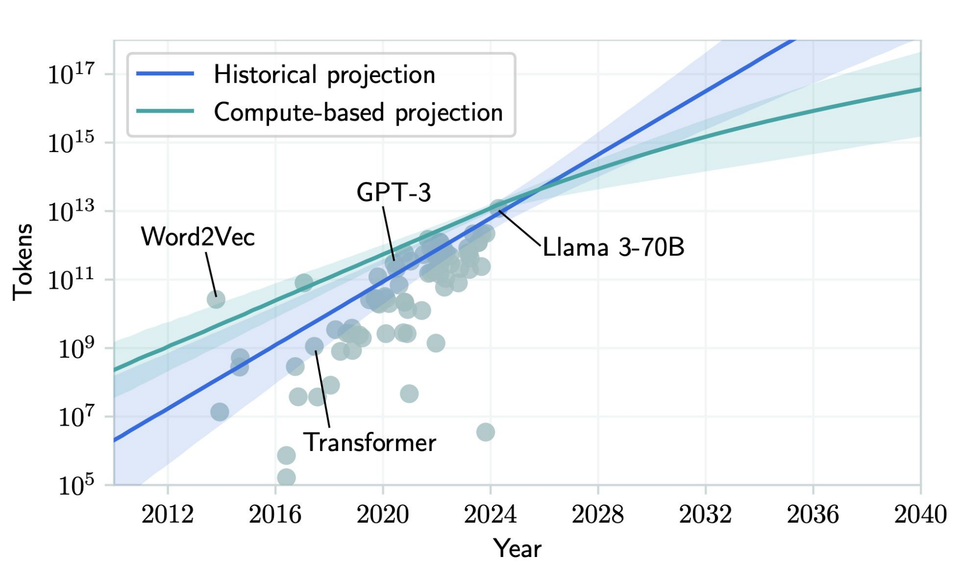
NOMIC

Explainable & Accessible AI

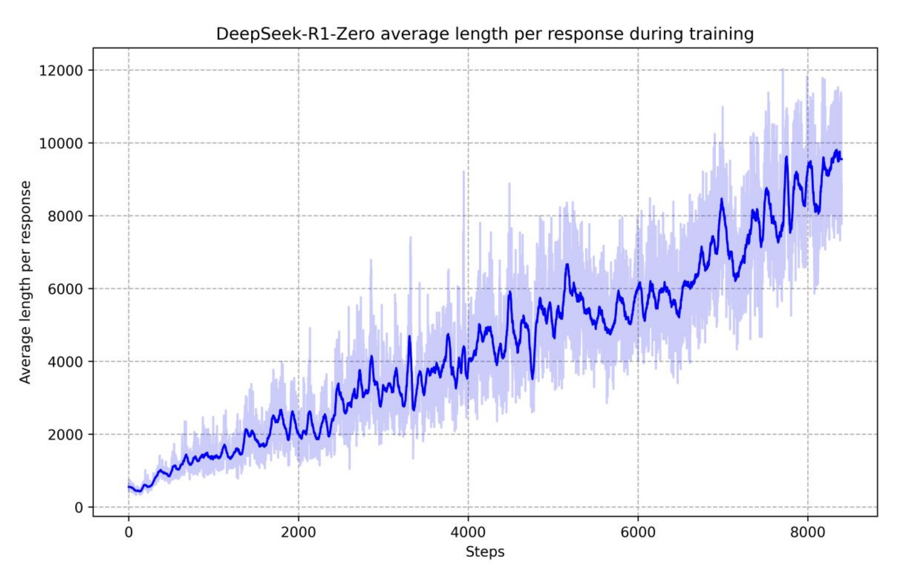
Outline

1. NOMAD Projection
2. Interpreting Projections
3. Questions

Datasets are getting Bigger!

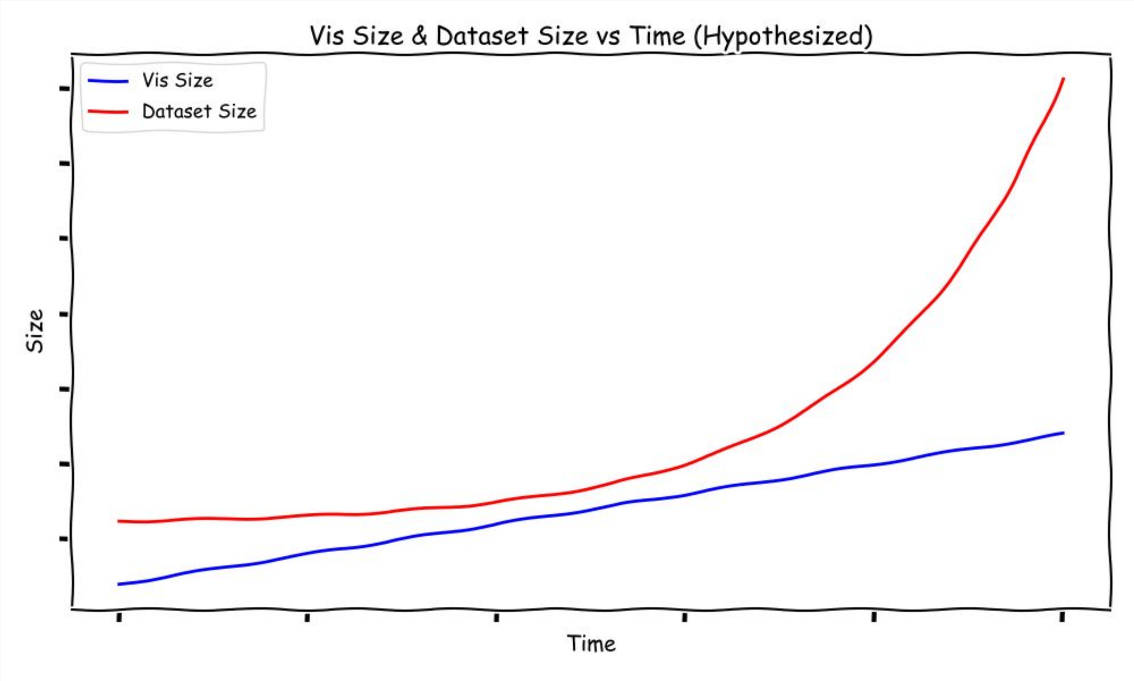


https://arxiv.org/abs/2211.04325

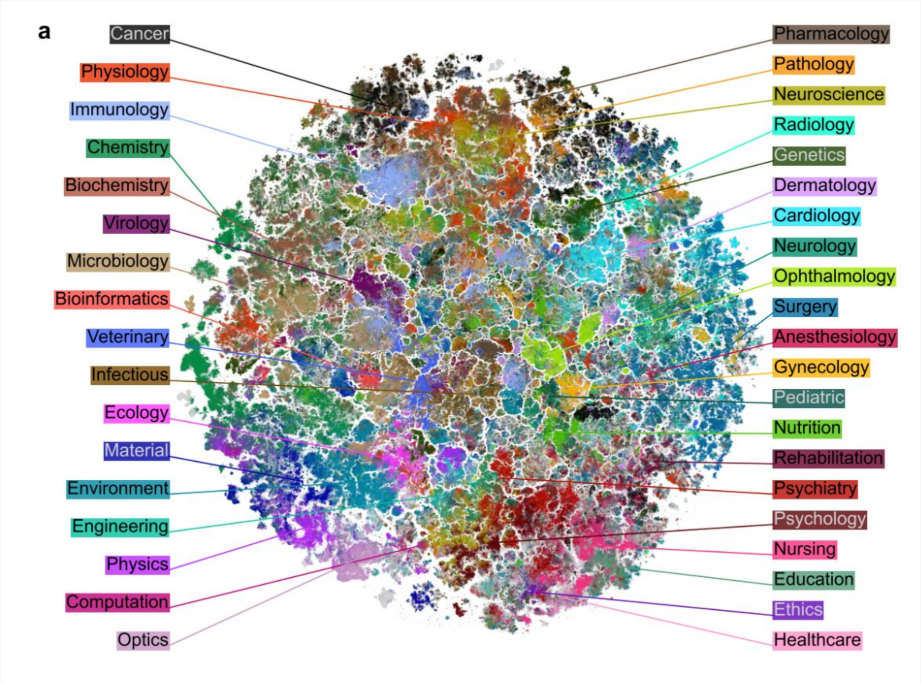


https://arxiv.org/pdf/2501.12948

Viz Methods Aren't Keeping Up!



