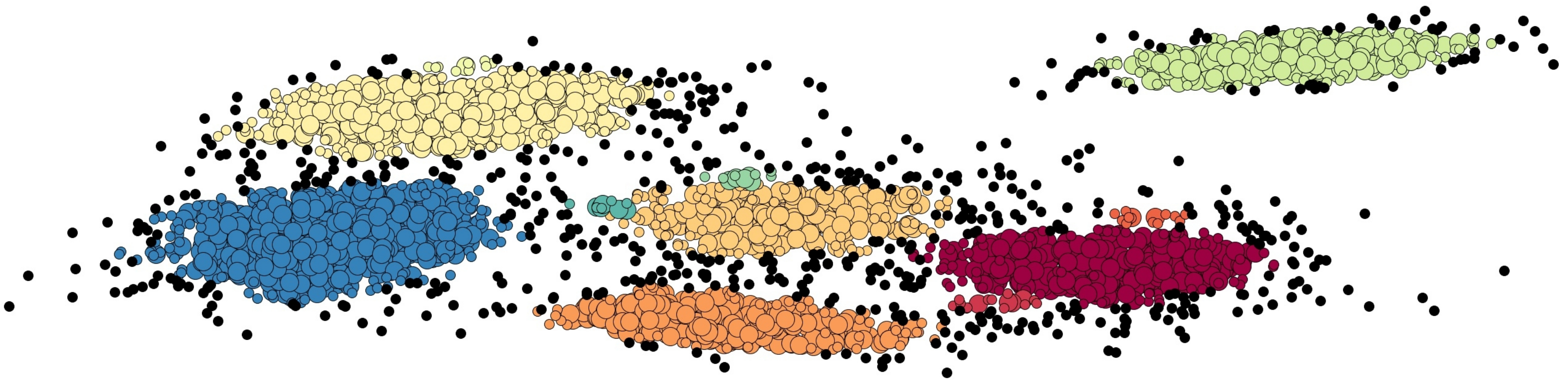
technische universität dortmund

lehrstuhl physik e5



Search for Sources of Rare Types by Cluster Analysis

Simone Mender

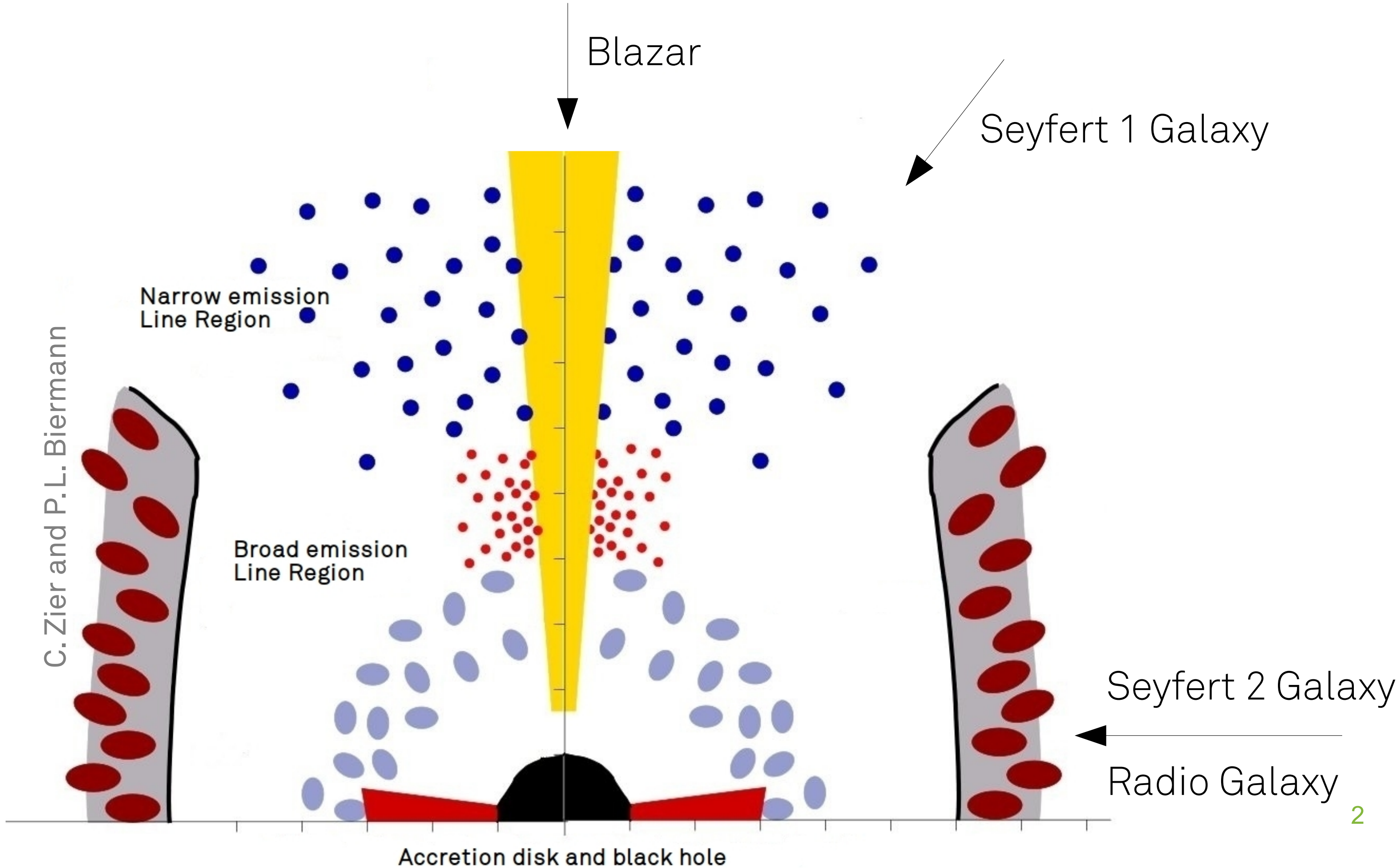
School for Astroparticle Physics

October 4-12, 2017

Motivation

- Multimessenger analysis
- Different classification tasks
- Examination of theoretical models

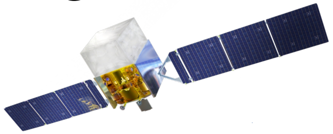
- Improve the comprehension of extragalactic source types
- Search for rare source types

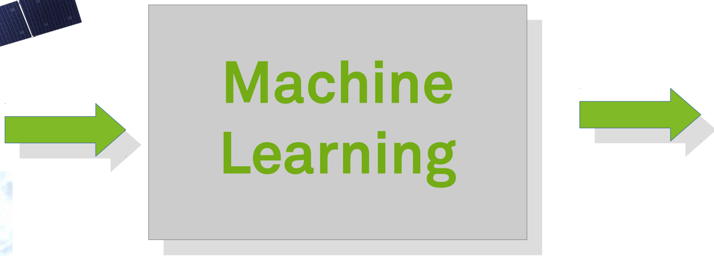


Classification through Machine Learning

Multiwavelength data

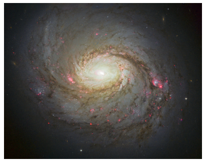

arXiv:1410.5073


https://science.nasa.gov/toolkits/spacecraft-icons

