

Universidade Federal do Rio Grande
Instituto de Matemática, Estatística e Física
Programa de Pós-Graduação em Física

Rosiane Carneiro da Rosa

**Estudo de hidrocarbonetos aromáticos
policíclicos em galáxias dominadas por
Starbursts: modelagem não-supervisionada
usando análise de correspondência**

Rio Grande

2022

Rosiane Carneiro da Rosa

**Estudo de hidrocarbonetos aromáticos
policíclicos em galáxias dominadas por
Starbursts: modelagem não-supervisionada
usando análise de correspondência**

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande - FURG, como requisito parcial para a obtenção do título de Mestre em Física.

Área de Concentração: Astronomia

Orientadora: Prof.^a Dr.^a Dinalva A. Sales

Rio Grande

2022

Ficha Catalográfica

R788e Rosa, Rosiane Carneiro da.

Estudo de hidrocarbonetos aromáticos policíclicos em galáxias dominadas por *Starbursts*: modelagem não-supervisionada usando análise de correspondência / Rosiane Carneiro da Rosa. – 2022. 141 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Física, Rio Grande/RS, 2022.

Orientadora: Dra. Dinalva Aires de Sales.

1. Astroquímica 2. Hidrocarbonetos policíclicos aromáticos
3. Modelagem não-supervisionada 4. Análise de correspondência
I. Sales, Dinalva Aires de II. Título.

CDU 547.555

Catlogação na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

Estudo de hidrocarbonetos aromáticos policíclicos em galáxias dominadas por Starbursts: modelagem não-supervisionada usando análise de correspondência

Rosiane Carneiro da Rosa

Orientadora:

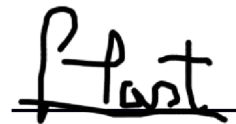
Prof^a. Dr^a. Dinalva Aires de Sales

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Física no Curso de Mestrado em Física, como parte dos requisitos necessários à obtenção do título de Mestre em Física.

Aprovada por:



Profª Drª. Dinalva Aires de Sales



Profª Drª. Miriani Pastoriza



Prof. Dr. Fabrício Ferrari

Rio Grande
Fevereiro de 2022

*“Aos esfarrapados do mundo e aos que neles se descobrem e, assim descobrindo-se,
com eles sofrem, mas, sobretudo, com eles LUTAM.” (Paulo Freire)*

Agradecimentos

Primeiro de tudo, agradeço aos Orisàs pela graça da vida e da oportunidade de evolução neste plano. Agradeço à minha família, pela inspiração em seguir firme e forte. Henrique meu amor, grata por ser meu porto seguro nos momentos de tempestade, tu e a Chocolate foram indispensáveis pra chegar até aqui.

Aos amigos de longe e de perto que com conversas, chamadas de vídeo e memes me ajudaram a lidar com a distância e a saudade. Pessoas do IFFar, em especial professor Mairon, Marta e sua “gangue da biblioteca” que incentivaram, acreditaram e apoiaram meu sonho, muito obrigada. Agradeço aos meus companheiros de orientação e de reuniões, Brenda, William e Douglas, vocês foram essenciais nessa caminhada. Aos colegas da pós Saiane, Solange e Vagner que mesmo em tempos de EAD, se fizeram presentes em vários momentos.

Aos colegas do AstroEsul, obrigada pela troca de conhecimento, foi um privilégio ter conhecido todos vocês. Aos professores do PPG Física, Matheus, Jorge, João Thiago, Fabricio e Stavros que deram o seu melhor nas aulas online. Professor Carlos Augusto do PGO que nas aulas de análise de dados, me ajudou a sair do vale de Dunning-Kruger.

Não poderia deixar de agradecer à minha orientadora Dinalva, por toda a paciência e por ter acreditado em mim, sem você esse sonho não teria se realizado. Obrigada

À PROPESP e CAPES, pelo apoio financeiro.

Às Instituições

“Chegar aqui de onde eu vim, é desafiar a lei da gravidade.”

Djonga

“O que dizemos ao Deus da Morte? - Hoje não!”

Arya Stark

Resumo

Hidrocarbonetos Aromáticos Policíclicos (PAHs) representam cerca de 20% do carbono existente no meio interestelar e também podem ser indicadores de formação estelar. No entanto, há discussões na literatura quanto à sua sobrevivência em galáxias ativas (AGNs). O objetivo deste trabalho foi efetuar uma análise estatística multivariada com modelagem não-supervisionada de dez propriedades físico-químicas das fontes de ionização e PAHs, em uma amostra de 148 galáxias dominadas por starburst. Foram realizadas análises de componentes principais (PCA), correspondência (CA), agrupamento hierárquico e a matriz de correlação de Pearson. Foi possível concluir que galáxias com domínio de fonte estelar apresentam correlação com PAHs pequenos (< 225 átomos de carbono), enquanto, AGNs apresentam correlação com moléculas de PAHs grandes (> 226). O resultado da PCA permite concluir que as três primeiras dimensões representam $\sim 58\%$ da variância amostral, sendo que grande parte desse valor é atribuído aos PAHs muito pequenos, pequenos e muito grande, além das fontes ionizantes rAGN e rPDR. A análise de agrupamentos infere a existência de 3 subgrupos na amostra, onde o Grupo 1 e 2 mostram significativa contribuição da componente estelar. Enquanto o Grupo 3, mostra dominância do núcleo ativo. A análise CA mostra que as três primeiras dimensões representam $\sim 66\%$ da variância amostral, onde grande parte desse valor é atribuído à largura equivalente do PAH em $6.2\mu\text{m}$ e à propriedade rAGN. Isso reforça a composição da amostra que são galáxias dominadas por starbursts, mas com parcelas de contribuição de AGNs. Os algoritmos de modelagem não-supervisionadas, além de validar os resultados publicados anteriormente eles também se mostraram ferramentas importante para a nova era da astroquímica onde será necessário trabalhar com um grande volume de dados como os que serão disponibilizados pelos grandes telescópicos, ex. “*JWST*, *LSST*, *ELT*” e etc.

Abstract

Polycyclic Aromatic Hydrocarbons (PAHs) represent about 20% of the carbon presented in the interstellar medium and can also be used as indicators of star formation. However, there are discussions in the literature as to their survival in active galaxies' nuclei (AGNs). The objective of this work was to perform a multivariate statistical analysis with unsupervised modeling of ten physico-chemical properties of ionization sources and PAHs. In a sample of 148 galaxies dominated by starburst. The principal components' analysis (PCA), the correspondence analysis (CA), the hierarchical clustering and the Pearson correlation matrix matrix were performed on the sample. This way, it was possible to conclude that galaxies dominated by stellar source correlate with small PAHs (<225 carbon atoms), while AGNs correlate with large PAHs molecules (>226). The PCAs allow us to conclude that the first three dimensions represent $\sim 58\%$ of the sample variance. With a large part of this value being attributed to very small, small and very large PAHs, in addition to the ionizing sources rAGN and rPDR. The clustering analysis infers the existence of 3 subgroups in the sample, where Group 1 and 2 show a significant contribution of the stellar component. While Group 3 shows dominance of the active core. The CAs analysis shows that the first three dimensions represent $\sim 66\%$ of the sample variance, where a large part of this value is attributed to the equivalent width of the PAH at $6.2\mu\text{m}$ and the rAGN property. This reinforces the composition of the sample, which are galaxies dominated by starbursts, but with AGN contribution shares. The non-supervised modeling algorithms, in addition to validating the previously published results also proved to be important tools for the new era of astrochemistry where it will be necessary to work with a large volume of data such as those that will be made available by large telescopes, e.g. “*JWST*, *LSST*, *ELT*” and etc.

Lista de Figuras

1.1	Classificação morfológica de Hubble, à esquerda estão as galáxias elípticas, dispostas na horizontal. Ao lado, a classe das lenticulares. No topo, galáxias espirais normais e abaixo as espirais barradas. À direita, as galáxias irregulares. Fonte: Gastão (2020)	22
1.2	A linha preta sólida representa o contínuo típico de “ <i>radio louds</i> ”, que são poderosas fontes de radio. A linha rosa tracejada retrata AGNs “ <i>radio quiets</i> ”, menos intensos que os anteriores. O pontilhado verde representa o contínuo de uma galáxia normal. Fonte: Gastão (2020)	23
1.3	Esquema do Modelo Unificado de AGNs. Em cinza com manchas vermelhas, se encontra o torus molecular ao redor do SMBH central. Em vermelho, o disco de acreção circundante e os jatos perpendiculares, em amarelo. São indicadas as localizações da BLR e NLR, bem como a direção do observador para identificar os objetos. Fonte: Adaptada de Zier e Biermann (2002). . .	26
1.4	Estruturas moleculares, com pericondensados à esquerda e catacondensados na direita. Fonte: Adaptado de Tielens (2005)	29
1.5	Espectro IR de PAHs. No topo e no centro da imagem são espécies neutras. Na parte debaixo, a molécula é ionizada. Fonte: Adaptado de PAHdb. . . .	30
2.1	Telescópio Espacial Spitzer. Fonte: https://www.spitzer.caltech.edu/ . . .	35
2.2	Espectros de 5 - 15 μm dos modelos AGN, HII e PDR utilizados na decomposição espectral. Também apresenta o modelo de emissão de silicato. Fonte: Hernán-Caballero e Hatziminaoglou (2011).	37

2.3	Espectros das 148 galáxias dominadas por Starbursts utilizadas neste trabalho estão representados pela cor rosa. O espectro na cor preta representa o espectro médio da amostra de 257 galáxias dominadas por Starburst derivado por Hernán-Caballero e Hatziminaoglou (2011). Todos os espectros foram normalizados em 7 μm . Fonte: a autora	38
2.4	Espectro IR no intervalo entre 5 e 15 μm de espécies moleculares disponíveis no PAHdb. Fonte: https://www.astrochemistry.org	39
2.5	Detalhamento dos PAHs no banco de dados teóricos, versão 2.00. Na horizontal, se encontra a quantidade de átomos de carbono e na vertical, o número de espécies moleculares. As cores indicam a composição das moléculas: em azul, aparecem as espécies “puras”, contém apenas carbono e hidrogênio; em vermelho, possuem adição de magnésio/ferro; em amarelo, adição de oxigênio e, em verde, adição de nitrogênio. As áreas hachuradas indicam adições e remoções entre versões 1.10 e 2.00. Fonte: (Boersma et al., 2014)	40
2.6	Detalhamento dos PAHs no banco de dados teóricos, versão 2.00. Na horizontal, se encontra a quantidade de átomos de carbono e na vertical, o número de espécies moleculares. As cores indicam a carga das moléculas: verde, indica cátions; em amarelo são os ânions e, em vermelho, neutros. As áreas hachuradas indicam adições e remoções entre versões 1.10 e 2.00. Fonte: (Boersma et al., 2014)	40
2.7	Uma estrutura dendrograma na qual o grupo inicial é subdividido em vários subgrupos. Na direção vertical, a similaridade é expressa de acordo com a distância euclidiana e na direção horizontal, o valor de observação. Fonte: https://support.minitab.com/	42
3.1	Matriz de Correlação de Pearson entre propriedades da amostra. Os rótulos indicam o grau de correlação, positivamente correlacionadas em verde e, em roxo, negativamente correlacionadas.	56
3.2	Porcentagem da contribuição de cada componente principal para a amostra. Cinco PCs concentram mais de 80% de toda a variabilidade encontrada na amostra.	57

3.3	Porcentagem da contribuição das propriedades originais para a composição dos componentes principais, obtida com a PCA. O tamanho e a transparência do círculo roxo é proporcional ao valor desta contribuição.	58
3.4	Histograma com o número ideal de clusters, o eixo horizontal mostra a quantidade de grupos sugeridas, indo de 0 a 6. No eixo vertical, quantos índices sugerem essa quantidade. De acordo com a regra da maioria, o melhor número de grupos é 3.	61
3.5	Dendrograma gerado pelo método Ward. Foi realizado um corte próxima a distância euclidiana 3, delimitando os três grupos: em rosa, o Grupo 1; em azul, o Grupo 2 e em verde o Grupo 3. O eixo horizontal apresenta a amostra total, colorida por grupos e nos rótulos são apresentados os nomes das galáxias.	66
3.6	Porcentagem da variância para cada dimensão, obtida com a CA. O tracejado azul indica o autovalor médio, mostrando que as três primeiras dimensões concentram a dessemelhança de toda amostra.	69
3.7	Porcentagem da contribuição das propriedades originais para cada dimensão. O tamanho e a transparência do círculo roxo é proporcional ao valor desta contribuição.	70
3.8	Percentil da contribuição das galáxias para a Dimensão 1, o autovalor médio é indicado pelo tracejado vermelho.	72
3.9	Percentil da contribuição das galáxias para a Dimensão 2, o autovalor médio é indicado pelo tracejado vermelho.	74
3.10	Percentil da contribuição das galáxias para a Dimensão 3, o autovalor médio é indicado pelo tracejado vermelho.	75
3.11	Biplot das propriedades das galáxias das Dimensões 1 e 2. As propriedades são representadas por triângulos pretos e as galáxias, coloridas de acordo com a contribuição para a composição das dimensões	76
3.12	Biplot das propriedades das galáxias das Dimensões 2 e 3. As propriedades são representadas por triângulos pretos e as galáxias, coloridas de acordo com a contribuição para a composição das dimensões.	77