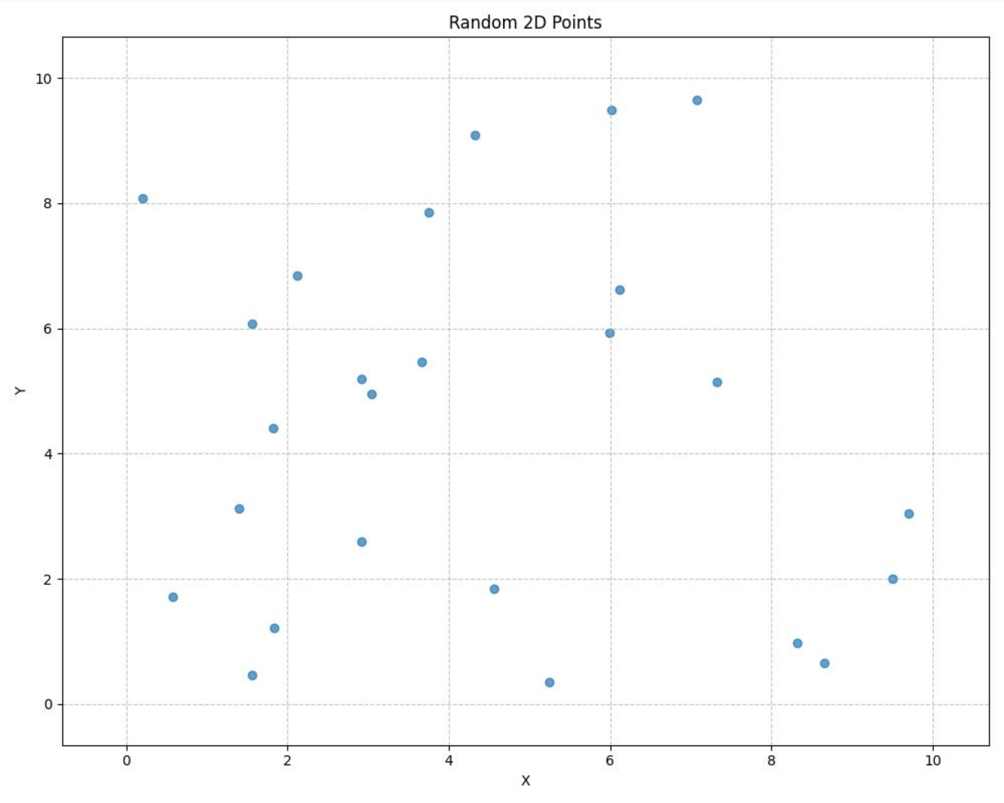
Viz Methods Aren't Keeping Up!



https://www.biorxiv.org/content/10.1101/2023.04.10.536208v1

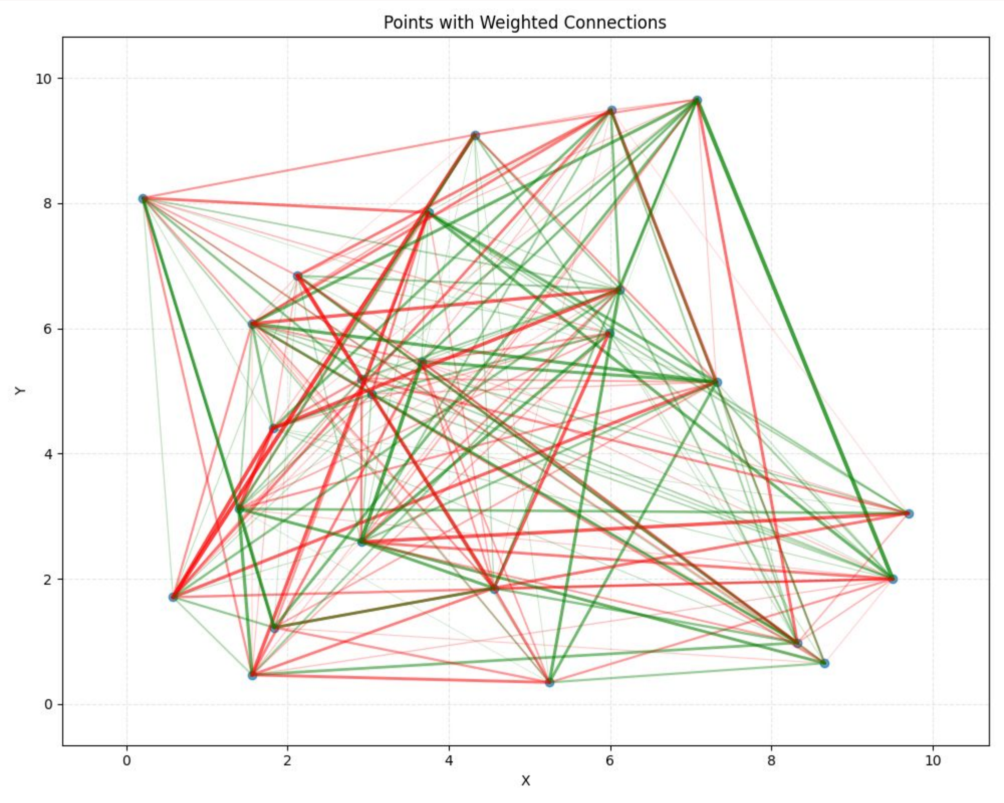
- Published in 2024
- 21M Points
- ~8 Hours Computation on CPU via OpenTSNE

Quick Review: Force Directed Layouts



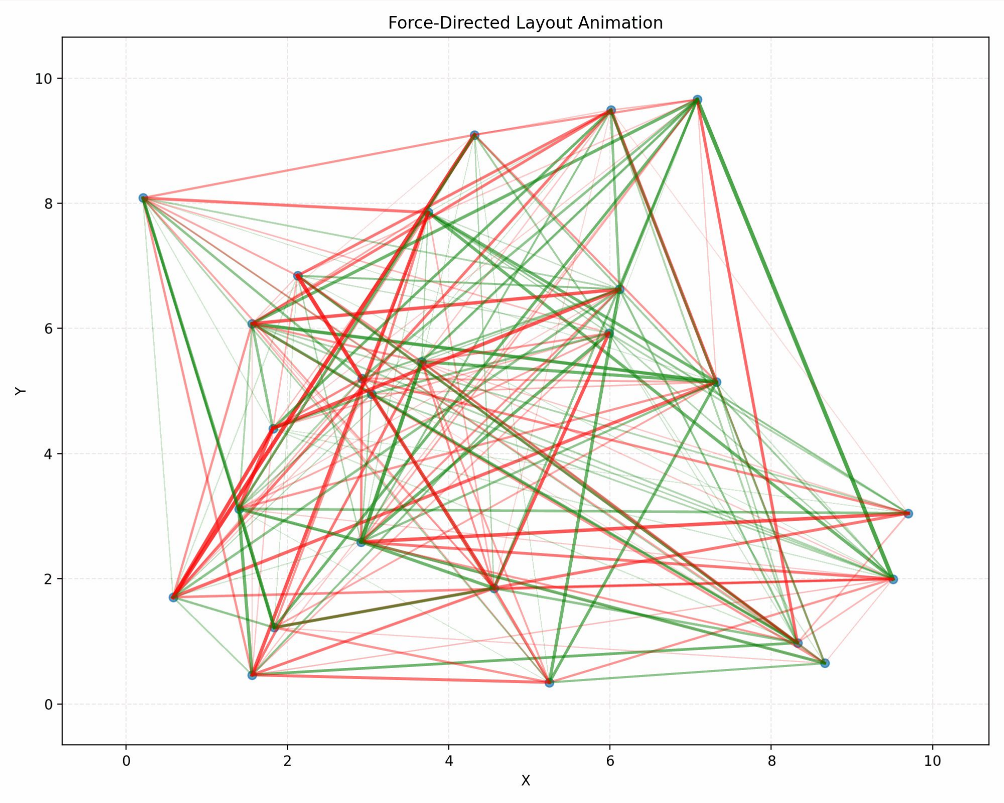
- Start with some data!

Quick Review: Force Directed Layouts



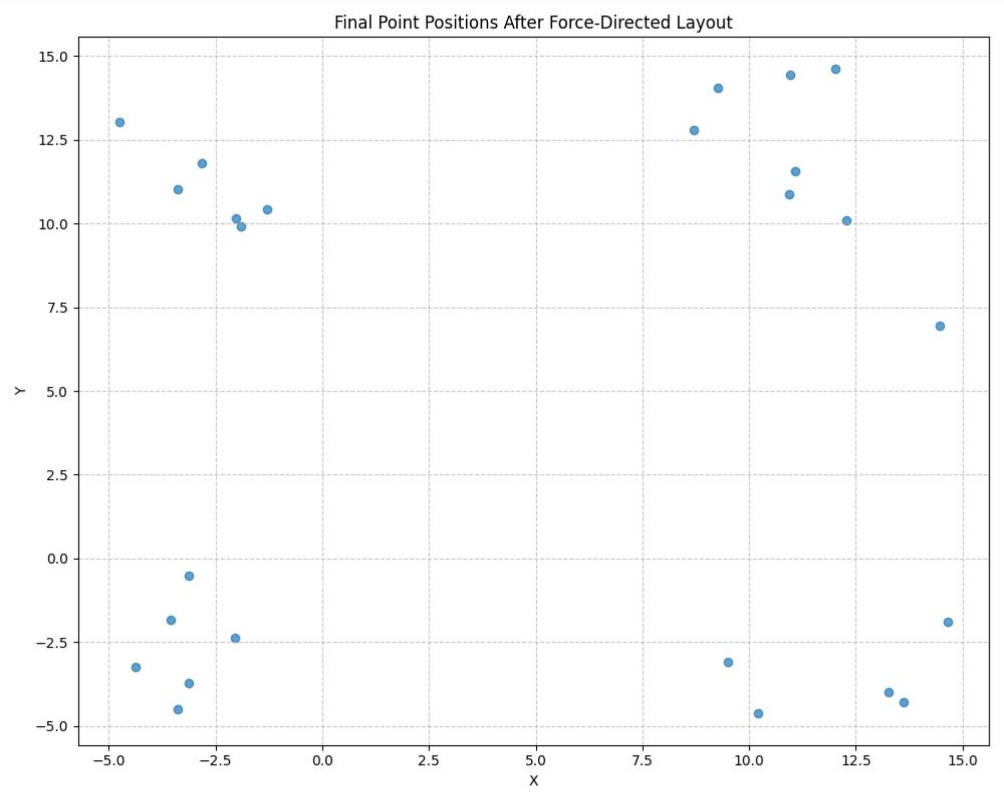
- Build a proximity structure (🚩 quadratic)
- Usually done with embedding inner products and knn (🚩 memory intensive)
- 25M points in 768d = 76 GB VRAM (H100 has 80GB VRAM)

Quick Review: Force Directed Layouts



- Run an iterative optimizer!
(🚩 lots of cycles)
- Usually compares low d proximity to high d proximity
(🚩🚩 quadratic in cycle!)
- Think of this like
a spring system

Quick Review: Force Directed Layouts



- Points cluster together!

