Meme Review
Dead or Alive?

Predicting Reader Response to Timely Memes using Machine Learning

Phase 2

Motivation & Problem Definition Recap

2 out 5 Facebook posts are memes

Memes are a great marketing tool

Millennials LOVE memes

Online Businesses and Influencers

Massive Data dump

Background

What we set out to learn

How the internet affects many variants of people

How internet memes gained and maintained popularity

How images are shared in internet media

How to automatically find classifiable features

How to cluster similar looking images

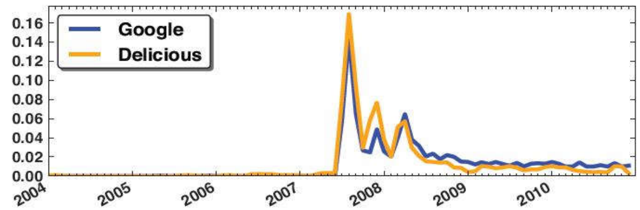
“Insights into Internet Memes”

It is commonly assumed that Internet memes spread virally but scientific evidence as to this assumption is scarce.

Investigate the epidemic dynamics of 150 famous Internet memes.

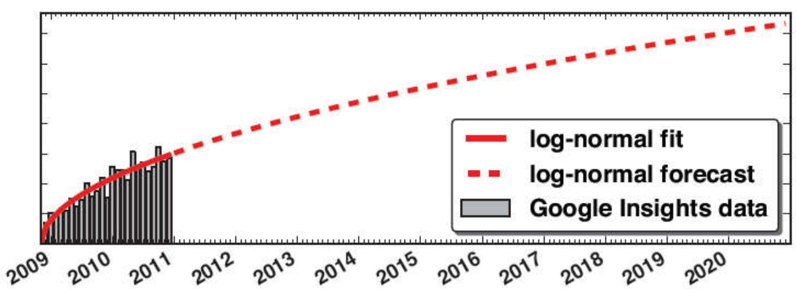
The analysis is based on time series data that were collected from Google Insights, Delicious, Digg, and StumbleUpon.

(a) instances of the "chocolate rain" meme



(b) two time series (retrieved from Google Insights and Delicious) reflecting the rise and decline in popularity of this Internet meme

Differential equation models from mathematical epidemiology + simple log-normal distributions = a good account of the growth and decline of memes.

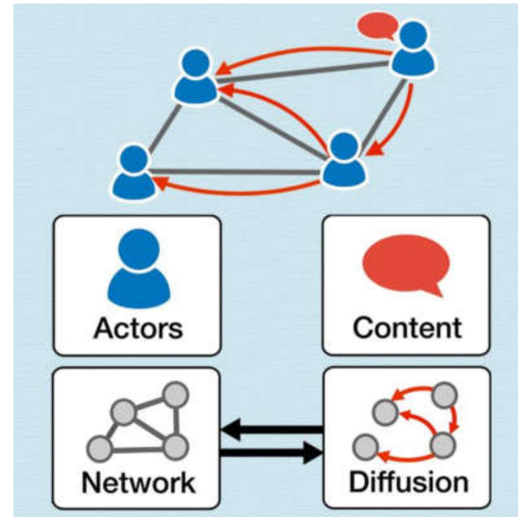


Traditional compartment models might be good but lack flexibility.

Log-normal distributions are more practical.

Thus, concluding by saying memes might be viral but only on social media platforms and not the rest of the internet.

"Predicting the Content Diffusion Path in Online Social Networks"



"Marketing with Twitter: Challenges and Opportunities"

255 million monthly active users

Specific features may have more overall effect in the final decision of a user response

Holds attention + memorable

Upvotes, Author, Dates

“A Study of Meme Propagation”

Memes gain momentum and then lose their popularity over time

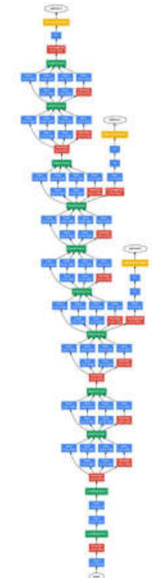
Categorized the Memes, i.e. Political memes vs General Memes

“Going Deeper with Convolutions”

- Google’s own research paper proposing the Deep Convolutional Inception Neural Net
  - By C. Szegedy & 8 others from Google Inc, University of North Carolina and
  - Published to Cornell’s ArXiv on the 17th of September 2014
- ImageNet Large-Scale Visual Recognition Challenge 2014 1st Place Winner for Classification and Detection.

