Discovery From Hyperspectral ALMA Imagery With NeuroScope

Erzsébet Merényi[1,3], Andrea Isella[2], Josh Taylor[1]

[1]Department of Statistics
[2]Department of Physics and Astronomy
[3]Department of Electrical and Computer Engineering

Rice University, Houston, Texas

Support by

ALMA Cycle 5
Development Study

Goal: New analysis capabilities to map source regions with distinct kinematic and compositional properties

- Exploit the richness of hyperspectral ALMA (VLA, and other) data deeper than current capabilities; delineate spectrally homogeneous regions in more detail for discovery of relevant physical processes
- Visualize in one integrated view

Tool: NeuroScope, our “data scoping” computational instrument

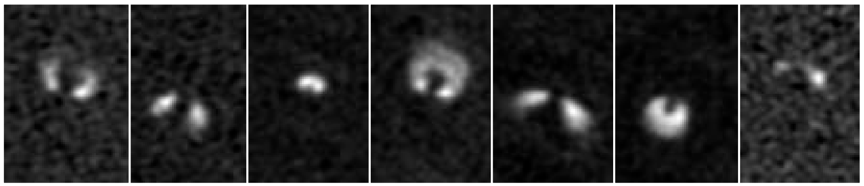
- collection of *neural map based machine learning methods* (clustering and classification) and related tools geared for high-D data with complex structure

Make it suitable for pipeline processing

- Automated
- Fast

Example: ALMA hyperspectral image – spectral variations

Image planes from ALMA Band 7, protoplanetary disk HD 142527

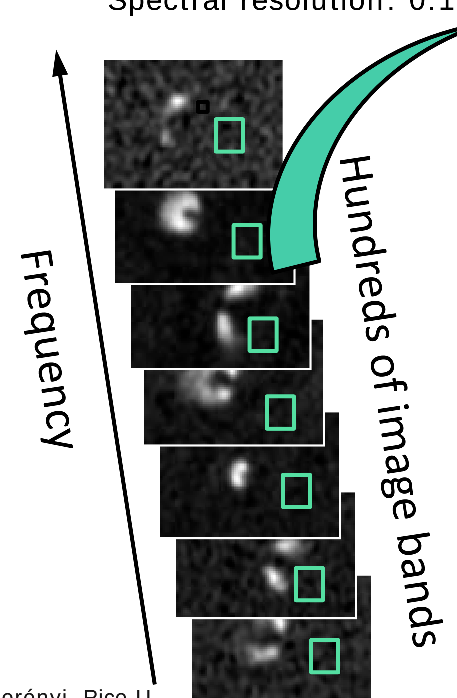


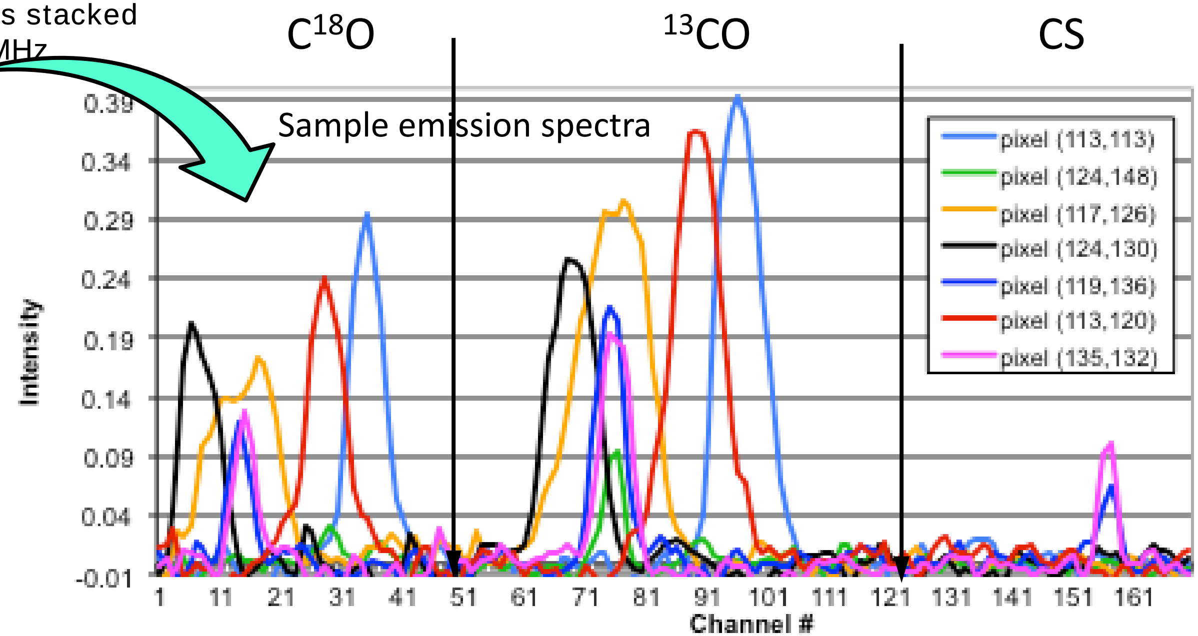
GHz 329.299-329.305 | 330.555 – 330.564 | 342.850-342.856

Ch 1 … 50 | 51 … 120 | 121 … 170

Cluster the spectral signatures to map regions of distinct kinematic and compositional behavior.

170 channels: $C^{18}O$, ^{13}CO, CS lines stacked
Spectral resolution: 0.122 MHz





ALMA spectra from combined $C^{18}O$, ^{13}CO, CS lines, showing differences in composition, Doppler shift, temperature

(Data credit: JVO, project 2011.0.00318.5)

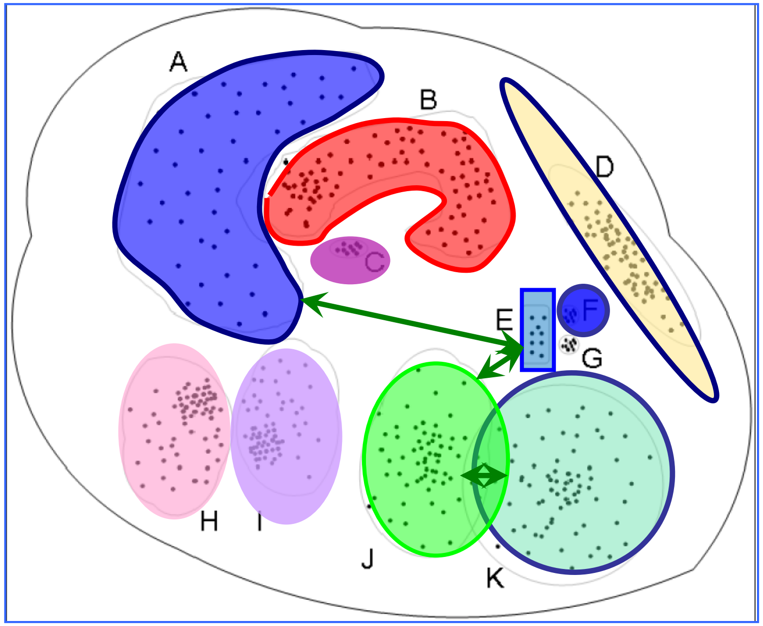
E. Merényi, Rice U
erzsebet@rice.edu

The richness is key to discovery – but creates complexity hard to exploit

- Discrimination of *__many relevant spectral types__* is expected
- Interesting phenomena may manifest in *__subtle spectral differences__*
- *__Interesting regions may be very small__* (few samples, poor statistics); many methods may miss the discovery.

