

ChromaClust: Latent Color Topic Modeling for Images

Nan Xiao^{1,*}¹Department of Human Genetics, The University of Chicago^{*}me@nanx.me

Image Data Retrieval

5,102 movie poster images scraped from 4 Tumblr blogs in February 2016.

Feature Extraction

Convert RGB to HSV color space. Compute the color histogram ($10 \times 10 \times 10 = 1000$ dimension vector) as features for each image.

Color Topic Model

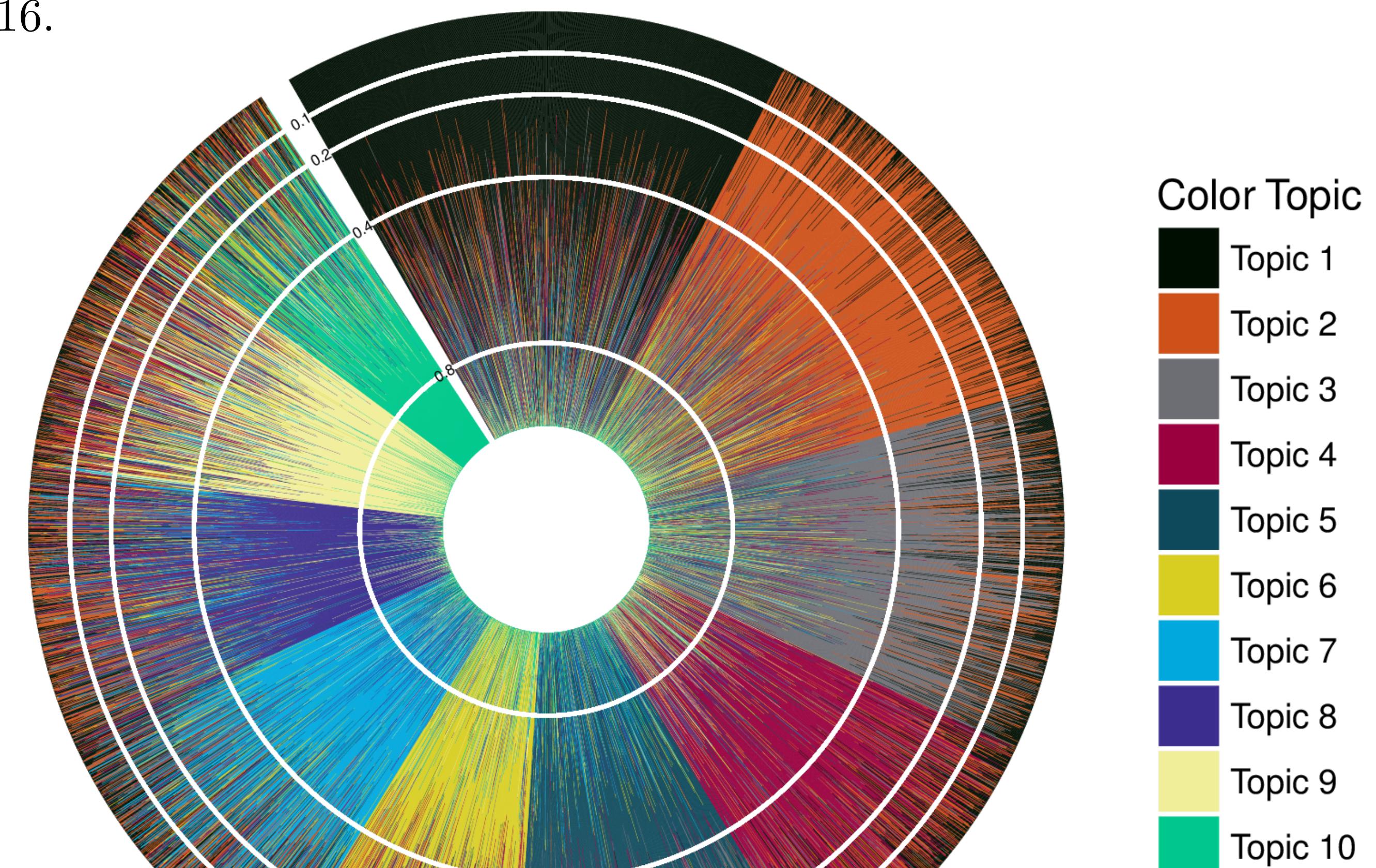
Assume the following generative process for each image:

- $\psi_k \sim \text{Dirichlet}(\beta)$: Draw color distributions ψ_k for each color topic k .
- $\theta_d \sim \text{Dirichlet}(\alpha)$: Draw color topic distributions θ_d for each image d .
- $n_d \sim \text{Poisson}(\gamma)$: For each image d , draw its number of pixels n_d

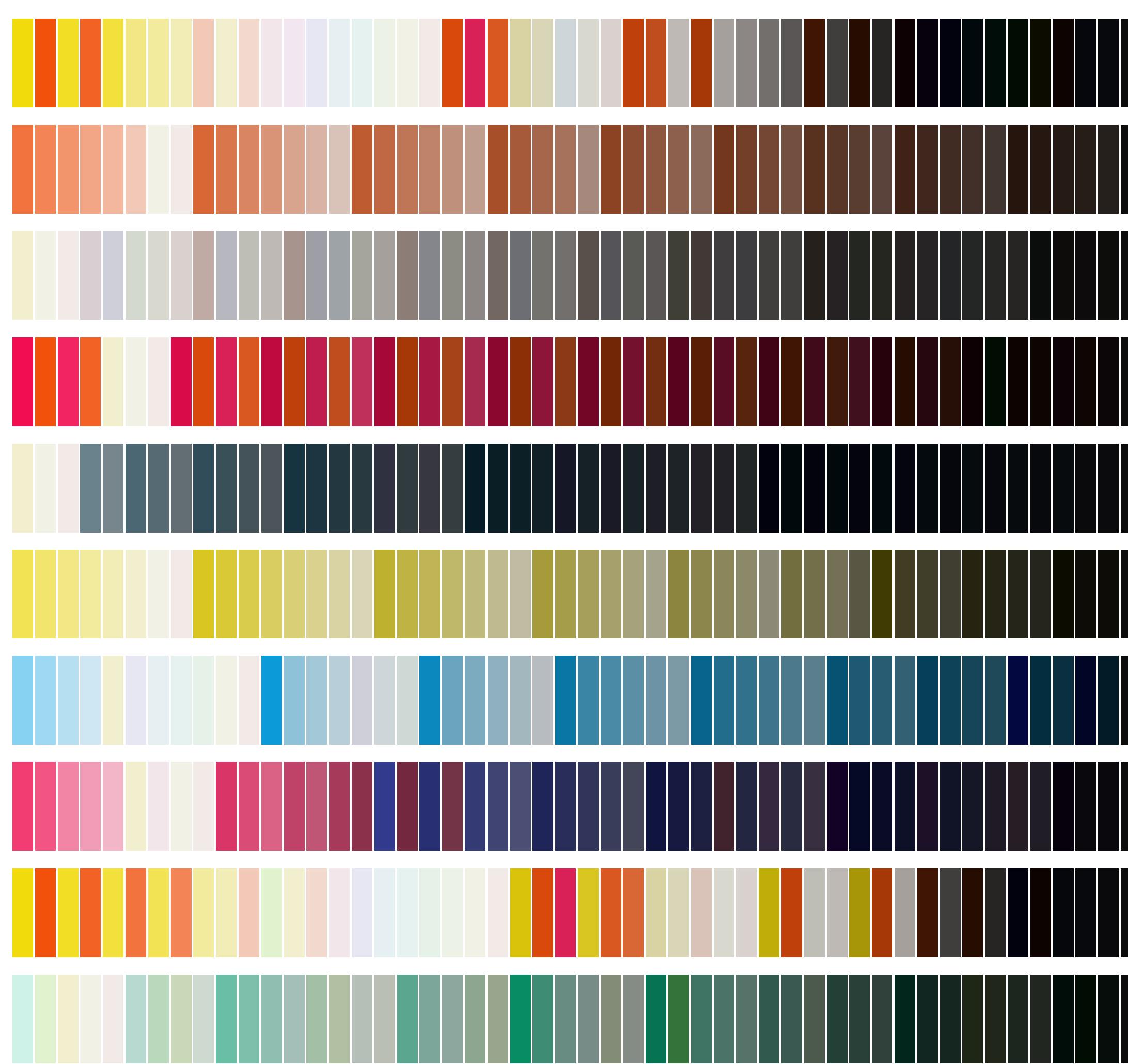
For each color $i \in \{1, 2, \dots, n_d\}$ in image d :

- $z_{di} \sim \text{Multinomial}(\theta_{di})$: Draw the pixel's color topic.
- $w_{di} \sim \text{Multinomial}(\psi_{z_{di}})$: Draw the pixel's color.

