0.12
pioneered format count	0.36	0.13	0.25	0.88	0.93	-0.04
w out degree/w in degree	-0.04	-0.04	-0.02	-0.06	-0.04	0.08
w out degree/media count	-0.05	-0.12	-0.10	-0.03	0.04	0.05

Meme Formats - Ranking

1. Yes Chad

45 users (49.45%)

191 memes (0.05%)

2. Drake

34 users (37.36%)

238 memes (0.07%)

3. Galaxy Brain

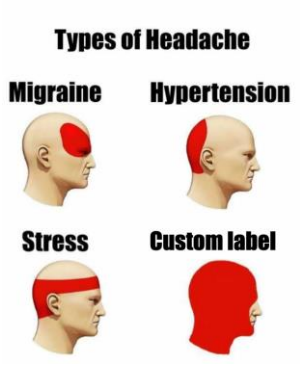
33 users (36.26%)

143 memes (0.04%)

4. Swole Doge

30 users (32.96%)

85 memes (0.02%)



5. Headaches

26 users (28.57%)

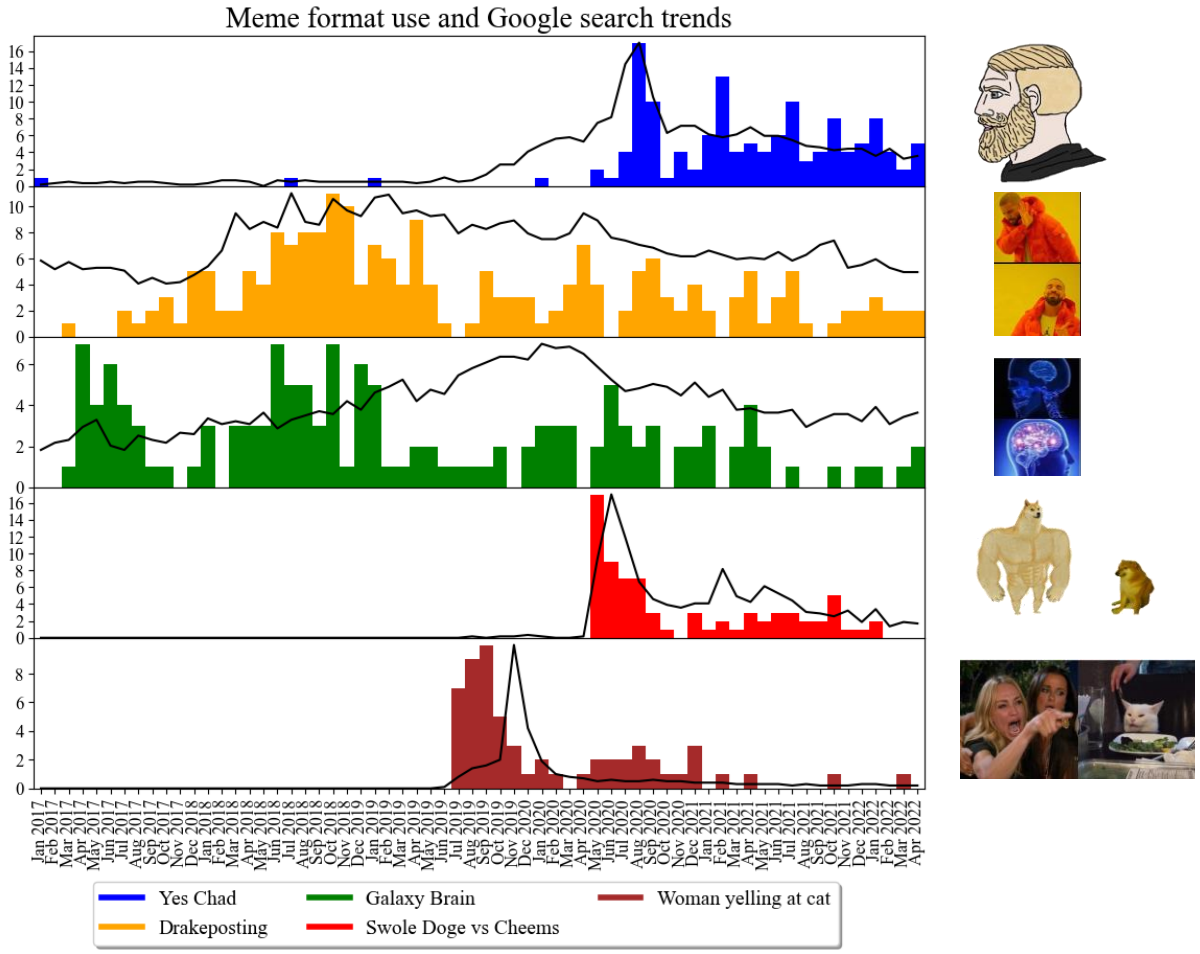
70 memes (0.02%)