Imagine in 100 dimensions!

- *__Highly structured feature space:__* many clusters of widely varying shapes, sizes, densities, … non-linear reparability, etc.



No statistical models

To faithfully learn data relations, no (or least) assumption should be made about the structure. **Let the data speak.**

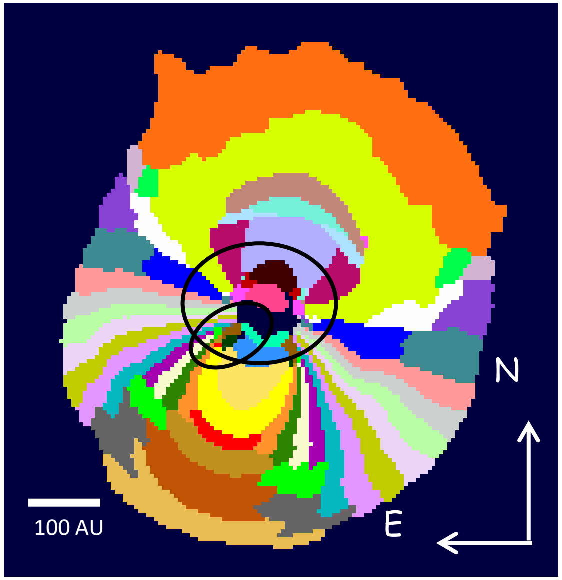
Many clustering / manifold learning methods fail to express structure faithfully.

Ex: K-means is tuned to capture spherical / ellipsoidal clusters. Can't capture irregulars.

E. Merényi, Rice U
erzsebet@rice.edu

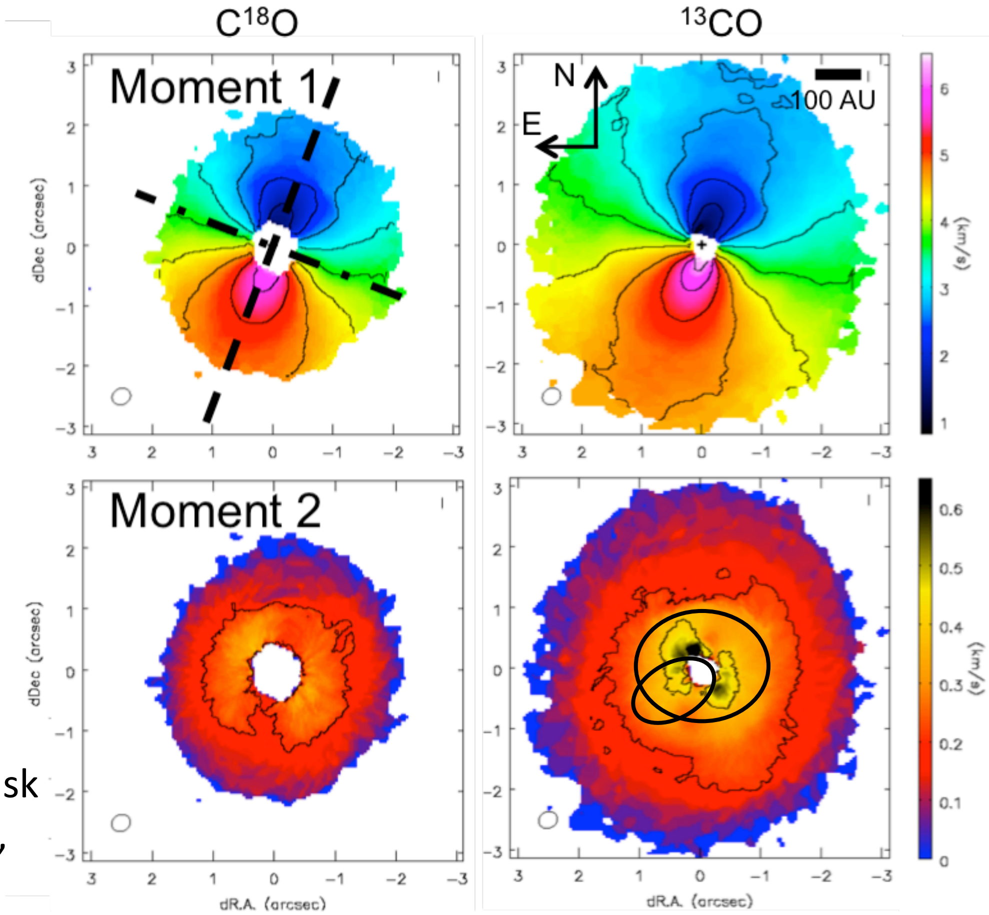
NeuroScope structure discovery from ALMA data
HD 142527 protoplanetary disk (data: Isella 2015)

NeuroScope cluster map from stacked $C^{18}O$, ^{13}CO lines, 100 + 100 channels as input feature vectors



The emerging structure of the protoplanetary disk based on all channels of two molecular tracers, visualized in one 2-D view

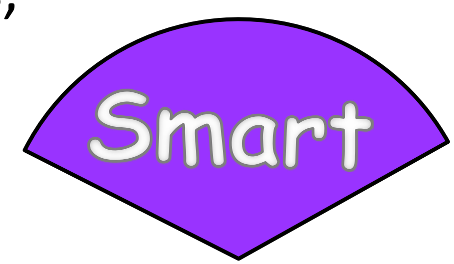
Coloring of clusters is arbitrary, not a heat map!



E. Merényi, Rice U
erzsebet@rice.edu

The NeuroScope approach

I. **Learn the data structure well** (find clusters, extract salient details) to enable discoveries



- No assumption about data distribution
- No prior dimension reduction (use all frequency channel) – to keep the discovery potential
- Using data straight out of the ALMA data reduction pipeline