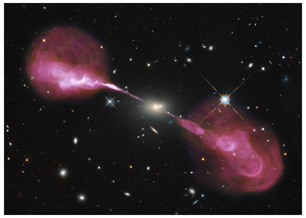


Classification

http://www.spacetelescope.org/news/heic1305/



- Seyfert galaxy
- Dust obscured blazar
- Radio Galaxy
- FSRQ
- BL Lac
- ...


http://www.spacetelescope.org/images/opo1247a/

Classification through Machine Learning

Multiwavelength data → **Machine Learning** → Classification

Supervised machine learning

- Training of model with labeled data, e.g. Decision Tree
- Data into trained model
  → estimation of a label

→ Need labeled data to train the model

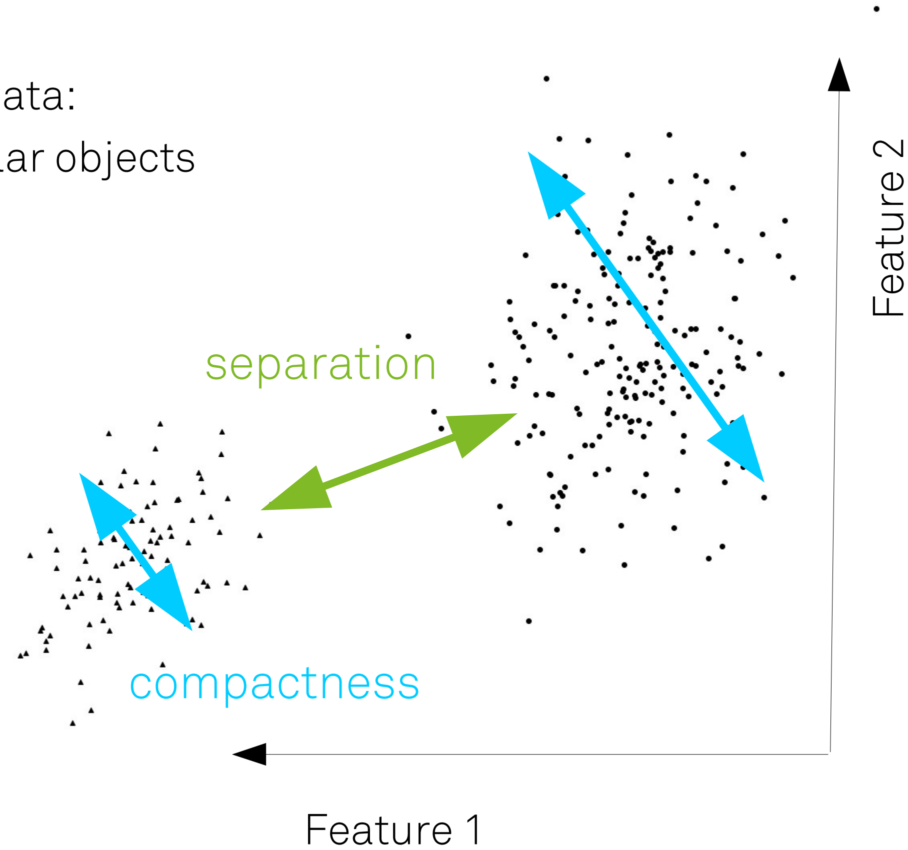
Unsupervised machine learning

- Detection of patterns in the input data
- Example: Cluster Analysis
  Creation of groups/clusters of similar objects