T-SNE as a Force Directed Layout

Algorithm 1: Simple version of t-Distributed Stochastic Neighbor Embedding.

Data: data set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$,
cost function parameters: perplexity $Perp$,
optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$.
Result: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, \ldots, y_n\}$.
begin
  compute pairwise affinities $p_{j|i}$ with perplexity $Perp$ (using Equation 1)
  set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$
  sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, \ldots, y_n\}$ from $\mathcal{N}(0, 10^{-4}I)$
  for t=*1* **to** T **do**
    compute low-dimensional affinities q_{ij} (using Equation 4)
    compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 5)
    set $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t) \left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)}\right)$
  end
end

Proximity Structures

https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

T-SNE as a Force Directed Layout

Algorithm 1: Simple version of t-Distributed Stochastic Neighbor Embedding.

Data: data set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$,
cost function parameters: perplexity $Perp$,
optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$.
Result: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, \ldots, y_n\}$.
begin
  compute pairwise affinities $p_{j|i}$ with perplexity $Perp$ (using Equation 1)
  set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$
  sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, \ldots, y_n\}$ from $\mathcal{N}(0, 10^{-4} I)$
  for *t=1* **to** T **do**
    compute low-dimensional affinities q_{ij} (using Equation 4)
    compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 5)
    set $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t) \left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)} \right)$
  end
end

Iterative Optimizer

https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

T-SNE as a Force Directed Layout

Algorithm 1: Simple version of t-Distributed Stochastic Neighbor Embedding.

Data: data set $\mathcal{X} = \{x_1, x_2, ..., x_n\}$,
cost function parameters: perplexity $Perp$,
optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$.
Result: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, ..., y_n\}$.
begin
compute pairwise affinities $p_{j|i}$ with perplexity $Perp$ (using Equation 1)
set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$
sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, ..., y_n\}$ from $\mathcal{N}(0, 10^{-4} I)$
for *t=1* **to** T **do**
compute low-dimensional affinities q_{ij} (using Equation 4)
compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 5)
set $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t) \left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)} \right)$
end
end

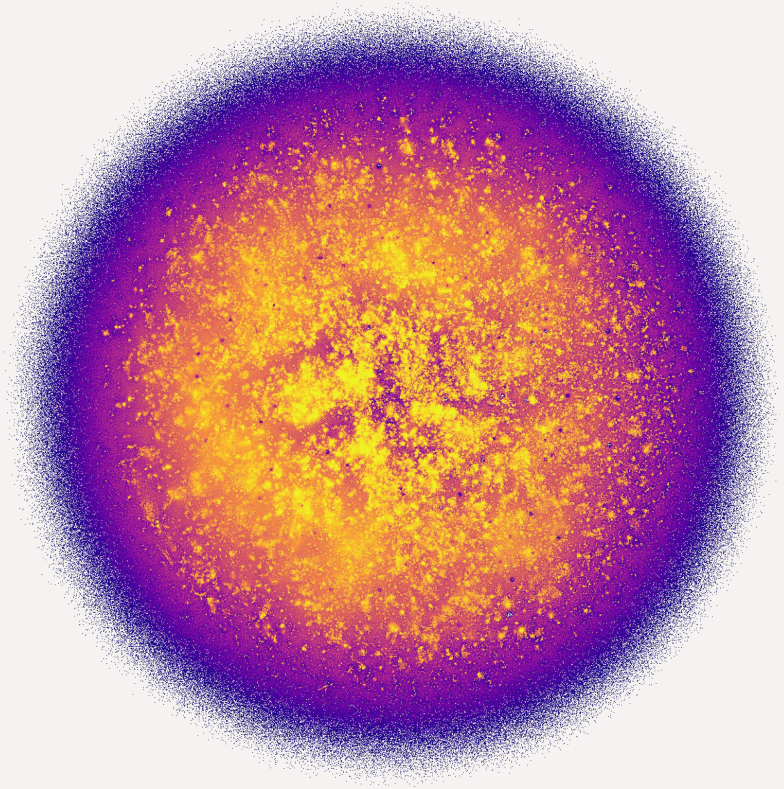
Springs

https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

Q: How to Unlock Scaling for FDL?

- Sub-Quadratic layout algorithm
  - Ideally linear!
- Multi-GPU implementation
  - Handle VRAM bottleneck
  - Handle interconnect bottleneck
- Theoretical relation to existing methods
  - Situate it in broader literature

NOMAD Projection



- **N**egative **O**r **M**ean **A**ffinity **D**iscrimination
- Linear layout algorithm
- Multi-GPU implementation
  - Cleanly shards embedding matrix
  - Sends minimal data over interconnect
- Approximate upper bound on InfoNC-T-SNE
- Computed first map of Multilingual Wiki (61M)

InfoNC T-SNE

- Noise Contrastive Estimation (NCE) converts unsupervised density estimation problems into supervised learning problems
  - Main Idea: train a binary classifier to discriminate between data samples and noise samples
  - InfoNCE: train a multiclass classifier to discriminate between a data sample and several noise samples
- InfoNC T-SNE: Train a multiclass classifier to discriminate between a true proximities and noise proximities

InfoNC T-SNE

$$\mathcal{L} = -\mathbb{E}_{\substack{ij\sim P \\ M\sim\xi}}\left[\log\left(\frac{q(ij)}{q(ij)+\sum_{m\in M} q(im)}\right)\right]$$

http://arxiv.org/pdf/2206.01816

InfoNC T-SNE

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)}{\sum_{k\neq l}\exp\left(-\|x_k - x_l\|^2/2\sigma^2\right)},$$

$$q(\theta_i, \theta_j) = \frac{1}{1 + ||\theta_i - \theta_j||^2}$$

$$\mathcal{L} = -\mathbb{E}_{\substack{ij\sim P \\ M\sim\xi}}\left[\log\left(\frac{q(ij)}{q(ij) + \sum_{m\in M} q(im)}\right)\right]$$

Uniform!

http://arxiv.org/pdf/2206.01816

InfoNC T-SNE

Induces Positive Springs

$$q(\theta_i, \theta_j) = \frac{1}{1 + ||\theta_i - \theta_j||^2}$$

$$\mathcal{L} = -\mathbb{E}_{\substack{ij\sim P \\ M\sim\xi}}\left[\log\left(\frac{q(ij)}{q(ij) + \sum_{m\in M} q(im)}\right)\right]$$

Induces Negative Springs

http://arxiv.org/pdf/2206.01816

InfoNC T-SNE

$$\mathcal{L} = -\mathbb{E}_{\substack{ij\sim P \\ M\sim\xi}}\left[\log\left(\frac{q(ij)}{q(ij) + \sum_{m\in M} q(im)}\right)\right]$$

✅ Linear Negative Forces!

http://arxiv.org/pdf/2206.01816

InfoNC T-SNE

🤔 Linear Positive Forces if Sparse?

$$\mathcal{L} = -\mathbb{E}_{\substack{ij\sim P \\ M\sim\xi}}\left[\log\left(\frac{q(ij)}{q(ij)+\sum_{m\in M} q(im)}\right)\right]$$

http://arxiv.org/pdf/2206.01816

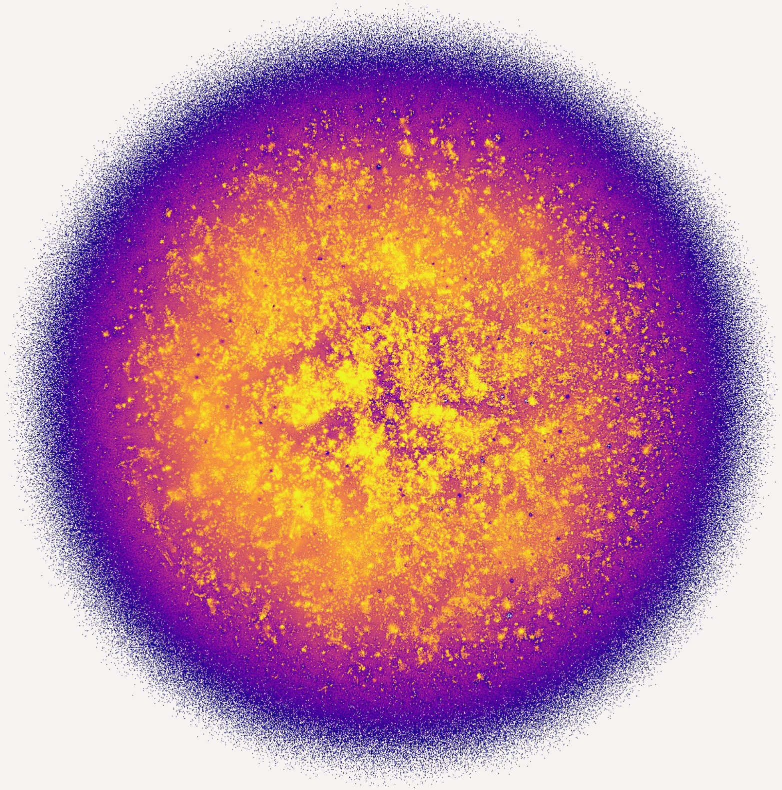
UMAP

3.1 Graph Construction

The first phase of UMAP can be thought of as the construction of a weighted k-neighbour graph. Let $X = \{x_1, \ldots, x_N\}$ be the input dataset, with a metric (or dissimilarity measure) $d : X \times X \to \mathbb{R}_{\geq 0}$. Given an input hyper-

https://arxiv.org/pdf/1802.03426

NOMAD Projection

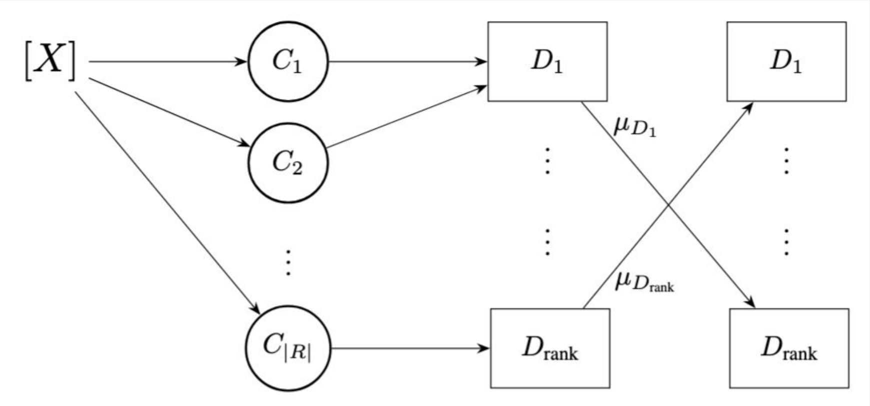


- Use KNN to retain a linear number of positive spring forces, and sampling to retain a linear number of negative spring forces