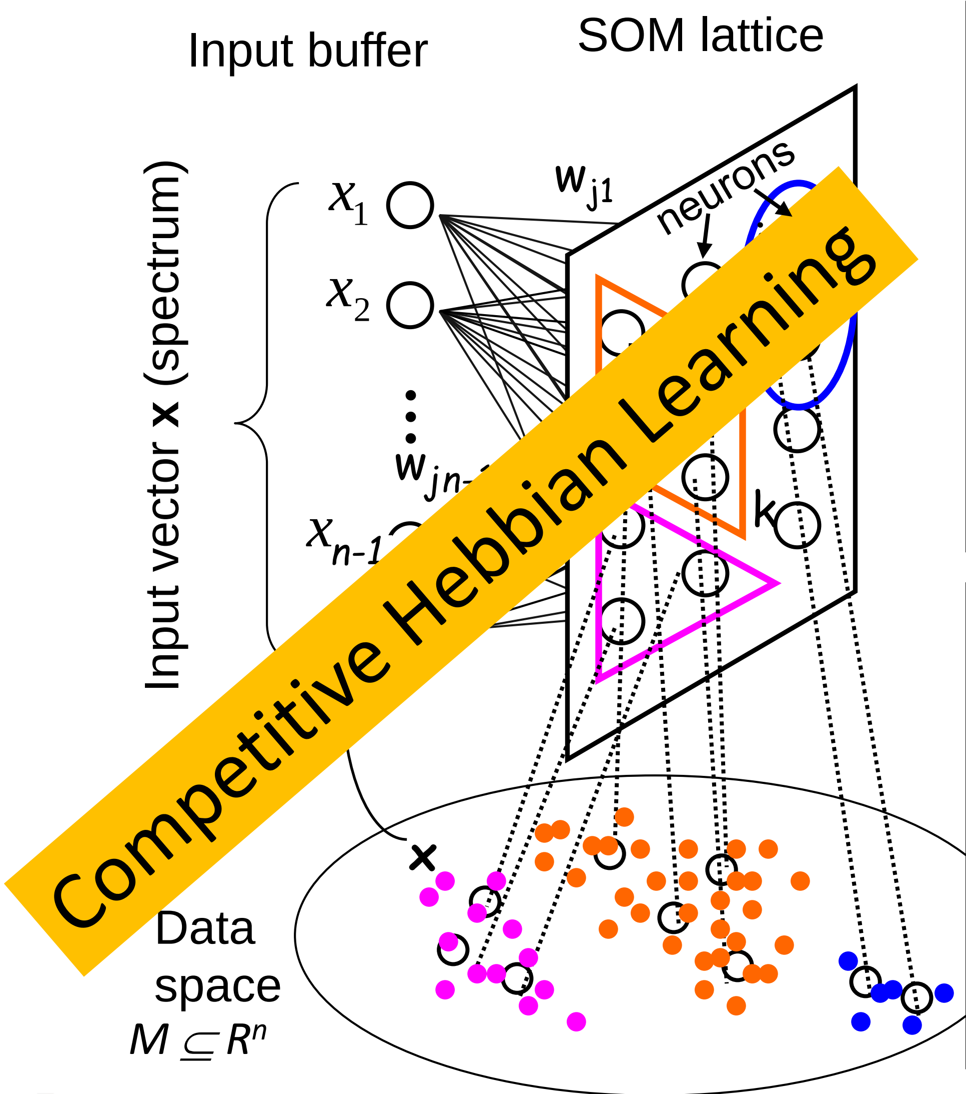
II. **Automate**

III. **Do it fast**

- to be suitable for pipelines / large archives

E. Merényi, Rice U
erzsebet@rice.edu

Part I. Learn the structure with Self-Organizing Maps
Machine learning analog of biological neural maps in the brain



Two simultaneous actions:

- Adaptive Vector Quantization (VQ), summarization of N data vectors by O(sqrt(N)) prototypes; *all neural maps do this*
- Ordering of the prototypes on the SOM grid according to similarities; *only SOMs do this*.

I.e., SOMs learns the structure (the distribution) AND expresses the topology (similarity relations) on a low-dimensional lattice.

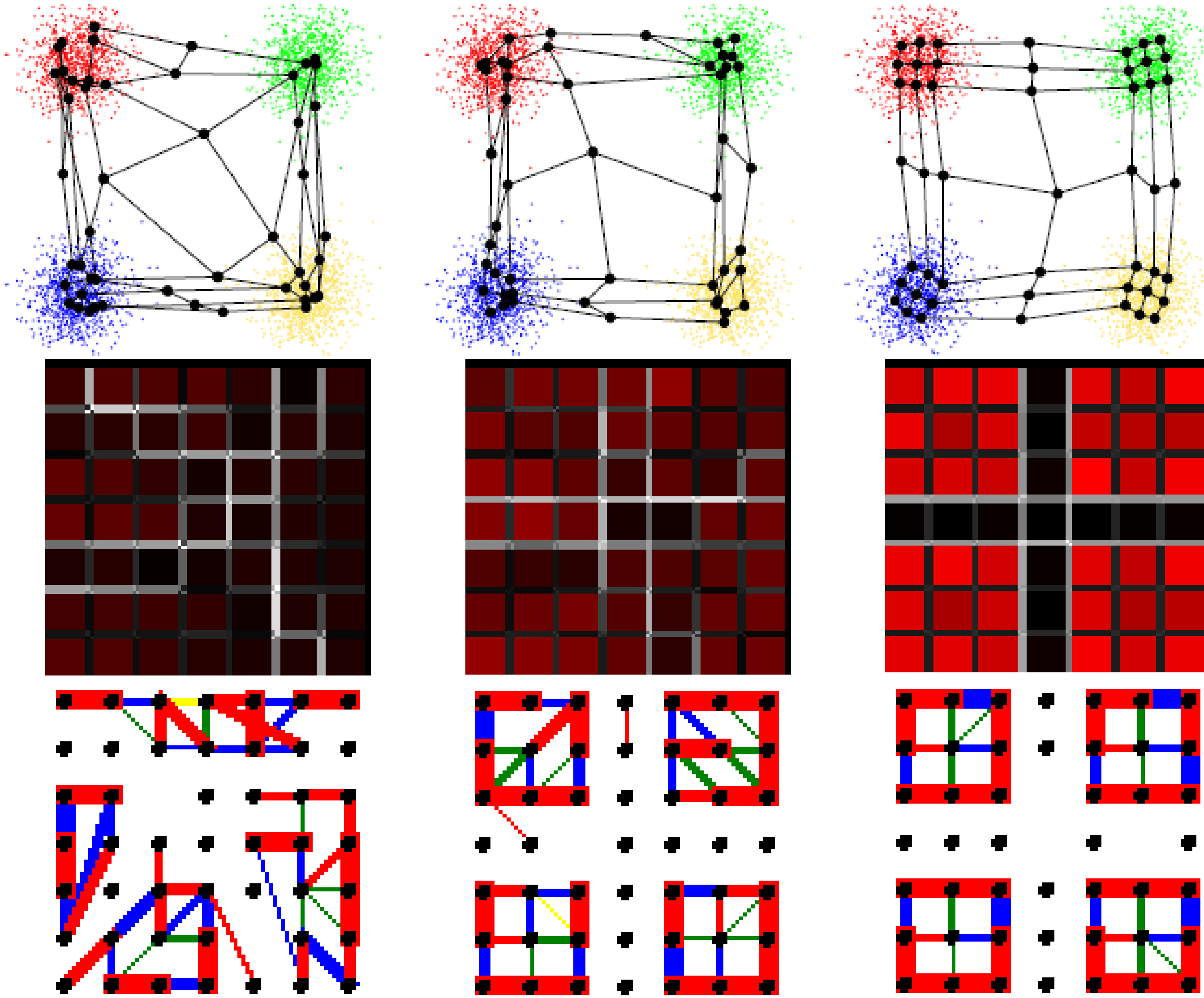
(Assuming learning goes correctly …)

Finding the prototype groups: post-processing

In SOM lattice

E. Merényi, Rice U
erzsebet@rice.edu

Toy example: unsupervised SOM learning of 4 Gaussian clusters Evolution of prototypes, and visualization



