

## AGRUPAMENTO ESTATÍSTICO DE GALÁXIAS A PARTIR DE PROPRIEDADES ESPECTRAIS NO INFRAVERMELHO MÉDIO

**JULIA DE CAMPOS FISCHER<sup>1</sup>; VITOR DAS NEVES AVELANEDA<sup>2</sup>; CARLA MARTINEZ CANELO<sup>3</sup>; DINALVA A. SALES<sup>4</sup>**

<sup>1</sup>*Universidade Federal do Rio Grande – juh.fischer@gmail.com*

<sup>2</sup>*Universidade Federal do Rio Grande – avelaneda.vitor@gmail.com*

<sup>3</sup>*Universidade Federal do Rio Grande – camcanelo@gmail.com*

<sup>4</sup>*Universidade Federal do Rio Grande – dsales@furg.br*

### 1. INTRODUÇÃO

O estudo e a classificação de galáxias evoluíram com o uso da espectroscopia, que permite caracterizar suas propriedades físico-químicas. O espectrógrafo infravermelho do telescópio Spitzer possibilitou avanços na faixa espectral de 5–37 $\mu$ m, revelando a composição da poeira e a presença de Hidrocarbonetos Aromáticos Policíclicos (PAHs), essenciais para compreender a interação entre radiação estelar e meio interestelar, além de distinguir ionização devido a formação estelar de núcleo ativo de galáxias (AGN, SALES; PASTORIZA; RIFFEL, 2010; SALES et al., 2013).

A poeira desempenha um papel central nesses ambientes, ao reprocessar a radiação estelar e emitir no infravermelho, sendo responsável por até 30% da energia óptica reirradiada (BERNSTEIN; FREEDMAN; MADORE, 2002). Galáxias com forte emissão no infravermelho médio (MIR) estão associadas a Starbursts ou AGNs, e suas assinaturas espectrais — como bandas de PAHs e absorção de silicatos em 9,7  $\mu$ m — são diagnósticos fundamentais (DRAINE, 2003; SALES; PASTORIZA; RIFFEL, 2010; SMITH et al. 2007).

Nesse contexto, SPOON et al. (2006) propuseram um esquema de classificação baseado na largura equivalente (EW) dos PAHs e na profundidade da absorção de silicatos, distinguindo galáxias Starburst, AGNs e objetos intermediários, além de subcategorias que consideram o grau de obscurecimento.

Embora eficaz, esse método pode ser refinado por abordagens quantitativas. A análise de componentes principais (PCA) é uma técnica estatística eficiente para reduzir a dimensionalidade dos dados espectrais e identificar padrões dominantes (RENCHER, 2003), permitindo complementar a separação proposta por Spoon. Neste trabalho, utilizamos dados disponíveis na base IDEOS<sup>1</sup> e aplicamos PCA seguida de clusterização hierárquica, construindo um dendrograma baseado nas principais componentes (JAFARZADEGAN; SAFI-ESFAHANI; BEHESHTI, 2019), a fim de verificar se a segmentação espectral obtida está alinhada com as classes estabelecidas por Spoon. Dessa forma, buscamos avaliar quantitativamente a robustez da separação entre AGNs (Classe A), Starbursts (Classe C) e objetos intermediários (Classe B).

Os resultados deste estudo contribuirão para o aprimoramento da modelagem espectral de galáxias infravermelhas e podem auxiliar na caracterização de populações obscurecidas no universo.

<sup>1</sup> <http://ideos.astro.cornell.edu/>

## 2. METODOLOGIA

Foram utilizados espectros da base IDEOS (HOUCK, 2004; WERNER, 2004; SPOON et al., 2022), cobrindo a faixa espectral de 5–37 $\mu$ m. Para garantir a qualidade da análise, aplicou-se um critério de seleção de razão sinal-ruído (S/N) na região de 6,6 $\mu$ m > 50, assegurando que apenas galáxias com alto sinal-ruído fossem consideradas na análise, eliminando aquelas com baixo nível de detecção ou grande incerteza. Após essa filtragem, dos 298 objetos restantes, foram selecionados apenas os que possuíam classificação no MIR de acordo com o diagrama diagnóstico de Spoon et al. (2006), especificamente as classes 1A, 1B e 1C, ou seja, galáxias não obscurecidas. Essa escolha inicial permite analisar melhor as emissões espectrais sem os efeitos da absorção pela poeira. A amostra final resultou em 226 galáxias, sendo 133 classificadas como 1A, 20 como 1B e 73 como 1C.

Para a aplicação da PCA, foi realizada a normalização das variáveis de fluxo, dividindo-as por seus valores máximos, a fim de evitar distinções artificiais causadas por diferenças de distância que tornam os fluxos de galáxias mais distantes relativamente menores. A PCA foi então aplicada para identificar padrões dominantes nos espectros das galáxias selecionadas, reduzindo a dimensionalidade dos dados ao encontrar combinações lineares das variáveis originais que explicam a maior variância (RENCHER, 2003). As variáveis de entrada incluíram a classificação no MIR, o redshift, os fluxos e EW das bandas de PAHs, a força de silicato em 9,7 $\mu$ m, os fluxos das linhas de emissão no MIR e as razões entre os valores do contínuo espectral em 15, 24 e 30 $\mu$ m em relação ao valor em 5,5 $\mu$ m.