Lista de Tabelas

1.1	Modos vibracionais no IR características dos PAHs. Fonte: Adaptado de Tielens (2005)	31
2.1	Amostra de Starbursts Dominated.	48
3.1	Galáxias pertencentes aos grupos 1, 2 e 3.	61
3.2	Medidas estatísticas dos Grupos 1, 2 e 3.	68

Sumário

1. Introdução	21
1.1 Galáxias com Núcleos Ativos	23
1.1.1 Galáxias Seyfert	25
1.1.2 Galáxias Starburst	27
1.2 Hidrocarbonetos Aromáticos Policíclicos (PAHs)	28
1.3 Motivação e Objetivos	30
2. Materiais e Métodos	35
2.1 Telescópio Espacial Spitzer	35
2.1.1 Câmera infravermelha - IRAC	36
2.1.2 Fotômetro de imagens multibanda - MIPS	36
2.1.3 Espectrógrafo infravermelho - IRS	36
2.2 Amostra de galáxias dominadas por Starburst	37
2.3 Biblioteca espectral das moléculas de PAHs	39
2.4 Métodos estatísticos multivariados	41
2.4.1 Análise de Agrupamento	41
2.4.1.1 Clustering Hierárquico com o Método Ward	43
2.4.2 Análise de Componentes Principais - PCA	44
2.4.3 Análise de Correspondência	45
2.5 Características gerais da amostra	45
3. Resultados e Discussões	55
3.1 Análise de Componentes Principais (PCA)	55
3.2 Análise de Agrupamento	60

3.2.1	Detalhamento dos Grupos	65
3.3	Análise de Correspondência (CA)	68
4.	<i>Conclusões</i>	79
	<i>Referências</i>	83
	<i>Apêndice</i>	89
A.	<i>Artigo Completo publicado na Revista CADERNOS DE ASTRONOMIA</i>	91
A.1	Rosa, Sales, Canelo & Matoso, v. 2, n. 2, p. 132, (2021)	91

Introdução

Conhecida por ser uma das ciências mais antigas, a Astronomia surgiu associada aos aspectos práticos da vida humana, como por exemplo, a medição da passagem do tempo, orientação em viagens e era primordial para a agricultura (Kepler e Saraiva, 2017). Telescópios mais potentes revelaram um céu desconhecido e a descoberta de outras faixas de radiação permitiu observações em toda a gama do espectro eletromagnético, apresentando o Universo além da luz visível. Dessa forma, a Astronomia impulsionou a evolução e o aperfeiçoamento de recursos tecnológicos ainda presentes no cotidiano.

A espectroscopia iniciou um novo campo de pesquisa, ao mostrar que a luz dos corpos celestes exibe não apenas cores, mas informações detalhadas sobre a composição química. Bem como o movimento de direção ou afastamento daquela fonte, por meio do efeito Doppler. Além disso, mostrou que quase todas as nebulosas espirais estão se afastando da Via Láctea em velocidades muito altas, o que gerou ainda mais polêmica sobre a natureza das nebulosas. Foi só em 1923 que Edwin P. Hubble usou estrelas variáveis Cefeidas para mostrar que as nebulosas são sistemas estelares tão grandes quanto a Via Láctea (Gastão, 2020).

Estima-se que existam cerca de 2 trilhões de galáxias no universo observável com tamanhos, formatos e luminosidades diferentes, classificadas em: elípticas (E); espirais subdivididas em normais (S) e barradas (SB); por fim as irregulares (I). Posteriormente, foi incluída uma classe intermediária, as lenticulares que podem ser normais (S0) ou barradas (SB0), como mostra a Fig.1.1. A classificação morfológica de Hubble ou diagrama do diapasão fornece informações à respeito da composição e complexidade das mesmas (Hubble, 1936).

Galáxias elípticas possuem diferentes graus de achatamento, populações estelares velhas

($> 7 \times 10^9$ anos) (Kepler e Saraiva, 2017), além de pouco gás e poeira. As galáxias espirais possuem nuvens moleculares, regiões de formação estelar e são ricas em gás e poeira. Podem ser classificadas pela razão entre tamanho do bojo/tamanho do disco e pelo formato dos braços espirais, onde predominam estrelas jovens ($< 7 \times 10^9$ anos). Já as estrelas mais velhas se encontram próximas ao núcleo. Já as galáxias irregulares não possuem formato definido, apresentam populações estelares jovens e velhas, para Gallagher e Hunter (1984) as estruturas irregulares podem ser resultado da interação gravitacional entre galáxias.

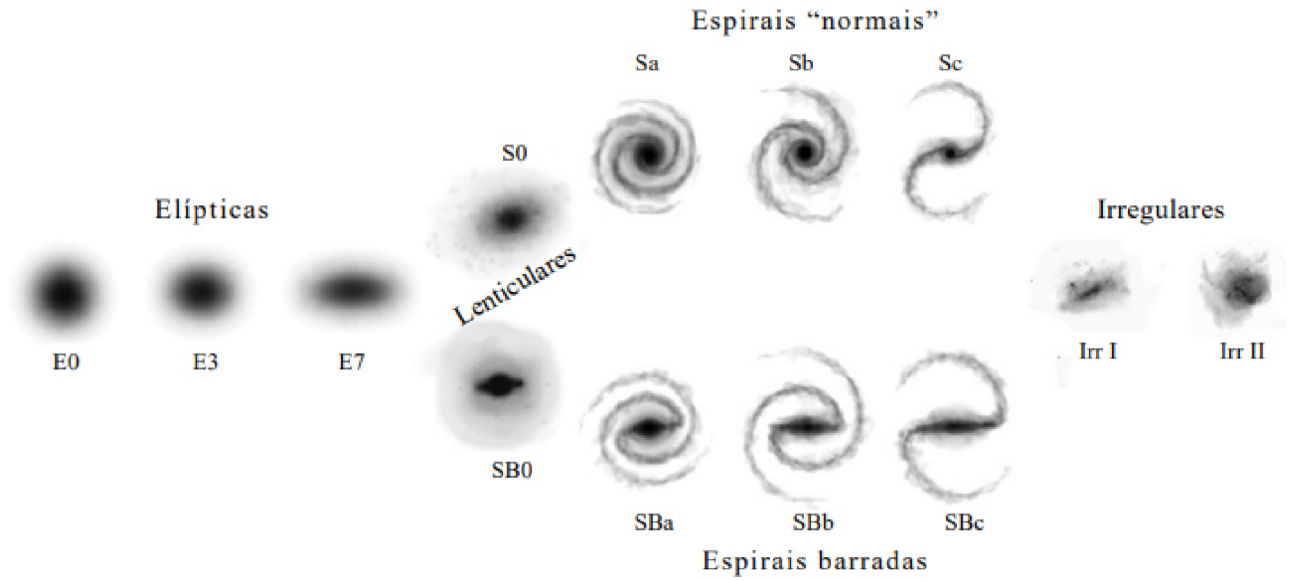


Figura 1.1: Classificação morfológica de Hubble, à esquerda estão as galáxias elípticas, dispostas na horizontal. Ao lado, a classe das lenticulares. No topo, galáxias espirais normais e abaixo as espirais barradas. À direita, as galáxias irregulares. Fonte: Gastão (2020)

O diagrama de Hubble classifica galáxias chamadas de “normais”, que apresentam espectros com linhas de absorção provenientes de estrelas e algumas linhas de emissão, provenientes do gás em regiões de formação estelar. Entretanto, existem galáxias com núcleos brilhantes e grande quantidade de linhas de emissão, não explicadas pelos processos térmicos estelares. Estes objetos com núcleos massivos e alta luminosidade são chamados galáxias com núcleos ativos (AGN, “*active galactic nuclei*”) (Peterson, 1997) que são objetos deste estudo e suas principais propriedades são descritas nas próximas seções.

1.1 Galáxias com Núcleos Ativos

Acredita-se que os AGNs são os objetos mais energéticos do Universo e podem emitir até $L \approx 10^{10} L_{\odot}$ (Peterson, 1997). Galáxias normais apresentam valores de luminosidade na ordem de $\sim 10^{38} \text{ erg s}^{-1}$, em contrapartida AGNs podem chegar à $\sim 10^{48} \text{ erg s}^{-1}$ (Beckmann e Shrader, 2012). A distribuição espectral de energia (SED, “*spectral energy distribution*”) dos AGNs difere de galáxias normais, como mostra a Figura 1.2 onde nota-se o grande intervalo de comprimentos de onda - desde raios gama até ondas de rádio. Bem como pela presença de duas regiões com fortes linhas de emissão contínua: uma no ultravioleta (UV, “*ultra-violet*”) e outra no raio-X (Schneider, 2015).

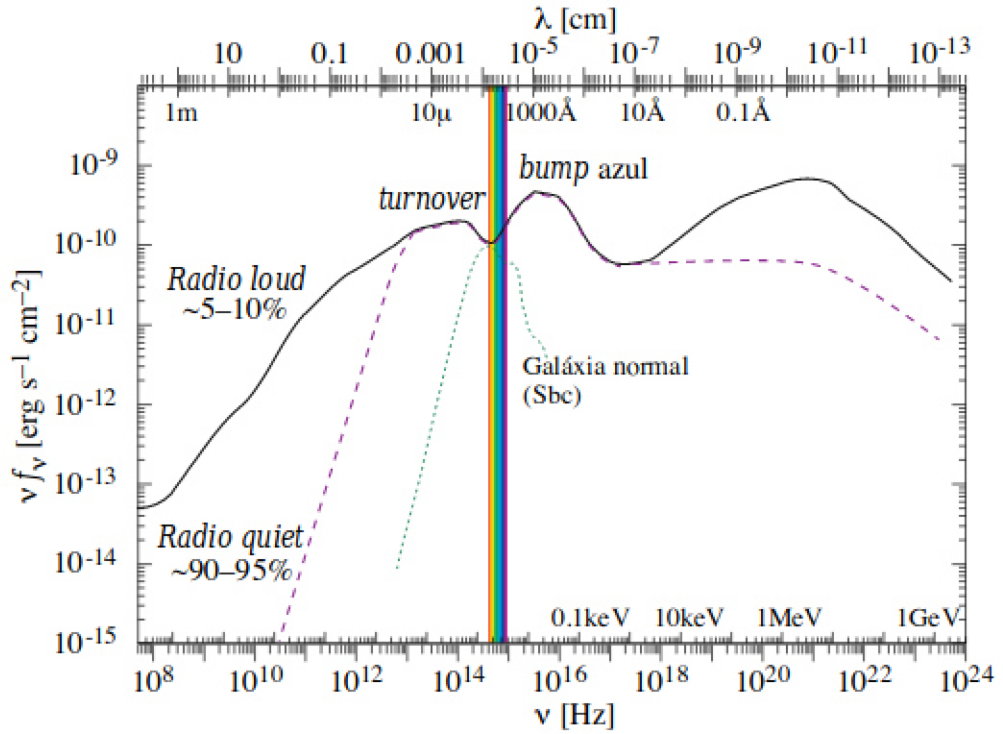


Figura 1.2: A linha preta sólida representa o contínuo típico de “radio louds”, que são poderosas fontes de radio. A linha rosa tracejada retrata AGNs “radio quiets”, menos intensos que os anteriores. O pontilhado verde representa o contínuo de uma galáxia normal. Fonte: Gastão (2020)

Podem apresentar linhas de emissão intensas e alargadas, com larguras à meia altura (FWHM, “*full width half maximum*”) de até 1000 km s^{-1} , além de variação no contínuo e nas linhas de emissão (Beckmann e Shrader, 2012). Uma vez que a energia dos AGNs não pode ser explicada por processos estelares, acredita-se que seja gerada pela transformação da energia potencial gravitacional da matéria, ao ser acretada a um Buraco-Negro Super-massivo (SMBH, “*super massive black hole*”, Antonucci e Miller, 1985). Essa proposta

soluciona a origem da intensa emissão de energia, bem como o pequeno tamanho da região emissora, ~ 100 pc que surge como fontes pontuais no espectro óptico Woltjer (1959).

O termo AGN abrange vários objetos astronômicos classificados de acordo com suas peculiaridades, como luminosidade e características espectrais. Os principais tipos são os quasares (“*quasi-stellar radio source*”), com luminosidade bolométrica entre $\sim 10^{44}$ e $\sim 10^{48} \text{ erg s}^{-1}$, considerados os mais luminosos e as Seyferts, cerca de 100 vezes menos luminosas que os quasares (Peterson, 1997). Outro exemplo de AGNs são as rádio galáxias, em sua grande maioria galáxias elípticas e como o nome indica são fortes emissoras no rádio, formando dois jatos relativísticos a partir núcleo, que se estendem por centenas de parsecs (Peterson, 1997).

O próximo grupo de AGNs são os blazares, extremamente luminosos, com alta variabilidade de fluxo no contínuo e forte emissão de rádio. São subdivididos em: galáxias variáveis ópticas violentas (OVVs, “*optically violently variable*”) e objetos BL Lac, que se destacam do primeiro pela ausência de linhas fortes, seja de emissão ou absorção (Schneider, 2015). Schneider (2015) cita que outros exemplos de AGNs são galáxias com regiões emissoras de linhas nucleares de baixa ionização (LINERs, “*low-ionization nuclear emission-line region*”), menos luminosas e com intensas regiões de absorção estelar.