Cannot be shown for >2-D

SOM prototypes (black dots) in data space

Can be shown for >2-D

SOM knowledge visualized

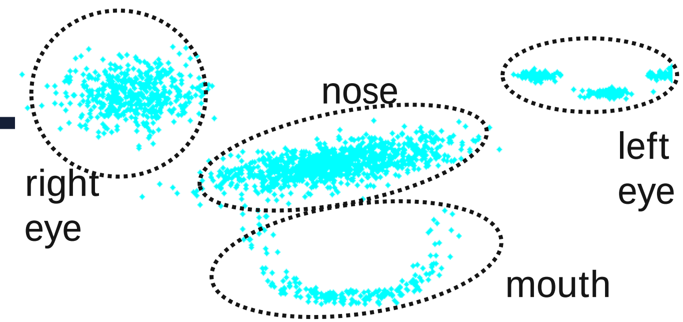
mU-matrix representation

CONN graph Representation

E. Merényi, Rice U
erzsebet@rice.edu

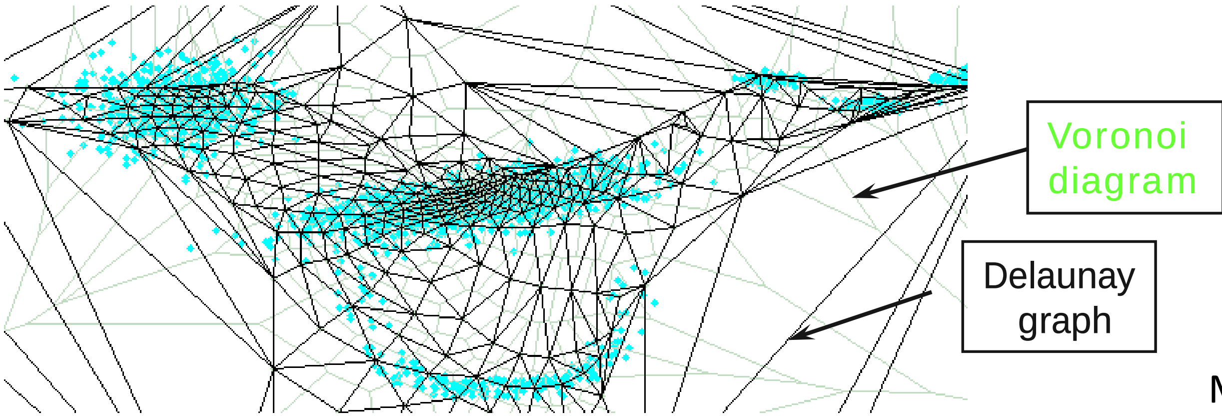
Graph representation of SOM knowledge: Induced Delaunay graph

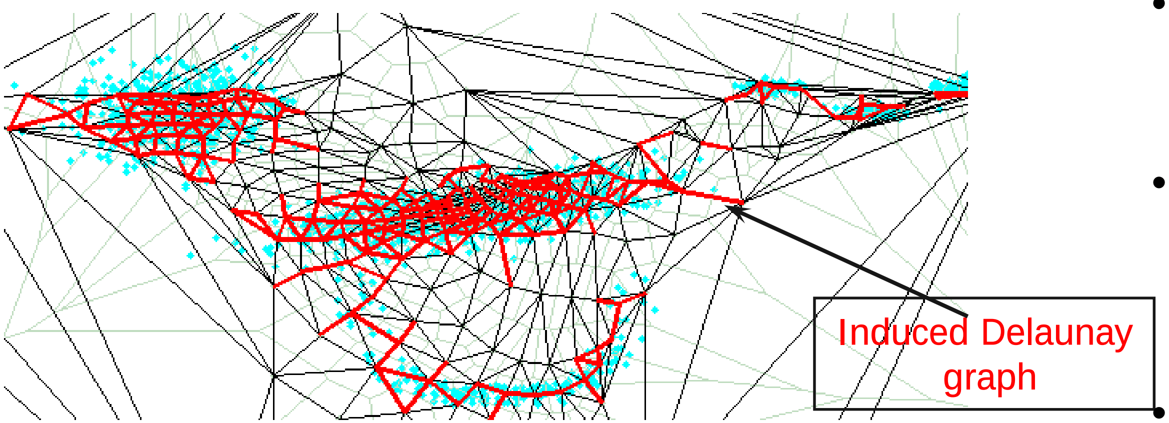
Well-learned SOM prototypes (black vertices), nicely follow the data distribution.
Placement of prototypes is crucial! (Assume correct learning.)



2-D “Clown” data

(Data: Vesanto and Alhoniemi, 2000)





(Figures from Taşdemir and Merényi, 2009)

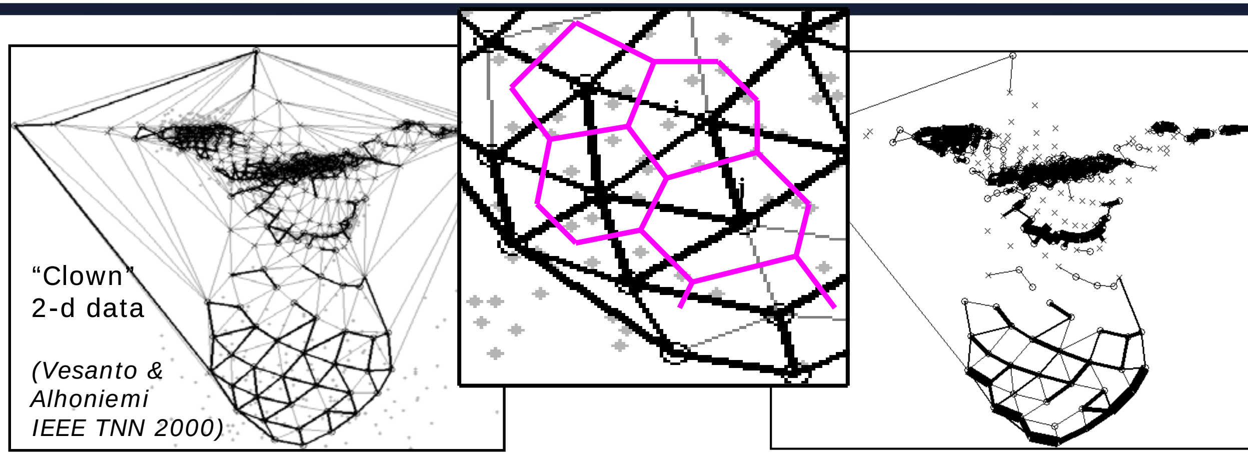
Martinetz and Schulten, 1994:

- The induced Delaunay graph perfectly represents topology - but how to get it in high-D space?
- Competitive Hebbian learning (neural maps) produces the induced Delaunay graph (with one mild condition)
- To get it: Connect two prototypes if they are closest and 2nd closest match for a data vector

Connectivity (CONN) similarity measure and graph

(Taşdemir & Merényi, IEEE TNN 2009)