6. Woman Yelling at Cat

25 users (27.47%)

76 memes (0.02%)

Meme Formats - Trends



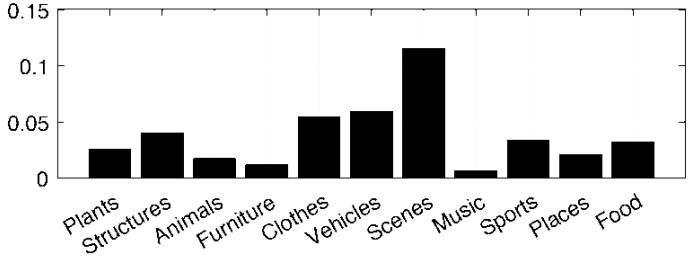
Conclusions

Our *hypothesis* has been **verified**.

- We have been able to:
  - ✓ **Formalize** our problem and define new concepts.
  - ✓ **Detail** a process for building and computing the definitions.
  - ✓ **Implement** a new meme detection algorithm.
  - ✓ **Extract** features with a Convolutional Neural Network.
  - ✓ **Cluster** memes according to their meme format.
  - ✓ **Interpret** the results of applying our methodology.

Related Work

Influence in Social Networks



Differences of social correlation between friends and non-friends [1].

Meme Detection and Clustering

- Memesequencer [2]
- Meme-Hunter [3]

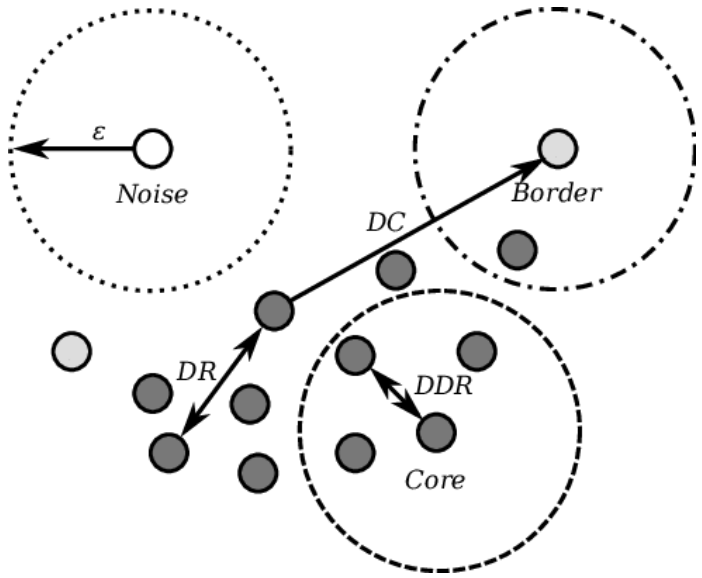
Meme-Hunter model features

[1] Quanzeng You et al. “Cultural diffusion and trends in facebook photographs”. In: *International AAAI Conference on Web and Social Media*. Vol. 11. 1. 2017, pp. 347–356.

[2] Abhimanyu Dubey et al. “Memesequencer: Sparse matching for embedding image macros”. In: Proceedings of the World Wide Web Conference. 2018, pp. 1225–1235.

[3] David M Beskow, Sumeet Kumar, and Kathleen M Carley. “The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning”. In: Information Processing & Management 57.2 (2020), p. 102170.

