



Fermi

Gamma-ray Space Telescope



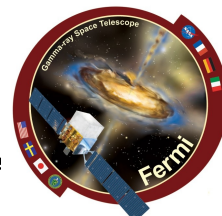
A Statistical Approach to Recognizing Source Classes for Unassociated Sources in the Second Fermi-LAT Catalog

**Maria Elena Monzani
and Nicola Omodei
on behalf of the Fermi-LAT
Collaboration**

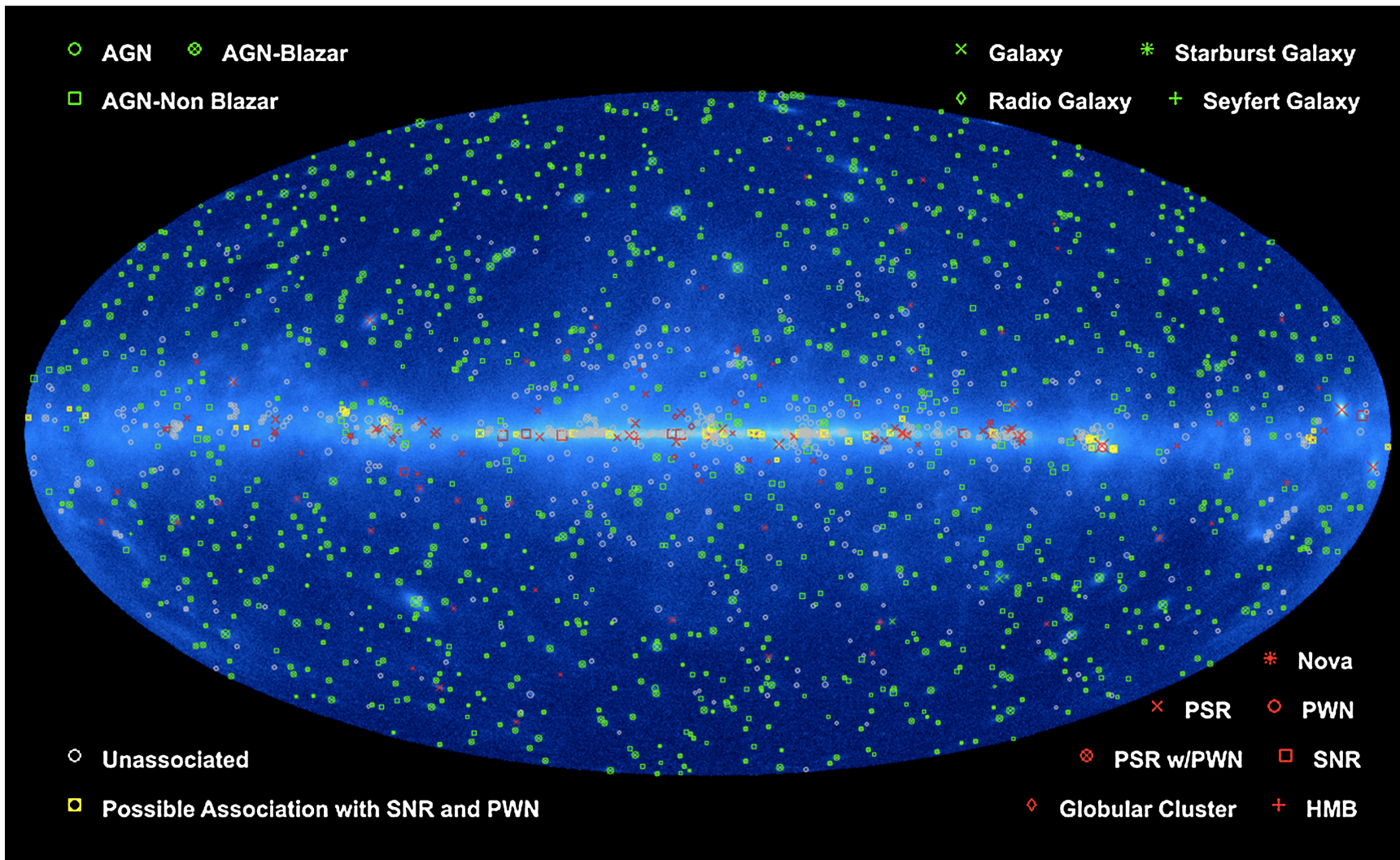


219th AAS Meeting - Jan 10, 2012 - Austin, TX
<http://fermi.gsfc.nasa.gov/ssc/library/austinaas2012.html>