Induced
Delaunay
graph
- binary



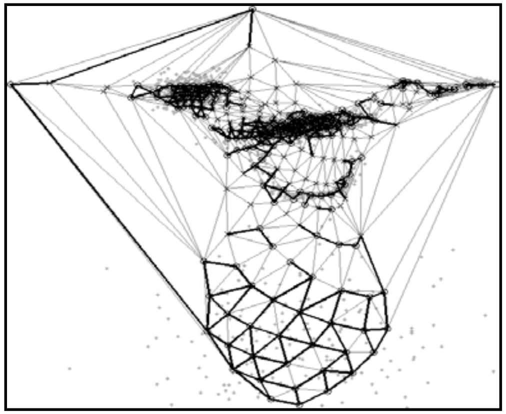
Adjacency

Connectivity

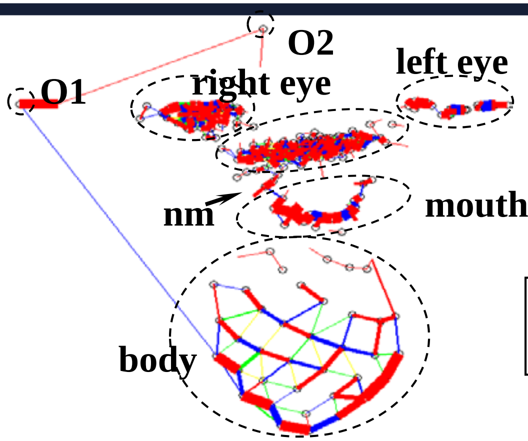
CONN graph:
Weighted
induced
Delaunay
graph

Connectivity (CONN) similarity measure and graph

(Taşdemir & Merényi, IEEE TNN 2009)

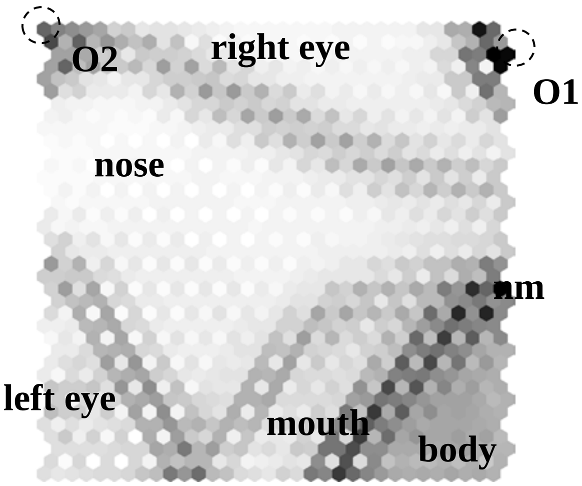


Adjacency



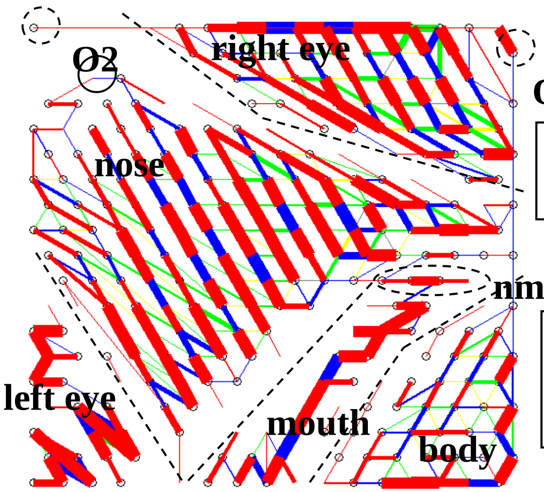
Connectivity

Cannot be shown for data dim > 2



U-matrix ($\Sigma\| w_i - w_j \|$) overlain the SOM grid

Figure adapted from Vesanto & Alhoniemi IEEE TNN 2000



CONNectivity matrix draped over SOM grid: The SOM / CONN portrait of the Clown

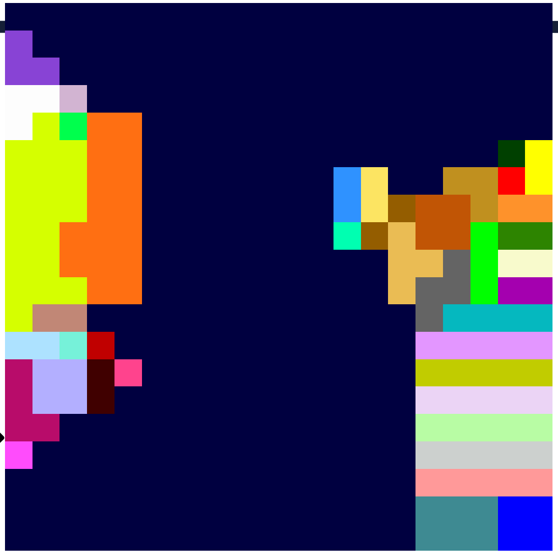
Can be shown For data dim >2

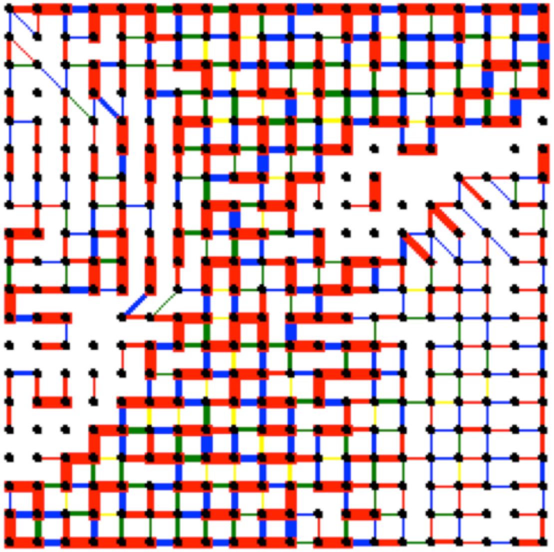
Bonus: CONN shows topology violations

Part I recap: NeuroScope approach to structure discovery

Step 1: Learn the data manifold with SOMs - easy, reliable, little tuning needed, automatic, unsupervised.

- Use all input features – keep the discovery potential
- Use Conscience SOM (CSOM) for maximum entropy learning (best matching of the data distribution)





CSOM / CONN portrait of the ALMA cube of HD142527

Step 2: Segment the SOM (cluster the SOM prototypes) – can be hard

- Need good knowledge representation, sensitive similarity measure, like the CONN graph, and visualization.
- Interactive cluster extraction is best so far.

