# A Semantic Landscape of the Last.fm Music Folksonomy

**Joseph Biberstine, Russell J. Duhon, Elisha Allgood, Katy Börner**
*Cyberinfrastructure for Network Science Center,*
*School of Library and Information Science,*
*Indiana University*
**André Skupin**
Department of Geography,
*San Diego State University*

INDIANA UNIVERSITY

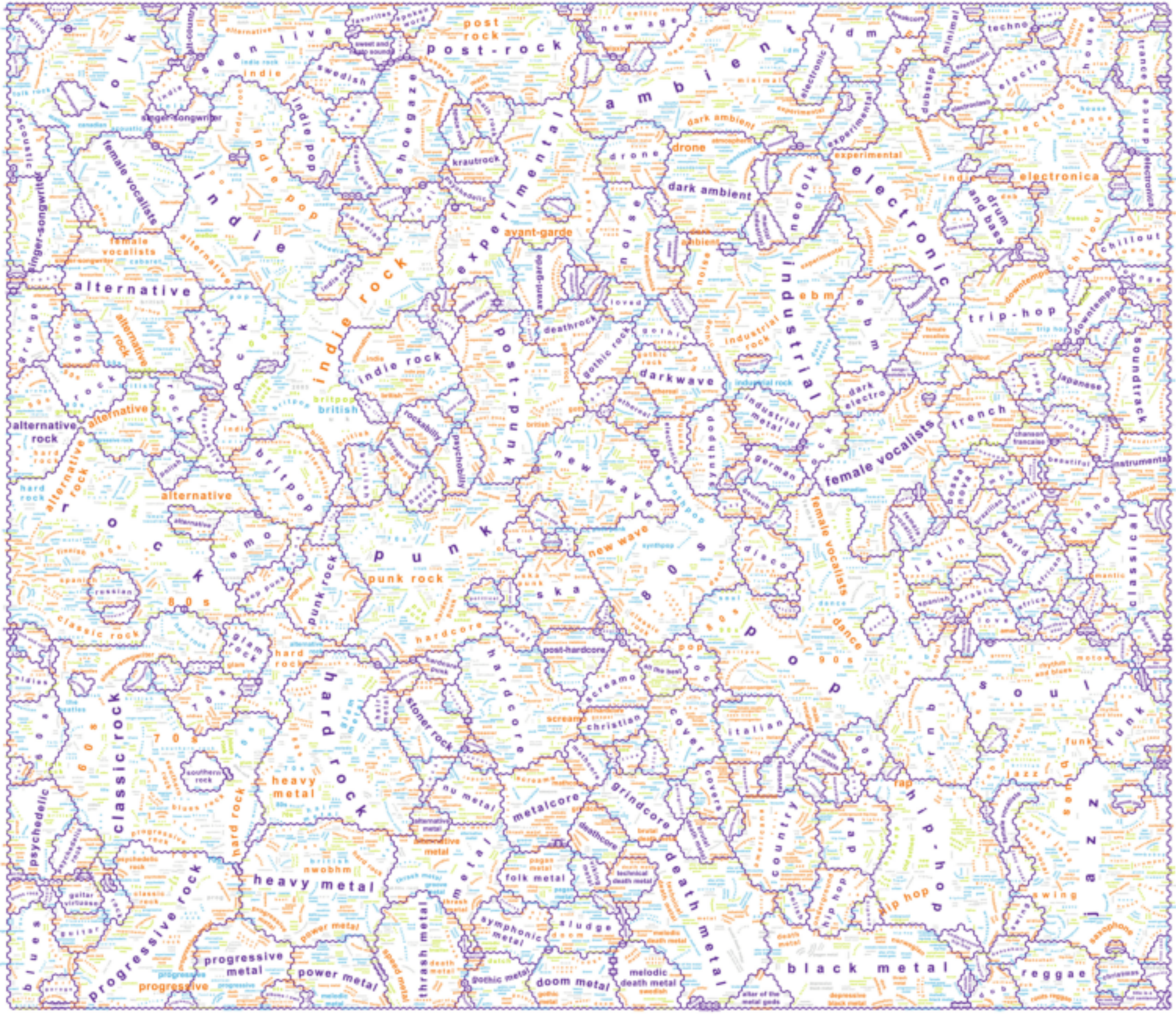
cyberinfrastructure for NETWORK SCIENCE CENTER
cns.iu.edu

SAN DIEGO STATE UNIVERSITY

# Motivation

- Domain
    - How is the world of music and music experience organized?
        - What kinds of themes emerge in this domain and what is their structure?

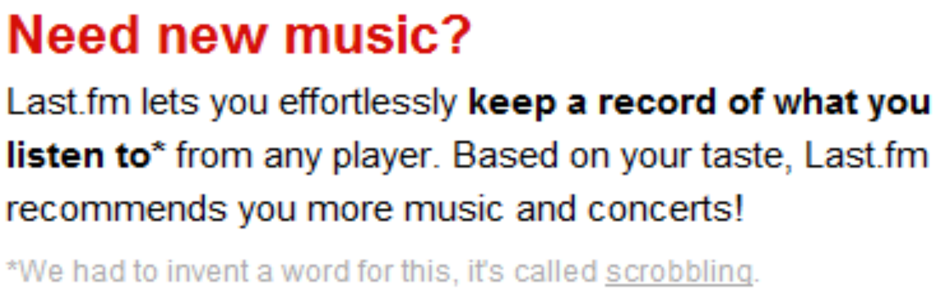
- Challenges
    - Collect and prepare high-dimensional social data
    - Create a model large enough to faithfully represent the domain
    - Train a model of this substantial size
    - Design a visualization that does justice to the richness of the model