- Approximate KNN graph so that each component is local to one device

- Approximate cross-device negative spring forces with weighted cluster means to minimize data interchange

Partitioning Strategy



- First, cluster a sample of input data
  Then, compute exact nearest neighbors within each cluster
- TADA! You now have an approximate nearest neighbor index that shards cleanly by cluster
- Removes the need for exchanging data related to positive forces
- Enables shared storage of embedding matrix

Sampling Strategy

$$\mathcal{L} = -\mathbb{E}_{i \sim P_i} \left[\sum_j p(j|i) \log \left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}} \right) \right]$$

$$\mathcal{L} = -\mathbb{E}_{\substack{ij \sim P \\ M \sim \xi}} \left[\log \left(\frac{q(ij)}{q(ij) + \sum_{m \in M} q(im)} \right) \right]$$

Sampling Strategy

$$\mathcal{L} = -\mathbb{E}_{i \sim P_i}\left[\sum_j p(j|i) \log\left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}}\right)\right]$$

$$\mathcal{L} = -\mathbb{E}_{\substack{ij \sim P \\ M \sim \xi}}\left[\log\left(\frac{q(ij)}{q(ij) + \sum_{m \in M} q(im)}\right)\right]$$

Sampling Strategy

$$\mathcal{L} = -\mathbb{E}_{i \sim P_i}\left[\sum_j p(j|i) \log\left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}}\right)\right]$$

$$\mathcal{L} = -\mathbb{E}_{\substack{ij \sim P \\ M \sim \xi}}\left[\log\left(\frac{q(ij)}{q(ij) + \sum_{m \in M} q(im)}\right)\right]$$

Sampling Strategy

$$\mathcal{L} = -\mathbb{E}_{i \sim P_i}\left[\sum_j p(j|i) \log\left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}}\right)\right]$$

Cluster Means as Negative Samples

Normal Negative Samples

Sampling Strategy

$$\mathcal{L} = -\mathbb{E}_{i \sim P_i}\left[\sum_j p(j|i)\log\left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}}\right)\right]$$

- Let R be a partition of the ANN graph
- Let $\tilde{R}$ be the partition cells that we wish to approximate

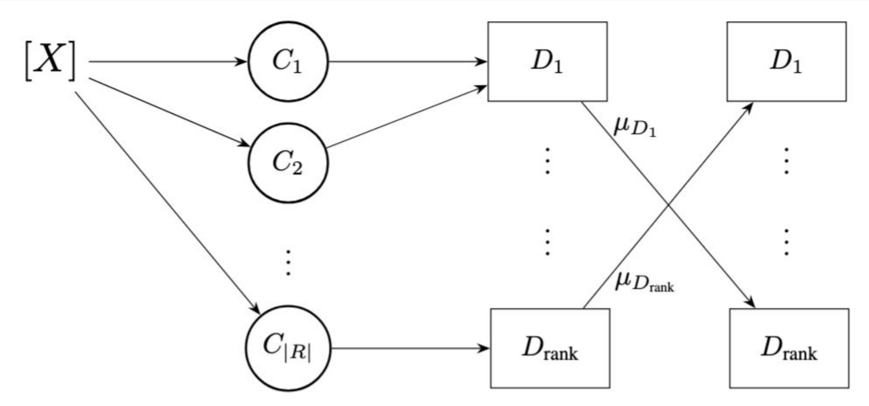
Sampling Strategy

$$\mathcal{L} = -\mathbb{E}_{i \sim P_i} \left[\sum_j p(j|i) \log \left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}} \right) \right]$$

$$\tilde{\mathcal{M}} = |M| \sum_{r \in \tilde{R}} p(m \in r) q(i\mu_r)$$

$$\mathcal{M} = \sum_{r \in R \setminus \tilde{R}} \mathbb{E}_{M \sim \xi} \left[\sum_{m \in M_r} q(im) \right]$$

Sampling Strategy

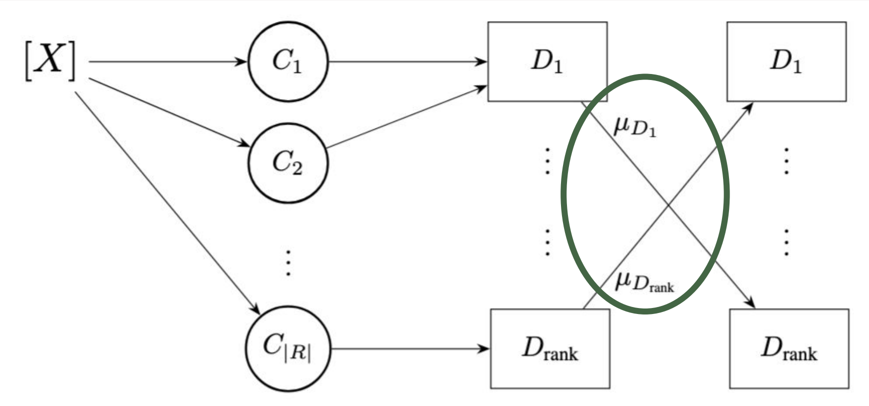


$$\mathcal{L} = -\mathbb{E}_{i\sim P_i}\left[\sum_j p(j|i)\log\left(\frac{q(ij)}{q(ij)+\tilde{\mathcal{M}}+\mathcal{M}}\right)\right]$$

$$\tilde{\mathcal{M}} = |M|\sum_{r\in\tilde{R}} p(m\in r)q(i\mu_r)$$

$$\mathcal{M} = \sum_{r\in R\setminus\tilde{R}} \mathbb{E}_{M\sim\xi}\left[\sum_{m\in M_r} q(im)\right]$$

Sampling Strategy



$$\mathcal{L} = -\mathbb{E}_{i \sim P_i} \left[\sum_j p(j|i) \log \left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}} \right) \right]$$

$$\tilde{\mathcal{M}} = |M| \sum_{r \in \tilde{R}} p(m \in r) q(i\mu_r)$$

$$\mathcal{M} = \sum_{r \in R \setminus \tilde{R}} \mathbb{E}_{M \sim \xi} \left[\sum_{m \in M_r} q(im) \right]$$

Relationship to InfoNC-T-SNE

$$\mathcal{L}^I = -\mathbb{E}_{\substack{ij \sim P \\ M \sim \xi}} \left[\log \left(\frac{q(ij)}{q(ij) + \sum_{m \in M} q(im)} \right) \right]$$

Relationship to InfoNC-T-SNE

$$\mathcal{L}^I = -\mathbb{E}_{\substack{ij \sim P \\ M \sim \xi}} \left[\log \left(\frac{q(ij)}{q(ij) + \sum_{m \in M} q(im)} \right) \right]$$

Jensen's Inequality
Taylor Expansion
Definition of Expectation

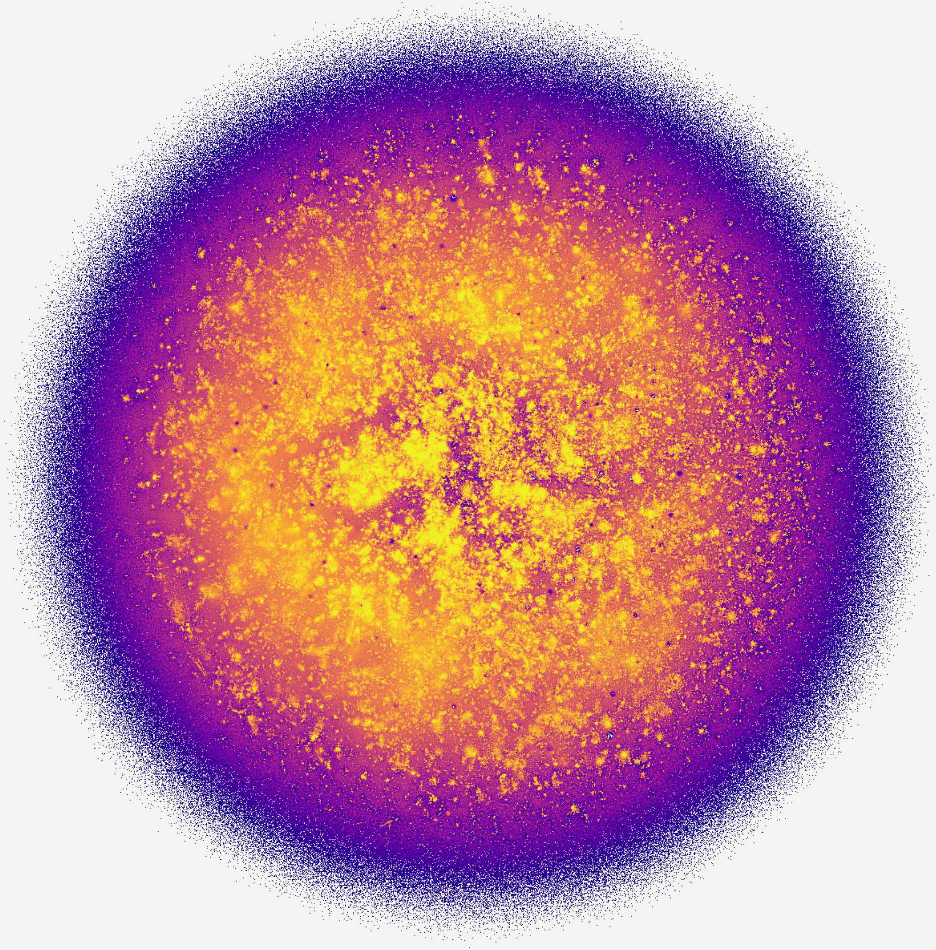
Relationship to InfoNC-T-SNE

$$\mathcal{L}^I = -\mathbb{E}_{\substack{ij \sim P \\ M \sim \xi}} \left[\log \left(\frac{q(ij)}{q(ij) + \sum_{m \in M} q(im)} \right) \right]$$

