Motif Mining: Finding and Summarizing Remixed Image Content

William Theisen, Daniel Gonzalez Cedre, Zachariah Carmichael, Daniel Moreira, Tim Weninger, Walter Scheirer
Contact: wtheisen@nd.edu

COMPUTER VISION @ND

Motivation

We aim to provide a robust, accessible and fast tool for exploring large datasets of unsorted and unlabelled images. By discovering emerging trends in online social media artifacts human rights watchers can more quickly address and refute false and misleading narratives.

Data set

- Telegram posts related to the Ukraine-Russo War
- Accounts identified by subject matter experts
- Now over 4.5M posts spanning 2015 - Present
- Collection on-going, a cleaned snapshot is in preparation
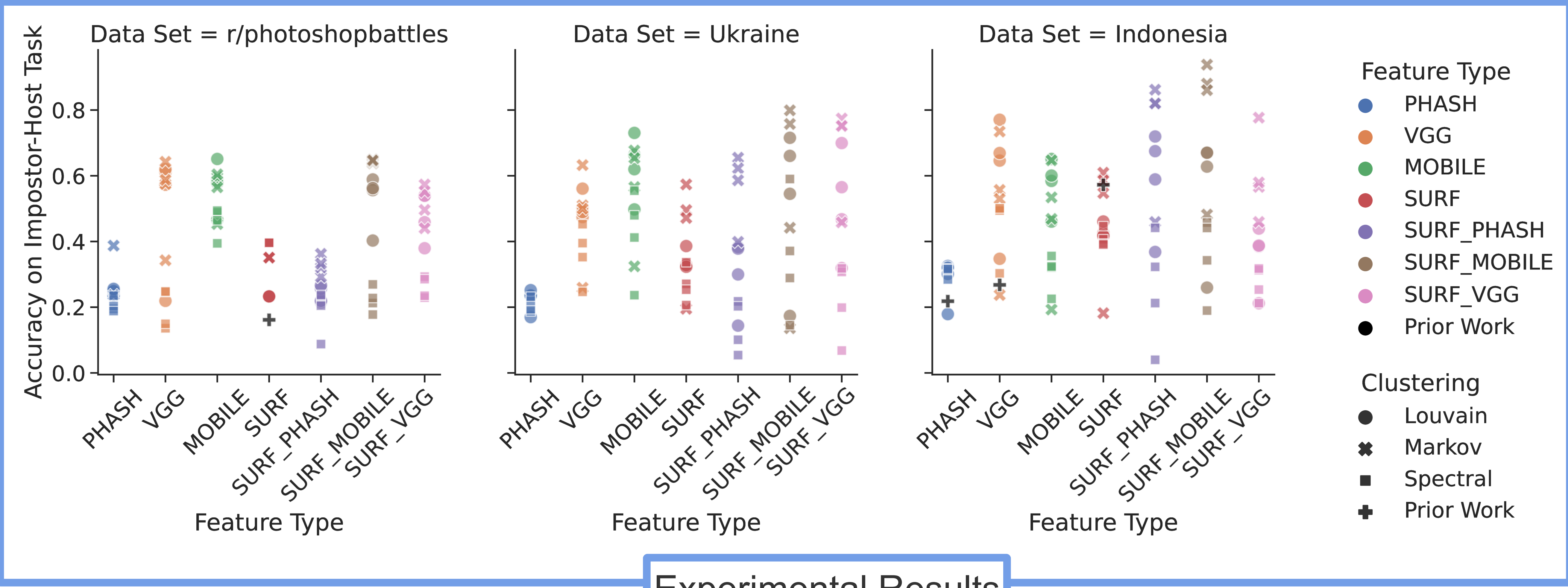