folk
seen live
post rock
post-rock
new age
idm
techno
trance
swedish
shoegaze
ambient
singer-songwriter
indie
indie pop
krautrock
experimental
dark ambient
drone
electronic
minimal
dubstep
electro
house
electronica
drum and bass
chillout
female vocalists
avant-garde
noise
neofolk
industrial
ebm
trip-hop
downtempo
alternative
indie rock
deathrock
gothic rock
darkwave
post-punk
industrial rock
industrial metal
dark electro
soundtrack
japanese
french
chanson francaise
instrumental
alternative rock
britpop
british
rockabilly
psychobilly
new wave
synthpop
german
female vocalists
rock
emo
punk
punk rock
ska
80s
disco
dance
latin
world
bossa nova
piano
classical
spanish
arabic
african
russian
80s
classic rock
hardcore
post-hardcore
screamo
christian
covers
pop
90s
soul
funk
rhythm and blues
glam rock
hard rock
stoner rock
60s
70s
southern rock
heavy metal
nu metal
metalcore
grindcore
italian
country
rap
hip-hop
hip hop
jazz
swing
saxophone
psychedelic
progressive rock
nwobhm
alternative metal
metal
folk metal
pagan metal
deathcore
technical death metal
death metal
blues
guitar
guitar virtuoso
progressive metal
progressive
power metal
speed metal
thrash metal
symphonic metal
gothic metal
doom metal
sludge
melodic death metal
black metal
reggae

# Raw Data - Source

- Last.fm is a social Internet radio site
  - Users share information about songs they are listening to
  - They can also tag songs
    - With any strings of text they like

last.fm



**Need new music?**

Last.fm lets you effortlessly **keep a record of what you listen to*** from any player. Based on your taste, Last.fm recommends you more music and concerts!

*We had to invent a word for this, it's called scrobbling.

# Raw Data - Summary

- Gathered during the first half of 2009
- 99,405 registered users
  - 52,452 active
- 281,818 tags
- 1,393,559 songs

- 10,936,545 annotations
  - An annotation is a (user, tag, song) triple, a tagging event

Data originally collected for:
Schifanella, R., Barrat, A., Cattuto, C., Markines, B., and Menczer, F. (2010). Folks in Folksonomies: Social Link Prediction from Shared Metadata. Proc. 3rd ACM International Conference on Web Search and Data Mining (WSDM).

# Top Tags

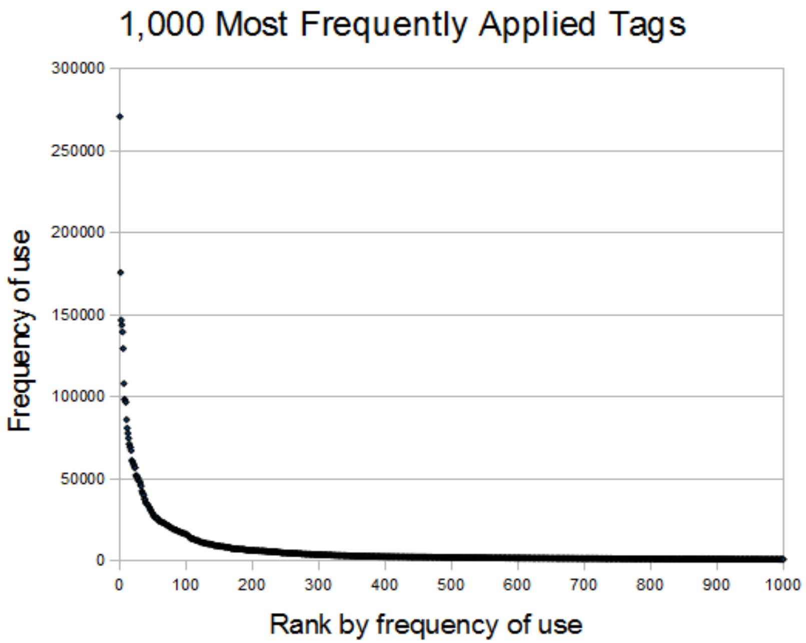
*rock*
*electronic*
*seen live*
*indie*
*alternative*
*pop*
*female vocalists*
*jazz*
*classic rock*
*experimental*
*ambient*
*metal*
*alternative rock*
*singer-songwriter*
*80s*
*folk*
*hard rock*
*progressive rock*
*indie rock*
*electronica*
*punk*
*instrumental*
*soul*
*black metal*
*industrial*
*death metal*
*heavy metal*
*chillout*
*dance*
*british*
*90s*
*psychedelic*
*blues*
*hip-hop*
*post-rock*
*new wave*
*soundtrack*
*classical*
*00s*

# Tags Are More Than Just Genres

- Intensional
    - From recognized genres to simple objective facts
        - *rock* (rank 1)
        - *electronic* (2)
        - ..
        - *female vocalists* (7)
            - *female vocalist* (64)
        - *acoustic* (51)
        - ..
        - *title is a full sentence* (101)
- Extensional
    - A mix of social signals, properties of the user-song experience, and aides to personal categorization
        - *seen live* (3)
        - *beautiful* (48)
        - *favorites* (54)
        - *albums i own* (97)
        - *altar of the metal gods* (58)
            - A case of graffiti?