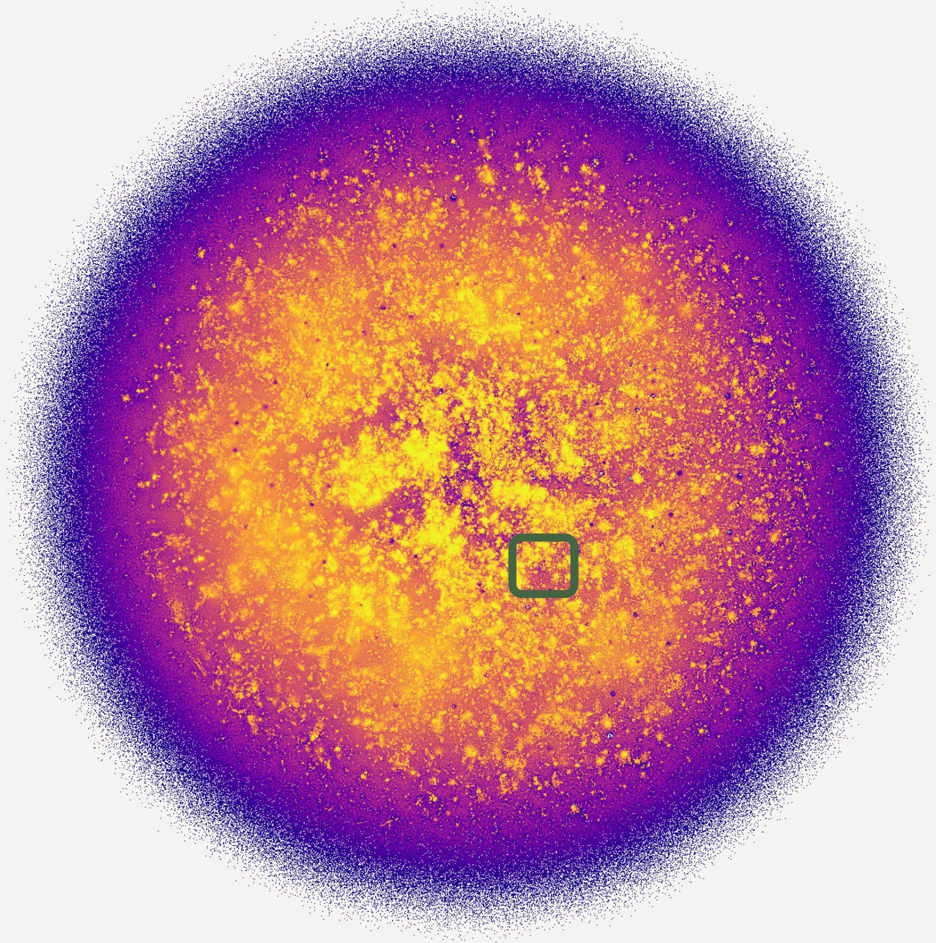
Jensen's Inequality
Taylor Expansion
Definition of Expectation

$$\mathcal{L}^I \lesssim -\mathbb{E}_{i \sim P_i} \left[\sum_j p(j|i) \log \left(\frac{q(ij)}{q(ij) + \tilde{\mathcal{M}} + \mathcal{M}} \right) \right]$$

