Data Mining in Nearby Galaxies

S. Rahmani[1], H. Teimoorinia[2], E. Peeters[1,3,4], P. Barmby[1,3]

[1] Department of Physics and Astronomy, Western University, London, Ontario; [2] Department of Physics and Astronomy, University of Victoria, Victoria, British Columbia;
[3] Centre for Planetary Science and Exploration, Western University, London, Ontario; [4] SETI Institute, Mountain View, California

Introduction

- Artificial neural networks are very powerful tools to use in data processing and pattern recognition problems.
- They use training methods to learn about non-linear and complex relations between input and output data and how to apply it to new sets of data.
- The Kohonen Self organizing map (or self organizing map, SOM) is an unsupervised neural network for mapping and visualizing a complex and non-linear high dimension data while preserving topological features of the original data (Kohonen, 1998, Neurocomputing, 21, 1).
- Galaxies are very complex systems; therefore, a complete understanding of them cannot be achieved only from studying linear or logarithmic correlations between their properties and various waveband data.
- The vast availability of data for nearby galaxies makes them suitable targets for exploratory data analysis.
- Spatially resolved maps provide us with a unique view of the inside of galaxies.
- In this project, we utilized the SOM method to study data from two well studied galaxies, M31 and M101.
- We are particularly interested in understanding the relationships between star formation indicators, stellar mass, and PAH emission.

Method

Self Organizing maps (SOMs)

- The SOM is a clustering method which reduces the dimension of data while preserving topological features.
- Results of SOM contains nodes (usually hexagonal ones) arranged in 1D or 2D arrays.
- Each nodes may contain one or more samples from input data and the distance between nodes represents similarity or dissimilarity of underlying samples.
- In SOMs, the purple hexagonal shapes represent the neurons.
- The distances in a distance map are shown by grey cycle colours between the neurons.
- White colour represents a high amount of the similarity between neurons, and the darker colour shows more differences.
- The key feature to SOMs is that the topological features of the original input data are preserved, e.g. all the regions in a galaxy with high amount of star formation rate (SFR) and low dust mass will appear in neurons in the same area of the SOM (Fig. 1).

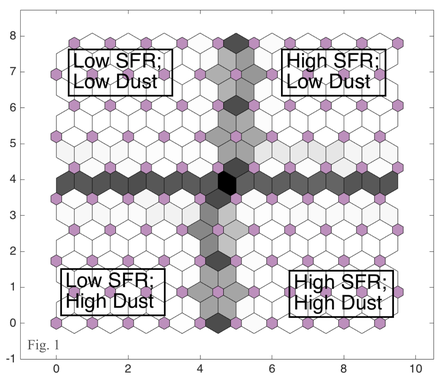


Fig. 1

Data

- We gathered data from 10 regions in M31 and 8 regions in M101 as inputs for SOMs.
- These regions were chosen based on the availability of mid-infrared spectroscopy data.

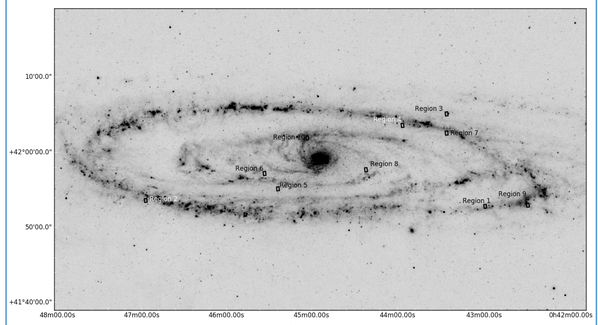


Fig. 2: Position of the 10 regions in M31 from Hemachandra et al., 2015, MNRAS, 454, 818

	Type of Data
M31	PAHs, narrow band and broad band photometry, total IR luminosity, dust mass and luminosity, stellar mass, gas mass, SFR, metallicity, and radiation hardness index (RHI)
M101	PAHs, broad band photometry, total IR luminosity, stellar mass, SFR, metallicity, and RHI

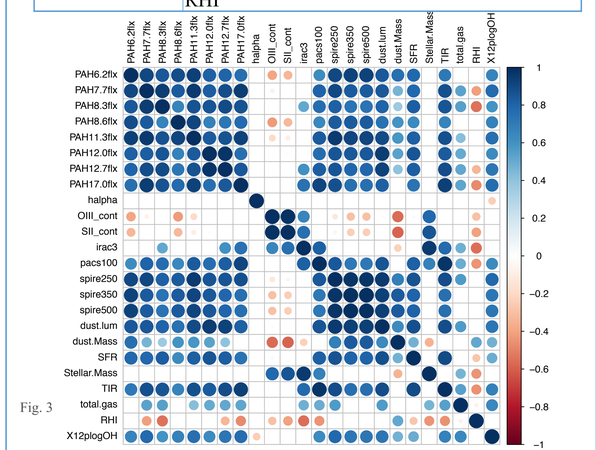


Fig. 3

- Pearson correlation coefficients with a confidence level of 95% for all the M31 data (Fig. 3).
- The non-significant correlations were left empty.

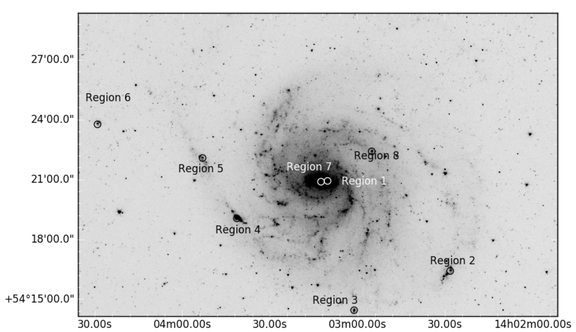


Fig 4: Position of the 8 regions in M101 from Gordon et al. , 2008, ApJ, 682, 336

Results

1D SOMs

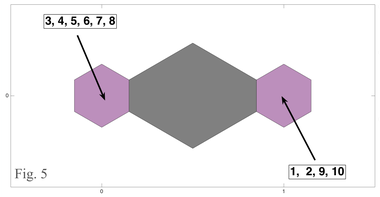


Fig. 5

- Clustering M31 data in 2 groups (Fig. 5).
- Regions 1, 2, 9, and 10 become separated from the other regions.
- These regions have higher PAH flux, dust luminosity and SPIRE emission than the other group.