Calculating Confusing Convolutions

type	patch size/ stride	output size	depth
convolution	7×7/2	112×112×64	1
max pool	3×3/2	56×56×64	0
convolution	3×3/1	56×56×192	2
max pool	3×3/2	28×28×192	0
inception (3a)		28×28×256	2
inception (3b)		28×28×480	2
max pool	3×3/2	14×14×480	0
inception (4a)		14×14×512	2
inception (4b)		14×14×512	2
inception (4c)		14×14×512	2
inception (4d)		14×14×528	2
inception (4e)		14×14×832	2
max pool	3×3/2	7×7×832	0
inception (5a)		7×7×832	2
inception (5b)		7×7×1024	2
avg pool	7×7/1	1×1×1024	0
dropout (40%)		1×1×1024	0
linear		1×1×1000	1
softmax		1×1×1000	0



Stats and Scores

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Table 2: Classification performance.

Team	Year	Place	mAP	external data	approach
UvA-Euvision	2013	1st	22.6%	none	Fisher vectors
Deep Insight	2014	3rd	40.5%	ImageNet 1k	CNN
CUHK DeepID-Net	2014	2nd	40.7%	ImageNet 1k	CNN
GoogLeNet	2014	1st	43.9%	ImageNet 1k	CNN

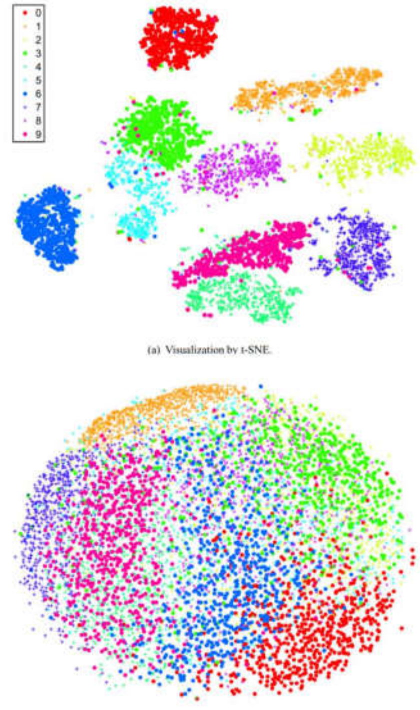
Table 4: Comparison of detection performances.

“Visualizing Data using TSNE”

- T-Distributed Stochastic Neighbor Embedding
- Published by Laurens van der Maaten of Tilburg University and Geoffrey Hinton of University of Toronto
- First published to The Journal of Machine Learning Research on the 9th of November 2008

Dimension Warping

- TSNE is a Dimensionality Reduction technique
- TSNE Projection can map and cluster High Dimensional Data to a 2D or 3D Space by representing noticeable features as vectors
- Similar to PCA (Principal Component Analysis)



(a) Visualization by t-SNE.

(b) Visualization by Sammon mapping.

Figure 2: Visualizations of 6,000 handwritten digits from the MNIST data set.

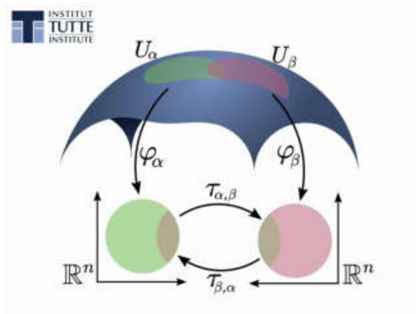
- Easily Separable Classifications
- Optimized for learning and preserving non-linear patterns in HD datasets
- Focuses on local structure rather than original global structure

“UMAP for Dimension Reduction”

- Uniform Manifold Approximation & Projection
- Published by Leland McInnes and John Healy from Tutte Institute for Mathematics and Computing, Canada
- Is a fairly recent technique, 13th February 2018, that is rivaling the otherwise state-of-the-art TSNE.