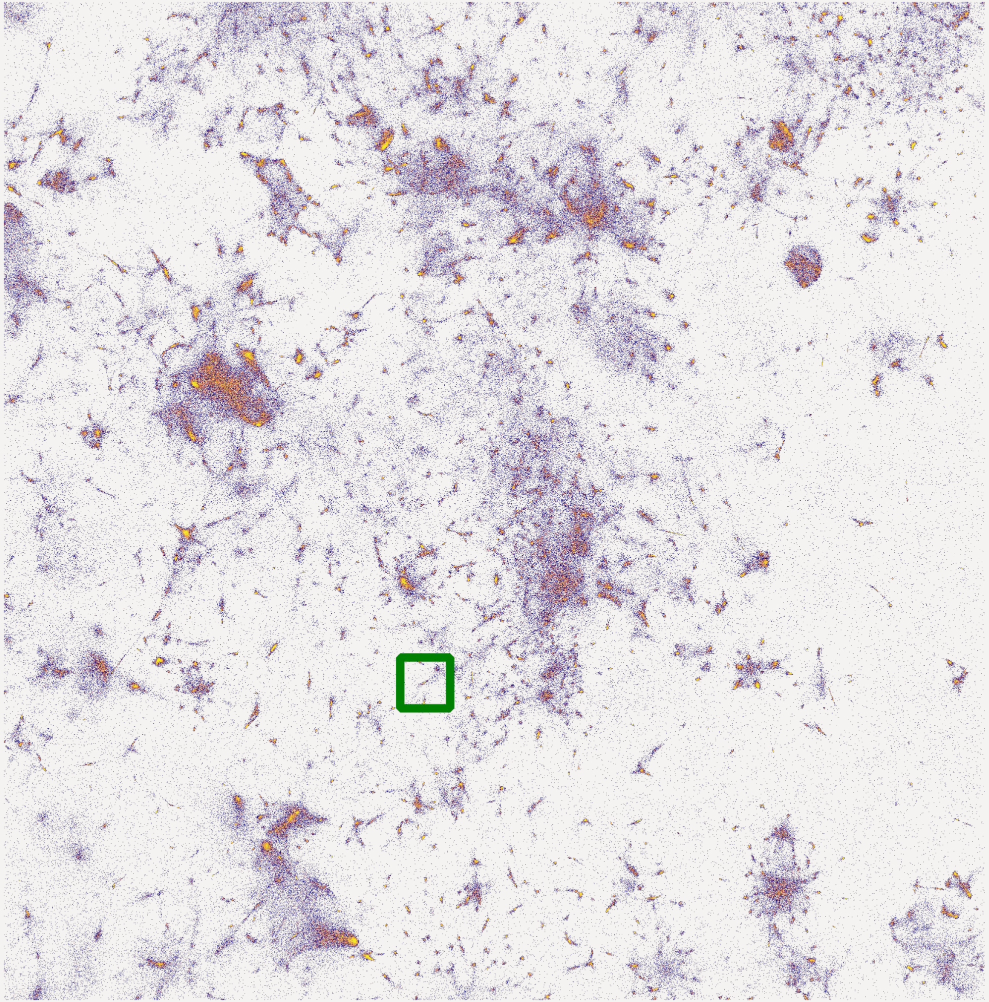
Wikipedia Map



Let’s zoom in

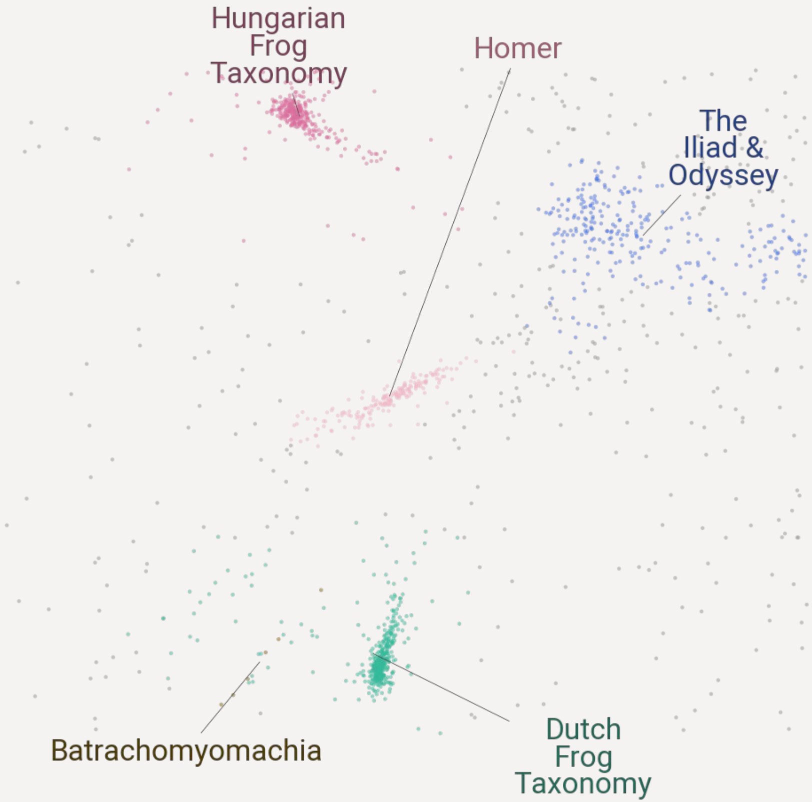


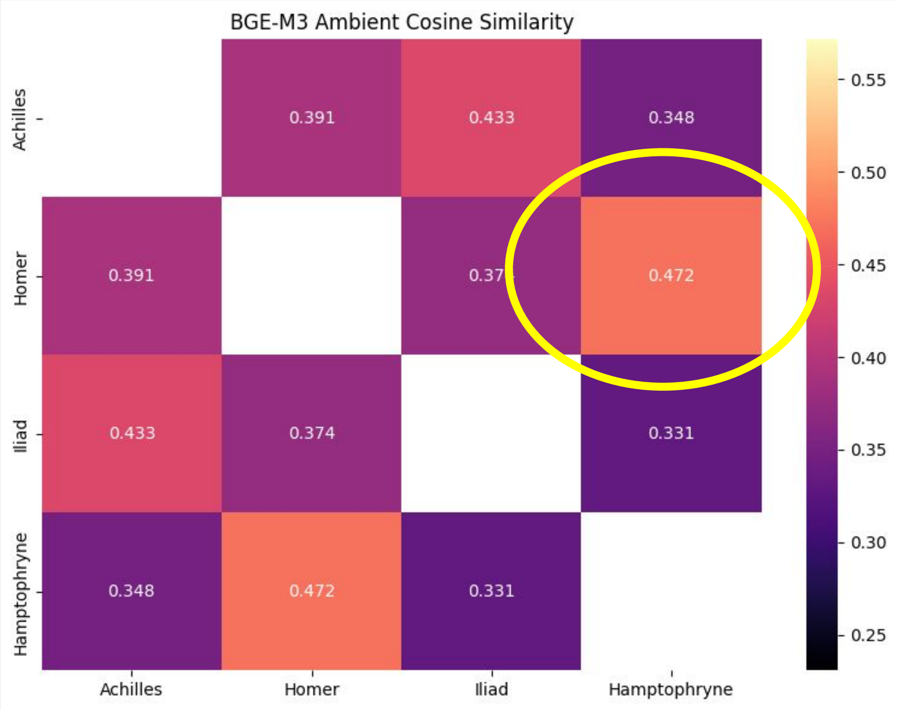
Again!

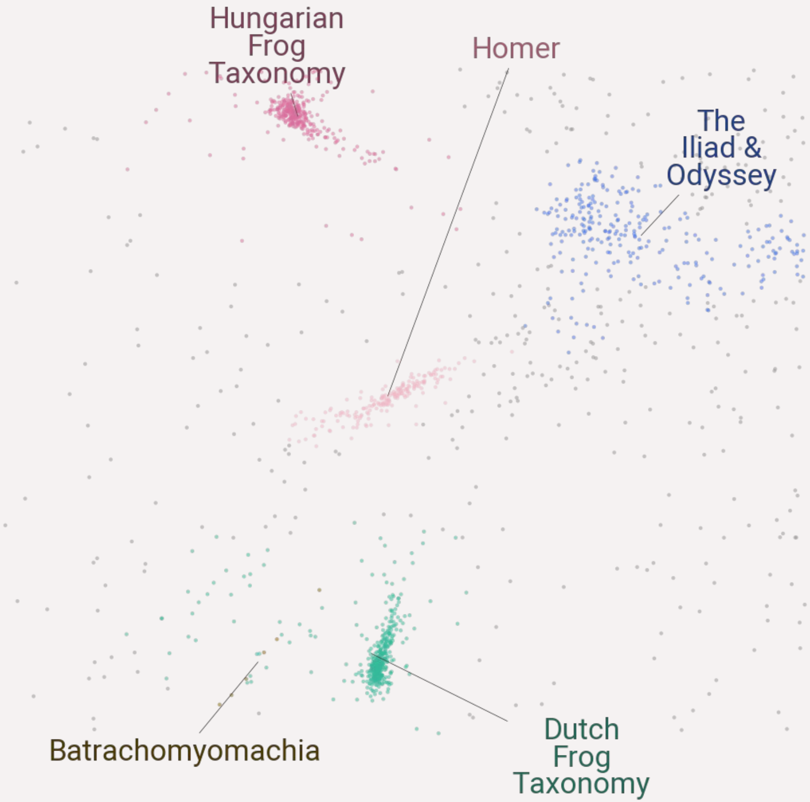


Here are the articles about homer

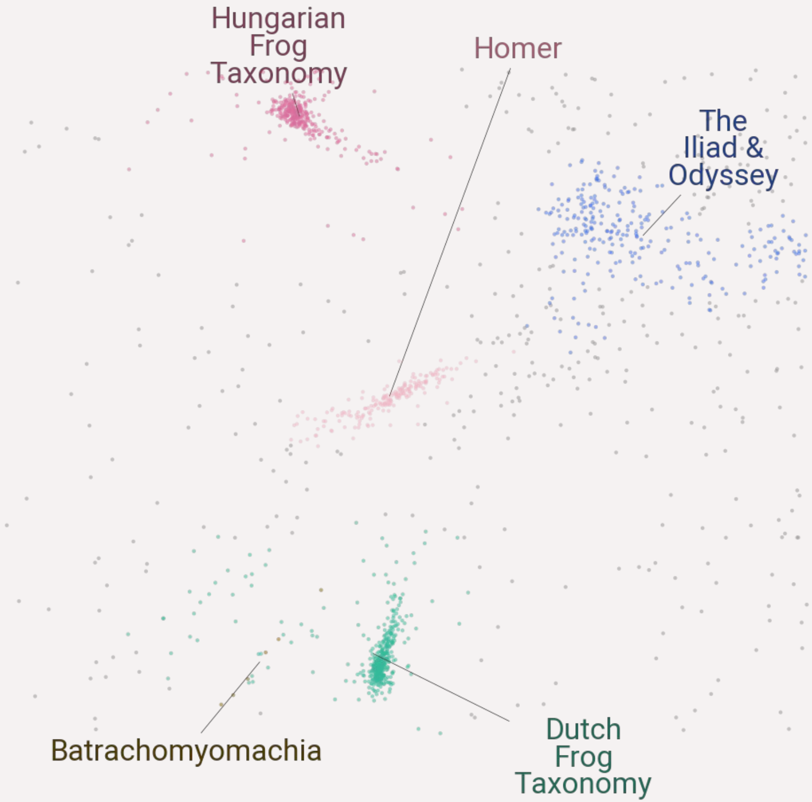
What are these frogs doing near Homer?




BGE-M3 Ambient Cosine Similarity
Achilles
Homer
Iliad
Hamptophryne
0.391
0.433
0.348
0.391
0.37
0.472
0.433
0.374
0.331
0.348
0.472
0.331
0.55
0.50
0.45
0.40
0.35
0.30
0.25


Hungarian Frog Taxonomy
Homer
The Iliad & Odyssey
Batrachomyomachia
Dutch Frog Taxonomy

What is Batrachomyomachia?



Batrachomyomachia

28 languages

Article Talk Read Edit View history

From Wikipedia, the free encyclopedia

"Frog-mouse war" redirects here. For the 20th-century controversy in the foundations of mathematics, see Brouwer–Hilbert controversy.

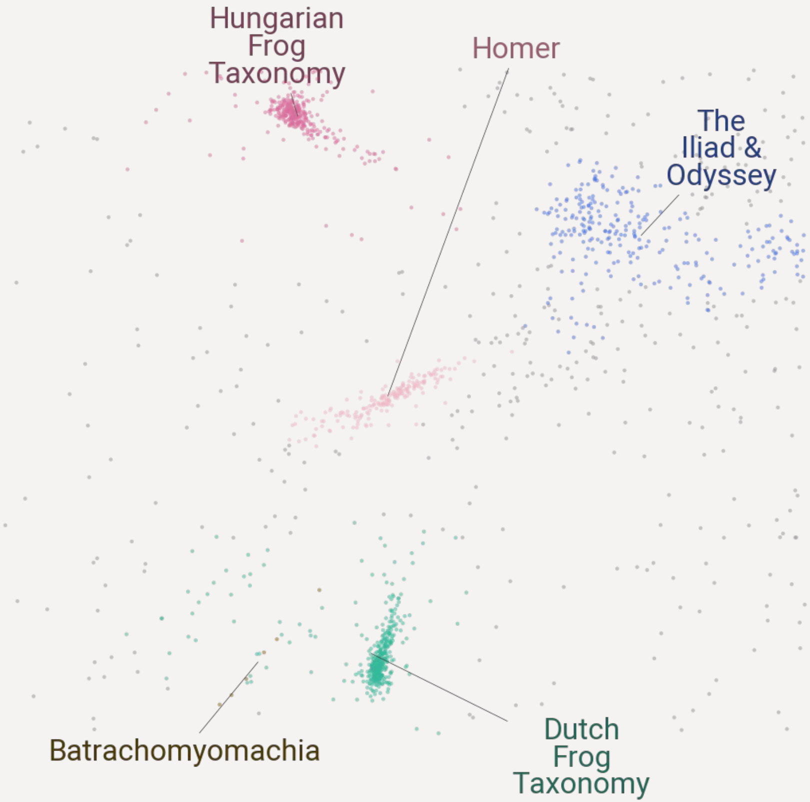
The ***Batrachomyomachia*** (Ancient Greek: Βατραχομυομαχία, from βάτραχος, "frog", μῦς, "mouse", and μάχη, "battle") or ***Battle of the Frogs and Mice*** is a comic epic, or a parody of the *Iliad*.