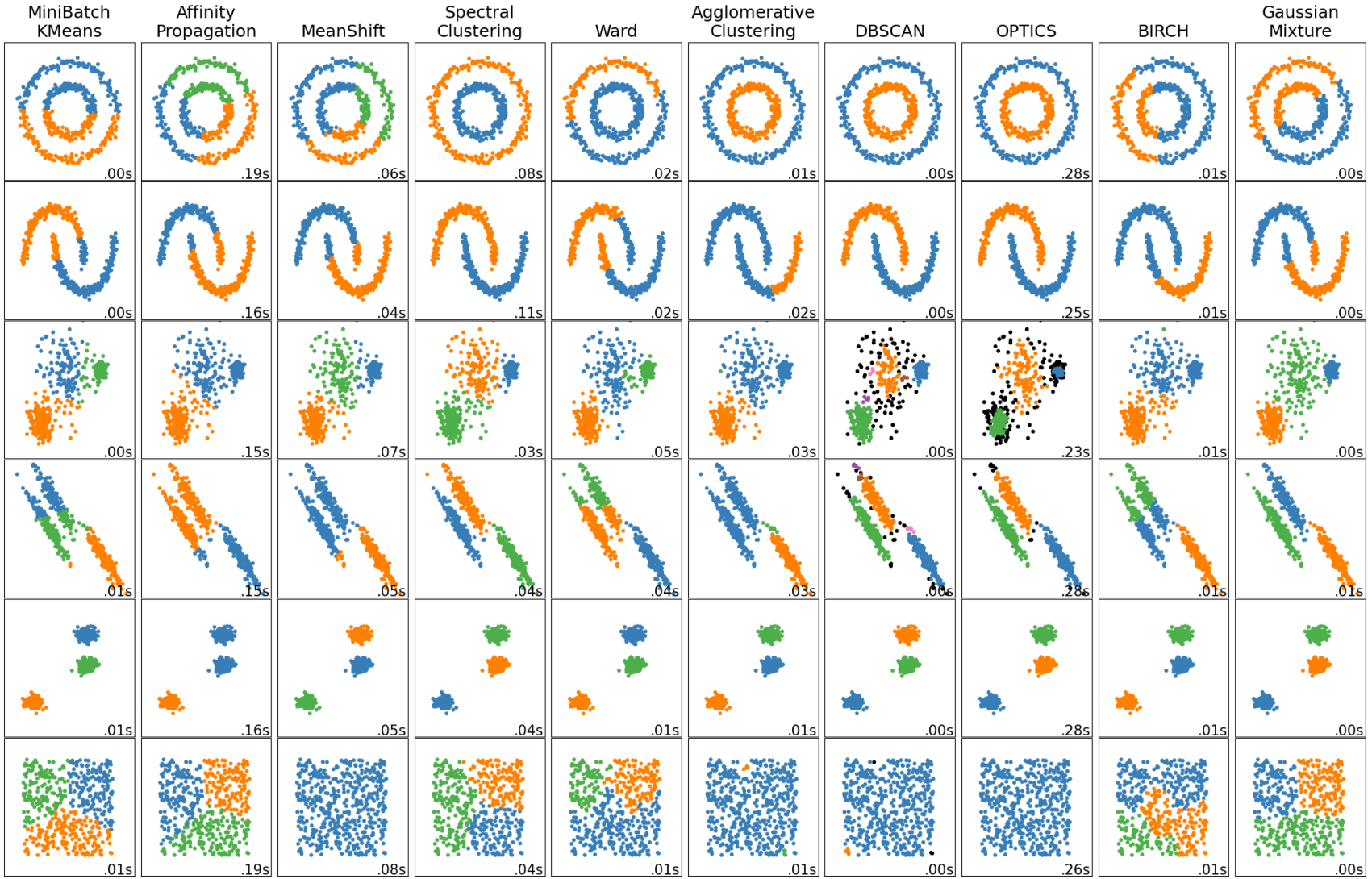
DBSCAN



1. A point p is a **core point** if there are more than *minPts* points within an ε radius.
2. A point q is **directly density reachable** (DDR) from a core point p if the distance between q and p is lower than ε.
3. A point q is **density reachable** (DR) from another point p if there is a path $p = p_o, p_1, \ldots, p_n = q$ where each $p_{i+1}$is directly density reachable from p_i.
4. Two points p and q are **density-connected** (DC) if there exists a point o such that p and q are reachable from o.
5. All other points are **outliers**.

*Clusters defined as groups of points where all the points are mutually **density-connected** and contain every **density-reachable** point from another point in the cluster.*

DBSCAN


MiniBatch KMeans
Affinity Propagation
MeanShift
Spectral Clustering
Ward
Agglomerative Clustering
DBSCAN
OPTICS
BIRCH
Gaussian Mixture
.00s .19s .06s .08s .02s .01s .00s .28s .01s .00s
.00s .16s .04s .11s .02s .02s .00s .25s .01s .00s
.00s .15s .07s .03s .05s .03s .00s .23s .01s .00s
.01s .15s .05s .04s .04s .03s .00s .28s .01s .01s
.01s .16s .05s .04s .01s .01s .00s .28s .01s .00s
.01s .19s .08s .04s .01s .01s .00s .26s .01s .00s