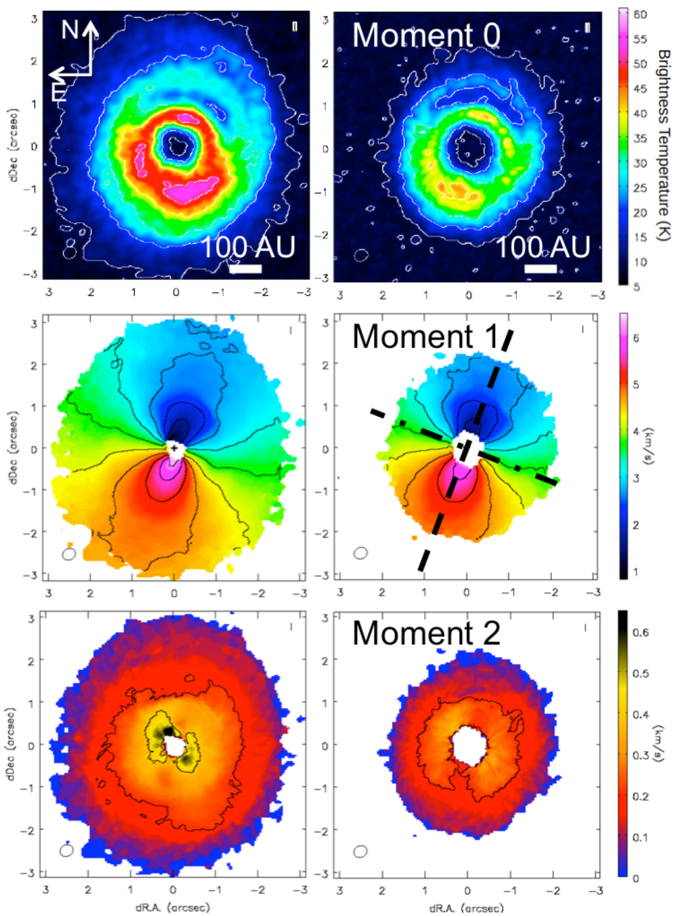
E. Merényi, Rice U
erzsebet@rice.edu

Clusters found in HD142527

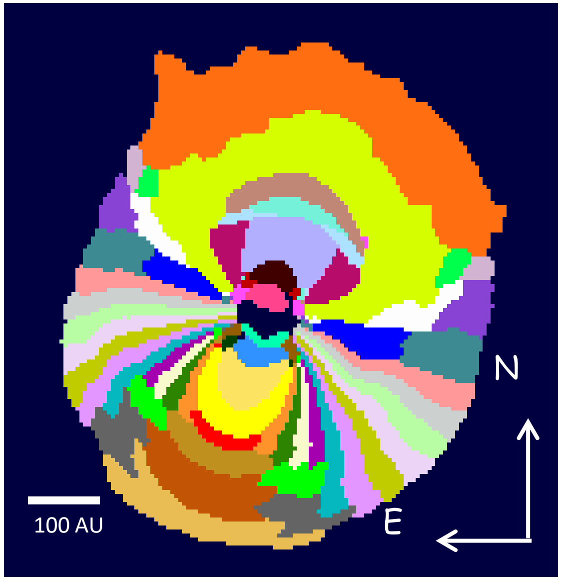
Data: ALMA image cube of HD142527 (Isella, 2015)

Moment maps
Each from 100 channels

^{13}CO J=3-2 $\quad$ $C^{18}O$ J=3-2



SOM / CONN cluster map from stacked $C^{18}O$, ^{13}CO lines, 100 + 100 channels as 200-D input feature vectors



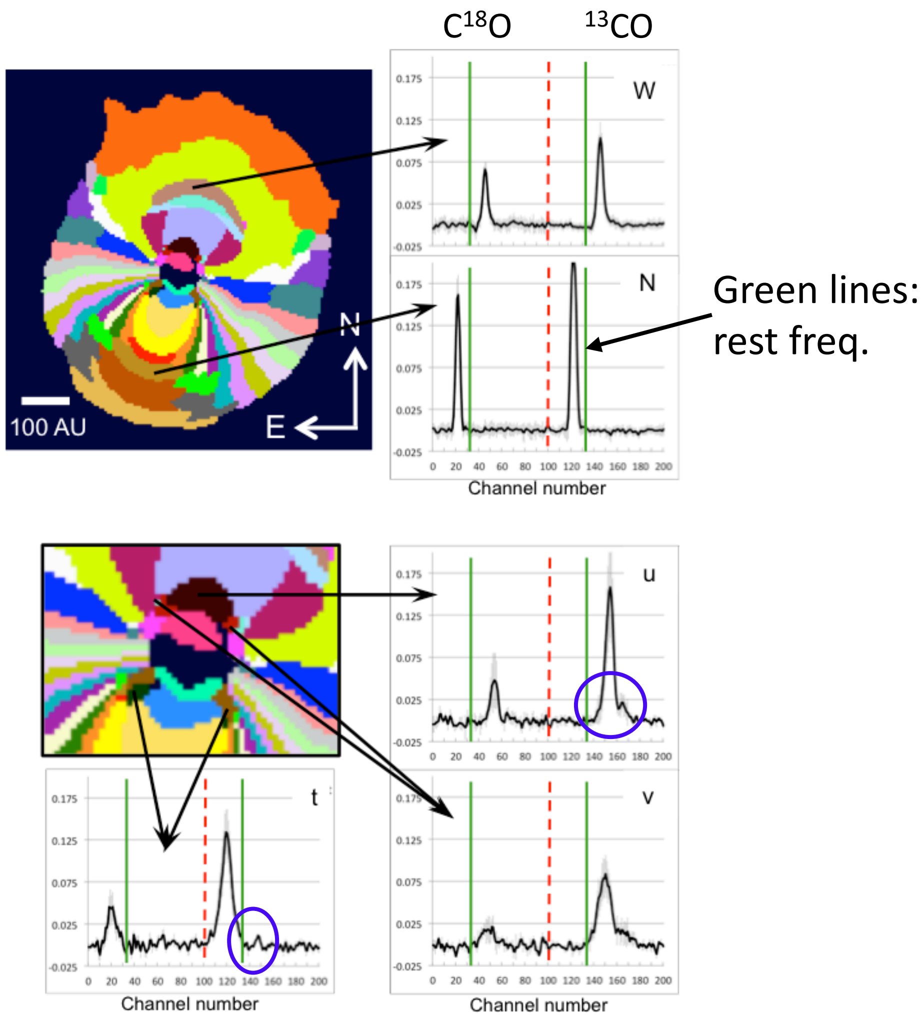
The emerging structure of the protoplanetary disk based on all channels of two molecular tracers, visualized in one 2-D view (40 clusters)