Embora existam diversas classificações, os AGNs compartilham várias características observacionais, como a intensa luminosidade e variabilidade. Dessa forma, supõe-se que sejam o mesmo fenômeno visto por diferentes ângulos de observação. O Modelo Unificado proposto por Antonucci e Miller (1985) presume que a fonte de radiação dos AGNs seja um SMBH central, de massa $M \gtrsim 10^8 M_{\odot}$ e envolto em um toroide molecular com dimensões de até centenas de parsecs, composto de gás e poeira.

Desse modo, os AGNs são classificados em dois grupos (tipo 1 e tipo 2), uma vez que a orientação espacial do toroide determina quais características serão observadas. A região mais compacta e próxima ao disco de acreção ($r < 1$ pc), chamada região de linhas largas (BLR, “*broad line region*”), possui nuvens de gás com alta densidade eletrônica ($n_e \sim 10^{11} \text{ cm}^{-3}$) e temperaturas até 20000 K. Isto ocasiona linhas com larguras típicas de 5000 km s^{-1} até 30000 km s^{-1} em algumas galáxias, além da ausência de linhas proibidas por conta da alta densidade eletrônica (Antonucci, 1993).

A região de linhas estreitas (NLR, “*narrow line region*”) se encontra além do toro molecular (de algumas dezenas até ~ 1 Kpc), é mais extensa e composta por gás menos

denso ($n_e \sim 10^3 \text{cm}^{-3}$) e mais frio. Nessas regiões surgem a emissão de linhas com larguras entre $200 - 1000 \text{ km s}^{-1}$, originadas por transições permitidas e proibidas (Antonucci, 1993). Além disso, no espectro de AGNs existem linhas coronais, alargadas e deslocadas para o azul (“*blueshift*”), sugerindo uma possível região intermediária entre a BLR e NLR: a região de linhas coronais (CLR, “*coronal line region*”, Rodríguez-Ardila et al., 2002). Dessa forma, o Modelo Unificado de AGNs classifica os objetos em duas categorias que são:

- AGNs do tipo 1: o ângulo de visada do observador com o núcleo da galáxia é de frente (“*face-on*”), permitindo visualizar as assinaturas espectrais da BLR e NLR.
- AGNs do tipo 2: o ângulo de visada do observador com o núcleo da galáxia é de perfil (“*edge-on*”), onde, a visão da BLR é obscurecida pelo toro molecular e somente a assinatura espectral das NLR são visualizadas.

Esse modelo foi proposto para solucionar o paradigma das galáxias Seyferts. A Figura 1.3, apresenta uma esquematização do modelo, onde as setas indicam as classes de AGNs, de acordo com o ângulo formado entre a linha de visada com o toro molecular e o jato. Os objetos que possuem jatos e BLRs obscurecidas são as rádio galáxias de linhas estreitas (NLRGs, “*Narrow Line Radio Galaxies*”) e são AGNs do tipo 2. Caso não possuam, são rádio galáxias de linhas largas (BLRGs, “*Broad Line Radio Galaxies*”) e são AGNs do tipo 1. Nas direções polares, com ausência de emissão rádio são chamadas quasares rádio silenciosos (“*radio-quiet*”).

1.1.1 Galáxias Seyfert

A primeira galáxia de núcleo ativo identificada, foi a NGC 1068 por Fath (1909) e diferente das demais, possuía fortes linhas de emissão e absorção, características de regiões altamente ionizadas. Somente mais tarde, Seyfert (1943) identificou espectros de outros objetos que apresentavam características semelhantes, como linhas de He I e He II alargadas e linhas de excitação média, como [O III] $5007\text{\AA}$. Chamadas de Seyferts, essas galáxias são os tipos mais comuns de AGNs. Posteriormente, o trabalho de Khachikian e Weedman (1974) as classificou em dois grupos, baseados na presença ou ausência de linhas largas no espectro óptico.

- Seyfert tipo 1 (Sy 1) - estas galáxias apresentam linhas permitidas alargadas, indicando altas velocidades desde 10^3 km s^{-1} até 10^4 km s^{-1} e originadas em um

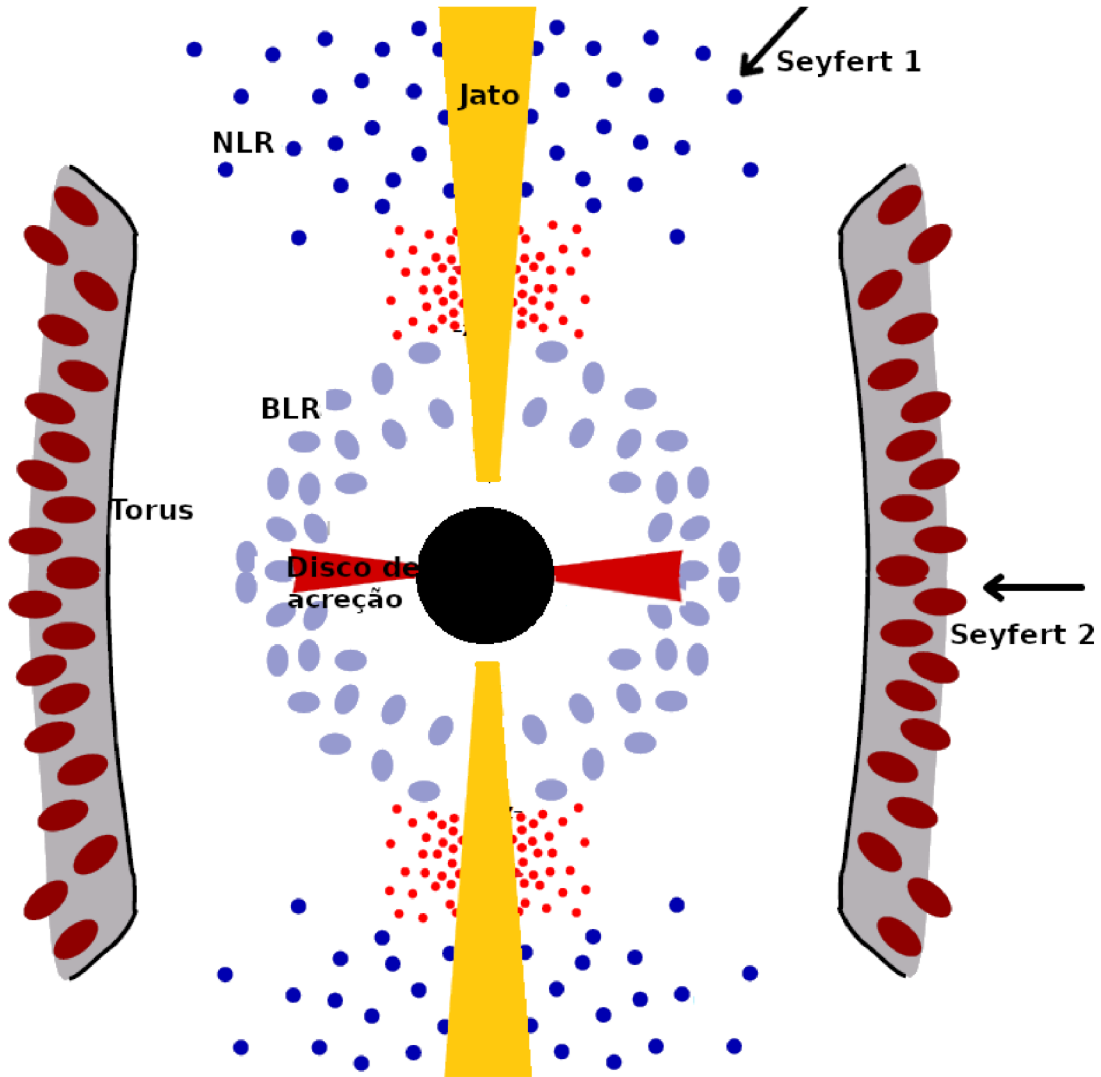


Figura 1.3: Esquema do Modelo Unificado de AGNs. Em cinza com manchas vermelhas, se encontra o torus molecular ao redor do SMBH central. Em vermelho, o disco de acreção circundante e os jatos perpendiculares, em amarelo. São indicadas as localizações da BLR e NLR, bem como a direção do observador para identificar os objetos. Fonte: Adaptada de Zier e Biermann (2002).

gás denso com $n_e \approx 10^3 - 10^6 \text{ cm}^{-3}$. Apresentam linhas estreitas relacionadas às transições permitidas e também às chamadas proibidas, que recebem esse nome por ocorrerem somente em densidades muito baixas como o espaço ou camadas superiores da atmosfera. As linhas estreitas apontam a presença de gás ionizado com baixa densidade eletrônica ($n_e \approx 10^3 - 10^6 \text{ cm}^{-3}$). Em torno de 30% das Seyferts são do tipo 1, que se destaca ainda pelo contínuo comportar-se como uma lei de potência.

- Seyfert tipo 2 (Sy 2) - estas galáxias apresentam linhas de emissão estreitas, tanto permitidas quanto proibidas. Cerca de 60% das seyferts fazem parte do tipo 2, que se diferencia do primeiro tipo por acrescentar um componente estelar ao contínuo.

Entre esses dois extremos, existem objetos intermediários com linhas permitidas largas e estreitas. Além de outras características, dentre elas: a forte emissão de raio-X nas Seyferts tipo 1.5; a presença de linhas largas de $H\alpha$, mas não a de $H\beta$ nas Seyfert tipo 1.9 ou as Seyfert tipo 1.8 que apresentam linhas largas de $H\alpha$ e $H\beta$. Galáxias Seyferts se destacam não apenas por propriedades espectrais, mas também pela luminosidade do núcleo que deixa visível o restante da galáxia hospedeira e pelo fato de grande parte das Seyferts habitam galáxias espirais.

1.1.2 Galáxias Starburst

O termo Starburst é usado para caracterizar galáxias que experimentam curtos períodos de intensa formação estelar, normalmente na região central (Leitherer et al., 2001). Apesar de se concentrarem em pequenas regiões entre 10 - 1000 kpc, em torno de 1 - 10%, possuem uma taxa de formação estelar (SFR, “*star formation rate*”) acima da soma total do restante da galáxia hospedeira. Galáxias como a Via Láctea possuem uma SFR de $\approx 2 M_{\odot}/\text{ano}$ com um raio em torno de 10 kpc, enquanto uma típica starburst a SFR fica entre 10 - 300 $M_{\odot}/\text{ano}$, em um raio de 0,5 kpc (Schneider, 2015). A atividade starburst é breve em relação à idade da galáxia hospedeira, cerca $\sim 10^8$ anos, diferente das galáxias comuns que mantém a formação estelar constante ao longo de sua existência.

As regiões starburst se destacam pela presença de estrelas do tipo espectral O e B, que emitem grande parte da radiação em UV e com luminosidade bolométrica entre $10^7 - 10^8 L_{\odot}$. Bem como pela abundância de gás interestelar e poeira, necessários para formação estelar, a radiação dessas estrelas ioniza gás presente na região entre $6 < E < 13.6$ eV, produzindo fortes linhas de emissão de H (Tielens e Hollenbach, 1985). Já os grãos de poeira absorvem nos demais comprimentos de onda, aquecem e resfriam emitindo no infravermelho médio (MIR, “*middle infrared*”) e distante (FIR, “*far infrared*”), em torno de $\lambda \sim 30 - 300 \mu\text{m}$.

Galáxias starburst e AGNs possuem características espectrais similares. Entretanto, podem ser observadas diferenças como a ionização por estrelas jovens nas starbursts e em AGNs, pelo disco de acreção no núcleo. Porém a intensidade de algumas linhas, como $[\text{OII}]\lambda 5007/H\beta$, $[\text{NII}]\lambda 6583/H\alpha$, $[\text{SiII}]\lambda\lambda 6716, 6731/H\alpha$ e $[\text{OI}]\lambda 6300/H\beta$ permitem a distinção entre ambas (Baldwin et al., 1981). Esses autores sugeriram um gráfico, chamado Diagrama BPT (Baldwin, Phillips & Terlevich) com base na razão $[\text{OII}]\lambda 5007/H\beta$ em função da

razão $[\text{NII}]\lambda 6583/\text{H}\alpha$ para classificar as linhas de emissão espectrais. E embora existam galáxias com formação estelar e núcleo ativo de forma simultânea, além de propriedades bastante similares às das regiões HII, chamadas de Sérsic-Pastoriza (Sérsic e Pastoriza, 1967; Pastoriza, 1975). Ainda não se sabe ao certo a conexão entre ambos objetos (Peterson, 1997).

O processo que dá origem as starbursts ainda não foi totalmente compreendido. Contudo, duas vertentes tentam explicar a formação desse fenômeno. Em galáxias espirais barradas o processo ocorre nas regiões centrais, aproximadamente $r \lesssim 1$ kpc. Devido à presença das barras, o gás se acumula no centro e alimenta o processo. A outra hipótese seria em relação às colisões galácticas, causando a fusão das galáxias ou formação de caudas de maré. Caso as galáxias em interação sejam ricas de gás, pode surgir uma região com intensa formação estelar.