Unassociated Sources in 2FGL



1873 sources in 2FGL; 573 unassociated after all association efforts (~30%)



See Elizabeth C. Ferrara, poster 149.19 and <http://arxiv.org/abs/1108.1435>

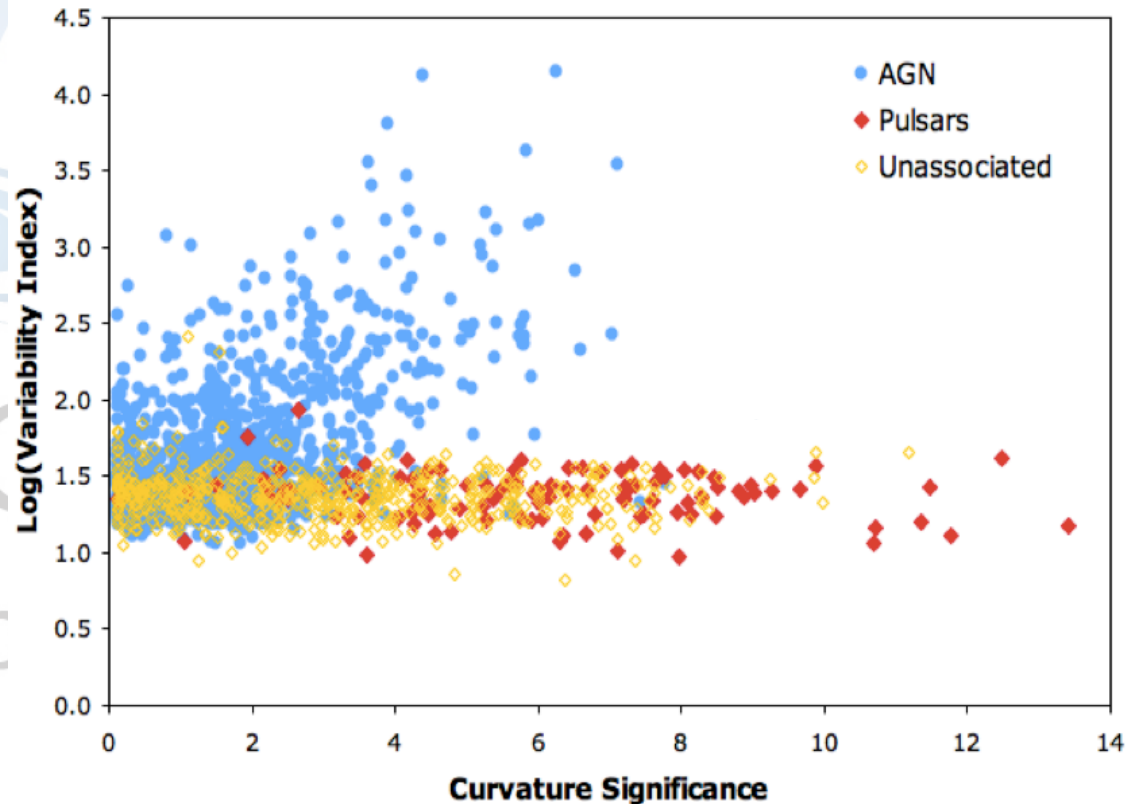


- **Implement statistical methods to determine likely source classifications for 2FGL unassociated sources**
 - **goal: predict the likely classification of Fermi sources based solely on their observed gamma-ray properties**
 - **principle: use the properties of known objects to implement a classification analysis which provides the probability for an unidentified source to belong to a given astronomical class**
 - **examples: Classification Trees (this work) and Logistic Regression**
 - **input sample: all the associated AGN and blazars (1077 sources, 60% of total); all the associated/identified pulsar and pulsar-like objects (includes SNR and potential associations: 180 sources, 10% of total)**
- **Classification Trees are a well-established class of algorithms in the general framework of data mining and machine learning**
 - **definition: Classification Trees are built through a process known as binary recursive partitioning, an iterative process of splitting the data into partitions using *if-then* logical conditions**
 - **advantage: Classification Trees are especially flexible in handling sparse or uneven distributions**