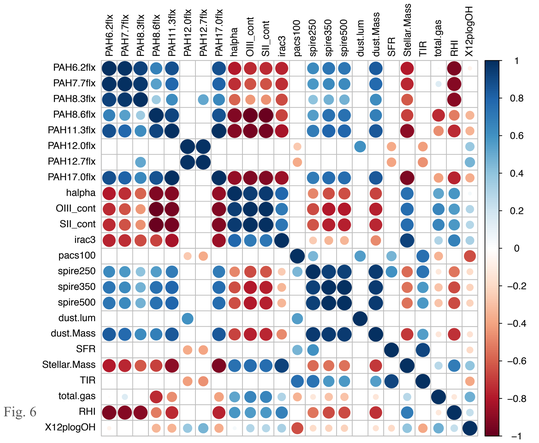
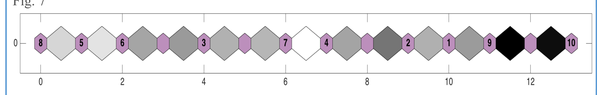


Fig. 6

- Pearson correlation coefficients for regions 3 to 8 in M31 (Fig. 6).
- The correlation coefficients for these 6 regions are different from the ones in Fig. 3.

Fig. 7



- SOM with 14 neurons is the first network that all the M31 regions become separated from each other (Fig. 7) .
- Region 10 is located on the bulge of M31 and became completely separated from other regions.
- Region 8 is located a little off the inner ring, and has the most differences with region 10.

2D SOMs

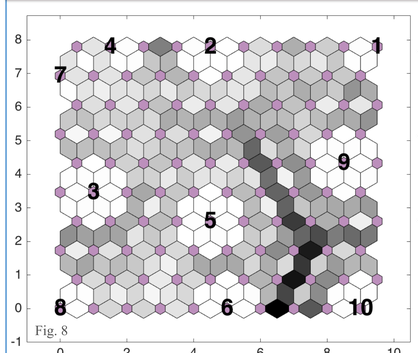


Fig. 8

- 2D networks provide a more complete picture of relations between data (Fig. 8).
- In this network regions again became separated from the other regions.
- Regions with locations close to each other in the galaxy,occupy neurons in the same area of the SOM.

2D SOMs (subsets)

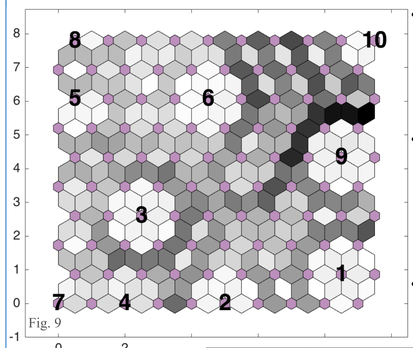


Fig. 9

- Creating SOMs from subsets of data helps us to understand the effect of each input in the position of the regions. in the SOM.
- E. g. if stellar mass were removed from the input data, the positions of regions 1 and 9 (with lowest stellar mass) move closer to position of region 10 (Fig. 9).
- Using subsets, we can predict the unmeasured quantities in a galaxy.

- To test the networks with data from other galaxies, we need to create a subset with exact same type of the data from those galaxies (Fig. 10).

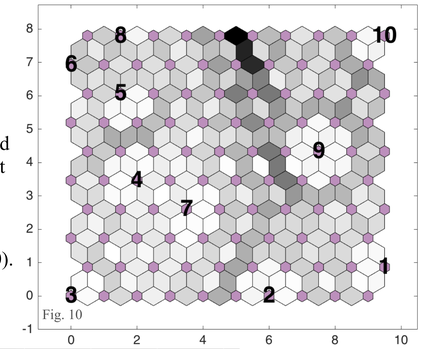


Fig. 10

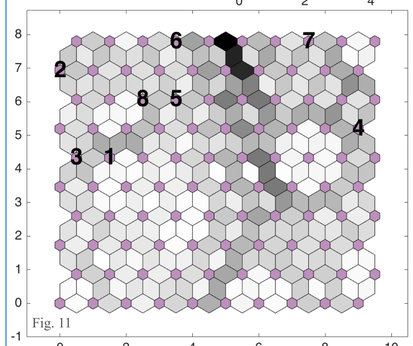


Fig. 11

- Giving the M101 data to network created with M31 data.
- Positions of the M101 data in the SOM, compared to those of the M31 data, tell us about the M101 regions' properties.
- E. g. region 7 in M101 is the nucleus of M101, and located near the location of region 10 in M31 (Fig. 11).

Summary

- We created various networks with different sizes/dimensions based on information we needed from the input data.
- Using smaller-sized SOMs, the data were clustered in 2 major groups; we found correlations that could not have otherwise been seen without clustering.
- In the maps with larger sizes, we created networks to illustrate the relative relations of the regions with one another.
- We applied the SOMs from M31 data to M101 data and found regions with similar properties in both galaxies placed in close regions in the SOMs.
- We created SOMs with different subsets, which give us the ability to predict or extrapolate the amount of various quantities based on their positions in SOMs.
- We can use these networks on the data from other galaxies and have a prediction about their properties very fast.

NSERC CRSNG

Email: srahma49@uwo.ca