UMAP Ur Own Clusters

- Isn't actually about maps. Shocker I know...
- Processes neighbors in HD space.
- Connects clusters with other clusters with fancy topological math and neighbor graphs.
- This forms a manifold with no distortion making it uniform



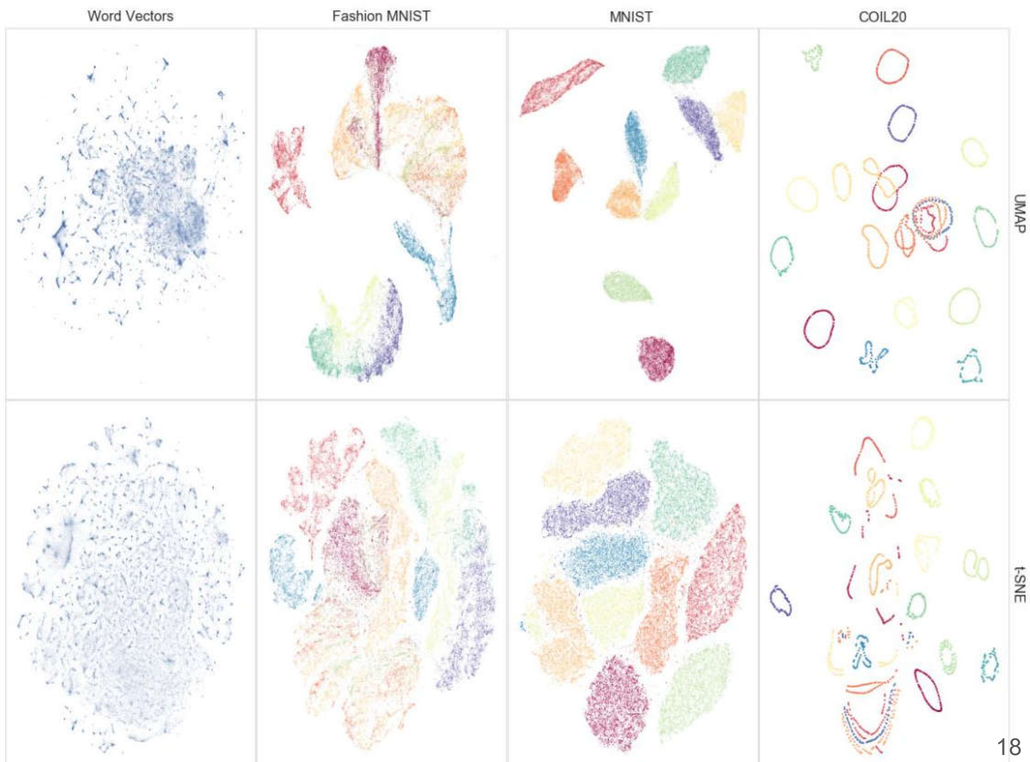
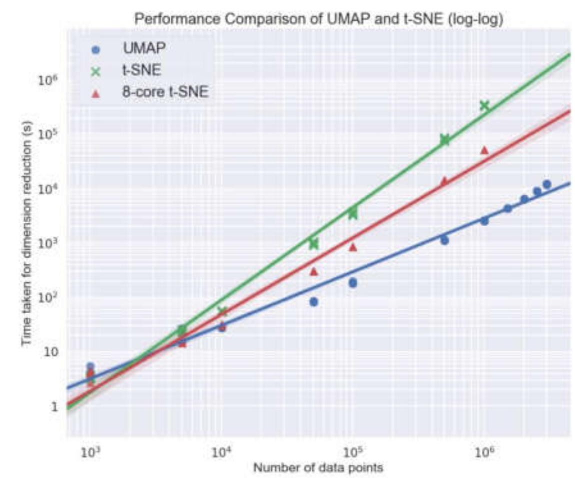


Table 1: Runtime of UMAP and t-SNE on various datasets

dataset	dataset size	t-SNE	UMAP
COIL20	1440x16384	20s	**7s**
COIL100	72000x49152	683s	**121s**
Shuttle	58000x9	741s	**140s**
MNIST	70000x784	1337s	**98s**
F-MNIST	70000x784	906s	**78s**
GoogleNews	200000x300	16214s	**821s**



Back to

Meme Review

What have WE done so far?