Selection of the training variables

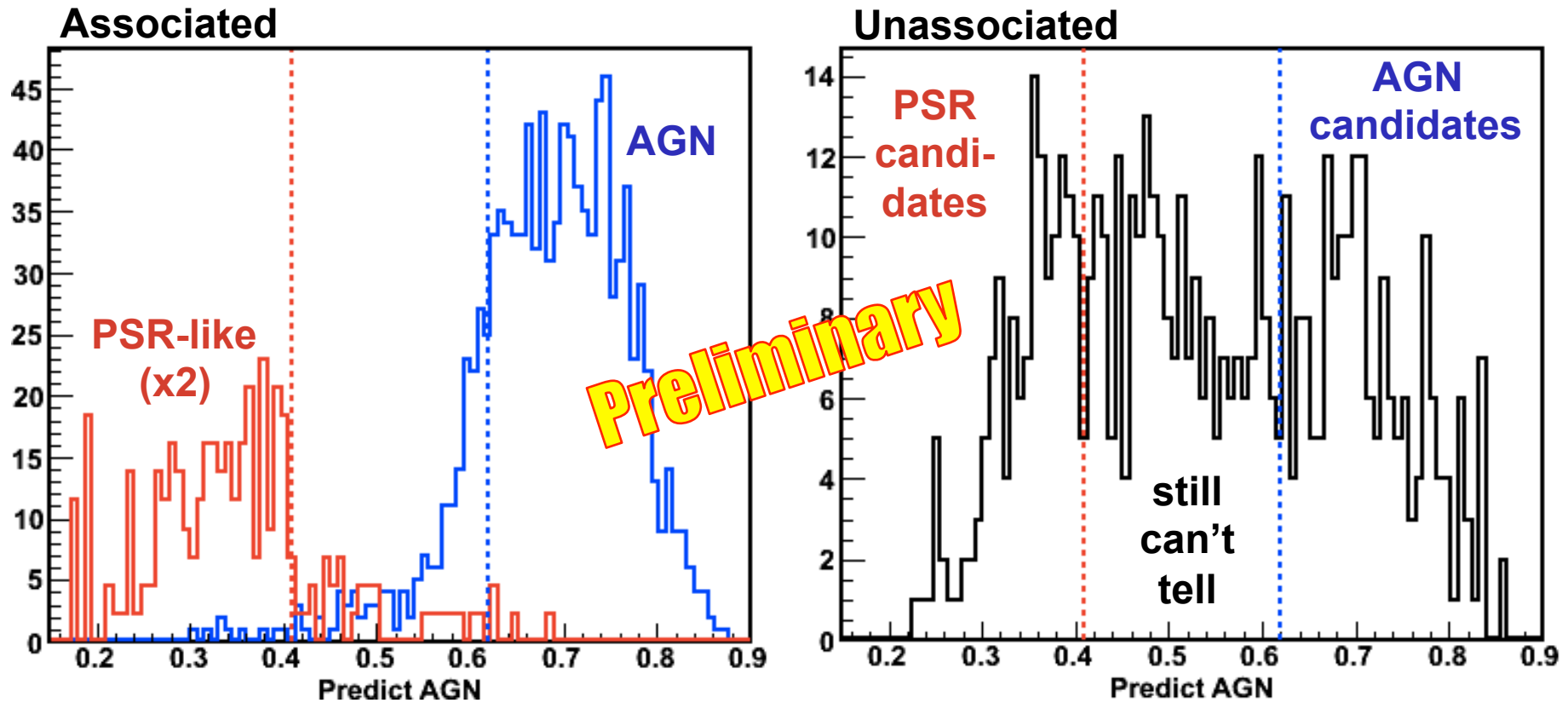


- This is a crucial step in the analysis:
 - physical considerations about the gamma-ray properties of each class
 - ensure that the selected variables are not dependent on the flux, the location or the significance of the source
 - avoid using the Galactic coordinates of the sources
- Ranking of the selected variables after training:
 - variability index (20%)
 - spectral index (16%)
 - curvature signif. (13%)
 - low energy flux (10%)
 - low and high energy hardness ratios (15%)
 - 3-band curvature (7%)
 - intermediate energy hardness ratios (10%)
 - 4-band fluxes (9%)