# Raw Data - Thresholding

- The self-organizing map (SOM) method will not scale to 280,000+ tags/dimensions in raw form
  - Not often used with more than hundreds of dimensions
- Consider only the 1,000 most frequently applied tags
  - Keep only songs annotated by some user with any of these tags



# Thresholded Data - Summary

| | **Raw** | **Thresholded** |
|---|---|---|
| Tags | 281,818 | 1,000 |
| Songs | 1,393,559 | 1,088,761<br>(78% of original) |
| Annotations per song<br>(average) | 7.8 | 6.8 |

# Approach

- Characterize each song as a vector over each tag dimension
    - Each coordinate is the number of annotations
        - Summed across users

- A song is a piece of tag relationship evidence

# Method - Background

- Self-organizing maps
  - Neural network training algorithm
  - Unsupervised
  - High-dimensional data

    ⇩

    Low-dimensional discrete geometric model

  - Goal:
    - Proximity in the input space

      ⇩

      Proximity on the map

# Self-Organizing Map Algorithm - Classical

1. Create a lattice of neurons
2. To each neuron assign an initial (often random) vector with as many dimensions as the training data
3. For each training vector:
    1. Identify the neuron of minimal distance according to the input space metric (the "best-matching unit")
    2. For each neuron:
        1. Pull this neuron's vector toward the training vector in proportion with this neuron's distance from the best-matching unit

# Self-Organizing Map Algorithm - Parallelized Implementation

- A previous project trained on twice as many data and twice as many dimensions
    - Completely intractable using widely available software
    - Created our own implementation
        - Divide the training data among multiple processes
            - Each process holds a complete copy of the map
            - Periodically synchronize process-local copies of the map to create a new process-global map

- Adapted with several project-specific optimizations from:
    - Lawrence, R.D., Almasi, G.S., Rushmeier, H.E. (1999). A Scalable Parallel Algorithm for Self-Organizing Maps with Applications to Sparse Data Mining Problems. Data Mining and Knowledge Discovery

# Training the Map

- 2D hexagonal lattice of neurons
  - 180 on either side = 32,400 altogether

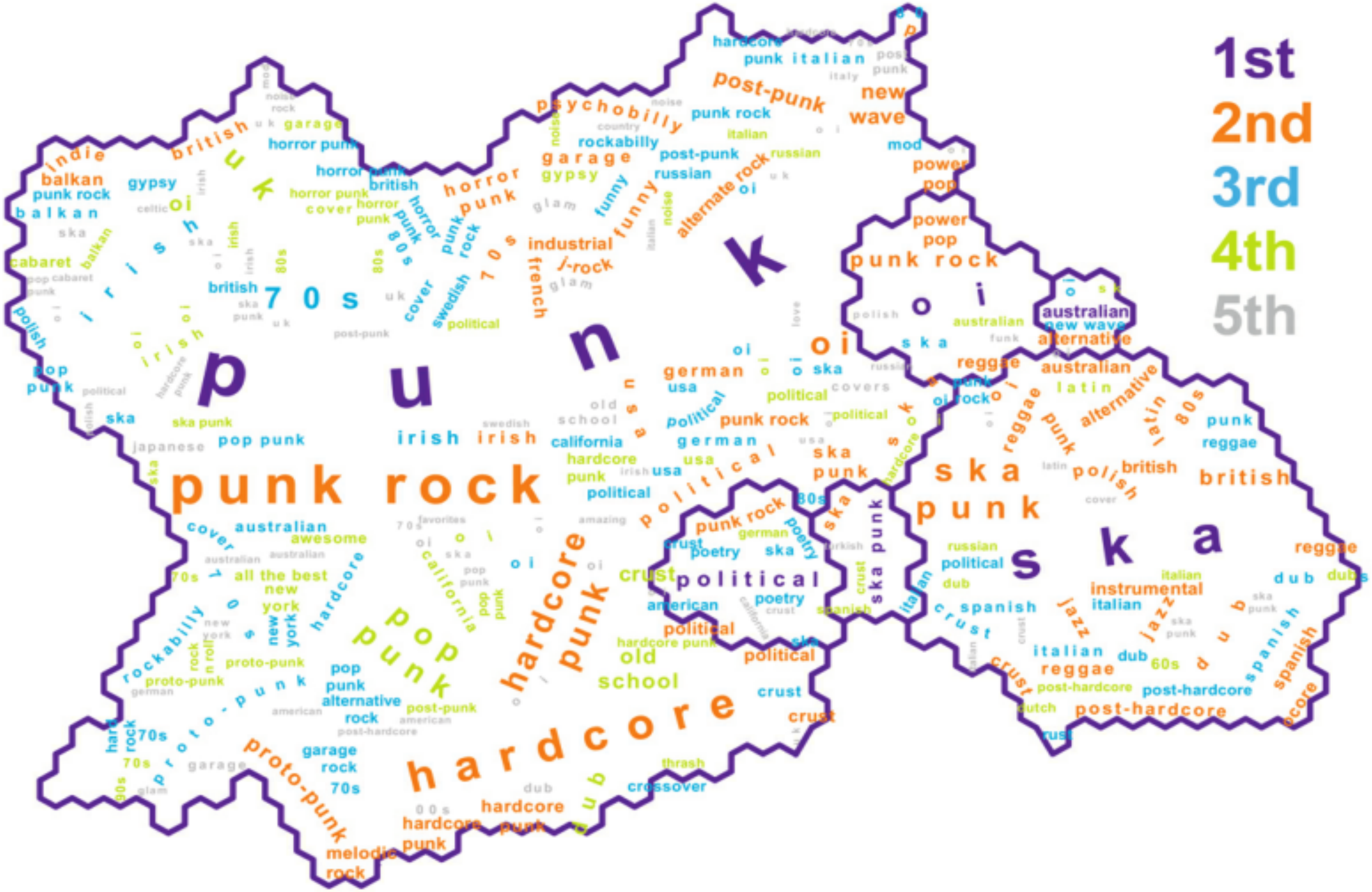
- Input space metric: **cosine similarity**
  - Induced interpretation: Each training vector (and so consequently each neuron vector) represents a direction in the 1,000-dimensional tag space
  - $$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