Color Topic Loadings

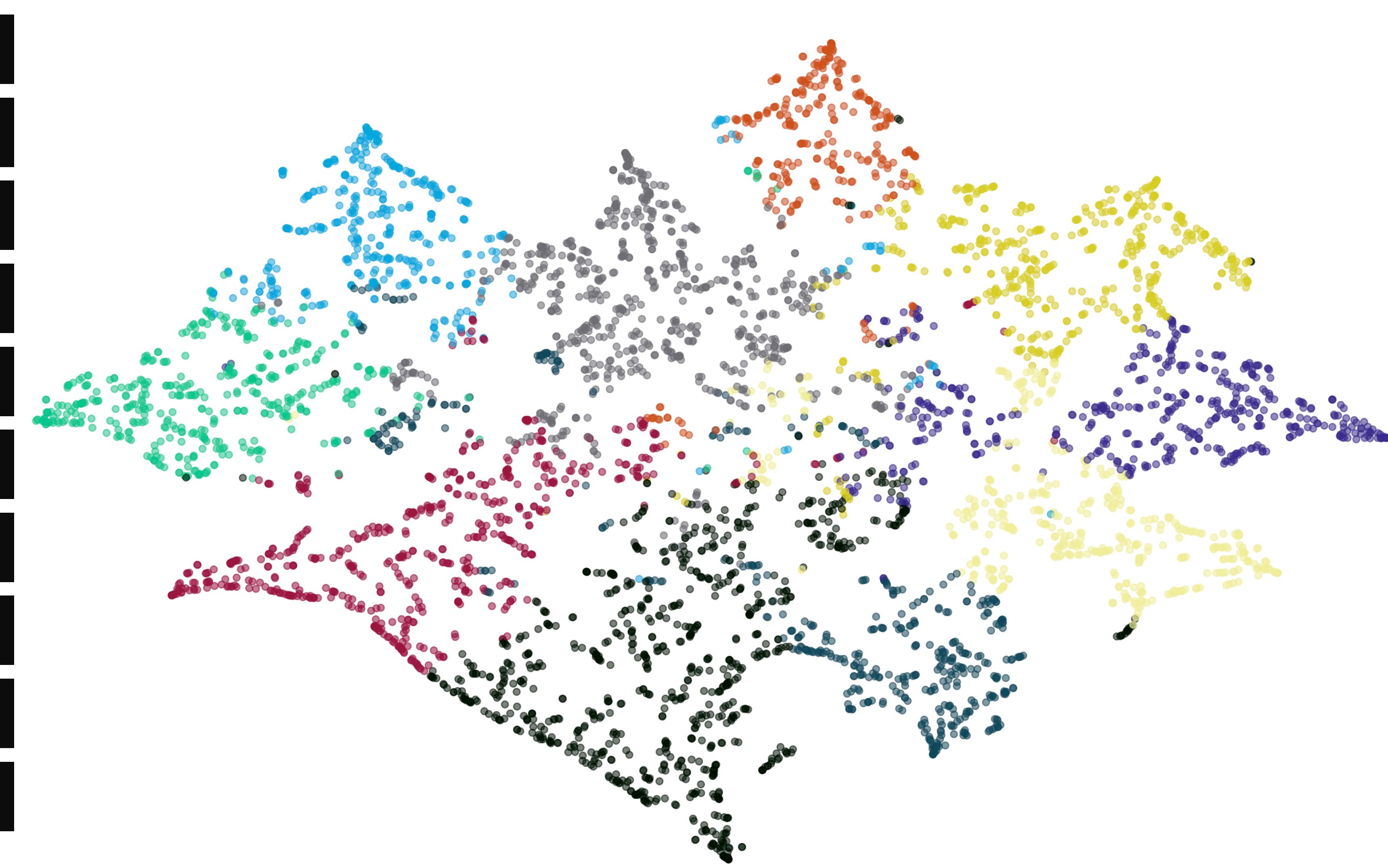


Latent Color Topics



Topic numbers K=10. K ranging from 5 to 60 were explored. The color palette displayed above has been reordered from the raw feature vectors for better perceptual interpretability.

t-SNE on Topic Loadings



t-SNE is used to project 5,102 images represented by the 10-dimensional topic loadings to a 2-dimensional plane.