The result of the training process is the Predictor, a parameter describing the probability for any given source to be either an AGN or a pulsar-like source



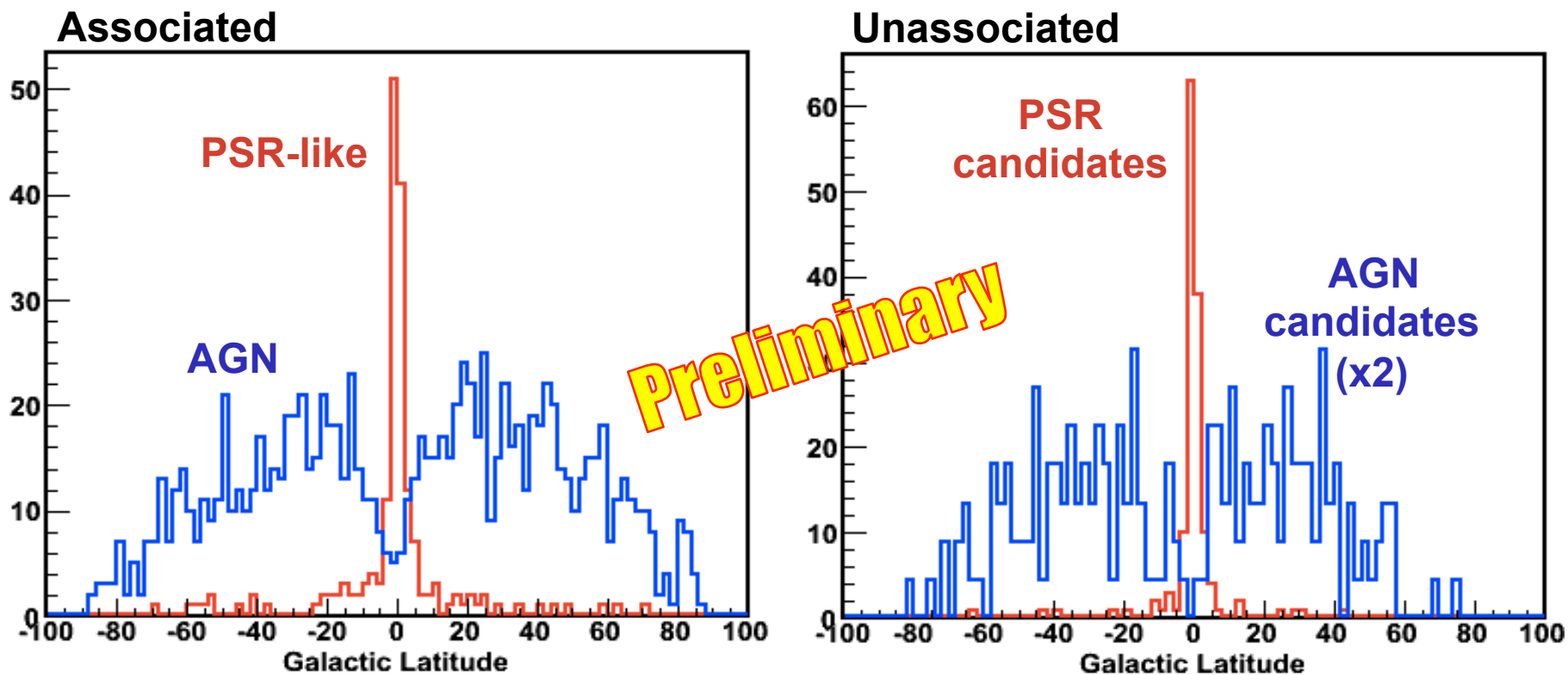
2 fiducial thresholds: **PSR candidates - $P < 0.41$** , **AGN candidates - $P > 0.62$**

fiducial regions: 82% efficiency and $< 5\%$ contamination on input samples

Validation of the Classification Analysis



- 30% of input sources, randomly selected from AGN and pulsar samples, were set aside for internal validation (KS test and efficiency comparisons)
- the Galactic latitude distribution for pulsar and AGN candidates mirrors the expected one (as observed for the Associated sources)