- 50 complete passes over the training data

# Computation

- 300 processes across 100 compute nodes of Big Red, a supercomputer at Indiana University
- Parallel runtime = 13 hours
    - Serial equivalent runtime = **5 months**

# Legend



1st
2nd
3rd
4th
5th

INDIANA UNIVERSITY

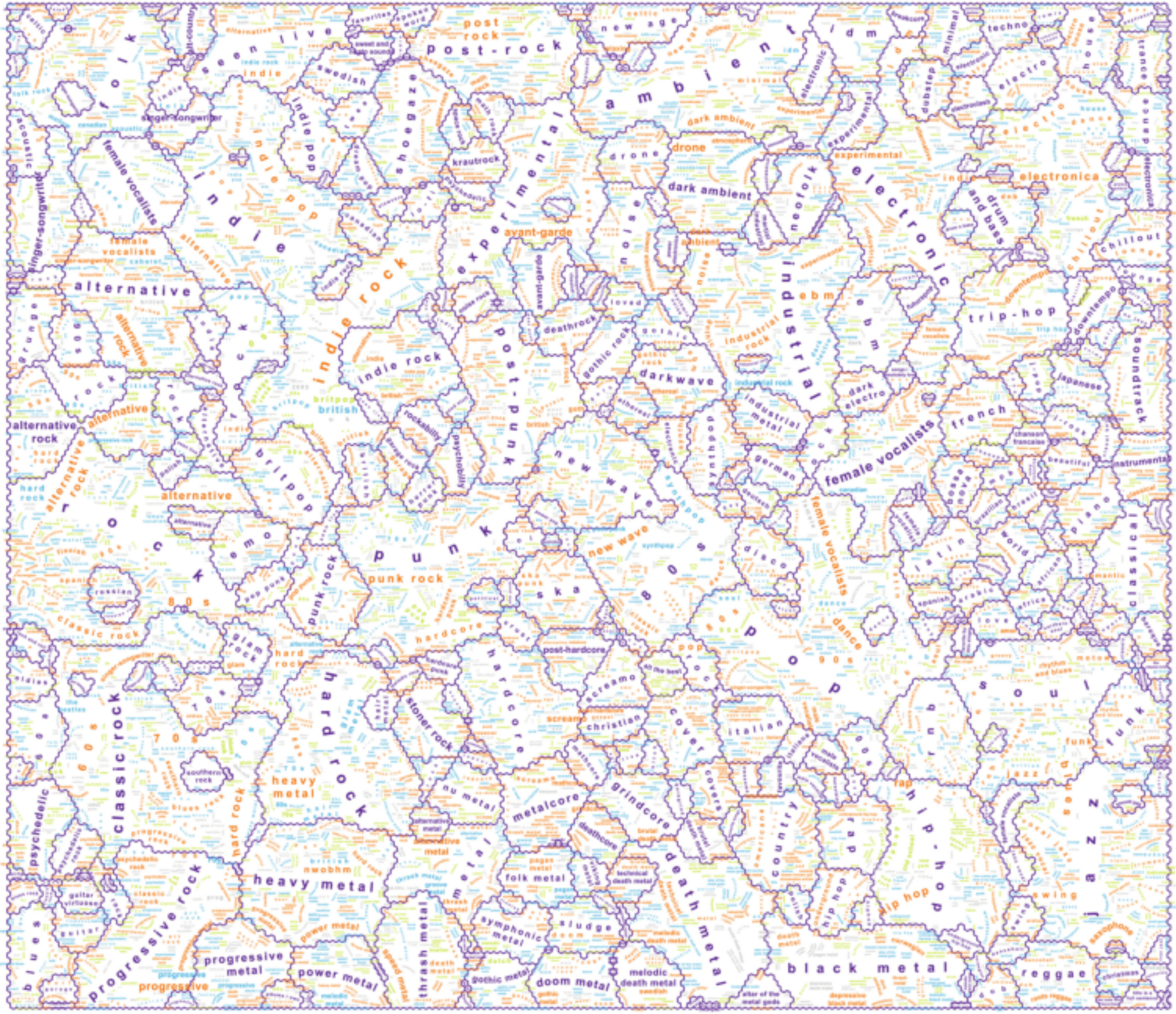
cyberinfrastructure for NETWORK SCIENCE CENTER
cns.iu.edu