→ No need for labeled data

Cluster Analysis

- Detection of patterns in the input data:
  Creation of groups/clusters of similar objects
- Different clusters should be
  - compact
  - separated from each other



Procedure of Cluster Analysis

1) Feature selection
2) Use of different algorithms
3) Validation of the results
4) Choose the best Cluster Analysis

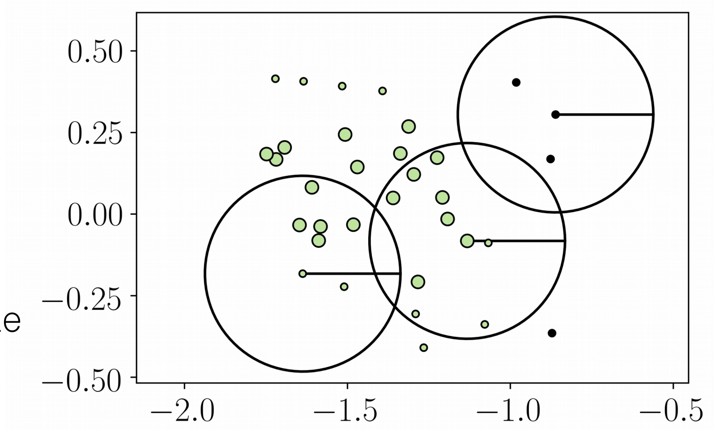
Different Methods of Cluster Analysis

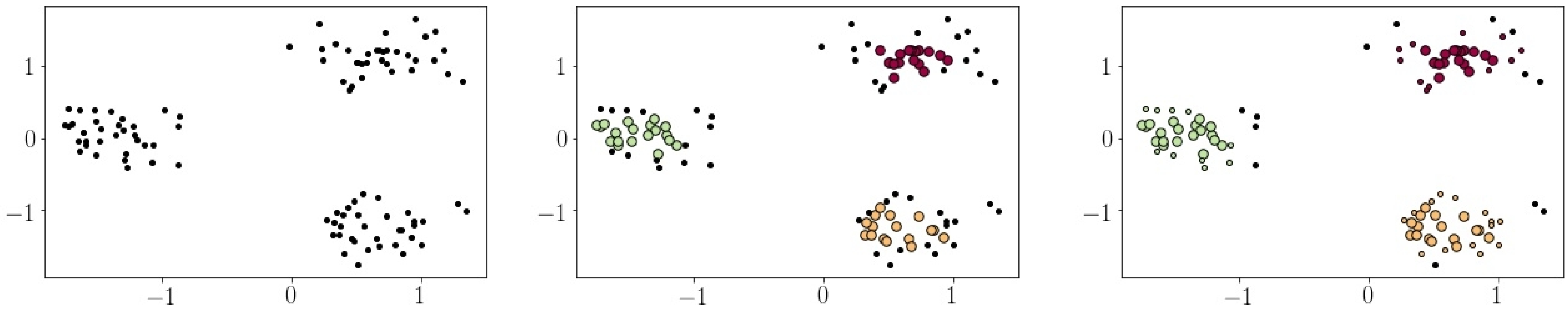
- Partitional Clustering: optimization of a certain criterion
- Hierarchical Clustering: seeks to build a hierarchy of clusters
- Density-based Clustering: separation due to dense and less dense regions

Density-Based Clustering

$\text{minPts} = 6 \quad \varepsilon = 0.3$

- Separation due to dense and less dense regions
- *DBSCAN*:
  - dependent on 2 parameters: $\text{minPts},\ \varepsilon$
  - divides the data points into core points, reachable points and outliers





Cluster Algorithms in Astroparticle Physics

A. Tramacere and C. Vecchio (2012):

- DBSCAN algorithm applied to *Fermi*-LAT data → detection of clusters in γ-ray data
- Method can be used to look for point-like sources and extended sources

