Mapping Visual Themes among Authentic and Coordinated Memes

Keng-Chi Chang · UC San Diego · kechang@ucsd.edu · kengchichang.com · 2022 PolMeth at WUSTL

Background & Research Question

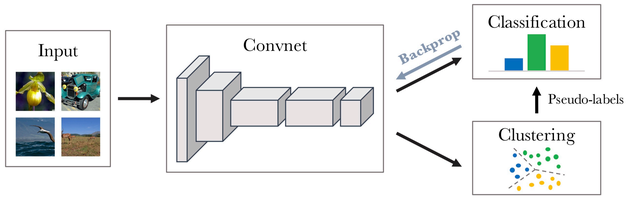
- Russian IRA shared 1.8M images on Twitter during 2016 election
- Validated US Twitter users, 19% tweets are memes, 30% political
- What distinguishes state-linked memes from authentic ones?

Data Collections

- 26K authentic memes from r/meme subreddit (authentic memes)
- 15K non-meme image-with-text data (COCO-Text) as negative sample for training (so not simply a classifier for images with or without text)
- 26K images from IRA on Twitter (coordinated memes if classified as meme)

Methods Overview

1. Classify IRA images into memes vs. non-memes (test accuracy > 0.97)
2. Extract visual embeddings jointly for both authentic memes (Reddit) and coordinated memes (IRA) using DeepCluster (Caron et al. 2019)