SAN DIEGO STATE UNIVERSITY

# Visualization

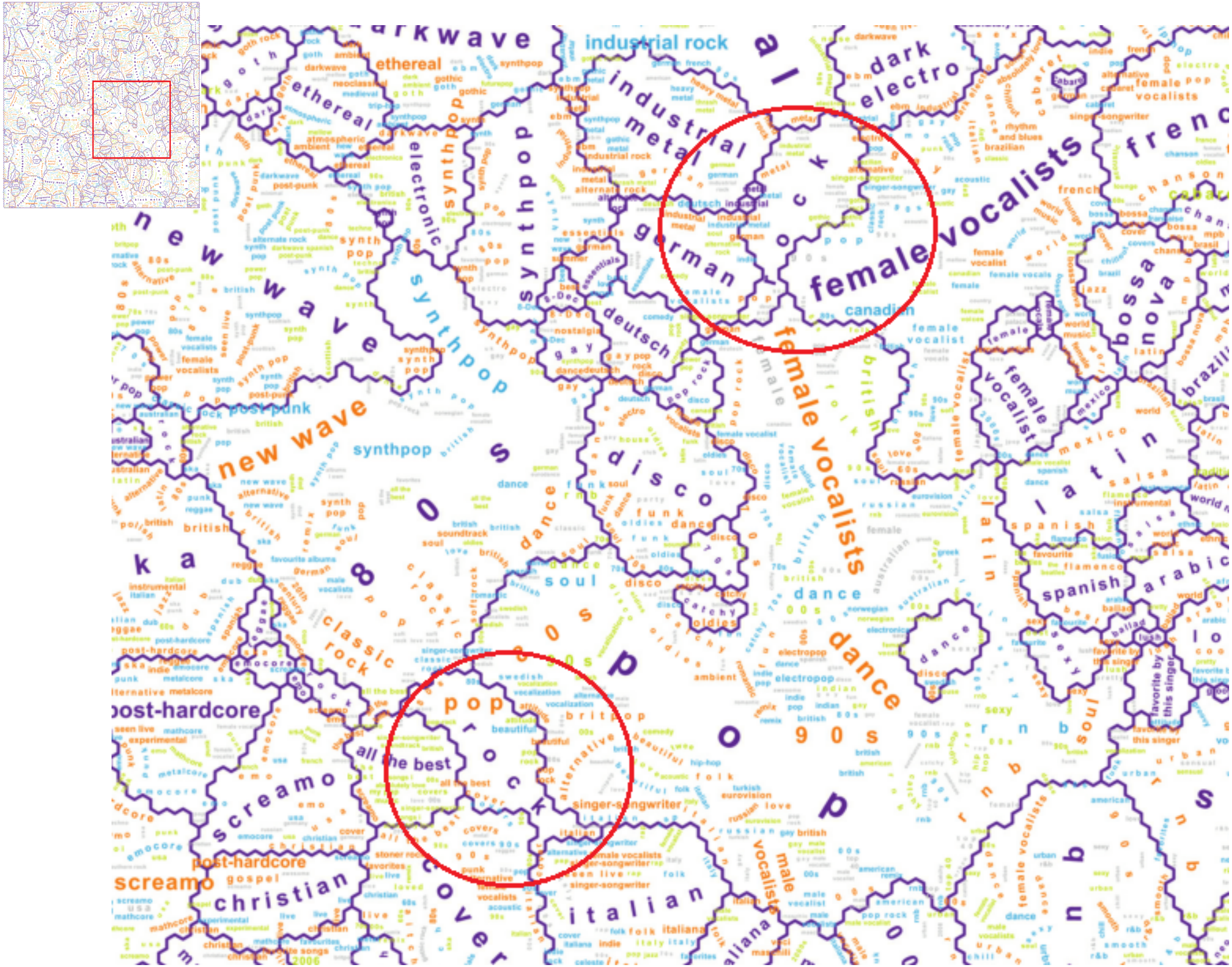
- □Recall there is a corresponding vector to each neuron which describes its position in the input space
- In other words, its position along each tag dimension