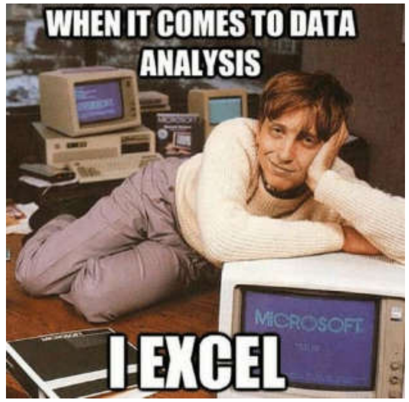
Data Source & Preprocessing Recap

Kaggle

Reddit meme dump

JSON metadata

- Image Preprocessing
  - Resize Resolution
  - Convert All images to JPG
- Process JSON metadata
  - Messy JSON Organization
  - UNIX-based UTC Timestamp



Neural Networks Recap

Inception

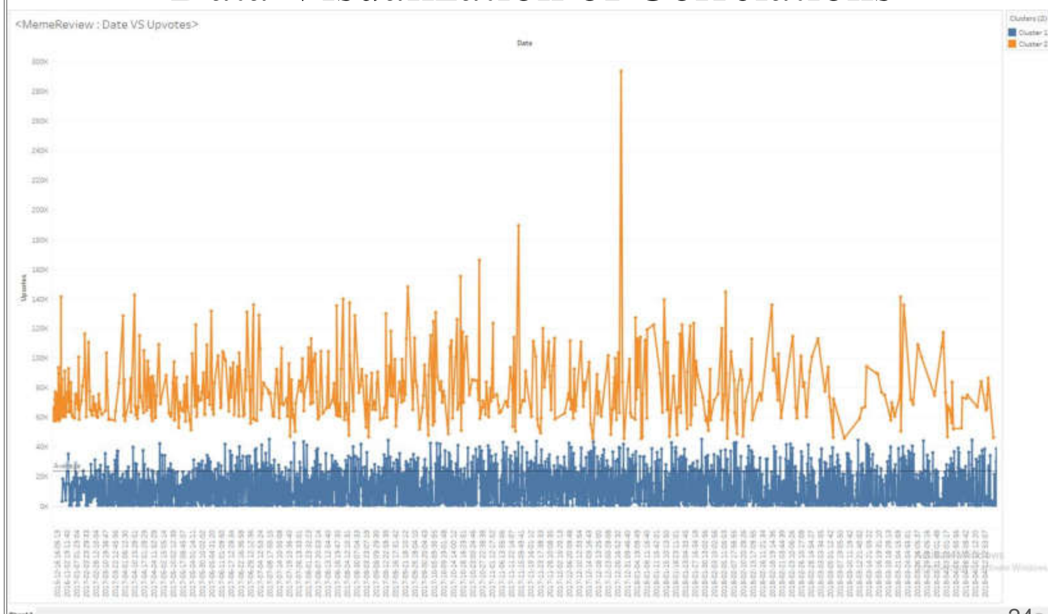
- Used before the TSNE Cluster to produce classifiable feature vectors