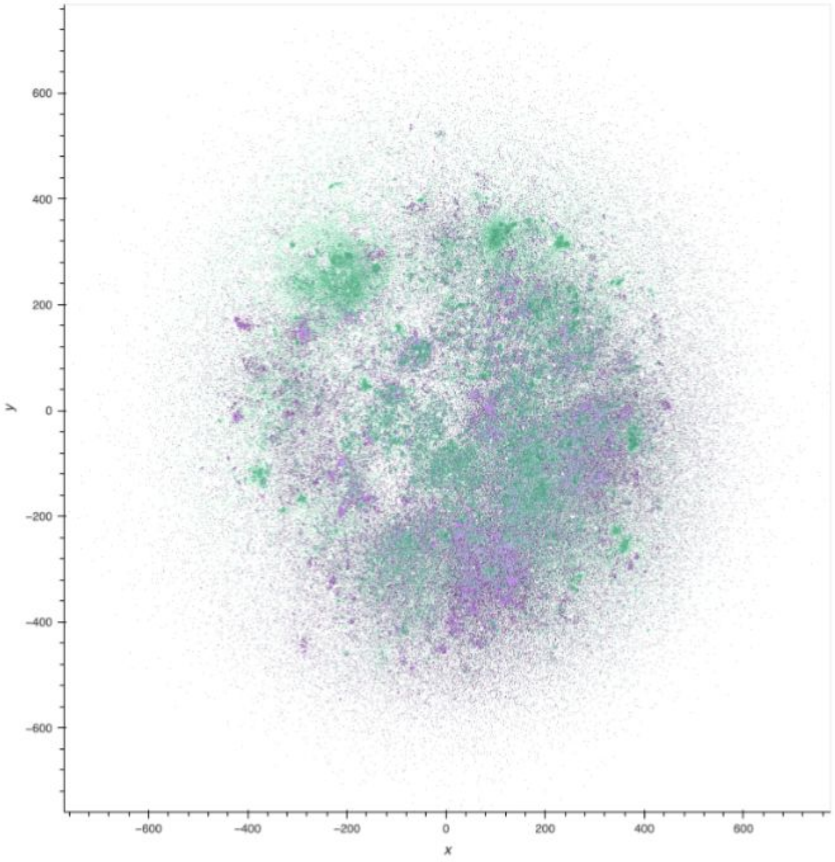
The word *batrachomyomachia* has come to mean "a trivial altercation". Both the Greek word and its German translation, *Froschmäusekrieg*, have been used to describe disputes such as the one between the ideologues and pragmatists in the Reagan administration.[1]

Illustration from an 1878 German edition of the *Batrachomyomachia*.

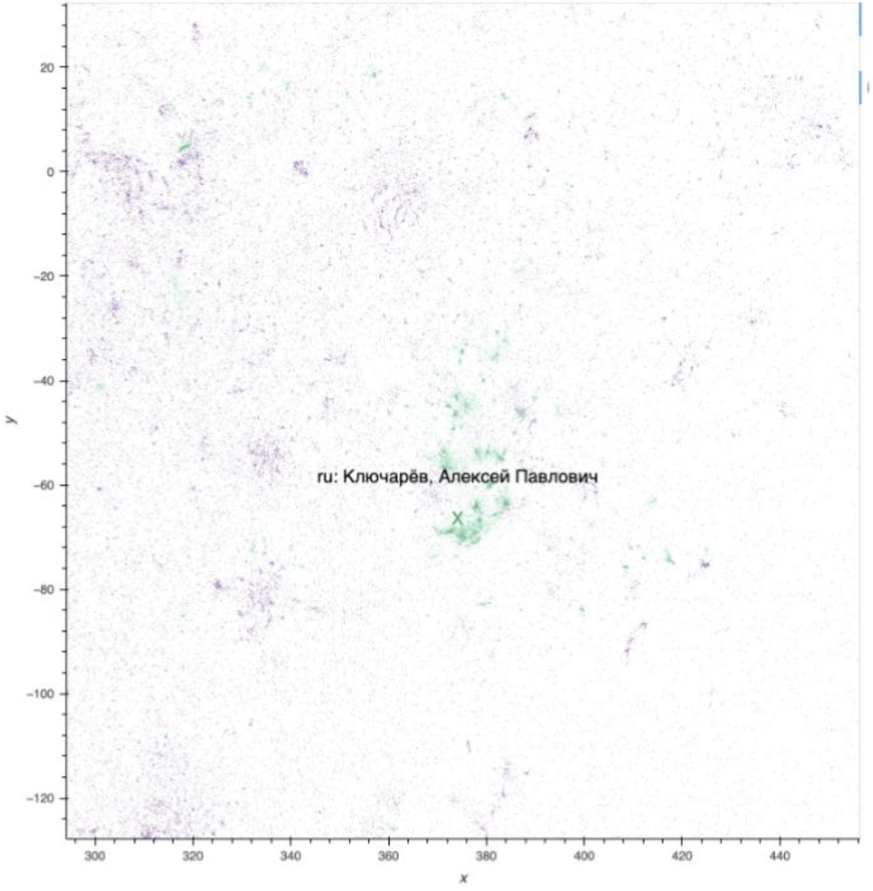
Plot [edit]

Psicharpax, the Mouse-Prince, having escaped a hunting cat, stops by the shore of a lake to drink, and encounters the Frog King Physignathus. Physignathus offers to show Psicharpax his kingdom, on the other side of the lake, and the Mouse agrees. Psicharpax climbs onto the Frog King's back, and Physignathus





600
400
200
0
-200
-400
-600
y
-600
-400
-200
0
200
400
600
x

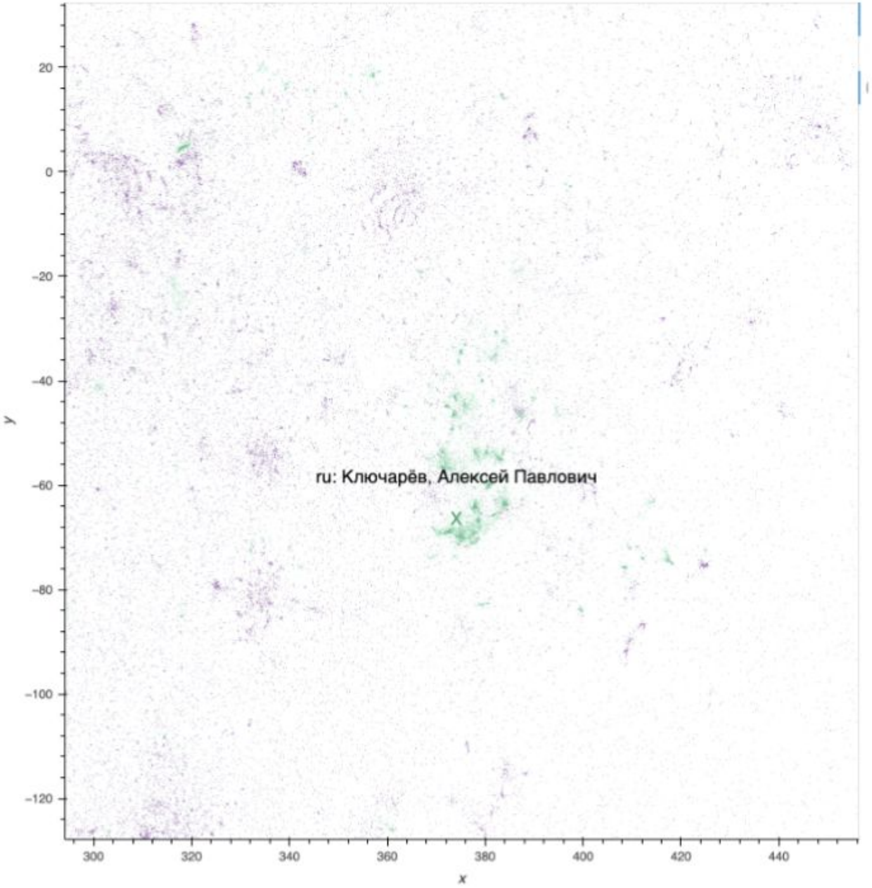


ru: Ключарёв, Алексей Павлович
X
20
0
-20
-40
-60
-80
-100
-120
y
300
320
340
360
380
400
420
440
x

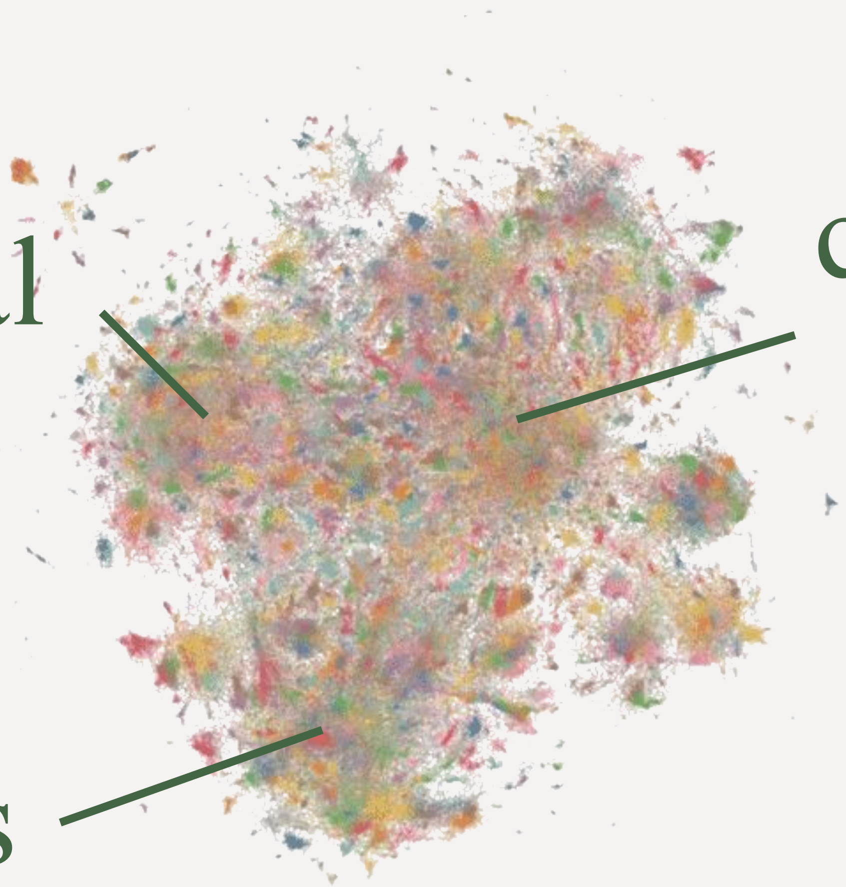
Алексей Павлович Ключарёв

Дата рождения	15 (28) сентября 1910
Место рождения	с. Соловьёво, Елецкий уезд, Орловская губерния
Дата смерти	24 июня 1997 (86 лет)
Место смерти	Харьков