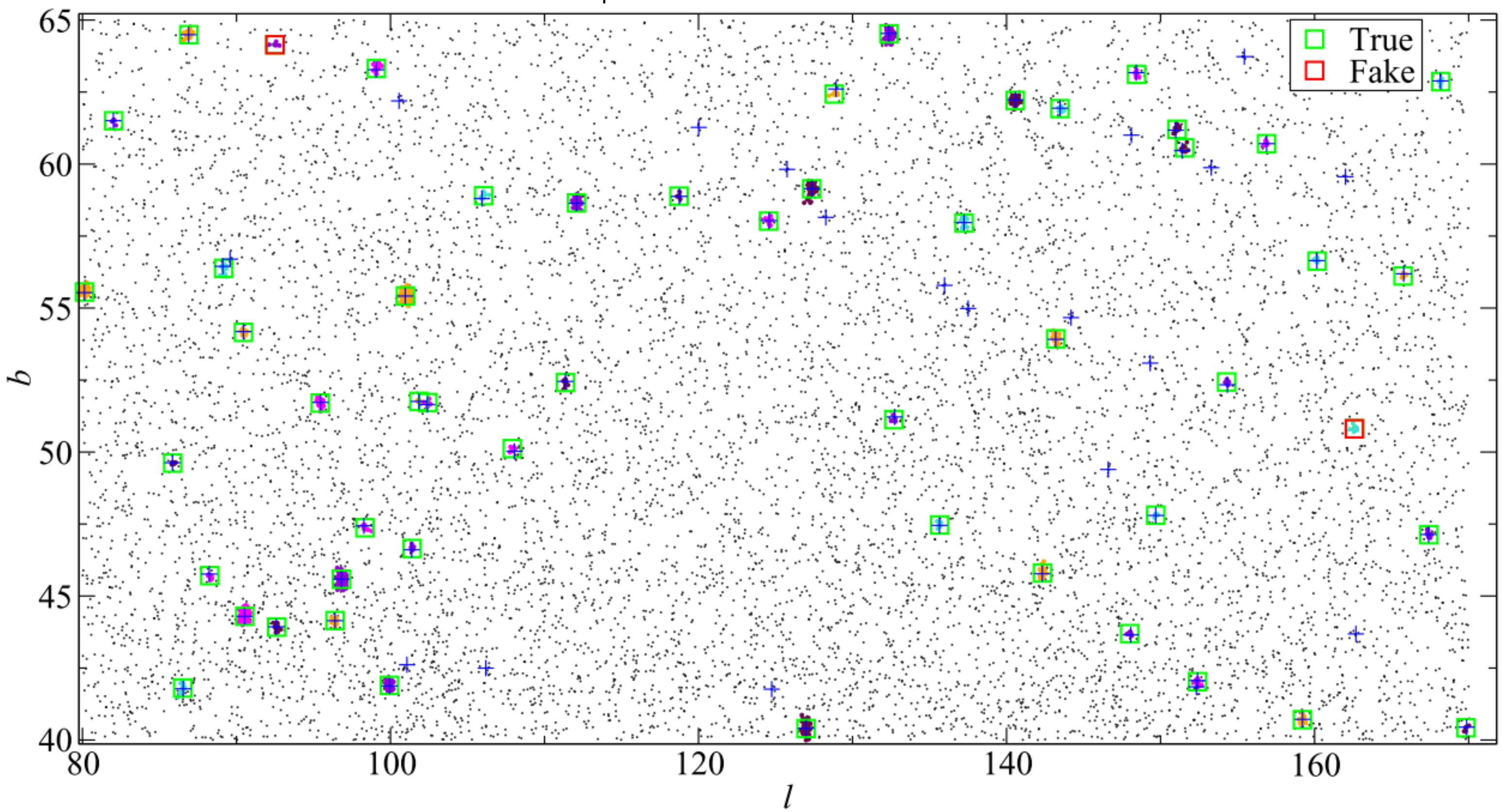
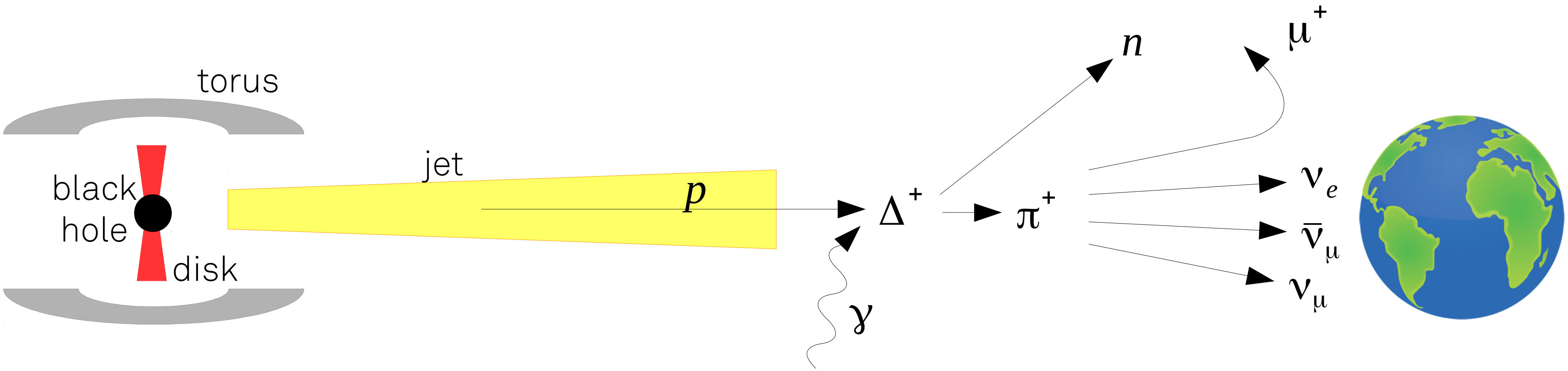


Fig. 1. Photon map for the *sky* test field 1, with the result of the γ-ray DBSCAN detection for $K = 5$ and $\varepsilon = 0.17$ deg. The blue crosses refer to the simulated sources, the green boxes to 51 detected *true* clusters, and the red boxes to the 2 *fake* ones. The black dots represent the background events, the remaining colors indicate cluster events.