AlexNet

VGGnet

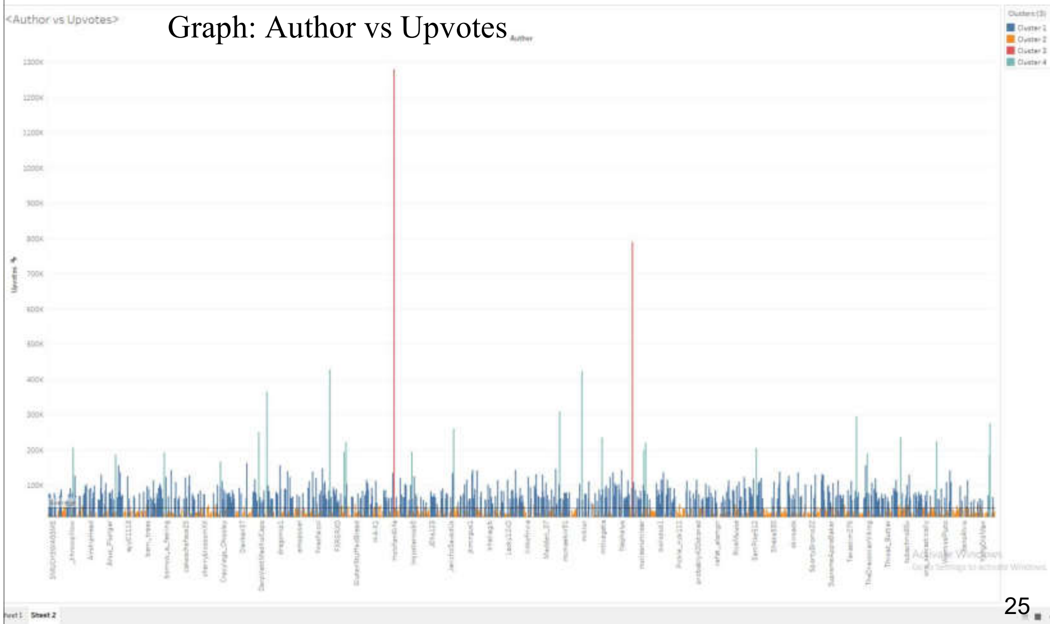
Data Visualization

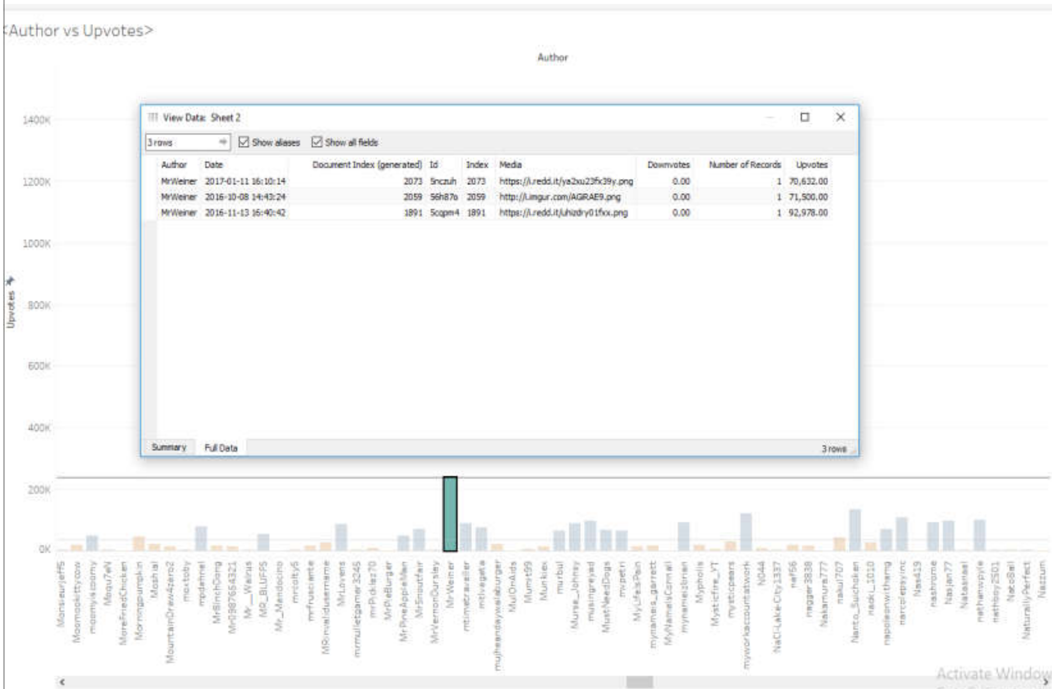
Data Visualization of Correlations



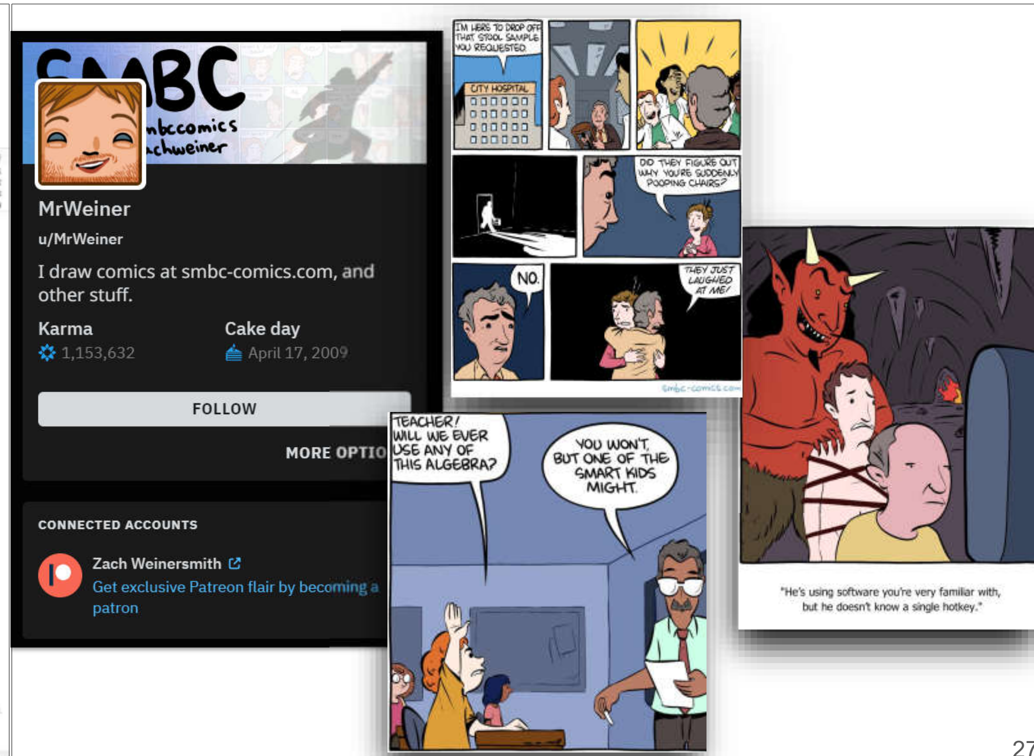
Data Visualization of Correlations

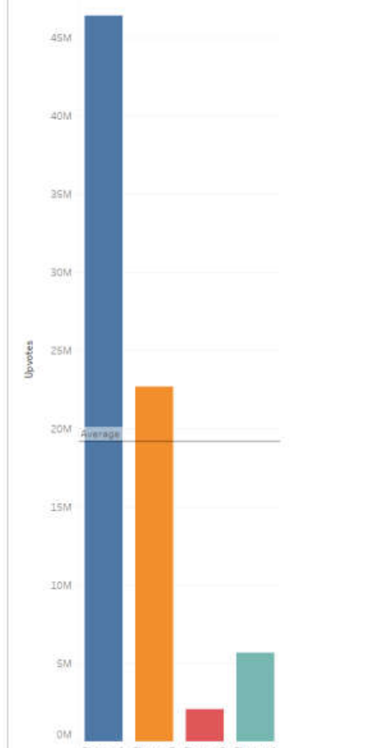
Graph: Author vs Upvotes





Author	Date	Document Index (generated)	Id	Index	Media	Downvotes	Number of Records	Upvotes
MrWeiner	2017-01-11 16:10:14	2073	5nczuh	2073	https://i.redd.it/ya2xu23fx39y.png	0.00	1	70,632.00
MrWeiner	2016-10-08 14:43:24	2059	56h87o	2059	http://i.imgur.com/AGRAE9.png	0.00	1	71,500.00
MrWeiner	2016-11-13 16:40:42	1891	5cqpm4	1891	https://i.redd.it/uhizdry01fxx.png	0.00	1	92,978.00