Os coeficientes das componentes principais foram analisados para identificar quais características espectrais desempenham papel mais significativo na diferenciação das galáxias. A partir dessas componentes, aplicou-se uma clusterização hierárquica para agrupar os objetos com base em suas semelhanças espectrais, construindo um dendrograma que permite visualizar as relações entre eles e verificar se a segmentação resultante coincide com a classificação no MIR. A métrica de similaridade utilizada foi a distância euclidiana, e o método de agrupamento escolhido foi a ligação completa, que minimiza a variabilidade dentro dos clusters. Para avaliar a qualidade da clusterização e determinar o número ideal de agrupamentos, foram aplicadas duas métricas: o índice de Silhueta e o índice Calinski-Harabasz (CHI). O maior valor do índice de Silhueta foi 0,49, sugerindo quatro clusters, enquanto o CHI atingiu 152,43, indicando melhor separação com três clusters. Considerando que a classificação de (SPOON et al., 2006) define três classes e que esse número foi reforçado pelas métricas adotadas, o dendrograma final foi construído com três agrupamentos principais.

Por fim, os agrupamentos obtidos foram comparados às classes de Spoon presentes no catálogo IDEOS, de modo a validar a robustez da classificação proposta.

## 3. RESULTADOS E DISCUSSÃO

A PCA evidenciou que as maiores contribuições para a variabilidade dos dados estão associadas às razões do contínuo, especialmente C30 $\mu$ m/C5,5 $\mu$ m (0,879) e C24 $\mu$ m/C5,5 $\mu$ m (0,466), seguidas em menor escala por C15 $\mu$ m/C5,5 $\mu$ m (0,100). As demais variáveis, como larguras equivalentes das bandas de PAHs,

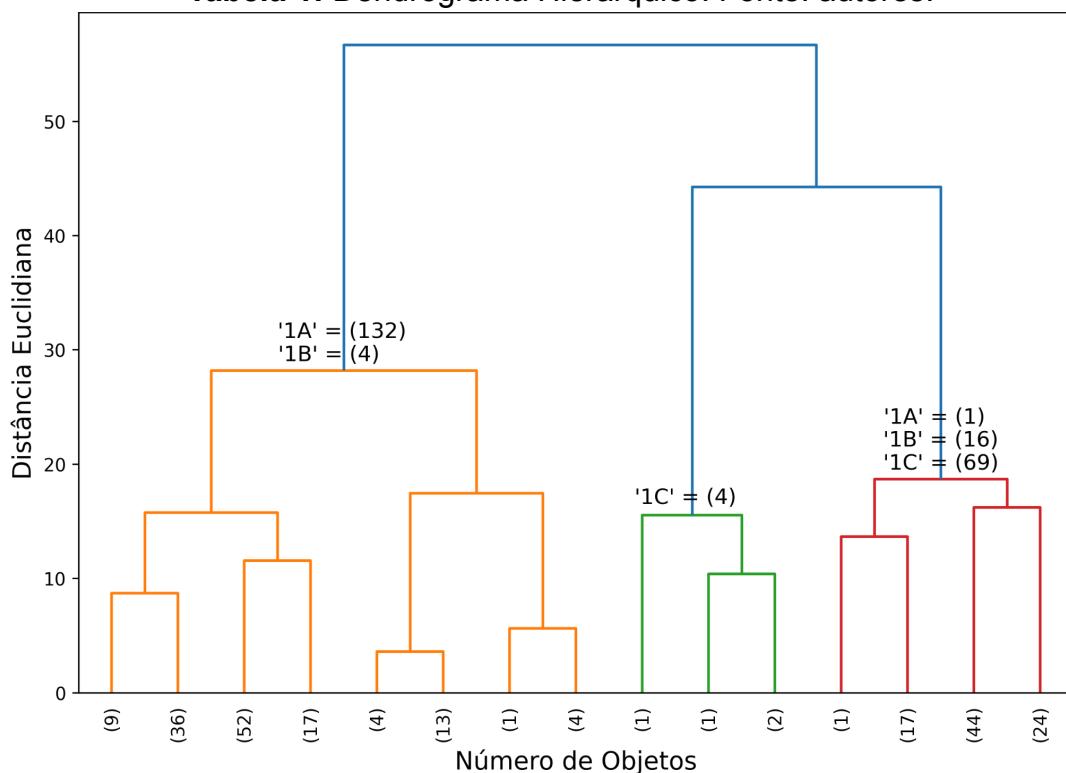
força de silicato e fluxos de linhas de emissão, apresentaram contribuições significativamente mais baixas.