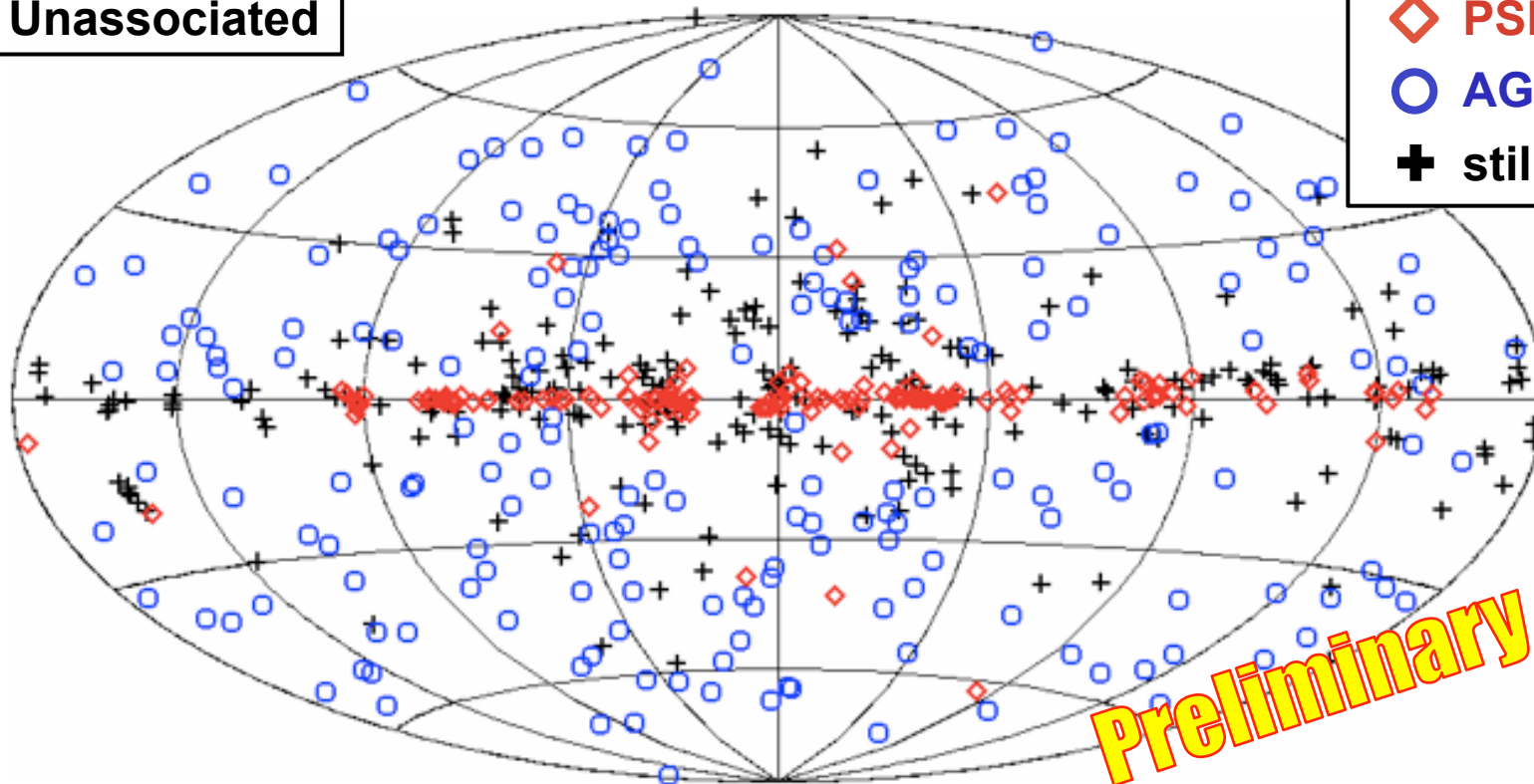


- further validation will be performed using input from multi-wavelength observations (now in progress; was successfully implemented for 1FGL)



- We implemented a method to predict likely source classifications for 2FGL unassociated sources, based solely on their gamma-ray properties
 - the performance of the method has been validated in several ways
 - this will allow to prioritize multi-wavelength observations, according to the predicted source classification

Unassociated



- ◇ PSR candidates
- AGN candidates
- ✚ still can't tell

Preliminary