"...in 1943-1944, he was the head of the physics department at the Kharkov Engineering and Technical Institute…"

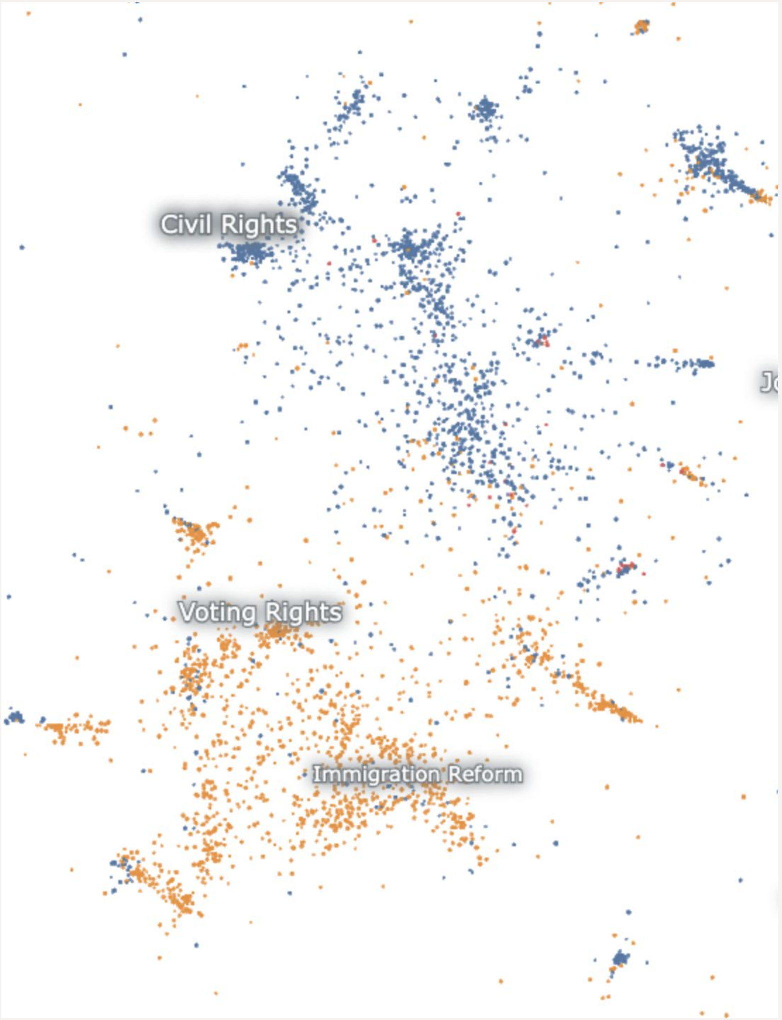


Twitter (prelon)

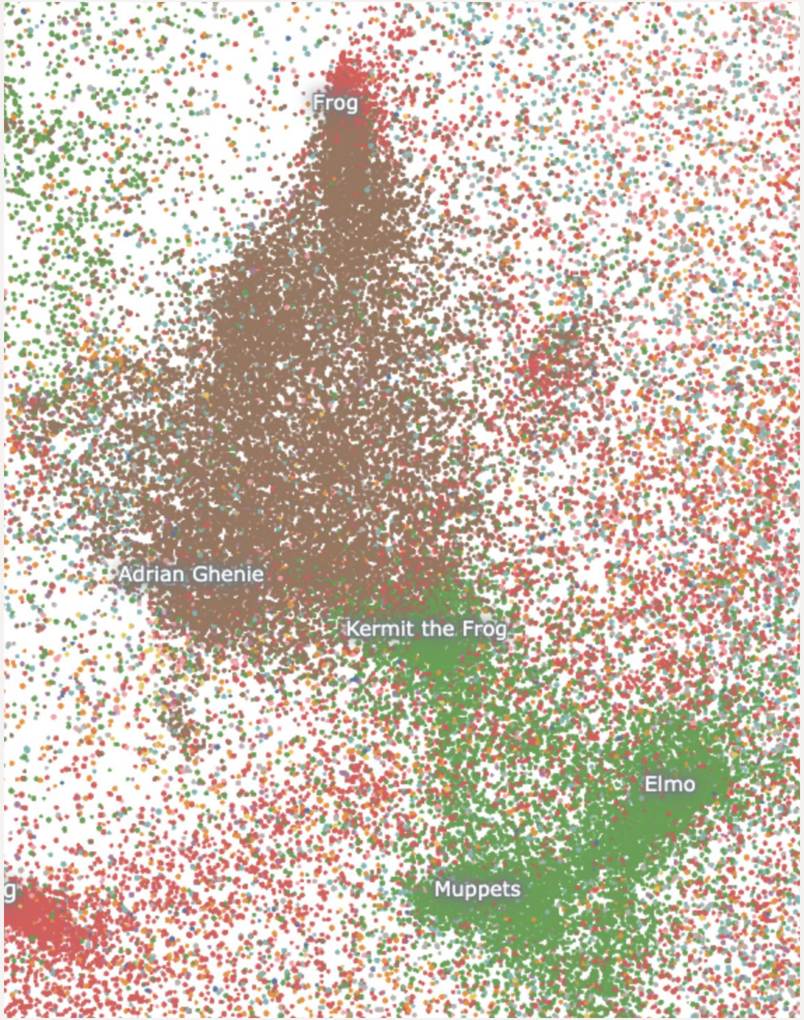

Inter-
personal
commercial
media
politics

Query Stream Analysis

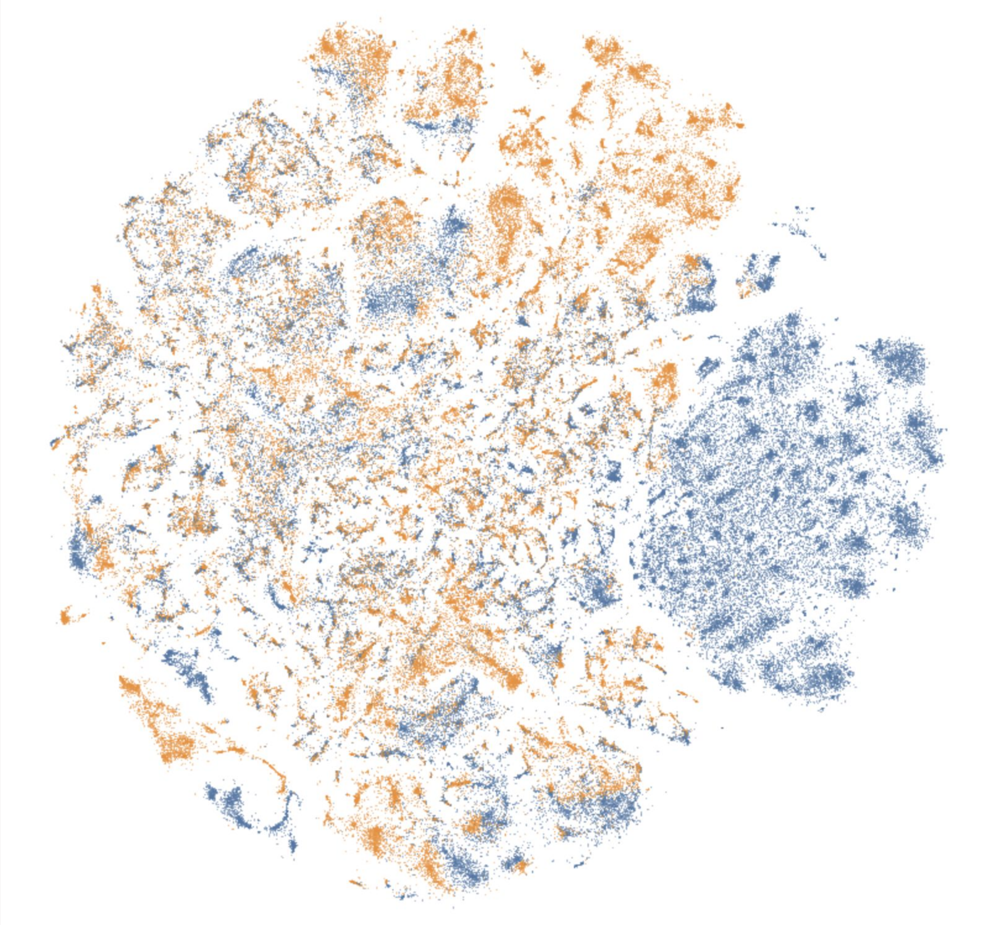
Axes of Variation



The Muppet Axis



Encoder Dependence



Questions?