1.2 Hidrocarbonetos Aromáticos Policíclicos (PAHs)

Os últimos anos foram abundantes na divulgação de pesquisas referentes aos fenômenos físico-químicos que ocorrem em ambientes astronômicos. Os avanços tecnológicos disponibilizaram grandes quantidades de dados observacionais. Atualmente, o censo de moléculas interestelares contém 241 espécies individuais, compostas por 19 elementos diferentes e possuem tamanhos desde 2 até 70 átomos (McGuire, 2021). Essas espécies contribuem para o ciclo de vida da matéria orgânica no meio interestelar (ISM, “*interstellar medium*”) e estão presentes em todas as fases (Tielens, 2013).

Observações nas décadas de 70 e 80 descobriram que alguns objetos apresentavam intensas bandas espectrais no IR, mas como a fonte dessas bandas não eram identificadas, foram chamadas de bandas de infravermelho não identificadas (UIRs, “*unidentified infrared*”) (Tielens, 2008). Somente mais tarde foram atribuídas aos hidrocarbonetos aromáticos policíclicos (PAHs, “*polycyclic aromatic hydrocarbon*” - De Becker, 2013). Os PAHs são compostos orgânicos arranjados em anéis benzênicos e que apresentam estruturas hexagonais, ligadas à sua estabilidade. Como visto na Figura 1.4, os catacondensados são lineares, enquanto os pericondensados são núcleos compactos de carbono, garantindo maior estabilidade a molécula.

Aproximadamente 20% ou até mais de todo carbono existente no ISM, encontra-se na

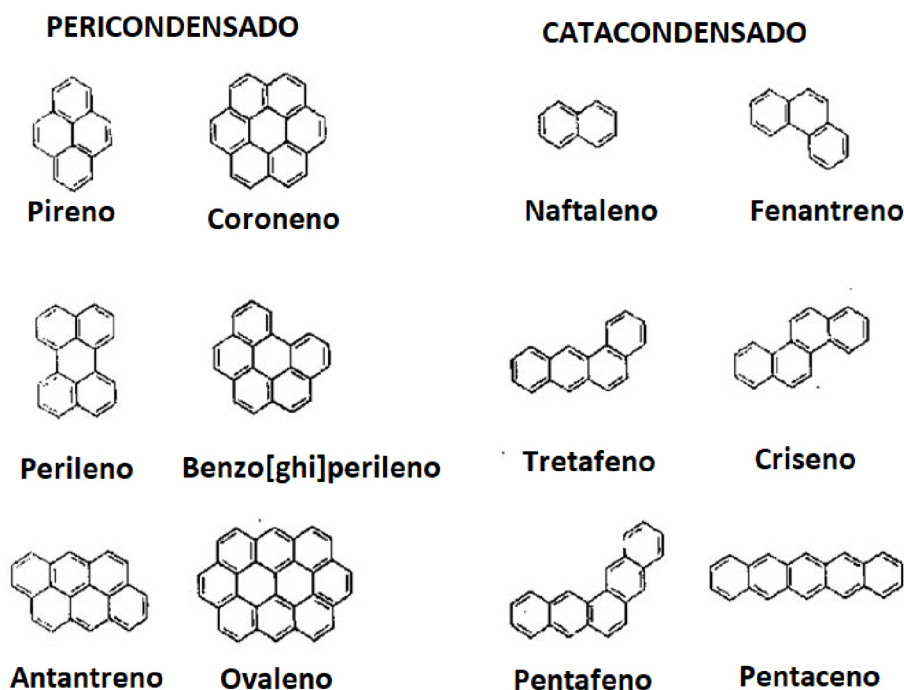


Figura 1.4: Estruturas moleculares, com pericondensados à esquerda e catacondensados na direita. Fonte: Adaptado de Tielens (2005)

forma de PAHs (Joblin et al., 1992) e podem desempenhar importante participação na origem da vida, pois são capazes de substituir alguns de seus átomos de carbono por nitrogênio ou oxigênio e (Ehrenfreund et al., 2006). Diversos objetos astronômicos como regiões H II e estrelas ricas em carbono possuem bandas de emissão atribuídas aos PAHs. Bem como nebulosas planetárias e AGNs. Isso infere que são materiais resistentes e abundantes no ISM (e.g. Tielens, 2013; Sales et al., 2010, 2013).

Os PAHs também afetam no aquecimento de gás neutro e no equilíbrio de ionização nas nuvens moleculares. Além disso, são importantes marcadores em regiões de formação estelar. A emissão nessas regiões surge de PAHs excitados por fótons UV, que ao absorverem a radiação ficam altamente excitadas, vibram e emitem no IR, durante o processo de desexcitação (Tielens, 2005). A Figura 1.5 apresenta espectros de pequenos PAHs, onde percebe-se a semelhança nas posições das assinaturas espectrais, uma vez que são construídos a partir da mesma unidade básica: o anel de benzeno.

Perto de 50% da luminosidade no MIR pode ser associada às bandas de emissão mais intensa dessas moléculas (Li, 2004). Estas variam de acordo com o tamanho, a geometria molecular e são características dos modos de excitação vibracional, *bending* (dobramento) e

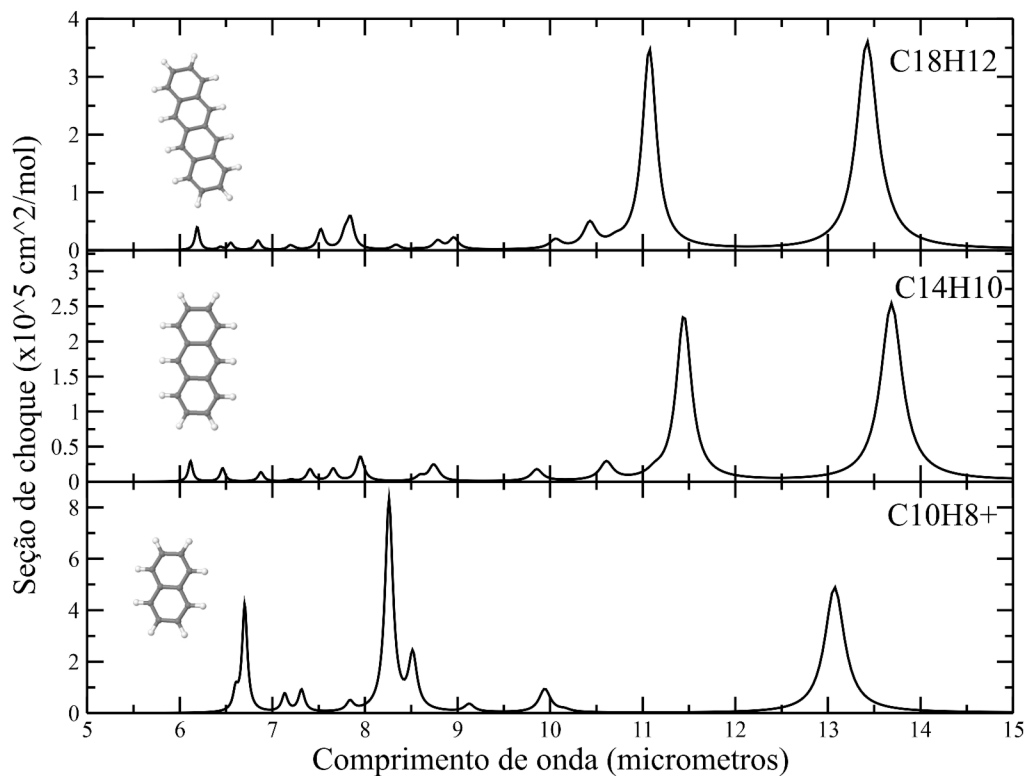


Figura 1.5: Espectro IR de PAHs. No topo e no centro da imagem são espécies neutras. Na parte debaixo, a molécula é ionizada. Fonte: Adaptado de PAHdb.

stretching (estiramento, Tielens, 2008). A Tabela 1.1 permite relacionar os modos vibracionais associados a cada banda de emissão características, esses valores são bem conhecidos na literatura (Tielens, 2005) e encontram-se entre 3-20 μm .

1.3 Motivação e Objetivos

Os AGNs são caracterizadas por sua intensa luminosidade. Estes objetos emitem cerca de 100 a 1000 vezes a energia de toda Via Láctea, $\sim 10^{48} \text{ erg s}^{-1}$ (Sales, 2012) e apresentam espectros dominados pela emissão infravermelha. Esse intervalo espectral permite compreender os processos físicos que levam a esses fenômenos. Além disso, a região do MIR contém linhas espectrais relacionadas a presença de espécies carbonáceas, mais especificamente, de PAHs. Hernán-Caballero e Hatziminaoglou (2011), apontam para a presença dessas moléculas em starbursts, desempenhando a função de estimar indiretamente a SFR.

Pesquisas conduzidas por Smith et al. (2007) e O'Dowd et al. (2009), mostraram que os PAHs não poderiam sobreviver nos AGNs, devido à fragilidade dessas moléculas. No

Tabela 1.1 - Modos vibracionais no IR características dos PAHs. Fonte: Adaptado de Tielens (2005)

Bandas de emissão	Modos de excitação
3.3 μm	alongamento de C-H
3.4 μm	alongamento da ligação alifática dos grupos metil
	alongamento de C-H em PANHs
5.2 μm	flexão de C-H
	alongamento de C-C
5.65 μm	flexão de C-H
	alongamento de C-C
6.0 μm	alongamento de C-O
6.2 μm	alongamento de C-C
6.9 μm	alongamento de C-H em ligação alifática
7.6 μm	alongamento de C-C
	flexão de C-H no plano da molécula
7.8 μm	alongamento de C-C
	flexão de C-H no plano da molécula
8.6 μm	flexão de C-H no plano da molécula
11.0 μm	flexão de C-H fora do plano da molécula, cátion
11.2 μm	flexão de C-H fora do plano da molécula, neutro
12.7 μm	flexão de C-H fora do plano da molécula, cátion
13.6 μm	flexão de C-H fora do plano da molécula
14.2 μm	flexão de C-H fora do plano da molécula
16.4 μm	flexão de C-C-C dentro e fora do plano da molécula no anel

entanto, o tamanho da amostra utilizada pelos autores era de tamanho reduzido, cerca de < 20 galáxias. Por outro lado, Voit (1992) propôs a existência de um material denso, que bloqueia a radiação vinda do núcleo e permite a presença de PAHs nos AGNs. Essa hipótese corrobora a proposta do modelo unificado de AGNs, que supõe um toro de poeira ao redor do núcleo ionizante e confirmado observacionalmente na literatura (e.g. Sales et al., 2010, 2013; Canelo et al., 2018; Ruschel-Dutra et al., 2014; Silva-Ribeiro et al., 2022; Alonso-Herrero et al., 2014).

Com base na pesquisa de Sales et al. (2013), Miranda et al. (2019) e Rosa et al. (2021) desenvolveram estudos de modelagem computacional para pequenos PAHs, a fim de investigar a presença de bandas de emissão molecular nos espectros de galáxias Seyferts e dominadas por starburst. Os resultados mostram que é possível inferir a presença dessas espécies aromáticas nos objetos extragalácticos analisados. Assim como a presença

de heterociclos policíclicos aromáticos nitrogenado (PANHs, “*polycyclic aromatic nitrogen heterocycles*”), que ocorrem quando os PAHs substituem alguns de seus átomos de carbono por nitrogênio, o que, para Canelo et al. (2018), representam uma das reservas de nitrogênio no Universo.

Utilizando uma amostra de 186 objetos, Sales et al. (2010) obtiveram resultados interessantes. Os autores verificaram que PAHs em AGNs são maiores (> 180 átomos de carbono), enquanto nas regiões HII e galáxias starburst, seu tamanho é menor (< 180 átomos de carbono). Também apontaram que moléculas presentes em AGNs eram altamente ionizadas, enquanto nas starburst, possuem baixo grau de ionização. Ainda, os espectros analisados mostraram bandas de emissão dos PAHs nas regiões formação estelar ao redor do núcleo. Isso indica que essas moléculas podem ser usadas como indicadores da SFR, bem como aponta para a sobrevivência de PAHs perto do núcleo ativo, uma vez que a emissão foi observada na parte central do AGN NGC1808 e outras galáxias Seyferts (ver Sales et al., 2013; Alonso-Herrero et al., 2014).

Dessa forma é fundamental expandir as pesquisas referentes aos PAHs em ambientes astrofísicos, em especial nas galáxias dominadas por starburst com coexistência de AGNs, a fim de entender quais estruturas permitem a presença abundante dessas espécies moleculares em condições inóspitas do espaço. Assim, avançar a compreensão das propriedades físicas e químicas dos PAHs, bem como possíveis relações entre a dimensão molecular e a principal fonte de ionização das galáxias. Portanto, os objetivos específicos desse trabalho são:

- Aplicar uma análise de componentes principais em 10 propriedades física-química de uma amostra de 148 galáxias dominadas por starburst para derivar as componentes principais dessa matriz de dados;
- Derivar a matriz de correlação de Pearson das 10 propriedades de uma amostra de 148 galáxias dominadas por starburst para encontrar correlações entre as variáveis;
- Fazer uma análise de agrupamentos hierárquico utilizando as 10 propriedades física-química de uma amostra de 148 galáxias dominadas por starburst para determinar grupos de galáxias semelhantes com método estatístico multivariado e supervisionado;

-
- Aplicar a técnica multivariada de *machine learning* não supervisionada, análise de correspondência, em 10 propriedades física-química de uma amostra de 148 galáxias dominadas por starburst para estudar eventuais associações entre variáveis e entre suas respectivas categorias e complementar com os resultados de agrupamentos encontrados com as componentes principais.