(Dust Obscured) Blazars: What makes them interesting?



IceCube Collaboration:

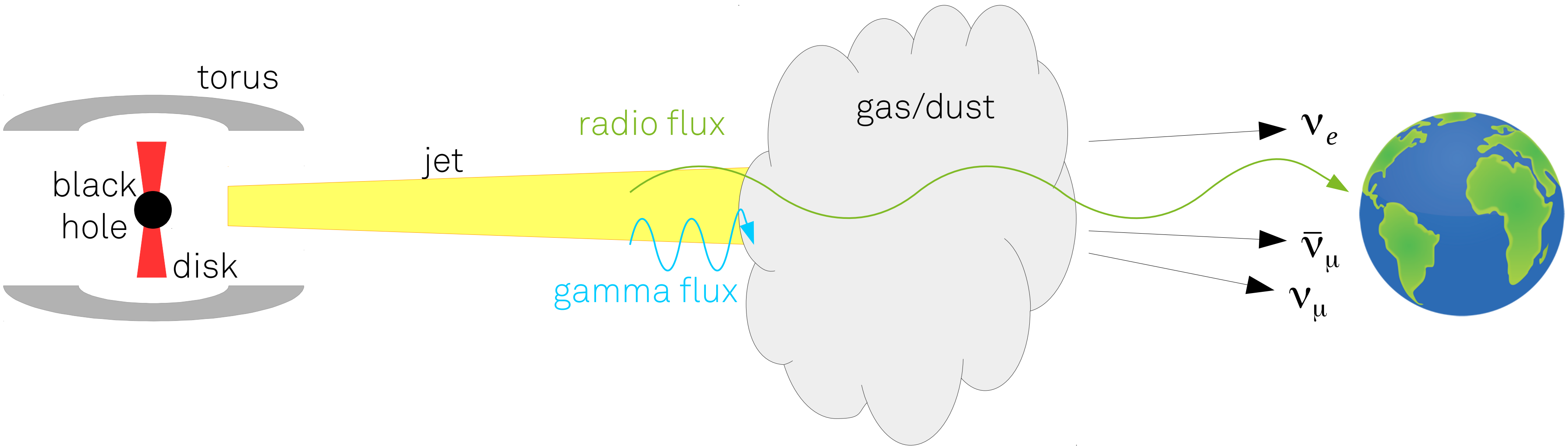
Analysis of the 4-year IceCube high-energy starting events (2016)

- 2010 – 2014: 52 neutrinos
  20 TeV > *E* > 2 PeV
- Source of high-energy neutrinos is still unknown

The Contribution of Fermi-2LAC Blazars to the diffuse TeV-PeV Neutrino Flux (2016)

- Blazars may be high-energy neutrino source
- 2LAC blazars can only explain 27 % of the high-energy neutrino flux

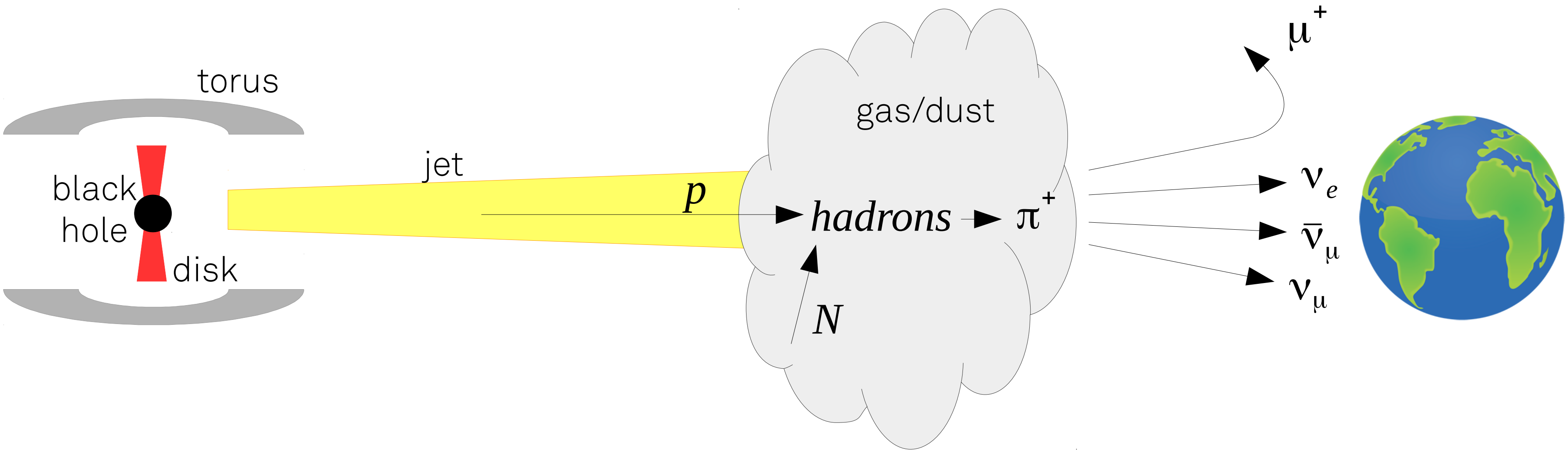
Dust Obscured Blazars: What do we expect?