As variações espectrais observadas foram dominadas pelas razões entre os valores do contínuo em diferentes regiões, refletindo efeitos associados às temperaturas do meio e ao avermelhamento, que diferenciam as galáxias Starburst de AGNs. Além disso, parâmetros como a EW das bandas de PAHs e a força de absorção dos silicatos também tiveram forte impacto, em concordância com o esquema de classificação proposto por Spoon. Em contraste, as linhas de emissão de enxofre, argônio e neônio mostraram menor relevância, enquanto apenas a banda de PAH em 11,3 $\mu$ m foi significativa nas componentes principais, possivelmente devido à forte extinção próxima à absorção do silicato em 9,7 $\mu$ m (AVELANEDA et al., 2024; HERNÁN-CABALLERO et al., 2020).

A clusterização hierárquica baseada nas 20 variáveis selecionadas pela PCA resultou em três agrupamentos principais, com distância euclidiana de separação de 44,25. O Cluster 1 concentrou 132 galáxias da classe 1A e 4 da 1B, o Cluster 2 reuniu 4 objetos da classe 1C, enquanto o Cluster 3 agrupou 69 galáxias 1C, 16 da 1B e uma 1A. Observou-se boa correspondência para as classes 1A (predominantes no Cluster 1) e 1C (concentradas no Cluster 3), mas as galáxias 1B, de natureza intermediária, foram distribuídas de forma menos consistente, refletindo a dificuldade do método hierárquico em separar variações espectrais de pequena escala. O Cluster 2, por sua vez, destacou um pequeno subconjunto de objetos da classe 1C com propriedades espectrais diferenciadas.

No total, 25 das 226 galáxias foram agrupadas de forma inadequada, correspondendo a um erro de 11%. Para maior robustez, recomenda-se identificar individualmente os objetos mal classificados, analisar seus espectros detalhadamente e correlacionar as componentes principais com parâmetros espectrais físicos, de modo a validar os padrões revelados pela PCA.

**Tabela 1:** Dendrograma Hierárquico. Fonte: autores.



## 4. CONCLUSÕES

Este estudo aplicou uma abordagem estatística à análise espectral de galáxias no MIR, mostrando que as classes 1A e 1C foram agrupadas com maior precisão, enquanto objetos com características de menor variabilidade exigem análises adicionais ou métodos de clusterização alternativos. Trabalhos futuros incluirão classes mais obscurecidas (2 e 3 de Spoon), visando refinar a modelagem espectral e ampliar a compreensão dos diferentes tipos de galáxias.

## 5. AGRADECIMENTOS

Os autores agradecem ao financiamento pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código Financeiro 001 e D.A.~Sales à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - Brasil (FAPERGS) - e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq).

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- AVELANEDA, V.; FISCHER, J.; SALES, D.; CANELO, C.; SILVA, E. Modelagem da extinção em galáxias Starburst: Impacto nas emissões de Hidrocarbonetos Aromáticos Policíclicos. In: **10º MCSUL**, Rio Grande, 2024. p.1-10.
- BERNSTEIN, R. A.; FREEDMAN, W. L.; MADORE, B. F. The first detections of the extragalactic background light at 3000, 5500, and 8000 Å. III. Cosmological implications. **ApJ**, v.571, n.1, p.107-120, 2002.
- DRAINE, B. T. Interstellar dust grains. **Annual Review of A&A**, v.41, n.1, p.241-289, 2003.
- HERNÁN-CABALLERO, A. et al. Extinction in the 11.2 µm PAH band and the low L11.2/LIR in ULIRGs. **MNRAS**, v.497, n.4, p.4614-4625, 2020.
- HOUCK, J. R. et al. The infrared spectrograph (IRS) on the Spitzer Space Telescope. **ApJS**, v.154, n.1, p.18-24, 2004.
- JAFARZADEGAN, M.; SAFI-ESFAHANI, F.; BEHESHTI, Z. Combining hierarchical clustering approaches using the PCA method. **Expert Systems with Applications**, Elsevier, v.137, p.1-10, 2019.
- RENCHER, A. C. Methods of Multivariate Analysis. **Wiley Series in Probability and Statistics**, New York: Wiley, 2003.
- SALES, D. A.; PASTORIZA, M. G.; RIFFEL, R. Polycyclic aromatic hydrocarbon and emission line ratios in active galactic nuclei and starburst galaxies. **ApJ**, v.725, n.1, p.605-620, 2010.
- SALES, D. A. et al. Polycyclic aromatic hydrocarbon in the central region of the Seyfert 2 galaxy NGC 1808. **MNRAS**, v. 429, n. 3, p. 2634-2642, 2013.
- SMITH, J. D. T. et al. The mid-infrared spectrum of star-forming galaxies: global properties of polycyclic aromatic hydrocarbon emission. **ApJ**, v.656, n.2, p.770-791, 2007.
- SPOON, H. W. et al. Mid-infrared galaxy classification based on silicate obscuration and PAH equivalent width. **ApJ**, v.654, n.1, p.L49-L52, 2006.
- SPOON, H. W. et al. The Infrared Database of Extragalactic Observables from Spitzer. II. The database and diagnostic power of crystalline silicate features in galaxy spectra. **ApJS**, v.259, n.2, p.37-55, 2022.
- WERNER, M. W. et al. The Spitzer space telescope mission. **ApJS**, v. 154, n. 1, p. 1, 2004.