Materiais e Métodos

Neste capítulo serão abordados os instrumentos e as base de dados utilizadas. Bem como os métodos e os procedimentos estatísticos e matemáticos, que possibilitaram a realização desta análise.

2.1 *Telescópio Espacial Spitzer*

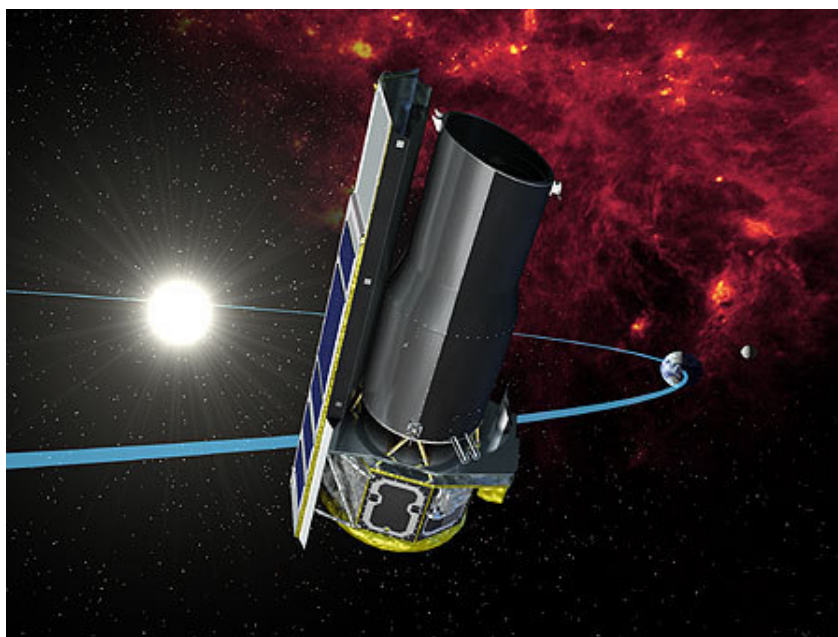


Figura 2.1: Telescópio Espacial Spitzer. Fonte: <https://www.spitzer.caltech.edu/>

Lançado inicialmente para uma missão de 2.5 anos, o Telescópio Espacial Spitzer, visto na Figura 2.1, permaneceu em operação por quase 17 anos. Projetado para investigar a região entre 3 a $180\ \mu\text{m}$ do IR, permitiu observar através de nuvens de gás e poeira. Também coletou dados de nuvens moleculares, discos protoplanetários, exoplanetas e de-

mais objetos astronômicos. Isso proporcionou maiores informações sobre a formação de galáxias, além disso muitas de suas contribuições servirão de guia para a missão do James Webb Space Telescope (JWST).

Dividido em dois compartimentos, na parte quente se encontram os sistemas de controle, comunicação e gerenciamento de comandos, bem como a base e painéis solares. Na parte fria ou criogênica se localizam o criostato que mantém o equipamento resfriado, o telescópio de 85 cm e as três ferramentas de observação: IRAC, MIPS e IRS, responsáveis pela coleta de dados. Essas componentes são isoladas termicamente e envoltos em uma cápsula feita de berílio.

2.1.1 Câmera infravermelha - IRAC

A IRAC (*"Infrared Array Camera"*) era uma câmera programada para detecções no infravermelho próximo e médio (3,6 - 8,0 μm). Seus quatro detectores, cada um com 256x256 pixels de dimensão captavam comprimentos de onda específicos: 3,6 μm ; 4,5 μm ; 5,8 μm e 8,0 μm respectivamente.

2.1.2 Fotômetro de imagens multibanda - MIPS

O MIPS (*"Multiband Imaging Photometer for Spitzer"*) era uma câmera para comprimentos de onda entre 24,7 - 160 μm e também realizava espectroscopia simples. Os instrumentos detectores para a região de 24 μm tinham dimensões de 128x128 pixels; na região de 70 μm , 32x32 pixels e a região de 160 μm , 2x20 pixels.

2.1.3 Espectrógrafo infravermelho - IRS

O espectrógrafo IRS (*"Infrared Spectrograph"*) viabilizou dados em alta e baixa resolução no infravermelho médio (5 - 40 μm). Cada um de seus quatro módulos, com dimensões de 128x128 pixels, atuavam em comprimentos de onda e resolução diferentes, além disso possuíam sua própria fenda para permitir a entrada das bandas infravermelhas: um módulo de baixa resolução e comprimento de onda curto para a região de 5,3 - 14 μm ; um módulo de baixa resolução e comprimento de onda longo para a região de 14 - 40 μm ; um módulo de alta resolução e comprimento de onda curto para 10 - 19,5 μm e, por fim, um módulo de alta resolução e comprimento de onda longo para observações na região entre 19 - 37 μm .

2.2 Amostra de galáxias dominadas por Starburst

O Projeto Spitzer/IRS ATLAS é uma coletânea de aproximadamente 750 espectros de diversas fontes astronômicas como galáxias obscuras e não obscurecidas, starbursts, galáxias infravermelhas luminosas e ultraluminosas, dentre outras obtidas pelas observações do IRS do Telescópio Espacial Spitzer. Além de espectros em baixa resolução ($R \sim 100$) e alta resolução ($R \sim 600$), disponibiliza catálogos com até 200 propriedades como resistência de silicatos, larguras equivalentes de PAHs e luminosidades, bem como modelos obtidos a partir do espectro médio de populações específicas (Hernán-Caballero e Hatziminaoglou, 2011). No acervo do projeto encontram-se classificações obtidas por meio da decomposição espectral, inferida através da equação:

$$F_{\lambda}(\lambda) = e^{-b\tau(\lambda)} \left(a_1 f_{AGN}(\lambda) + a_2 f_{HII}(\lambda) + a_3 f_{PDR}(\lambda) \right) \quad (2.1)$$

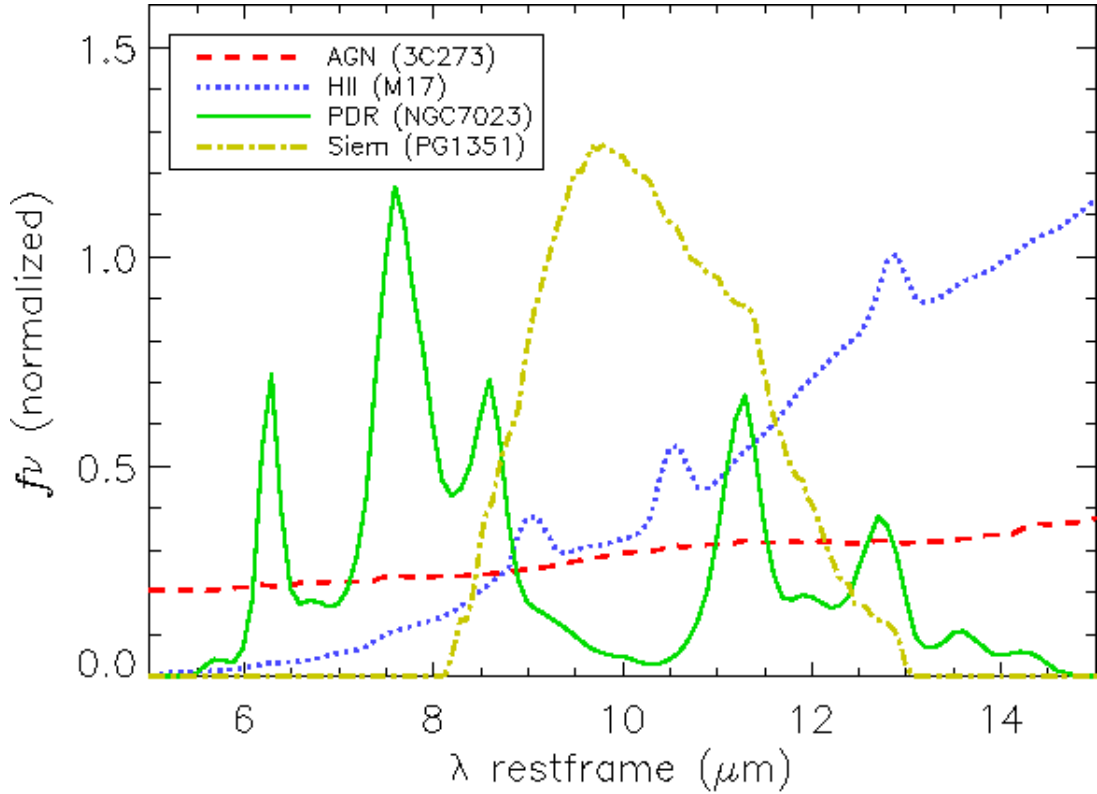


Figura 2.2: Espectros de 5 - 15 μm dos modelos AGN, HII e PDR utilizados na decomposição espectral. Também apresenta o modelo de emissão de silicato. Fonte: Hernán-Caballero e Hatziminaoglou (2011).

em que os parâmetros livres (b, a_1, a_2, a_1) são obtidos por meio do método de Levenberg-Marquardt ¹. Os termos da equação são representados pelos espectros de 3 objetos

¹ O algoritmo de Levenberg-Marquardt é um método de otimização computacional, que encontra

vistos na figura 2.2, a saber: o quasar 3C273 para f_{AGN} ; a região de fotodissociação (PDR, “*photodissociation region*”) na nebulosa de reflexão NGC 7023 para f_{PDR} e regiões próximas de estrelas OB na galáxia M17 para f_{HII} .

Dessa forma, foram quantificados os indicadores r_{HII} , r_{PDR} e r_{AGN} . Estes fornecem a razão da sua luminosidade integrada para a luminosidade total (r), no intervalo espectral de 5 - 15 μm , para os componentes HII, PDRs e AGNs (Hernan-Caballero et al., 2009). Uma vez que a maior parte da emissão IR de galáxias ativas vem dessas regiões, é relevante compreender o quanto cada componente contribui para a luminosidade vinda desses sistemas. Outro procedimento adotado pelos autores citados, foi a seleção dos seguintes parâmetros: a intensidade da absorção ou emissão de silicato em 10 μm (S_{sil}) em relação à largura equivalente da intensidade da banda de PAH em 6.2 μm (EW_{62}), também conhecido como diagrama Fork (Spoon et al., 2007). Esse diagrama permite diferenciar as fontes de emissão dominante em galáxias (Hernán-Caballero e Hatziminaoglou, 2011).

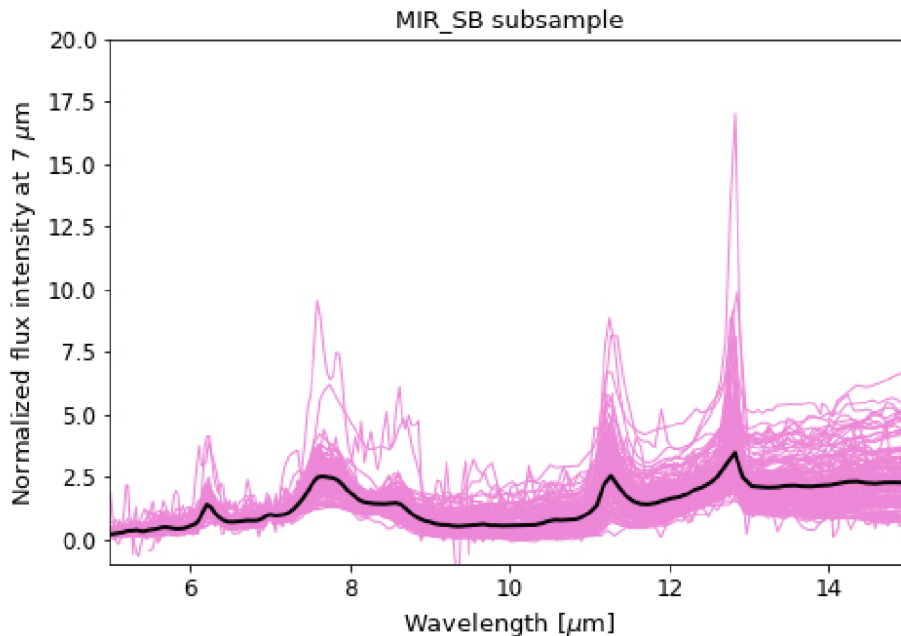


Figura 2.3: Espectros das 148 galáxias dominadas por Starbursts utilizadas neste trabalho estão representados pela cor rosa. O espectro na cor preta representa o espectro médio da amostra de 257 galáxias dominadas por Starburst derivado por Hernán-Caballero e Hatziminaoglou (2011). Todos os espectros foram normalizados em 7 μm . Fonte: a autora

A Figura 2.3 apresenta os espectros de galáxias utilizadas neste trabalho, a classe mínimos locais em uma função que converge mais rapidamente do que algoritmos conhecidos como genéticos.

dominadas por Starbursts (MIR_SB), composta por 257 objetos. Porém 109 espectros foram descartados por causa dos altos valores de ruído e/ou não possuem dados no intervalo de comprimento de onda entre 5 - 13 μm , de forma que a amostra total é composta por 148 galáxias. As seguintes propriedades das galáxias foram retiradas do trabalho de Hernán-Caballero e Hatziminaoglou (2011), são elas: contribuições das fontes de ionização (AGN, HII e PDR), assinatura de silicato e largura equivalente do PAHs em 6.2 μm .