Conclusions

1. Similar to population structure and languages, colors in images with certain themes (e.g. movie posters) have **color topics**, thus can be modeled with STRUCTURE-like topic models.
2. Most images in our dataset has a small number of color topics, which is a reflection of the principles in graphic design.

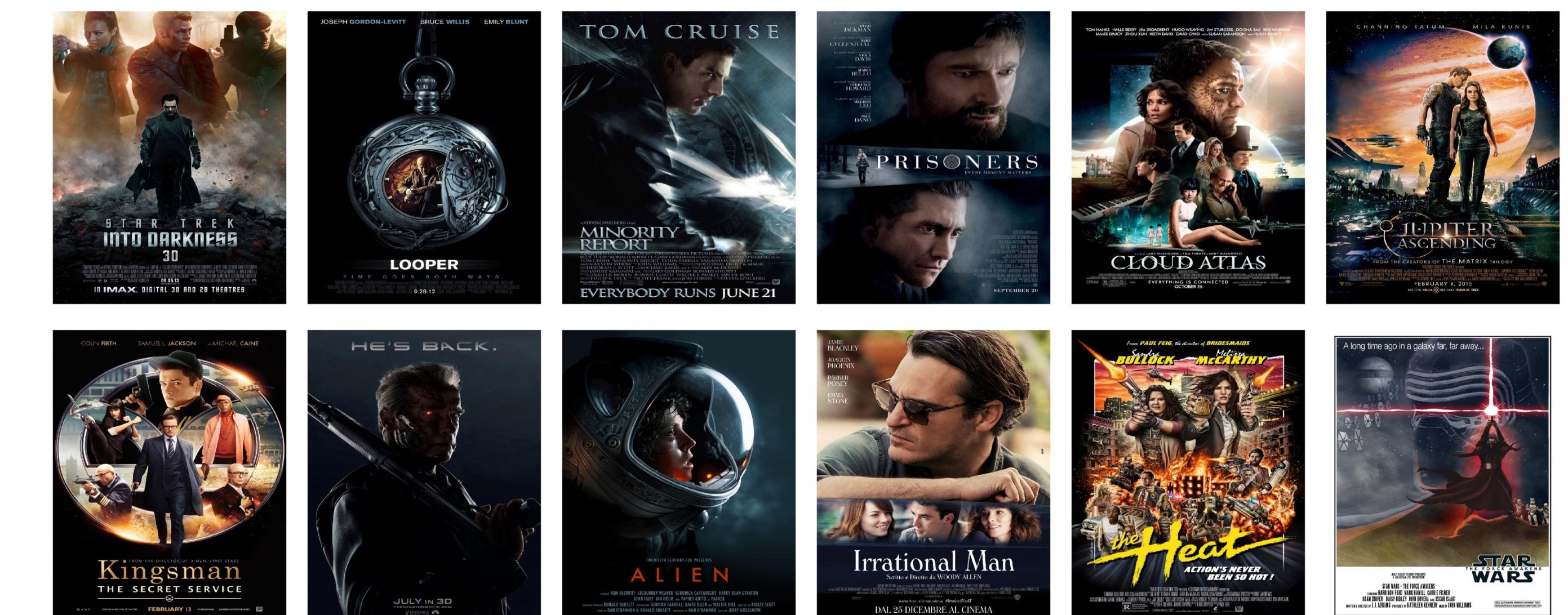
Potential Applications

1. Color topic loadings as features in recommender systems
2. Model-based image dominant color extraction for HCI design
3. Different features for image topic modeling with diverse purposes

Acknowledgements

We thank Prof. Matthew Stephens from Department of Human Genetics and Department of Statistics, Prof. John Novembre from Department of Human Genetics at University of Chicago for their wonderful stochastic modeling graduate course and their valuable inputs for making this project possible.

Exemplars in Clusters



Action, Sci-Fi (Dark Blue + Dark Green + Red)



Crime, Adventure, Drama (Light Green + Light Yellow + White)