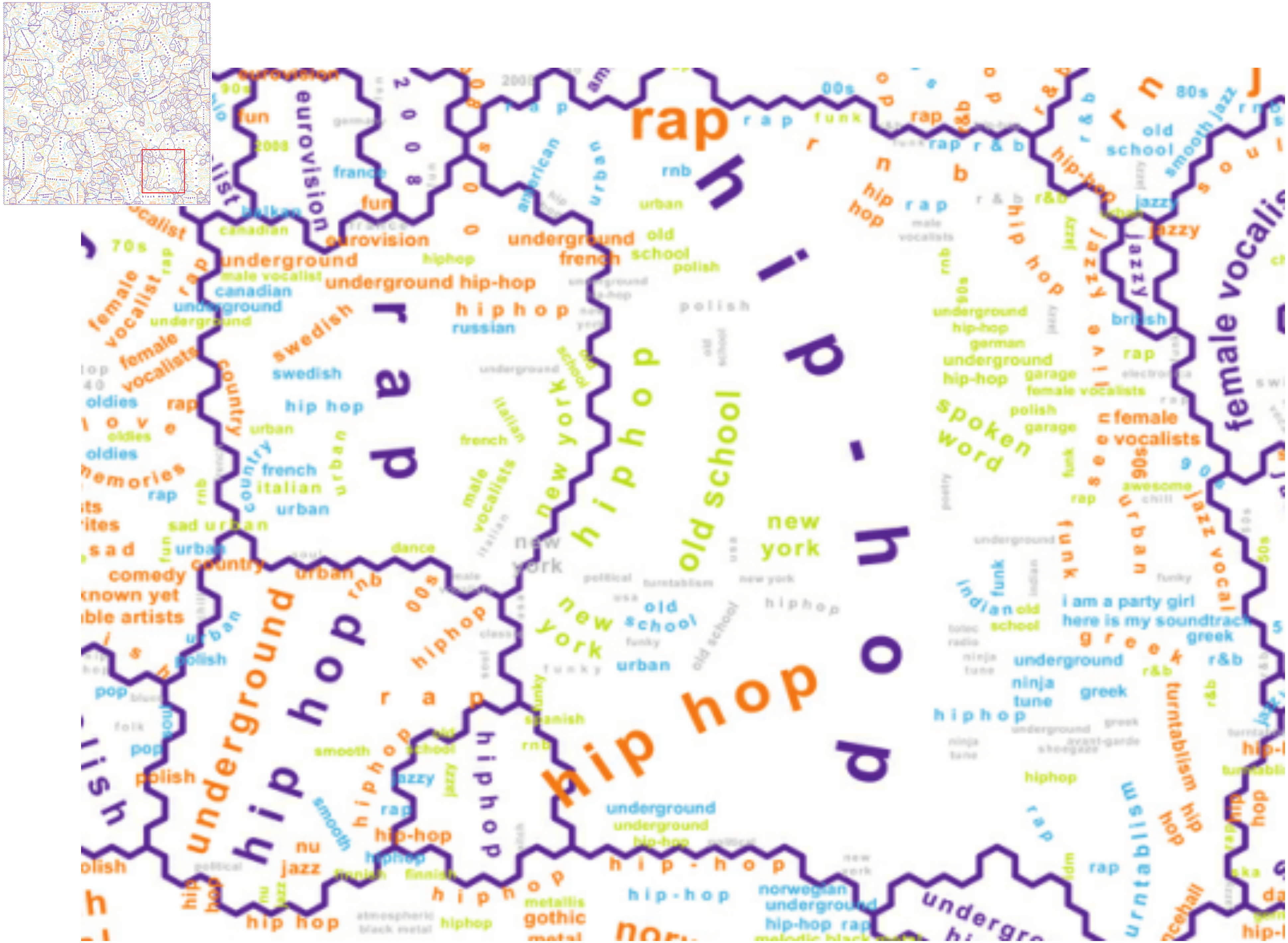
- Consider the nth strongest tag association of each neuron
- A contiguous swath of neurons sharing a common nth strongest tag association is termed a region
- As the map is trained over 1,000 tags, we have 1,000 distinct partitions of the map into such regions