2.3 Biblioteca espectral das moléculas de PAHs

O NASA Ames PAH IR Spectroscopic Database (PAHdb) reúne espectros IR de moléculas de PAH experimentais e modeladas computacionalmente, compreendendo o intervalo espectral de 2,5–4000 μm . A Figura 2.4 mostra exemplos de espectros disponíveis neste banco de dados. Também contém dados de PAHs puros, compostos apenas de carbono e hidrogênio, bem como os mistos com adição de nitrogênio, oxigênio ou magnésio, conforme mostra a Figura 2.5 e informações sobre moléculas neutras e ionizadas, visto na Figura 2.6.

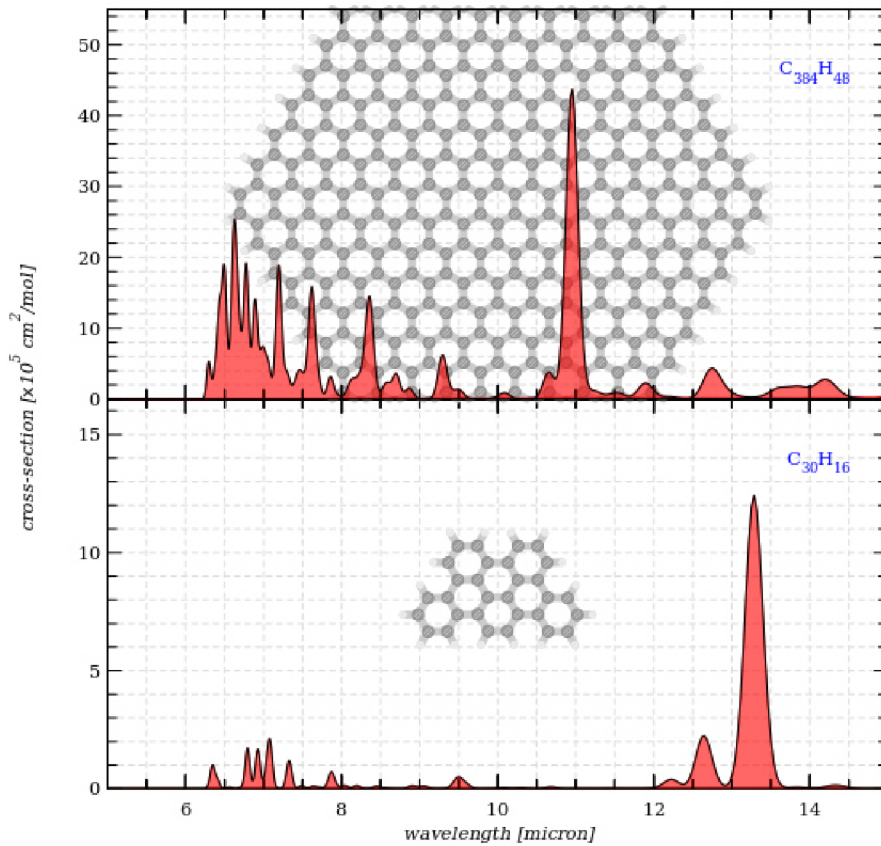


Figura 2.4: Espectro IR no intervalo entre 5 e 15 μm de espécies moleculares disponíveis no PAHdb. Fonte: <https://www.astrochemistry.org>

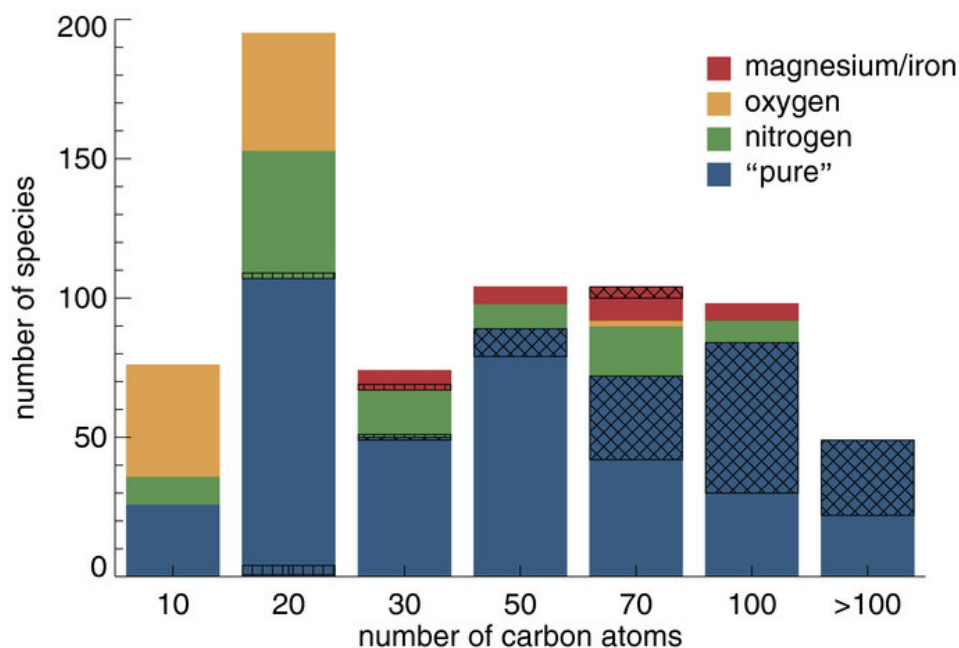


Figura 2.5: Detalhamento dos PAHs no banco de dados teóricos, versão 2.00. Na horizontal, se encontra a quantidade de átomos de carbono e na vertical, o número de espécies moleculares. As cores indicam a composição das moléculas: em azul, aparecem as espécies “puras”, contém apenas carbono e hidrogênio; em vermelho, possuem adição de magnésio/ferro; em amarelo, adição de oxigênio e, em verde, adição de nitrogênio. As áreas hachuradas indicam adições e remoções entre versões 1.10 e 2.00. Fonte: (Boersma et al., 2014)

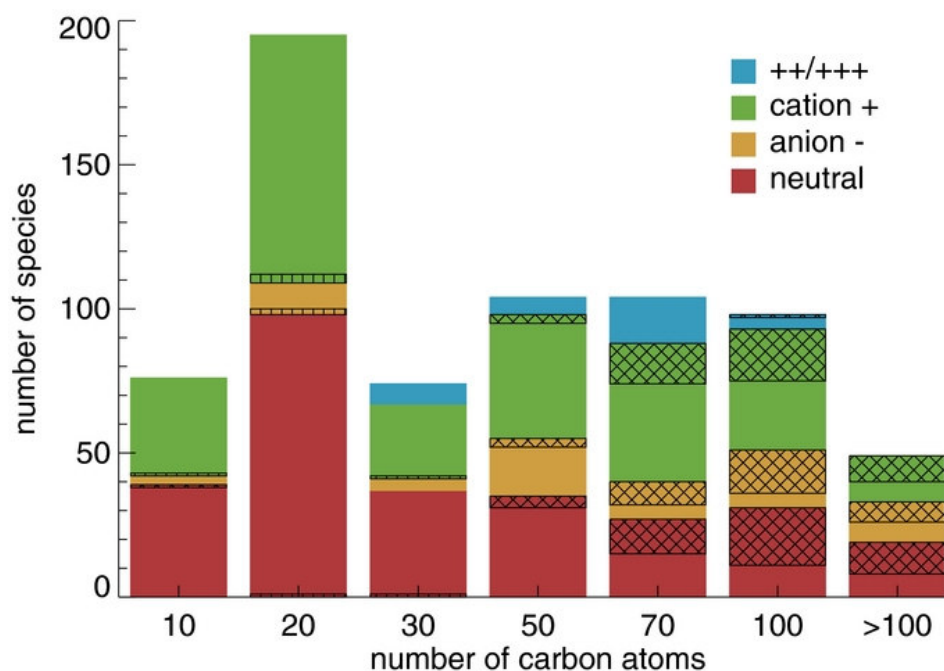


Figura 2.6: Detalhamento dos PAHs no banco de dados teóricos, versão 2.00. Na horizontal, se encontra a quantidade de átomos de carbono e na vertical, o número de espécies moleculares. As cores indicam a carga das moléculas: verde, indica cátions; em amarelo são os ânions e, em vermelho, neutros. As áreas hachuradas indicam adições e remoções entre versões 1.10 e 2.00. Fonte: (Boersma et al., 2014)

Além disso, o PAHdb fornece estruturas, transições vibracionais, conjuntos de até 5 temperaturas de excitação, como forma de comparação direta e quando disponível, referências da literatura, dentre outras propriedades. Esses dados podem ser baixados como XML ou tabelas ASCII, individuais ou grupos de moléculas. Também disponibiliza diferentes Interfaces de Programação de Aplicativos (APIs, do inglês “*Application Programming Interface*”), como AmesPAHdbIDLSuite, AmesPAHdbPythonSuite e uma versão simplificada desse conjunto de ferramentas, o pyPAHdb (Shannon e Boersma, 2018).

O PAHdb foi lançado em 2010 com 643 espectros, dentre teóricos e experimentais. Atualmente contém 4.233 espécies teóricas na versão 3.20 e 84 espécies experimentais na versão 3.00, além disso será de grande importância nas futuras observações do James Webb Space Telescope (JWST). Este trabalho, no entanto, fará uso da versão 2.00 que apresenta 700 espectros computacionais e 75 experimentais comparados (Boersma et al., 2014).

2.4 Métodos estatísticos multivariados

Existem vários métodos estatísticos, com finalidades bem diversas entre si. Técnicas que analisam, simultaneamente, vários dados sobre indivíduos ou objetos de pesquisa são chamadas de análise multivariada (Hair et al., 2005). E devido a capacidade de mensurar, interpretar e indicar relações entre variáveis, pode ser utilizada em diferentes áreas do conhecimento. Ainda segundo Hair et al. (2005), a análise multifatorial compreende métodos já estabelecidos na literatura e aqueles em ascensão.

Esta pesquisa fez uso dos seguintes métodos: (i) *análise de agrupamentos*, (ii) *análise de componentes principais (PCA)* e *análise de correspondência (CA)* que serão tratados a seguir.

2.4.1 Análise de Agrupamento

A análise de agrupamento é um grupo de técnicas multivariadas e são utilizadas nas mais diversas áreas, como ciências exatas e ciências sociais, possibilitando a redução de dados e a formulação de hipóteses. E tem por objetivo classificar indivíduos ou objetos em subgrupos menores e mutuamente excludentes, com base nas similaridades (Hair et al., 2005). O foco principal é maximizar a semelhança dentro dos grupo, enquanto se maximiza a diferença entre os grupos. De forma que os integrantes de um grupo compartilhem

características em comum, que também servem para diferenciar dos demais grupos (Galvão e Marin, 2009).

No método de agrupamento os clusters são definidos com base na medição de similaridades pelo algoritmo, de forma que realiza uma abordagem não supervisionada. As medidas de similaridades são determinadas pela distância entre dois pontos e expressam o quão semelhantes são os objetos. Quanto menor for a distância, mais semelhantes e permanecem no mesmo grupo. Da mesma forma, quanto maior a distância, mais diferentes e esses objetos estarão em grupos distintos. Existem diferentes métodos de agrupamento, de acordo com a abordagem utilizada, como por exemplo: hierárquicos, particionais, baseados em grade, baseado em densidade, redes neurais, etc. Neste trabalho optou-se por trabalhar com o agrupamento hierárquico.

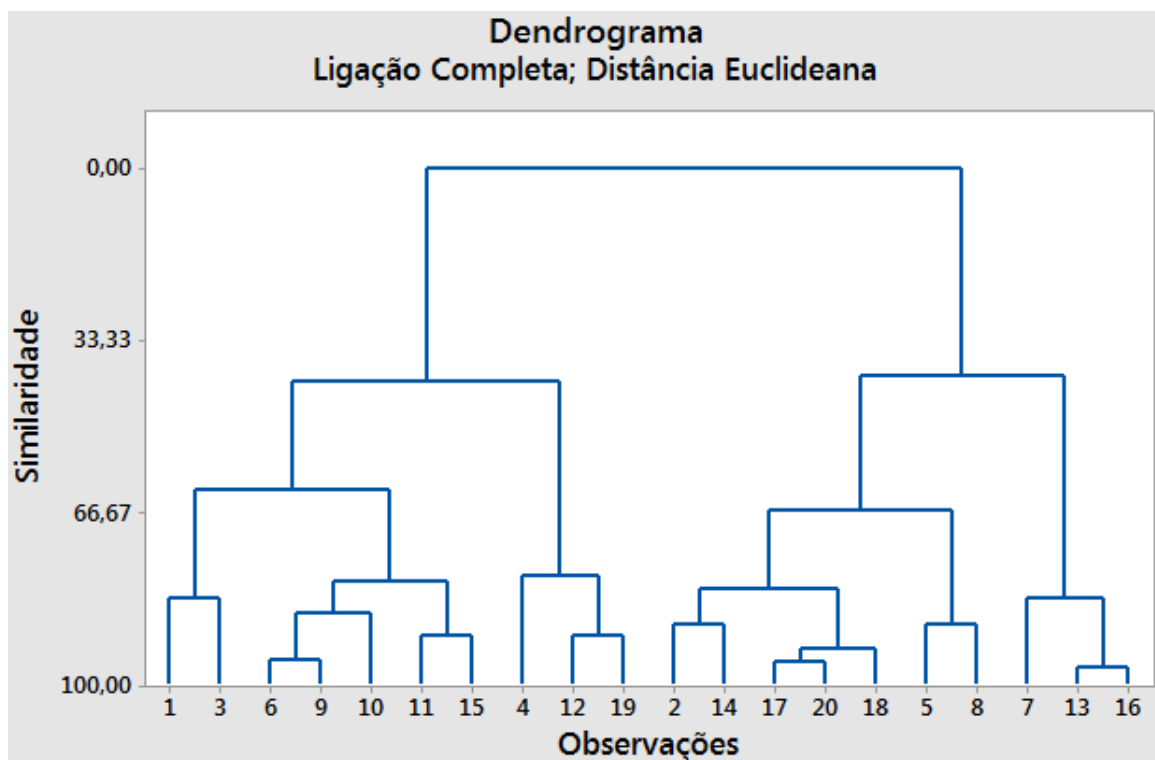


Figura 2.7: Uma estrutura dendrograma na qual o grupo inicial é subdividido em vários subgrupos. Na direção vertical, a similaridade é expressa de acordo com a distância euclidiana e na direção horizontal, o valor de observação. Fonte: <https://support.minitab.com/>

O algoritmo hierárquico compreende a divisão sucessiva dos objetos em agrupamentos, resultando em uma representação gráfica hierárquica ou dendrograma correspondente aos clusters (Everitt et al., 2001). Na Figura 2.7 é possível observar a sequência em que os dados foram agrupados, bem como a diferença de elementos entre os grupos. Nota-se que

as observações são representadas como galhos de uma árvore (dendro=árvore) e a medida de semelhança se dá na vertical, por meio da distância. Em outras palavras, quanto mais no topo do gráfico, menos semelhantes entre si os dados.