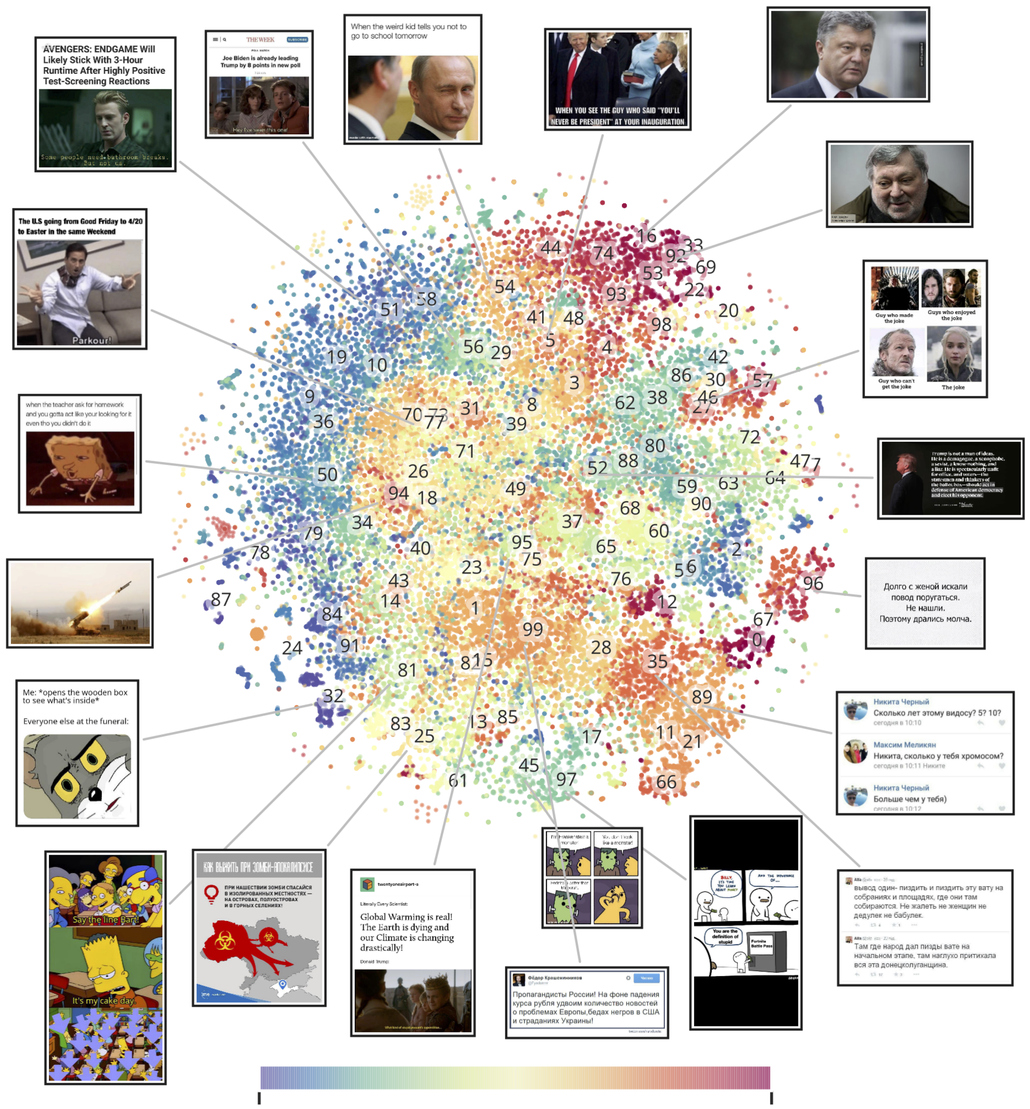


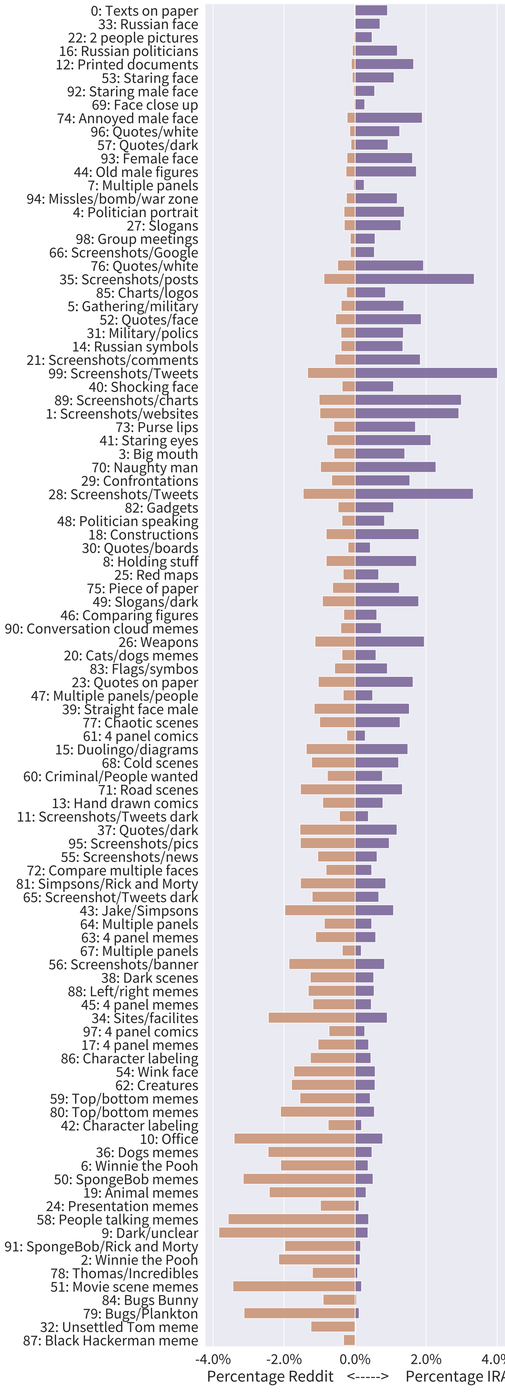
3. Cluster memes based on visual embeddings using K-means
4. Label the clusters and compare between authentic/coordinated

Summary of Findings

- Latent visual embeddings reveal similarity between memes
- Authentic and coordinated memes share most visual themes
- Coordinated IRA memes more military, gender, quotes
- Authentic Reddit memes more movie characters, comics
- IRA accounts are not widely utilizing popular meme schemes in the US
- Logistic regression on visual embedding discern IRA with $F_1 = 0.84$

Example Clusters





Next Steps

1. Compare with authentic memes on US Twitter or Russian social media
2. Use multimodal transformer (ex. VisualBERT) to extract embeddings that incorprate textual information and text-scene interactions
3. More flexible clustering models to incorporate tweet-level covariates