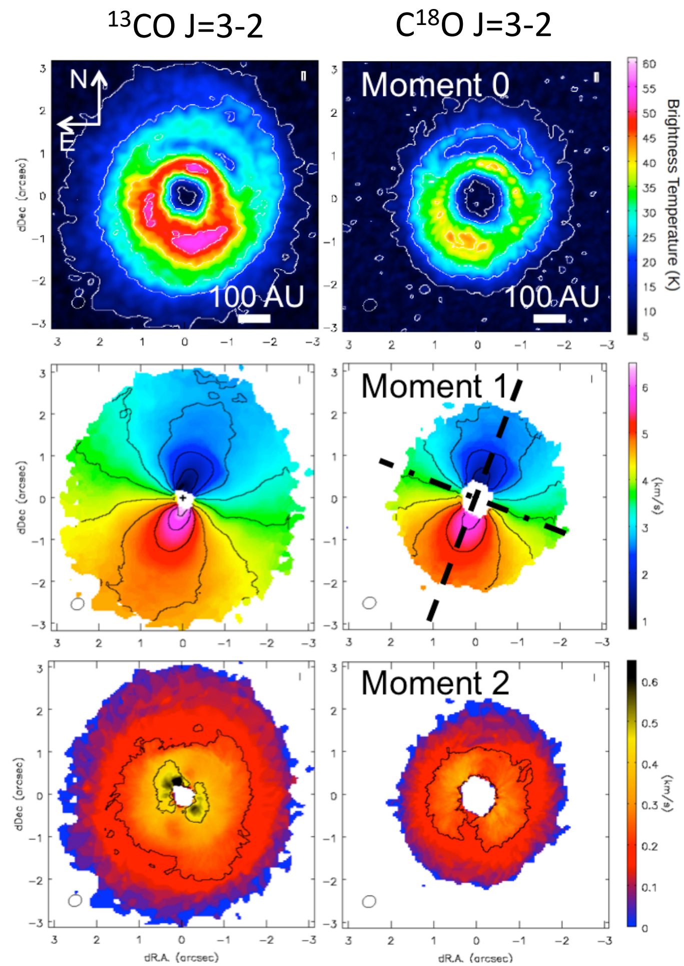
Coloring of clusters is arbitrary, not a heat map!

Input data: cleaned spectral cubes straight out of the ALMA data reduction pipeline, no additional pre-processing

Clusters found in HD142527

Data: ALMA image cube of HD142527 (Isella, 2015)





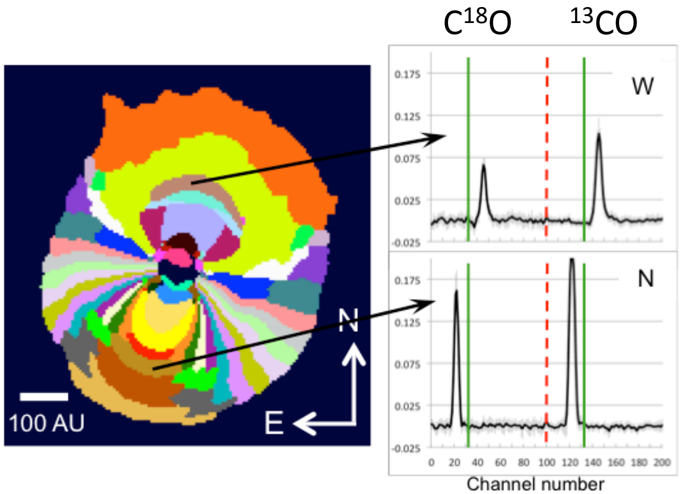
More discovery within one molecular line

More discovery from the combination of lines

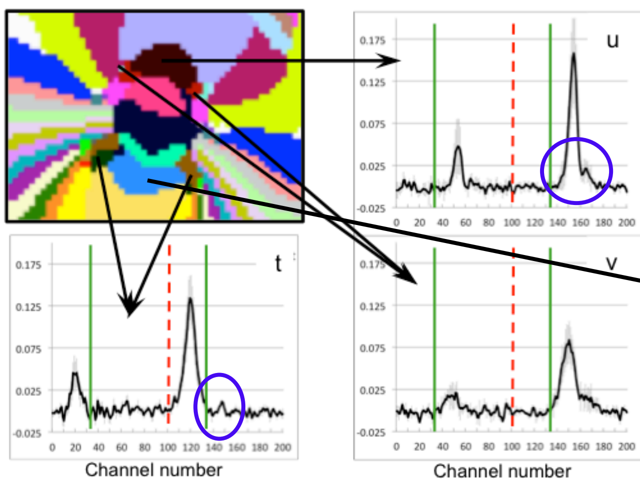
E. Merényi, Rice U
erzsebet@rice.edu

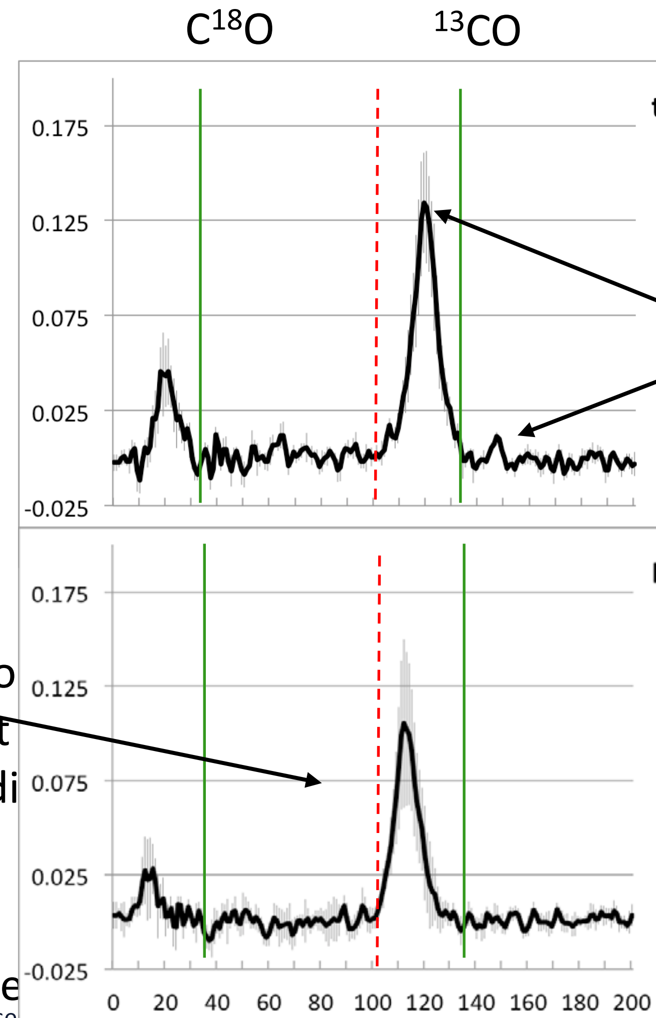
Clusters found in HD142527

Data: ALMA image cube of HD142527 (Isella, 2015)



Mean cluster signatures alert to interesting areas.





Two ... rom
rest ... moving
in di

More discovery within one molecular line

More discovery from the

E. Merényi, Rice U
erzsebet@rice.edu

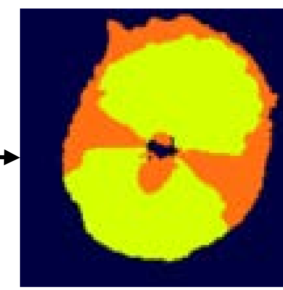
Part II: Automation for Step 2, cluster extraction from SOM

Graph-segmentation informed by SOM and CONN

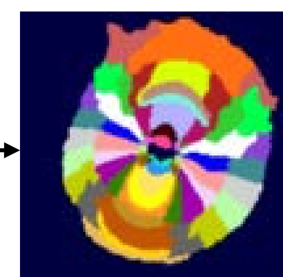
- ☺ Graph-cutting methods: automatic, only 1 or 2 parameters, some have none *
- ☹ Can't deal with many data points. N vectors => N^2 edges. For this small ALMA image (56,000 vectors), over 10^9 edges !!!
- ☺ ☺ Use the intelligently summarized data (SOM prototypes) as input
- ☺ ☺ ☺ Plus CONN similarity measure *(Merényi, Taylor, Isella, Proc. IAU 325, 2016)*

ALMA **data cube**
N = 56,000
N ^2=2.5*10^9
→ Graph-cutting alg using **Euclidean distance** → Prohibitively large time demand → ???