- Proton-proton interaction
  → increasing neutrino flux
- AGN with jet pointing towards us
  → high radio flux

- Attenuation of emission at high frequencies through the dust
  → low X-ray flux
  → no gamma-ray flux

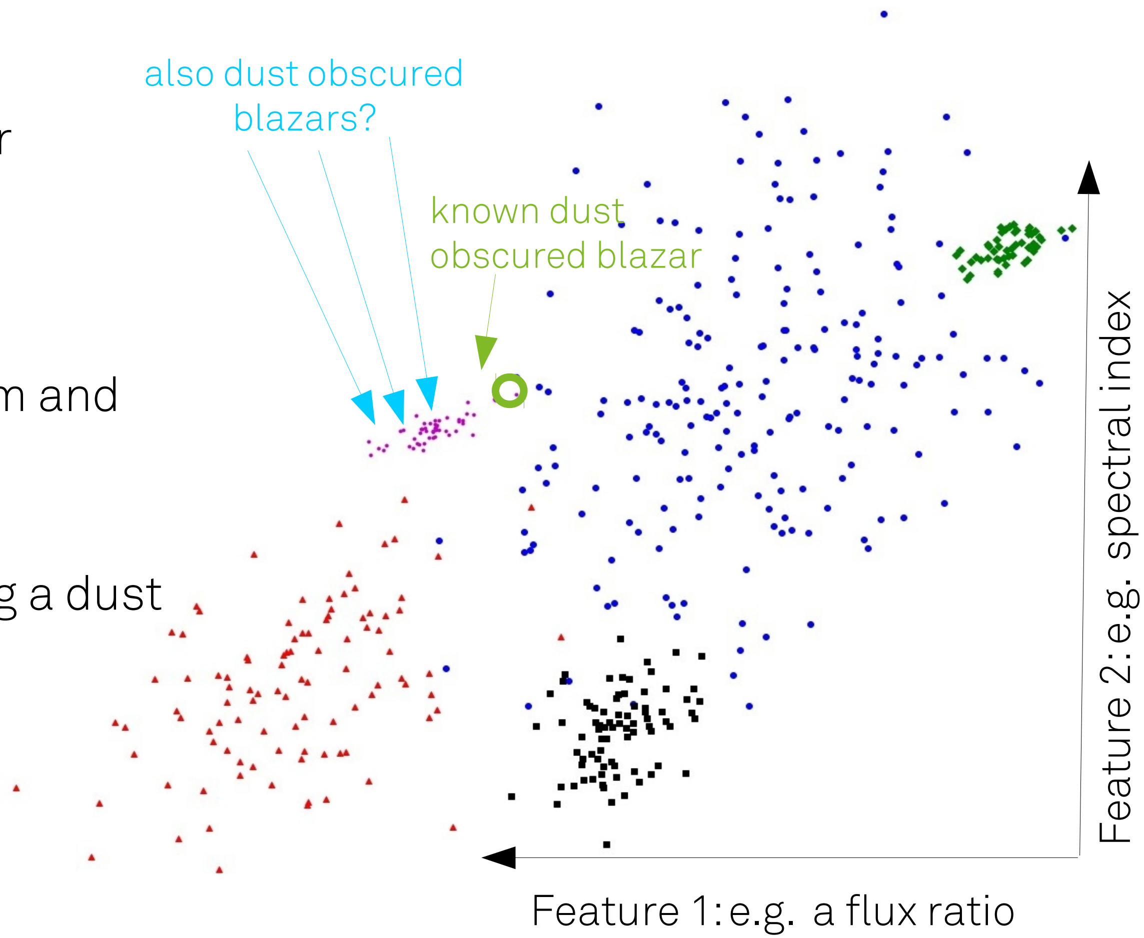
Dust Obscured Blazars: What makes them interesting?



Dust obscured blazars could be a source for high-energy neutrinos

Outlook: How to find Dust Obscured Blazars with Clustering?