Dentre as diversas maneiras de se calcular a distância entre grupos e sub-grupos, a utilizada mais frequentemente é a métrica Euclidiana. Inferida pela distância entre dois pontos de coordenadas (x_1, y_1) e (x_2, y_2) , conforme a seguinte equação, que faz uso do teorema de Pitágoras:

$$d_{euc} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2.2)$$

Os métodos hierárquicos não necessitam que seja definido um número ideal de clusters, no entanto é possível determinar esse valor por meio de critérios de validação. Kassambara (2017a) apresenta diferentes critérios para determinar o número ideal de clusters, uma vez que este depende da medida de semelhanças e do algoritmo utilizado. Os critérios de validação podem ser de três tipos: externo, que utiliza informações fornecidas anteriormente; interno, sem interferência externa e relativo, que avalia dois ou mais grupos e estima qual a melhor estrutura.

Dentre os parâmetros sugeridos pelo autor, está o pacote NbClust, que verifica mais de 30 índices ao mesmo tempo para identificar e determinar a quantidade apropriada de grupos, utilizando apenas uma função. Foi desenvolvido por Charrad et al. (2014), devido a literatura apresentar diversos índices de validação relativo, mas poucos programas para realizar os cálculos. Também fornece métodos de agrupamento e medidas de dessemelhança ou utiliza os parâmetros escolhidos pelo pesquisador.

2.4.1.1 Clustering Hierárquico com o Método Ward

Desenvolvido por Ward (1963), a análise hierárquica de agrupamentos ou método da variância mínima de Ward utiliza como critério a soma dos quadrados entre dois clusters somados sobre os demais e tem por objetivo minimizar a perda de informações. A equação que infere o valor da variância, é dada por:

$$ESS_k = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (2.3)$$

onde k é o agrupamento, n é a totalidade de objetos do agrupamento k e x_i o i -ésimo objeto desse agrupamento. No método de Ward, considera-se que cada objeto retrata um

grupo, que são continuamente ligados em pares até o fim do processo formando grupos cada vez maiores e observando a mínima variância interna. A cada passo os agrupamentos são realizados sempre que o algoritmo identifica a mínima variância entre pares de observações. De forma que os clusters tendem a ser mais homogêneos.

2.4.2 Análise de Componentes Principais - PCA

A análise de componentes principais ou PCA (*“Principal Component Analysis”*), introduzida por Pearson (1901) se tornou uma das técnicas mais utilizadas em análises de dados. A PCA transforma os dados disponíveis em novos conjuntos, não correlacionados e organizados de forma que as maiores variações, presentes no conjunto original, estejam nos primeiros elementos. Esses novos conjuntos recebem o nome de *“principal components”* (PCs) e são descritos conforme a seguinte equação:

$$Y_j = \sum_{i=1}^j a_{ij}x_i \quad (2.4)$$

em que j descreve a componente principal e a_{ij} o autovetor. Os demais componentes surgem da repetição deste método, levando em conta que o primeiro componente reúne a máxima variância das propriedades originais, o percentual de variância diminui ao longo de todo o processo. A variância total da amostra é dada pela equação a seguir:

$$\sigma^2(X_j) = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \quad (2.5)$$

tendo n como a totalidade da amostra e $\bar{X}_j$ como a média das propriedades encontradas em X .

Além disso, o coeficiente de correlação de Pearson tem por objetivo de indicar como as duas variáveis associadas estão entre si. Dessa forma, caso a correlação seja menor que zero, ou seja, correlação negativa, significa que as variáveis são inversamente relacionadas. No entanto, correlação não deve ser confundida com relação de causa e efeito (causalidade), uma vez que dificilmente pode-se afirmar quem varia em função de quem. Simplesmente pode-se dizer que há semelhanças entre a distribuição das variáveis.

2.4.3 Análise de Correspondência

Conforme Hair et al. (2005), trata-se de um método de interdependência, ou seja, que analisa todas as variáveis ao mesmo tempo para compreender a organização do conjunto. Trata-se de um método composicional, onde a representação gráfica se dá com base nas relações entre variáveis, bem como as relações entre elementos e variáveis. De certa forma, é semelhante à PCA, pois retém a variância total e a distribui em dimensões principais. Dessa forma reduz a dimensionalidade da amostra e permite a visualização de um grande volume de dados em gráficos bidimensionais (Kassambara, 2017b).

A CA dispõe os dados originais em uma matriz retangular $N(I, J)$, com as amostras representadas por linhas i e suas propriedades por colunas j . Onde o total das linhas é dado por $n_{i+} = \sum_j n_{ij}$, o total das colunas é dado por $n_{+j} = \sum_i n_{ij}$ e o total de linhas e colunas é dado por $n = \sum_j \sum_i n_{ij}$. Ao ser dividida pelo total de linhas e colunas n , a matriz original será convertida na matriz de correspondência $\mathbf{P} = \frac{1}{n}N(I, J)$.

Para determinar a medida de similaridade entre os dados, utiliza-se a métrica da Distância χ^2 ou Teste χ^2 de Pearson, pois graficamente, objetos próximos são mais semelhantes do que objetos distante. O χ^2 entre duas linhas pode ser inferido da seguinte maneira:

$$\chi^2 = \sqrt{\sum_{ij} \frac{n}{n_{+j}} \left(\frac{n_{ij}}{n_{i+}} - \frac{n_{i'j}}{n_{i'+}} \right)} \quad (2.6)$$

em que $n_{i'+}$ é qualquer elemento de linhas. Da mesma forma, pode se determinar a distância entre perfis de colunas. A variância entre os dados é descrita pela inércia, que é soma ponderada do quadrado da distância de um conjunto de pontos ao seu centro ou centroides. A inércia é determinada pela seguinte expressão:

$$\text{inércia} = \frac{\chi^2}{n} \quad (2.7)$$

2.5 Características gerais da amostra

Constituída por 148 galáxias, a amostra foi submetida à análise multivariada através dos métodos clustering, PCA e CA. Também foram escolhidas dez propriedades físicas que caracterizam as galáxias dominadas por Starbursts, são elas: (1) ionização por AGN

(rAGN); (2) ionização por HII (rHII); (3) ionização por PDR (rPDR); (4) assinatura de silicato (Ssil); (5) largura equivalente da banda $6.2 \mu\text{m}$ (EW62); (6) conjunto molecular muito pequeno (X0_30); (7) conjunto molecular pequeno (X31_150); (8) conjunto molecular médio (X151_225); (9) conjunto molecular grande (X226_330) e (10) conjunto molecular muito grande (maior_300).

Na Tabela 2.1, os valores exibidos em rAGN (1), rHII (2) e rPDR (3) são extraídos do trabalho realizado por (Hernán-Caballero e Hatziminaoglou, 2011) e referem-se às fontes predominantes da emissão em MIR de galáxias ativas e de formação estelar. Essas propriedades foram verificadas por meio da decomposição espectral, onde cada componente contribui com r para o fluxo de emissão total da galáxia. Cada contribuição foi obtida por meio da sobreposição dos espectros de galáxias com diferentes fontes de emissão: para AGN, o QSO 3C273; para PDR, a nebulosa de reflexão NGC 7023 e para HII, a nebulosa M17.

As categorias apresentadas em Ssil (4) e EW62 (5), na tabela 2.1, se referem ao segundo procedimento de classificação utilizado por Hernán-Caballero e Hatziminaoglou (2011). A coluna (4) apresenta a assinatura de silicato em $10\mu\text{m}$ (Ssil, “*silicate signature*”), um tipo de poeira abundante no ISM e também no torus do modelo unificado de AGN. Esse intervalo espectral caracteriza a intensidade de absorção ou emissão desse material e infere aspectos da distribuição de poeira, ao longo das regiões formadoras de estrelas ou AGNs.

A coluna (5) mostra valores da largura equivalente (EW, “*equivalent width*”) das bandas de PAH em $6.2 \mu\text{m}$. Essa propriedade é determinada por uma área retangular que possui o mesmo fluxo da linha espectral, sendo expressa pela seguinte equação:

$$W = \int_{\lambda_1}^{\lambda_2} \frac{[F(\lambda) - F_{cont}]}{F_{cont}} d\lambda \quad (2.8)$$

onde W representa a largura equivalente EW; $F(\lambda)$ representa o fluxo, ou a intensidade da linha espectral em relação ao continuum; F_{cont} é o fluxo do contínuo no centro da linha, resultante do espalhamento das linhas de emissão ou absorção e por fim, λ_2 e λ_1 , representam o intervalo dessa linha espectral. Essa propriedade indica a contribuição do gás em comparação com a fonte emissora e, uma vez que starburts possuem valores altos de EW, mas valores menores de Ssil. O limite de $EW62 \leq 0.2\mu\text{m}$ foi empregado para definir AGNs e $EW62 > 0.2\mu\text{m}$ para Starbursts (Magdis et al., 2013).

As colunas (6), (7), (8), (9) e (10) da Tabela 2.1 apresentam valores referentes aos

tamanhos de moléculas de PAHs neutras, extraídas do trabalho realizado por Espinoza (2020). Através da comparação pareada entre espectros de referência (neste caso, espectros moleculares de PAHs, disponíveis no PAHdb) e espectros de interesse (a amostra de galáxias dominadas por Starburts), o referido autor analisou a composição química dessas galáxias para identificar a presença e a contribuição de grupos moleculares diferentes.

Através da Transformada Rápida de Fourier (FFT, “*Fast Fourier Transform*”), os espectros de referência foram derivados não só na composição, mas também na intensidade dos componentes espectrais, assim os espectros de PAHs foram classificadas nos seguintes grupos: o muito pequenos, que são moléculas com até 30 átomos de carbono (X0_30); o grupo pequeno compreende PAHs entre 31 a 150 carbonos (X31_150); no grupo médio estão os que possuem entre 151 a 225 carbonos (X151_225); o grupo grande abrange moléculas entre 226 a 300 carbonos (X226_300) e, por fim, o grupo dos muito grandes que contém PAHs com mais de 300 carbonos (maior_300).

Nesse contexto, as propriedades físicas citadas permitem maior compreensão das emissões de MIR e também da evolução química de galáxias. O próximo capítulo faz uma análise estatística dessas propriedades e tenta inferir a relação entre tamanhos moleculares e indicadores de ionização.

Tabela 2.1 - Amostra de galáxias dominadas por Starbursts e suas propriedades.

