# A homogeneous analysis of proto-planetary disk spectra in the Spitzer Space Telescope archive through machine learning and visualization technologies

Main applicant: Prof. Rens Waters (Radboud Universiteit Nijmegen)

SURF advisor: Ben de Vries

## Introduction

Proto-planetary disks are disks around young and newly formed stars. From the dust and gas in these disks planets can form. Typical temperatures for these disks are low (several hundreds of Kelvin) and therefor these disks radiate strongly in the infrared. Studying this infrared radiation can help us understand many properties of the disk (disk mass, composition, disk structure to name a few) which is important to understand how and in what form planets are formed. Since infrared radiation does not penetrate our atmosphere, we use space telescopes to observe these objects. The Spitzer space telescope is an example of such a telescope. One of the observation modes of this telescope is a spectrometer in the mid-infrared (roughly light with a wavelength of 5-40 micrometer). The spectrum of proto-planetary disks typically peaks within this wavelength range, making Spitzer spectra ideal to study these disks. The Spitzer telescope is no longer active, but now we have access to a large archive of Spitzer spectra.

## Science rationale and the added value of SURF

The Spitzer database contains the largest sample of spectra of proto-planetary disks available. Probably some ~1000 spectra are of sufficient quality to derive the composition of the small dust grains in the disk surface. Many papers have described subsets of these data and have analysed the dust composition, using a variety of disk models and dust opacities. This heterogeneous approach hinders a full view on the dust composition and trends as a function of disk parameters, such as the mass of the star, the age of the star forming region, the total dust mass in the disk, and the disk geometry. All these parameters may have an effect on the resulting dust composition of the small dust grains in the inner, rocky planet forming regions of the disk. For instance, is there a trend that for older disks there is a change in dust composition, towards larger grains, more Fe-containing grains? What is the effect of changing dust composition on the appearance of gas bands (e.g. the HCN/C2H2 complex near 14 microns, or the bending mode of water at 6 microns).

For the first time we want to make a homogeneous classification of the whole Spitzer data set using state-of-the-art machine learning (ML) technologies. We want to know if there are trends/correlations in the full data set that have so far not been reported in the literature. These trends/correlations may be present in the Spitzer data themselves, or when combining the data with other parameters, such as the star forming region, central star luminosity, etc. This would require an unbiased search for trends/correlations by taking the full shape of the IRS spectrum into account. For instance, the slope of the spectrum, the strength of the silicate band, presence of PAHs, presence of crystalline silicate bands, gas line emission, etc. Sample spectra in which these different spectral signatures are identified can be used as a prior.

The classification can first be based on features from the raw data (infrared spectra) and extended with known properties of the star and its environment (luminosity, star forming region, etc). In a second step we want to model the spectra and add the results to the analysis. We will start with an already developed approach to obtain the dust composition of the small dust grains in the disk surface using stellar photospheric flux and Spitzer spectra as input data. The modeling results can be used to better understand the different classifications.

It is important to know how the classes found by the ML algorithm are related to different properties of the objects in the dataset and why the trained algorithm grouped the objects the way it did. For

this, visualizations that give insight into the working of the algorithm are essential. The interest in understanding ML results through visualizations is on the rise, but often these strategies are not generally applicable. Likely for this project we will also (partly) need to formulate and develop a visualization approach to understand our results.

To close, we would like to stress that this project will benefit greatly from SURF’s involvement since the ML and visualization expertise needed exceeds that present in the current research group. Enriching our research group at Nijmegen with experts in the ML and visualization teams at SURF will enable us to further our research ambitions.

## Planned outputs

* Paper
* Blog
* Open data and algorithm