- Find suitable catalogs for different wavelengths
- Apply different cluster algorithms, evaluate them and choose the best
- Look at clusters including a dust obscured bazar

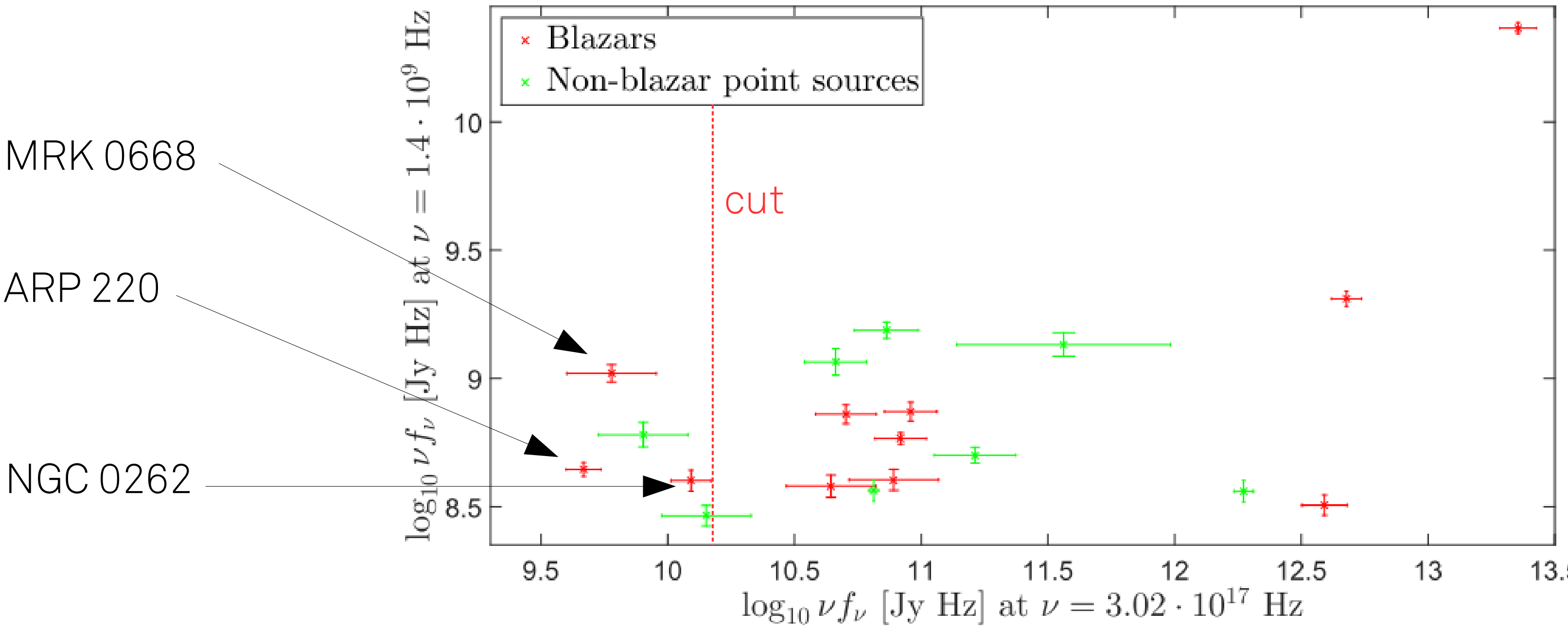


Backup

Dust Obscured Blazars

G. Maggi et al. (2016):

- Search for high-energy neutrinos from dust obscured Blazars
- Cut at an X-ray flux of $\log_{10} \nu F_\nu = 10.2\,\mathrm{Jy\,Hz}$
- Three interesting objects



Distance functions:

- Single Linkage:

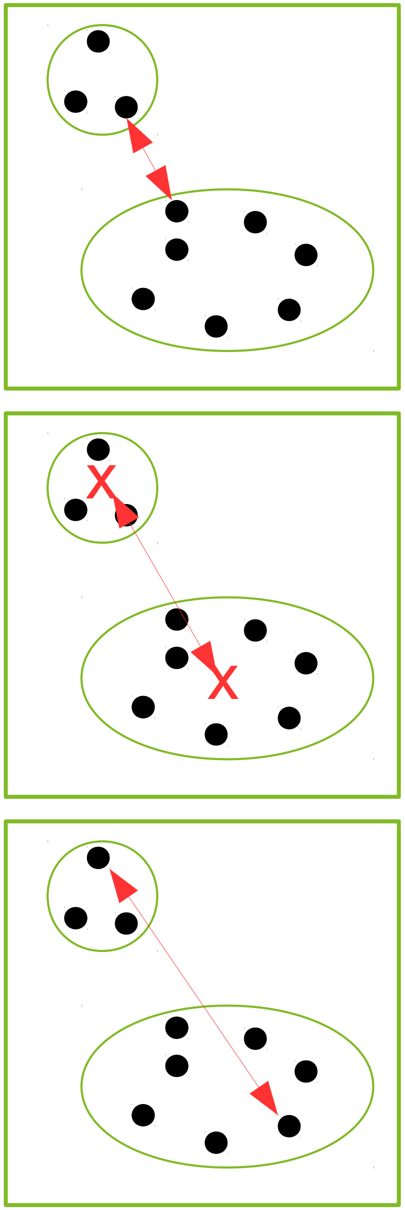
$$d_{\text{single-l}}(A,B) = \min_{a\in A, b\in B}\{d(a,b)\}$$

- Average linkage:

$$d_{\text{average-l}}(A,B) = \text{mean}\,(d(a,b))$$

- Complete Linkage

$$d_{\text{complete-l}}(A,B) = \max_{a\in A, b\in B}\{d(a,b)\}$$



Why Feature Selection?