SOM **prototypes** of ALMA cube
N=400, N^2=16,000
→ Graph-cutting alg using **Euclidean distance** → Very fast (< 1 second) →



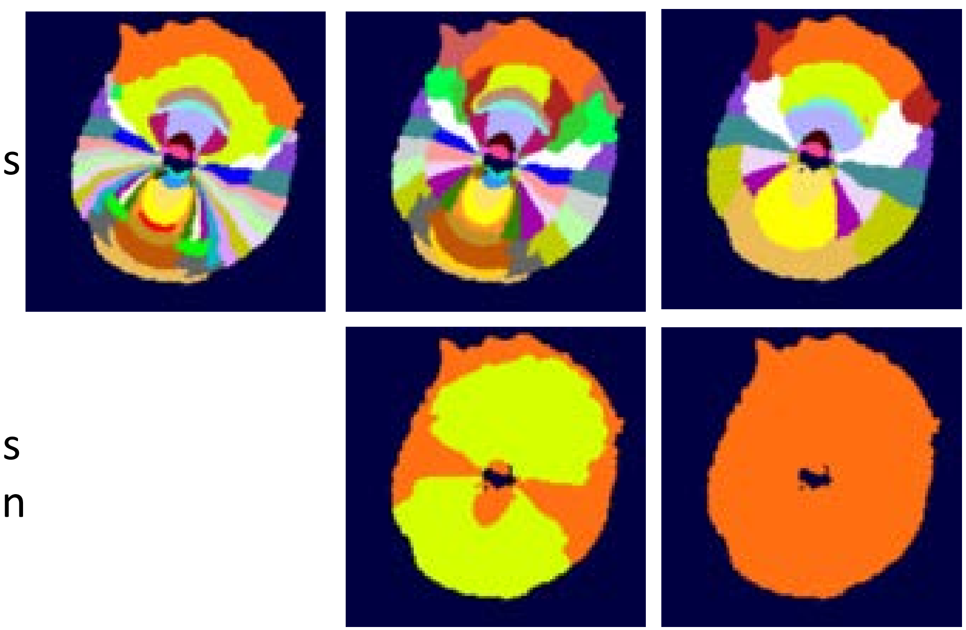
SOM **prototypes** of ALMA cube
N=400, N^2=16,000
→ Graph-cutting alg using our **CONN measure** → Very fast (< 1 second) →



* Review: Fortunato, 2010

Interactive vs automated results

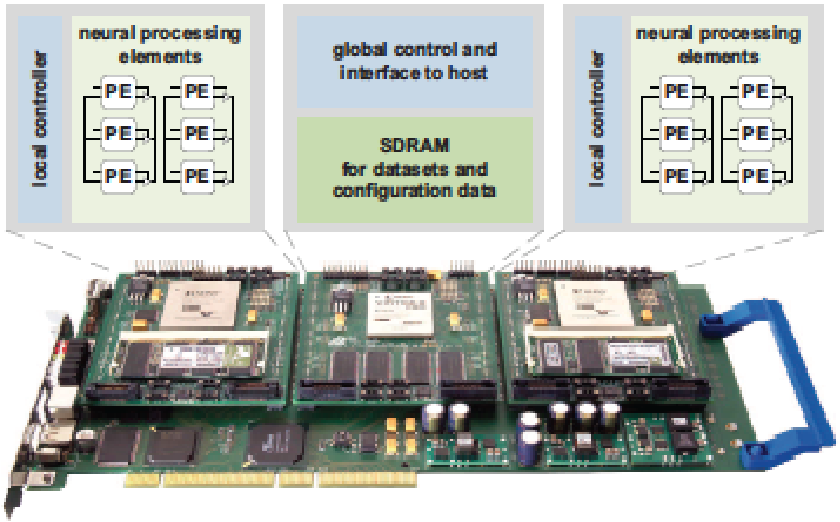
- Walktrap (Pons & Latapy, 2005) and Infomap (Rosvall & Bergstrom) – two best results with default setting (`igraph` package), 1 or 2 parameters.
- Details don't quite match, but differences reasonable. Graph-segmentation of SOM + CONN finds relevant structure, and FAST.



- Next: explore non-default parameters, to improve the graph-segmentation
- Interpret differences

Part III
Speed up Step 1: SOM learning in parallel hardware

SOM accelerator gNBXe *(Lachmair, Merényi, Porrmann, Rückert, Neurocomputing, 2013)*



- Custom-designed for optimal algorithm mapping
- We have been testing a mid-level FPGA-based model since summer, 2017;
- ~ 5 sec learning time for our ALMA cube

- KSOM and CSOM variants implemented, <u>reconfigurable, on-chip learning</u>
- <u>Large-scale computation:</u> handles hyperspectral imagery
- <u>Current:</u> FPGA-based prototype, ~ 12–25 x faster than Core-i7 PC, 4 threads, for large SOM / high-D data; consumes 80-90% less energy. Higher-end and next-gen versions 2-4 x faster.
- <u>Future:</u> ASIC implementation is expected to gain another factor of 10 (or more, depending on the nano-scale technology)

Mass-processing perspectives for pipelines

- Do SOM learning in parallel hardware : < 1 min
- Cluster the SOM prototypes automatically with SOM+CONN input to graph-segmentation alg: few seconds

=> Can map the structure of a protoplanetary disk and return the salient spectral properties of the clusters in a few minutes

Other benefits:

- Applicable to disparate data combined from different spectral windows or instruments
- Applicable to chaotic sources (GMCs)

Conclusions

- **Rich data** (e.g., spectral resolution for ALMA) **offer a magnifying lens for the underlying physical processes** (kinematics of atomic and molecular gas and the distribution of solid particles in the ALMA example).
- **Capabilities to exploit the richness and subtleties of features** (spectral details) **can enlarge the discovery space**.
- The NeuroScope approach provides some tools to achieve this.
- It also shows promise for large-scale, automated processing.

- Merényi, E., Taylor, J. and Isella, A. (2016), Deep data: discovery and visualization. Application to hyperspectral ALMA imagery. *Proc. International Astronomical Union*, *12*(S325), 281-290. doi:10.1017/S1743921317000175
- Merényi, E., Taylor, J. and Isella, A. (2016), Mining Complex Hyperspectral ALMA Cubes for Structure with Neural Machine Learning. *Proc. IEEE SSCI Symposium on Computational Intelligence and Data Mining*, Athens, Greece, Dec 6-9, 2016. 11pp. On-line: http://ieeexplore.ieee.org/document/7849952/ DOI: 10.1109/SSCI.2016.7849952
- Merényi, E., Taylor, J. (2017) SOM-empowered Graph Segmentation for Fast Automatic Clustering of Large and Complex Data. *Proc. 12th International Workshop on Self-Organizing Maps, WSOM+ 2017, Nancy, France, June 27-29, 2017. 9pp*