Methodolgy and Experiments

- Basic Pipeline
  1.) Extract local and global image features
  2.) Append single global 'tag' to each local feature
  3.) Feed images to indexing system (FAISS)
  4.) Use retrieval function to generate similarity graph
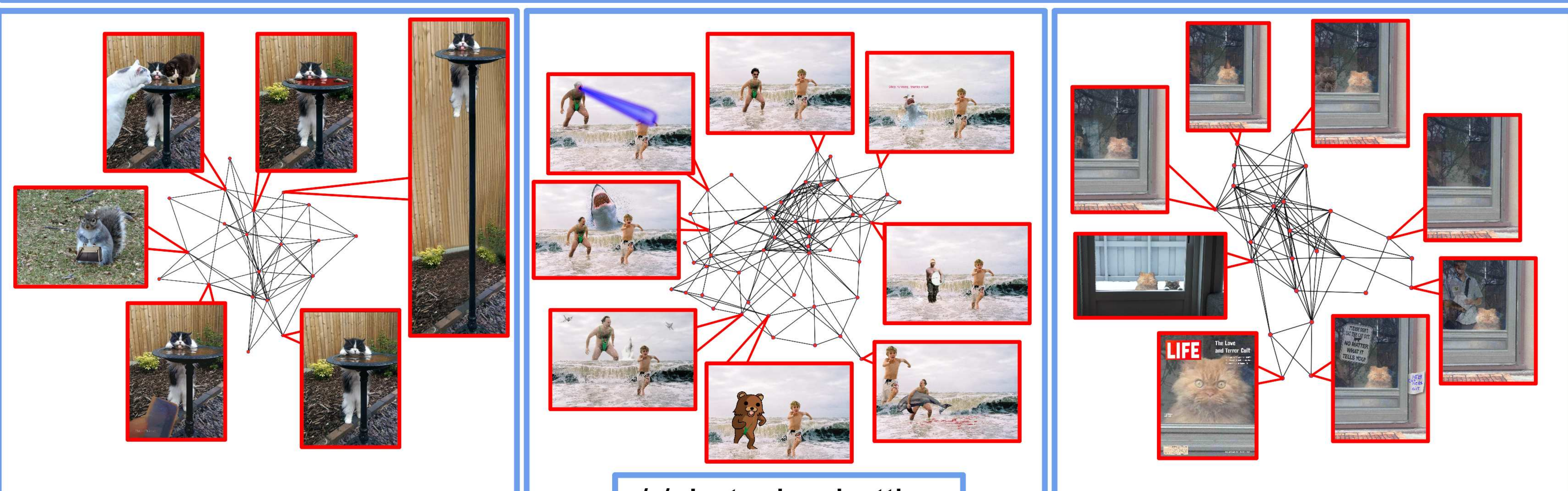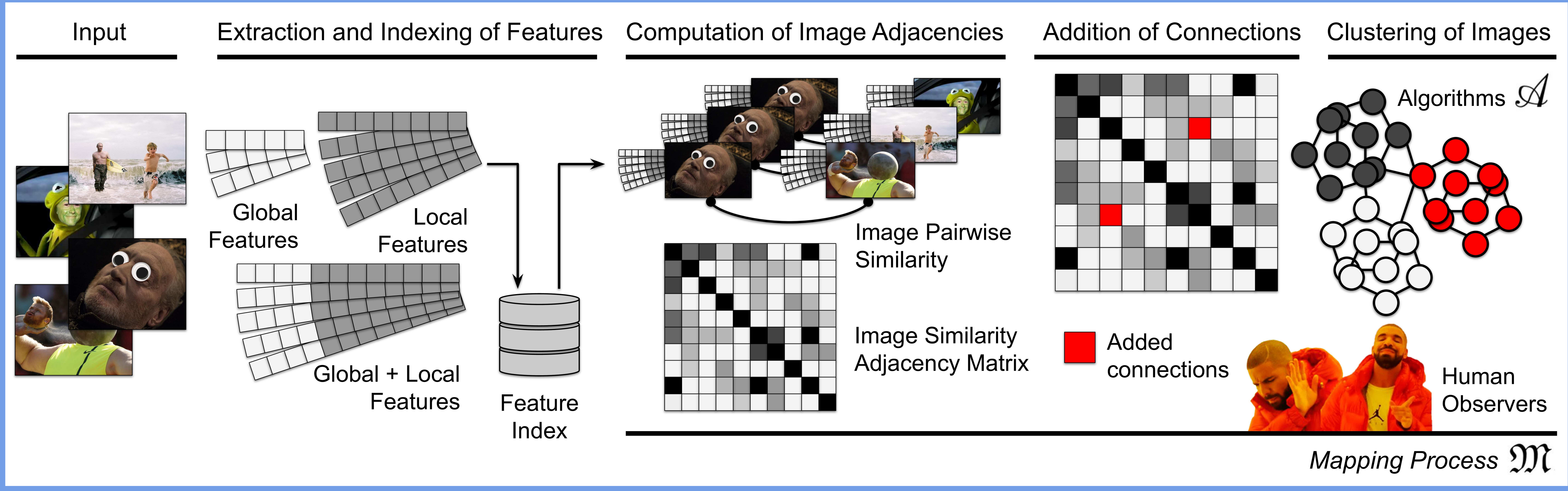 5.) Cluster similarity graph to discover visual motifs

- Use Amazon Mechanical Turk to perform 'imposter-host' test to measure motif similarity with a human feedback component. The imposter-host test ask humans to identify the image in a set of five that is from a different cluster from the other four. We use this as a proxy-measure for human-salient cluster inter-relatedness.

- Pipeline tested on 3 data sets, /r/photoshopbattles (10,588 images), Indonesia (44,612 images), and our Telegram new data set (16,433 images).

- State-of-the-art accuracy results were achieved on all three data sets.