- Curse of dimensionality

- 1 dimension: 100 data points

- Same density in 10 dimensions:
  100^{10} data points = 10^{20} data points

Pion Decay:

π^+ DECAY MODES

π^- modes are charge conjugates of the modes below.

For decay limits to particles which are not established, see the section on Searches for Axions and Other Very Light Bosons.

	Mode			Fraction (Γ_i/Γ)	Confidence level
Γ_1	$\mu^+\nu_\mu$		[a]	(99.98770 ± 0.00004) %	
Γ_2	$\mu^+\nu_\mu\gamma$		[b]	$(2.00 \pm0.25) \times 10^{-4}$	
Γ_3	$e^+\nu_e$		[a]	$(1.230 \pm0.004) \times 10^{-4}$	
Γ_4	$e^+\nu_e\gamma$		[b]	$(7.39 \pm0.05) \times 10^{-7}$	
Γ_5	$e^+\nu_e\pi^0$			$(1.036 \pm0.006) \times 10^{-8}$	
Γ_6	$e^+\nu_e e^+ e^-$			$(3.2 \pm0.5) \times 10^{-9}$	
Γ_7	$e^+\nu_e\nu\overline{\nu}$			$< 5 \times 10^{-6}$	90%
	Lepton Family number (*LF*) or Lepton number (*L*) violating modes				
Γ_8	$\mu^+\overline{\nu}_e$	*L*	[c]	$< 1.5 \times 10^{-3}$	90%
Γ_9	$\mu^+\nu_e$	*LF*	[c]	$< 8.0 \times 10^{-3}$	90%
Γ_{10}	$\mu^- e^+ e^+ \nu$	*LF*		$< 1.6 \times 10^{-6}$	90%

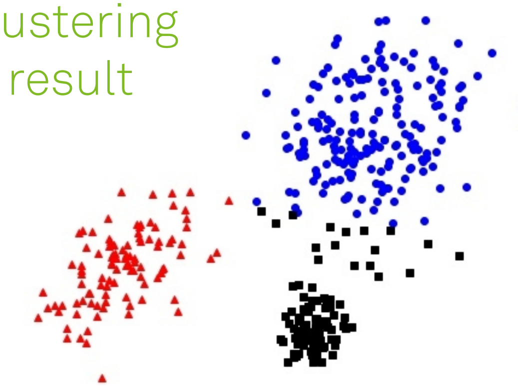
HTTP://PDG.LBL.GOV Page 3 Created: 8/21/2014 12:56

Partitional Clustering

data

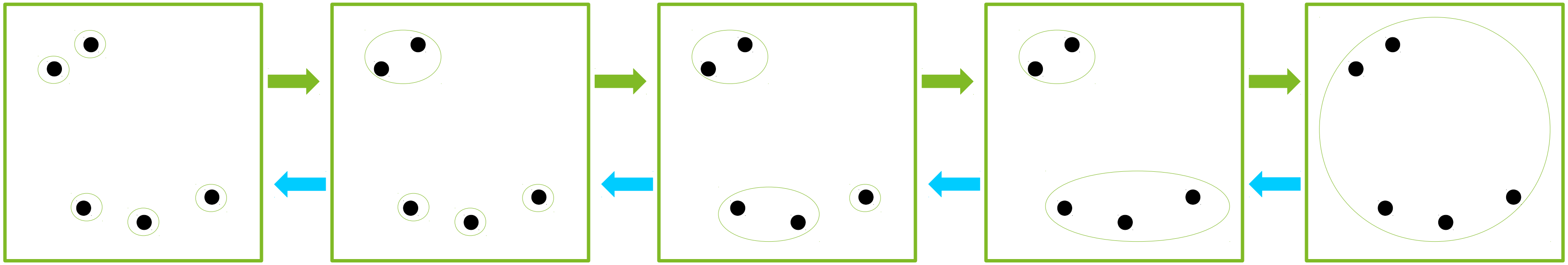
- Optimization of a certain criterion
- *k-means*
  1) Start with k cluster centers
  2) Iteration of
     Expectation-step:
     Calculation of all distances between data points and cluster centers
     Association of the data with the nearest cluster center
     Maximization-step:
     Calculate new cluster centers as average of the data points of each cluster

clustering result



Hierarchical Clustering

- Creation of a cluster hierarchy
- Strategies: agglomerative type and divisive type



Hierarchical Clustering

- Visualization with a dendogram
- Dependent on distance function
  here: single linkage