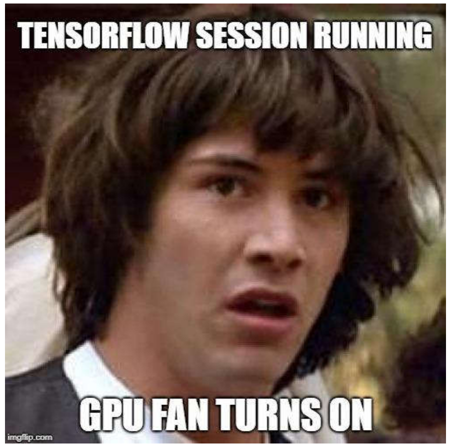
Author Clusters

4 Types of "Meme-creators"

See in Numbers

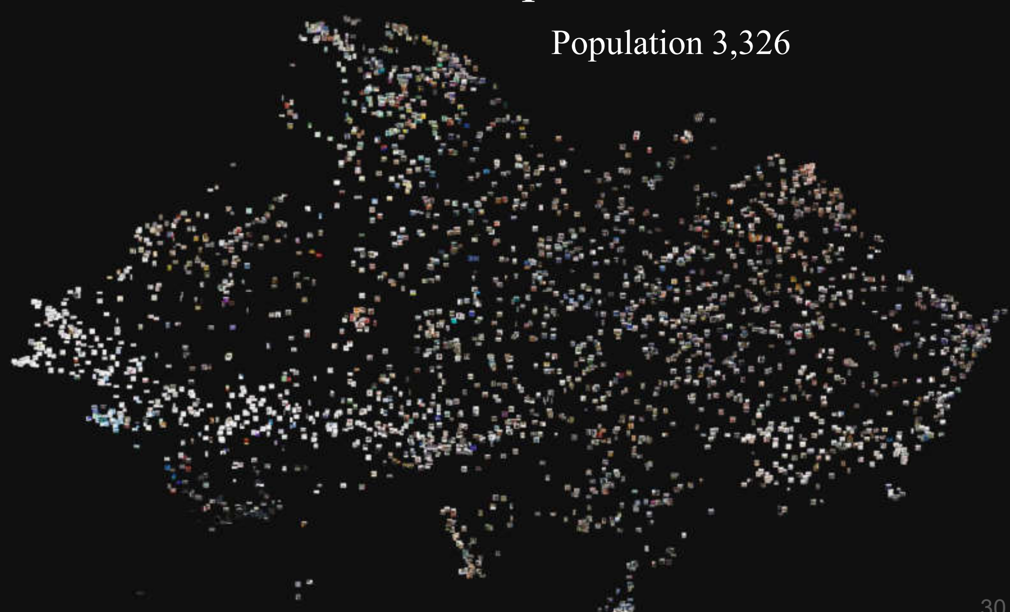
Produce image vectors from Inception Neural Net penultimate pool layer's weights using TensorFlow

Use previous Vectors for Clustering with UMAP then project onto an HTML Canvas running D3.js

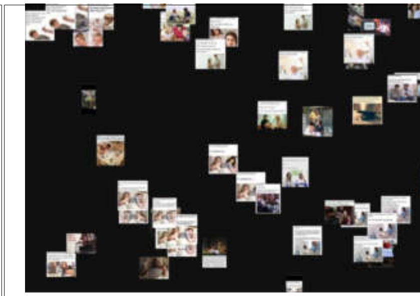


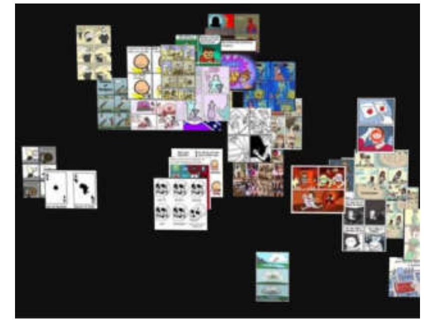
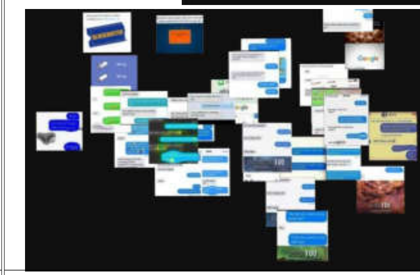
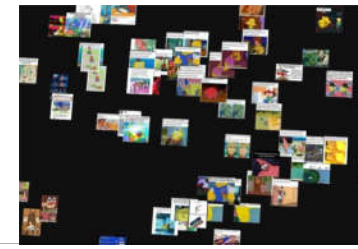
A UMAP Map of Memes