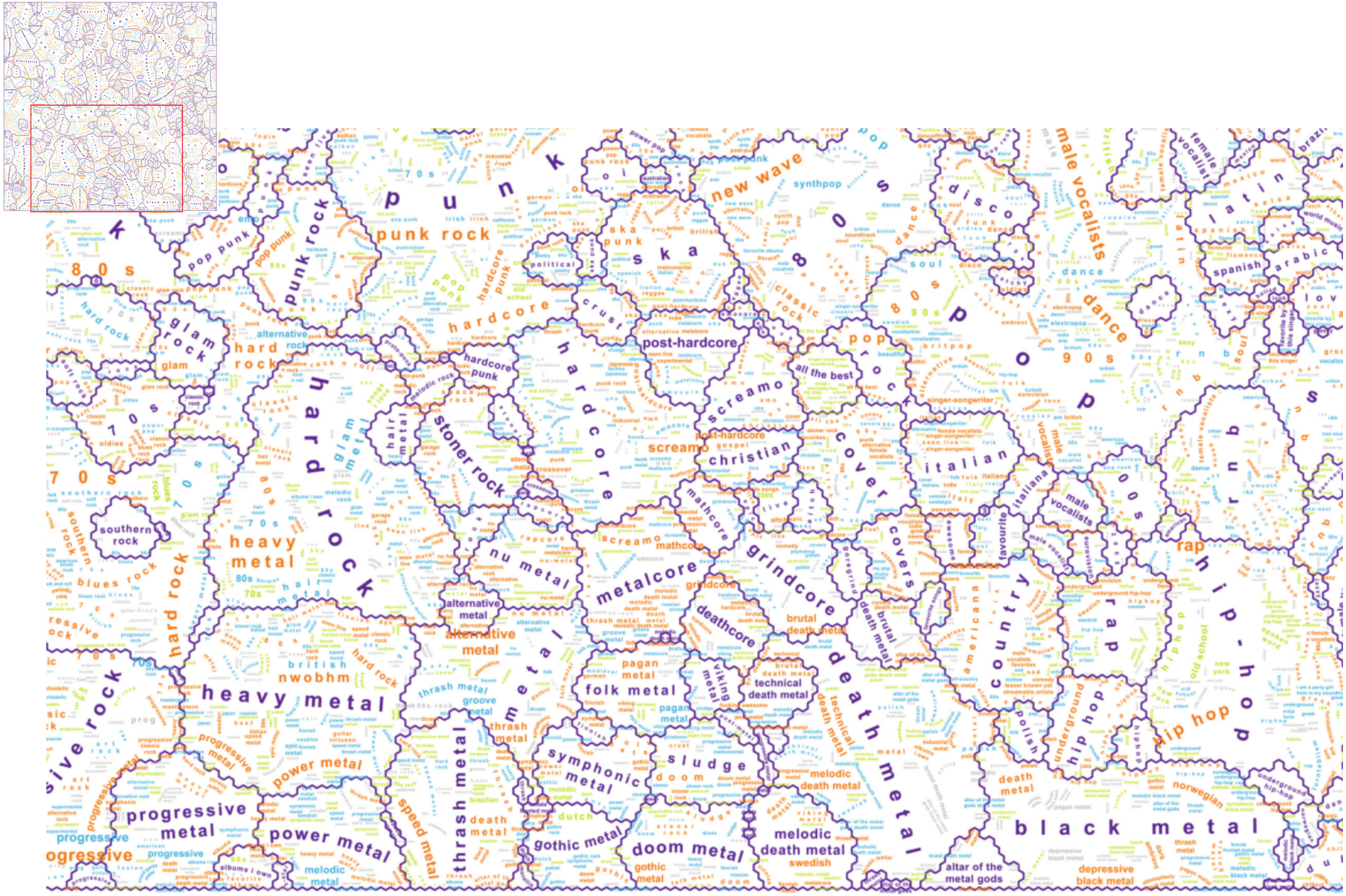

folk
seen live
indie
singer-songwriter
female vocalists
acoustic
alternative
alternative rock
post rock
post-rock
swedish
indie pop
shoegaze
krautrock
experimental
avant-garde
ambient
new age
drone
dark ambient
neofolk
noise
idm
electronic
minimal
dubstep
techno
electro
house
trance
electronica
drum and bass
chillout
trip-hop
downtempo
soundtrack
industrial
ebm
dark electro
indie rock
post-punk
deathrock
gothic rock
darkwave
industrial rock
industrial metal
japanese
french
chanson francaise
instrumental
female vocalists
rock
hard rock
grunge
90s
britpop
british
rockabilly
psychobilly
new wave
synthpop
german
punk
punk rock
emo
pop punk
ska
80s
disco
dance
pop
latin
world
spanish
africa
bossa nova
piano
classical
hardcore
post-hardcore
screamo
christian
cover
covers
90s
soul
funk
rhythm and blues
jazz
classic rock
60s
70s
oldies
the beatles
glam rock
southern rock
blues rock
heavy metal
stoner rock
nu metal
metalcore
grindcore
deathcore
brutal death metal
technical death metal
death metal
italian
country
rap
hip-hop
hip hop
swing
saxophone
reggae
black metal
depressive black metal
psychedelic
psychedelic rock
guitar virtuoso
guitar
blues
progressive rock
progressive metal
progressive
power metal
speed metal
thrash metal
nwobhm
british metal
metal
alternative metal
hair metal
symphonic metal
gothic metal
folk metal
pagan metal
sludge
doom metal
melodic death metal
altar of the metal gods