- Globally-tagged surf features (SURF-MOBILE) gave the best results on the Imposter-Host test with Louvain clustering (taking into account image distribution).
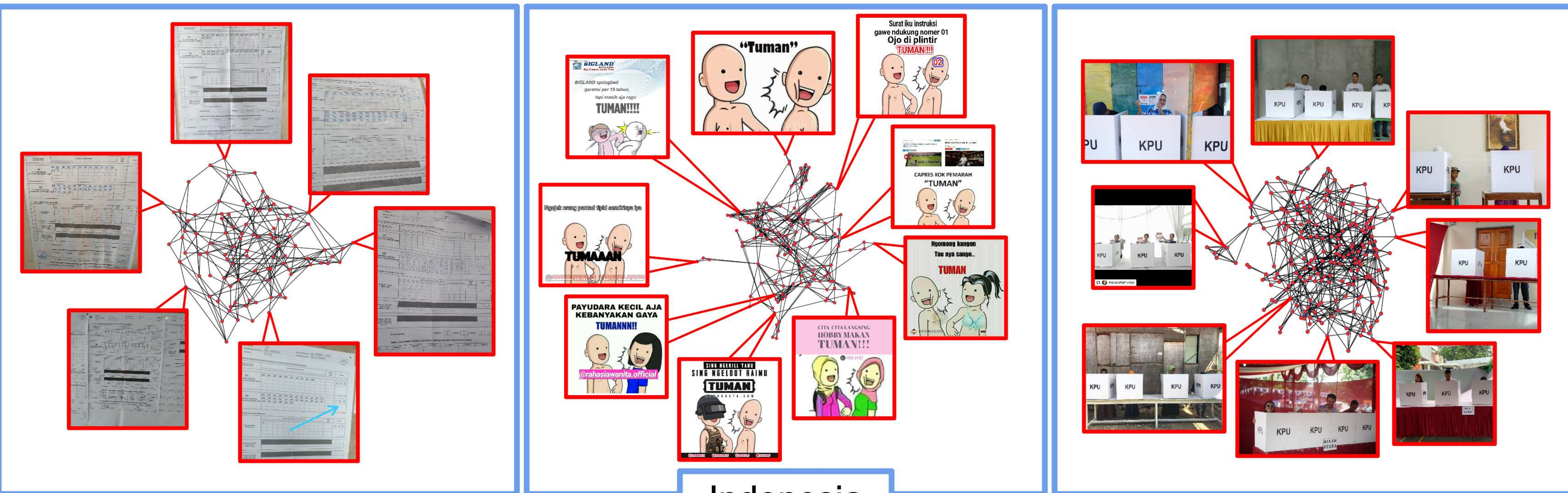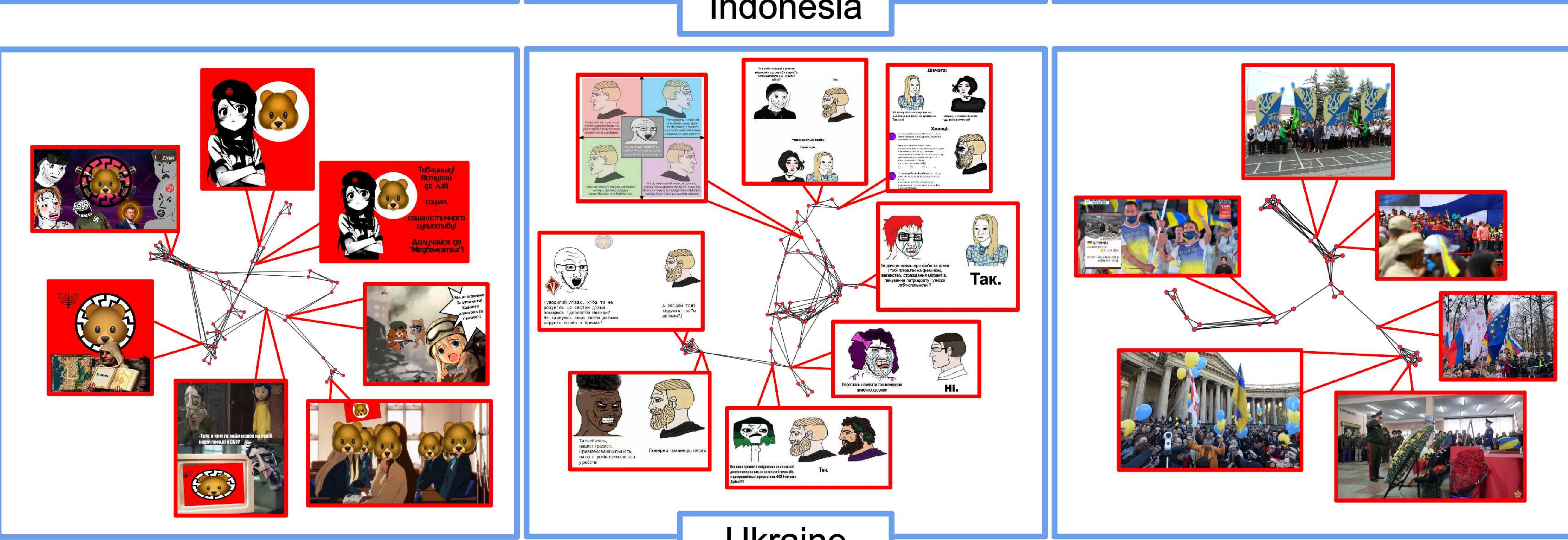
- Additionally near 10x speedups were seen.

Experimental Results

Methodology Pipeline

Input — Extraction and Indexing of Features — Computation of Image Adjacencies — Addition of Connections — Clustering of Images

/r/photoshopbattles

Indonesia

Ukraine