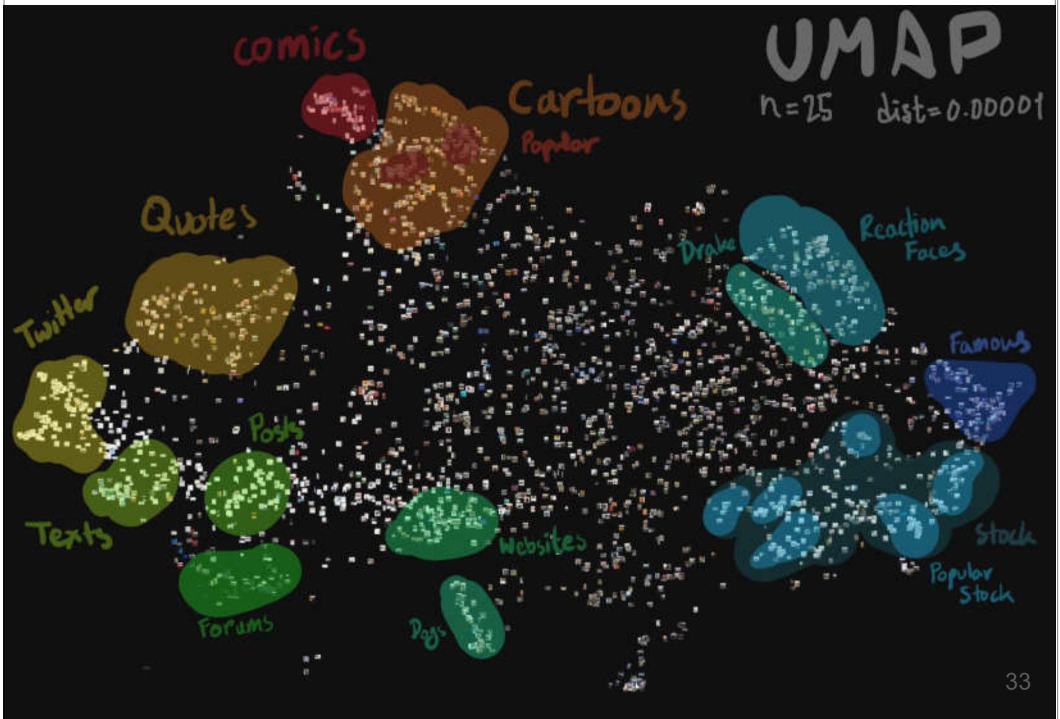
Population 3,326

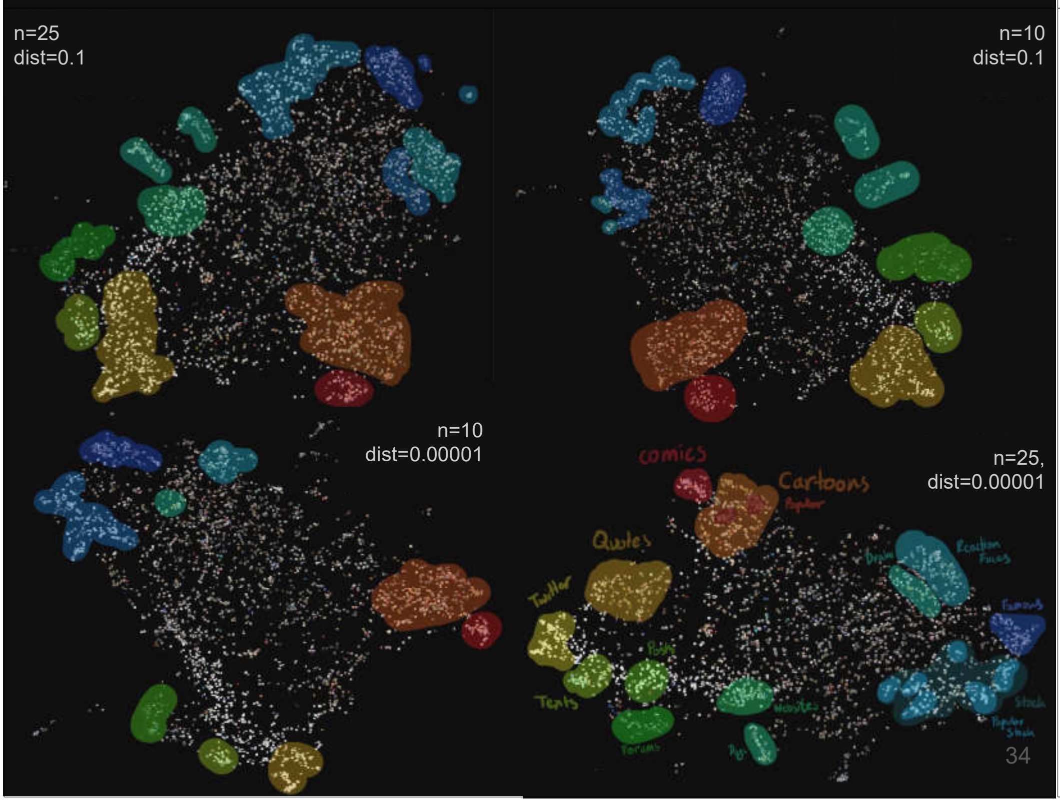


Clusters of Memes







Contributors

- **Abir Rahman**
  - *Marketing with Twitter*
  - *A Study of Meme Propagation*
- **Jawad Aziz Khan**
  - *Going Deeper with Convolutions*
  - *Visualizing Data with TSNE*
  - *UMAP for Dimension Reduction*
- **Taufiq Rahman**
  - *Insights into Internet Memes*
  - *Predicting the Content Diffusion*
  - *Path in Online Social Networks*

Any Questions?