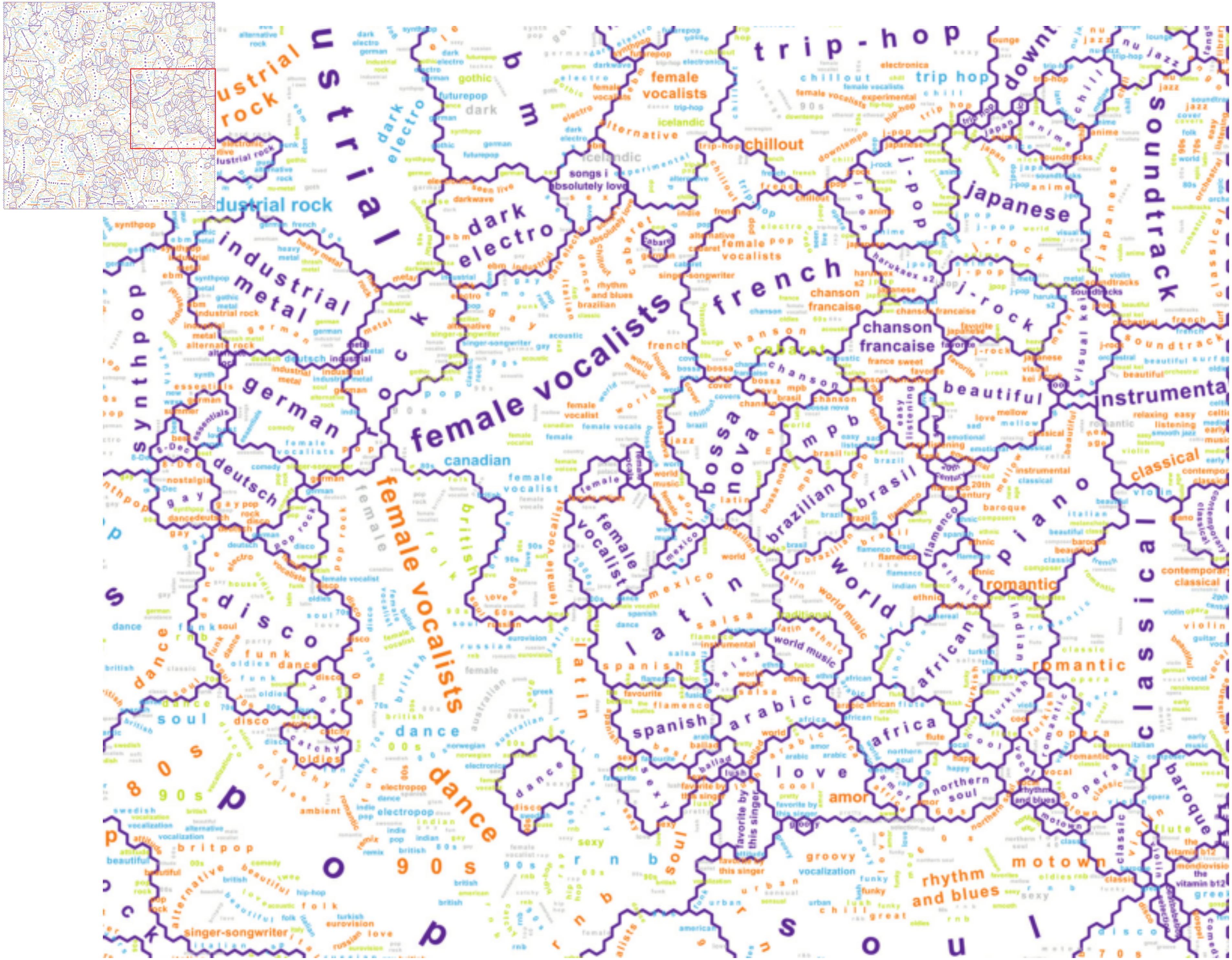
# Interpretation

- Interpreters report a mix of
  - Recognition
    - Patterns of hierarchical and neighborhood relationships among tags match expectations
  - Discovery
    - Opportunities to find new musical categories
  - Surprise
    - Relationship between *rock*, *blues*, and *jazz*


darkwave
industrial rock
goth
ethereal
dark electro
industrial metal
electronic
synthpop
french
female vocalists
german
rock
pop
female
canadian
new wave
synthpop
deutsch
chanson
bossa nova
post-punk
new wave
synthpop
female vocalists
british
female vocalist
disco
ska
80s
dance
soul
latin
spanish
mexico
salsa
arabic
classic rock
dance
electropop
90s
pop
00s
pop rock
all the best
singer-songwriter
beautiful
alternative
britpop
rnb
soul
post-hardcore
screamo
emocore
christian
cover
italian
male vocalists
italiana
rnb
latin
brazilian
world


eurovision
90s
fun
2008
france
balkan
canadian
2 0 0 8
rap
rnb
hip-hop
hip hop
hiphop
underground
underground hip-hop
male vocalist
canadian
female vocalist
female vocalists
swedish
russian
h i p h o p
old school
polish
new york
urban
country
french
italian
male vocalists
oldies
memories
love
70s
sad
comedy
r a p
r & b
r&b
00s
80s
jazz
jazzy
british
soul
female vocalists
spoken word
german
garage
funk
90s
awesome
chill
i am a party girl here is my soundtrack
greek
ninja tune
turntablism
indian
dance
smooth
jazzy
spanish
polish
pop
nu jazz
finnish
metallic
gothic metal
atmospheric black metal
hip-hop
norwegian
underground hip-hop
melodic black metal
dancehall
ska
h i p - h o p
underground


punk
punk rock
pop punk
punk rock
new wave
synthpop
ska
ska punk
political
80s
hardcore
hardcore punk
crust
post-hardcore
classic rock
disco
soul
dance
male vocalists
female vocalists
latin
spanish
arabic
80s
glam rock
hard rock
hard rock
glam metal
hair metal
stoner rock
hardcore
screamo
all the best
pop rock
pop
90s
singer-songwriter
rnb
70s
southern rock
blues rock
70s
heavy metal
christian
screamo
post-hardcore
cover
covers
italian
italiana
male vocalists
00s
rap
hip-hop
mathcore
nu metal
metalcore
grindcore
deathcore
brutal death metal
favourite
americana
country
alternative metal
alternative metal
british
nwobhm
heavy metal
thrash metal
groove metal
metal
pagan metal
folk metal
viking metal
technical death metal
death metal
technical death metal
polish
underground hip hop
rap
hip hop
hip hop
old school
progressive rock
progressive metal
power metal
power metal
speed metal
thrash metal
thrash metal
death metal
symphonic metal
sludge
doom
gothic metal
doom metal
gothic metal
melodic death metal
melodic death metal
swedish
altar of the metal gods
death metal
black metal
depressive black metal
norwegian
underground hip-hop
progressive
progressive


industrial
industrial rock
ebm
dark electro
trip-hop
trip hop
chillout
female vocalists
icelandic
alternative
songs i absolutely love
downtempo
j-pop
japanese
j-rock
anime
soundtrack
soundtracks
instrumental
classical
french
chanson francaise
chanson
cabaret
industrial metal
german
deutsch
synthpop
synth pop
disco
funk
dance
soul
80s
90s
pop
rock
female vocalists
female vocalist
female
canadian
british
bossa nova
mpb
brasil
brazilian
world
world music
latin
mexico
salsa
spanish
flamenco
arabic
africa
african
ethnic
indian
piano
romantic
baroque
opera
violin
visual kei
beautiful
love
amor
cool
northern soul
motown
rhythm and blues
sexy
rnb
groovy
soul
favorite by this singer
electropop
singer-songwriter
catchy
oldies
britpop
alternative
turkish
flute
contemporary classical
vocal

# Potential Applications

- Interactive music navigator and playlist generator
- Mapping portfolios as fields of neuronal activation
    - For the set of songs associated with any entity, we can see where in the world they belong
        - A user: Their favorite songs
        - A band: Their complete work
        - A group of users: What is their turf?
    - .. or look at the difference of any of these fields
        - What is the difference between The Who and The Guess Who?
        - How has this entity moved through the world of music over time?
        - Where have listeners like me headed next?

# Contributors

- Joseph Biberstine, Indiana University
    - Contact: jrbibers@indiana.edu
- André Skupin, San Diego State University
- Russell J. Duhon, IU
- Elisha Allgood, IU
- Katy Börner, IU

INDIANA UNIVERSITY

cyberinfrastructure for NETWORK SCIENCE CENTER
cns.iu.edu

SAN DIEGO STATE UNIVERSITY