Galáxias	(1) r_AGN	(2) r_HII	(3) r_PDR	(4) Ssil	(5) EW6.2	(6) X0_30	(7) X31_150	(8) X151_225	(9) X226_330	(10) maior_300
3C293	0,713	0,32	0,287	0,252	0,004	0,251	0,107	0,822	0,547	0,495
3C31	0,764	0,129	0,395	0,105	0,005	1	0,97	0	0	0
AGN15	0,407	0,789	0,24	0,296	0,004	0,267	0,107	0,823	0,543	0,488
Arp220	0,296	0,779	0,382	0,674	0,01	0,915	1	0	0	0
CGCG381-051	0,279	0,704	0,469	-0,073	0,321	0,414	0,484	1	0	0
E12-G21	0,646	0,206	0,468	0,015	0,01	0,444	0,483	0,986	0	0
EIRS-13	0,316	0,353	0,735	0,038	1	0,27	0,108	0,818	0,554	0,481
EIRS-14	0,277	0,456	0,692	0,199	0,021	0,542	0,491	0,926	0	0
EIRS-16	0,241	0,469	0,723	-0,067	0,022	0,546	0,491	0,924	0	0
EIRS-2	0,116	0,557	0,793	0,076	0,303	0,271	0,108	0,812	0,569	0,473
EIRS-21	0,767	0,253	0,281	0,654	0	0,549	0,468	0,955	0	0
EIRS-37	0,759	0,211	0,328	0,056	0,003	0,599	0,504	0,878	0	0
EIRS-41	0,544	0,493	0,337	-0,052	0,006	0,262	0,101	0,809	0,608	0,46
EIRS-55	0,717	0,335	0,268	0,323	0,002	0,525	0,567	0,824	0	0
EIRS-9	0,252	0,493	0,688	0,23	0,267	0,492	0,535	0,887	0	0
GN26	0,266	0,256	0,88	0,151	0,016	0,268	0,107	0,797	0,623	0,445
IRAS02021-2103	0,198	0,976	0,322	0,312	0,015	0,621	0,205	0,375	0,645	0,55
IRAS02480-3745	0,314	0,424	0,673	0,296	0,295	0,504	0,173	0,314	0,536	0,858
IRAS03209-0806	0,31	0,58	0,541	0,188	0,018	0,495	0,189	0,338	0,589	0,757
IRAS04074-2801	0,46	0,473	0,455	0,68	0,003	0,401	0,187	0,333	0,592	0,823
IRAS05020-2941	0,429	0,39	0,565	0,547	0,006	0,387	0,163	0,287	0,518	1
IRAS08474+1813	0,343	0,469	0,599	0,431	0,012	0,39	0,16	0,595	0,648	0,501
IRAS08591+5248	0,315	0,362	0,728	0,162	0,024	0,508	0,218	0,376	0,731	0,518
IRAS10594+3818	0,306	0,395	0,709	0,178	0,024	0,493	0,222	0,378	0,726	0,524
IRAS12447+3721	0,357	0,554	0,506	0,261	0,011	0,447	0,208	0,36	0,682	0,64
IRAS13106-0922	0,572	0,092	0,659	0,505	0,001	0,427	0,194	0,345	0,593	0,777
IRAS14121-0126	0,417	0,324	0,639	0,297	0,015	0,485	0,216	0,392	0,711	0,534

Continua na próxima página...

Tabela 2.1 - Continuação

Galáxias	(1) r_AGN	(2) r_HII	(3) r_PDR	(4) Ssil	(5) EW6.2	(6) X0_30	(7) X31_150	(8) X151_225	(9) X226_330	(10) maior_300
IRAS14197+0813	0,214	0,763	0,494	0,166	0,016	0,473	0,222	0,393	0,714	0,528
IRAS14202+2615	0,624	0,219	0,484	0,138	0,01	0,463	0,225	0,389	0,71	0,536
IRAS14485-2434	0,545	0,399	0,418	0,266	0,007	0,451	0,211	0,369	0,642	0,656
IRAS15043+5754	0,26	0,406	0,755	0,247	0,02	0,433	0,199	0,349	0,641	0,715
IRAS21477+0502	0,262	0,85	0,358	0,175	0,018	0,447	0,209	0,369	0,637	0,668
IRAS22088-1831	0,319	0,754	0,375	0,605	0,004	0,471	0,224	0,399	0,666	0,562
IRAS_00091-0738	0,292	0,812	0,357	0,824	0	0,254	0,112	0,797	0,625	0,445
IRAS_00456-2904	0,226	0,452	0,755	0,202	0,024	0,501	0,488	0,95	0	0
IRAS_00482-2721	0,374	0,622	0,427	0,453	0,014	0,281	0,106	0,762	0,714	0,399
IRAS_01166-0844	0,353	0,783	0,31	0,75	-0,001	0,256	0,101	0,703	0,886	0,338
IRAS_01199-2307	0,338	0,668	0,428	0,559	0,005	0,557	0,505	0,897	0	0
IRAS_01298-0744	0,486	0,432	0,459	1	-0,001	0,274	0,109	0,706	0,851	0,341
IRAS_01355-1814	0,249	0,955	0,281	0,547	0,009	0,315	0,121	0,757	0,689	0,379
IRAS_01494-1845	0,273	0,502	0,653	0,355	0,023	0,631	0,529	0,825	0	0
IRAS_01569-2939	0,379	0,662	0,385	0,579	0,002	0,296	0,117	0,728	0,6	0,517
IRAS_02411+0353	0,295	0,599	0,542	0,165	0,02	0,387	0,383	0,493	0,898	0
IRAS_03250+1606	0,274	0,515	0,642	0,3	0,265	0,308	0,13	0,763	0,65	0,397
IRAS_03521+0028	0,308	0,625	0,505	0,284	0,013	0,405	0,355	0,542	0,873	0
IRAS_08201+2801	0,41	0,501	0,489	0,53	0,005	0,422	0,296	0,524	1	0
IRAS_09039+0503	0,399	0,473	0,529	0,475	0,014	0,465	0,289	0,517	0,996	0
IRAS_09116+0334	0,238	0,436	0,755	0,209	0,023	0,469	0,31	0,521	0,944	0
IRAS_09463+8141	0,374	0,491	0,542	0,494	0,02	0,474	0,32	0,533	0,906	0
IRAS_09539+0857	0,355	0,639	0,434	0,745	0,007	0,516	0,327	0,538	0,858	0
IRAS_10035+2740	0,177	0,77	0,532	0,442	0,577	0,281	0,114	0,693	0,588	0,587
IRAS_10190+1322	0,242	0,382	0,797	0,213	0,253	0,311	0,126	0,789	0,628	0,39
IRAS_10485-1447	0,325	0,862	0,271	0,689	0,007	0,473	0,331	0,547	0,866	0
IRAS_10494+4424	0,323	0,33	0,746	0,324	0,023	0,33	0,124	0,768	0,619	0,416
IRAS_11028+3130	0,379	0,581	0,456	0,488	0,012	0,321	0,114	0,672	0,562	0,611

Continua na próxima página...

Tabela 2.1 - Continuação

Galáxias	(1) r_AGN	(2) r_HII	(3) r_PDR	(4) Ssil	(5) EW6.2	(6) X0_30	(7) X31_150	(8) X151_225	(9) X226_330	(10) maior_300
IRAS_11095-0238	0,425	0,691	0,304	0,873	0	0,425	0,326	0,547	0,91	0
IRAS_11130-2659	0,414	0,597	0,398	0,673	0,004	0,331	0,114	0,632	0,546	0,669
IRAS_11180+1623	0,386	0,432	0,581	0,436	0,005	0,4	0,326	0,565	0,902	0
IRAS_11387+4116	0,236	0,519	0,683	0,207	0,022	0,378	0,12	0,639	0,57	0,598
IRAS_11506+1331	0,466	0,517	0,409	0,532	0,011	0,46	0,123	0,568	0,529	0,666
IRAS_11582+3020	0,454	0,478	0,458	0,587	0,001	0,389	0,311	0,569	0,933	0
IRAS_12032+1707	0,405	0,643	0,371	0,64	0	0,429	0,308	0,57	0,912	0
IRAS_12112+0305	0,325	0,52	0,575	0,362	0,018	0,43	0,31	0,55	0,933	0
IRAS_12359-0725	0,365	0,847	0,24	0,491	0,01	0,392	0,307	0,567	0,942	0
IRAS_13335-2612	0,26	0,464	0,703	0,249	0,26	0,514	0,143	0,5	0,512	0,697
IRAS_13469+5833	0,382	0,535	0,494	0,39	0,014	0,418	0,318	0,584	0,88	0
IRAS_13509+0442	0,298	0,469	0,653	0,31	0,03	0,432	0,311	0,574	0,898	0
IRAS_13539+2920	0,289	0,378	0,745	0,29	0,024	0,485	0,159	0,485	0,524	0,693
IRAS_14060+2919	0,196	0,543	0,712	0,179	0,317	0,48	0,285	0,548	0,949	0
IRAS_14252-1550	0,327	0,538	0,556	0,256	0,019	0,528	0,291	0,54	0,916	0
IRAS_14348-1447	0,463	0,425	0,494	0,465	0,01	0,536	0,293	0,541	0,906	0
IRAS_15206+3342	0,374	0,717	0,341	0,126	0,014	0,444	0,166	0,489	0,549	0,681
IRAS_15225+2350	0,437	0,665	0,312	0,506	0,004	0,418	0,167	0,52	0,567	0,64
IRAS_16090-0139	0,57	0,37	0,415	0,54	0,002	0,406	0,161	0,534	0,586	0,623
IRAS_16300+1558	0,456	0,559	0,384	0,617	0,005	0,565	0,288	0,54	0,897	0
IRAS_16333+4630	0,373	0,444	0,585	0,296	0,017	0,611	0,292	0,51	0,898	0
IRAS_16468+5200	0,413	0,649	0,355	0,53	0,003	0,375	0,159	0,551	0,592	0,619
IRAS_16474+3430	0,29	0,465	0,666	0,282	0,248	0,384	0,161	0,582	0,637	0,529
IRAS_16487+5447	0,299	0,634	0,505	0,366	0,016	0,742	0,283	0,497	0,843	0
IRAS_17028+5817	0,254	0,438	0,732	0,308	0,253	0,653	0,176	0,342	0,557	0,706
IRAS_17068+4027	0,424	0,562	0,419	0,394	0,005	0,662	0,143	0,299	0,483	0,88
IRAS_20414-1651	0,277	0,564	0,595	0,407	0,022	0,685	0,146	0,299	0,518	0,827
IRAS_21208-0519	0,246	0,523	0,668	0,243	0,312	0,796	0,207	0,38	0,67	0,408

Continua na próxima página...

Tabela 2.1 - Continuação

Galáxias	(1) r-AGN	(2) r-HII	(3) r-PDR	(4) Ssil	(5) EW6.2	(6) X0.30	(7) X31.150	(8) X151.225	(9) X226.330	(10) maior.300
IRAS_21329-2346	0,399	0,433	0,563	0,524	0,008	0,761	0,212	0,391	0,699	0,381
IRAS_22206-2715	0,311	0,536	0,578	0,345	0,274	0,399	0,167	0,615	0,663	0,445
IRAS_22491-1808	0,213	0,755	0,502	0,297	0,018	0,408	0,172	0,648	0,698	0,356
IRAS_23129+2548	0,424	0,485	0,488	0,619	0,002	0,409	0,171	0,638	0,701	0,368
IRAS_23234+0946	0,321	0,71	0,412	0,433	0,017	0,609	0,188	0,337	0,586	0,689
M-2-33-34	0,416	0,662	0,341	0,082	0,005	0,463	0,184	0,405	0,724	0,579
M-2-40-4	0,837	0,211	0,234	0,033	0,002	0,462	0,22	0,396	0,735	0,516
M-5-13-17	0,553	0,478	0,338	0,05	0,005	0,433	0,168	0,616	0,682	0,403
MIPS15928	0,318	0,969	0,187	-0,028	0,008	0,419	0,165	0,577	0,647	0,498
MIPS8342	0,378	0,862	0,21	-0,015	0,008	0,499	0,202	0,417	0,792	0,444
MIPS8521	0,78	0,259	0,26	0,497	-0,001	0,419	0,165	0,577	0,647	0,498
Mrk1066	0,327	0,699	0,415	0,139	0,009	0,682	0,2	0,393	0,653	0,491
Mrk273	0,485	0,599	0,314	0,476	0,006	0,664	0,17	0,345	0,579	0,684
Mrk334	0,443	0,485	0,465	0,044	0,012	0,434	0,171	0,573	0,665	0,465
Mrk471	0,363	0,43	0,609	0,029	0,012	0,447	0,188	0,542	0,653	0,475
Mrk52	0,097	0,966	0,454	0,019	0,253	0,668	0,146	0,317	0,532	0,802
Mrk609	0,308	0,436	0,67	0,034	0,018	0,634	0,128	0,285	0,5	0,925
Mrk622	0,214	1	0,284	0,015	0,014	0,645	0,138	0,294	0,548	0,845
Mrk883	0,361	0,723	0,352	0,06	0,015	0,611	0,14	0,306	0,61	0,791
Mrk938	0,323	0,37	0,71	0,297	0,018	0,595	0,137	0,297	0,506	0,915
NGC1056	0,255	0,196	0,944	0,041	0,029	0,461	0,176	0,579	0,69	0,409
NGC1097	0,368	0,188	0,816	0,021	0,017	0,664	0,157	0,338	0,563	0,731
NGC1125	0,387	0,597	0,432	0,262	0,014	0,615	0,157	0,346	0,553	0,762
NGC1222	0,095	0,779	0,622	0,085	0,281	0,653	0,182	0,397	0,609	0,579
NGC1365	0,494	0,662	0,247	-0,007	0,004	0,634	0,2	0,406	0,63	0,527
NGC1566	0,564	0,166	0,602	0,043	0,01	0,586	0,2	0,395	0,619	0,581
NGC2146	0,102	0,415	0,936	0,214	0,303	0,488	0,171	0,56	0,641	0,469
NGC2273	0,482	0,586	0,328	0,073	0,005	0,569	0,181	0,413	0,61	0,613

Continua na próxima página...

Tabela 2.1 - Continuação

Galáxias	(1) r_AGN	(2) r_HII	(3) r_PDR	(4) Ssil	(5) EW6.2	(6) X0_30	(7) X31_150	(8) X151_225	(9) X226_330	(10) maior_300
NGC2623	0,215	0,605	0,631	0,385	0,019	0,548	0,179	0,419	0,6	0,631
NGC2639	0,689	0,09	0,521	0,097	0,01	0,468	0,164	0,527	0,608	0,566
NGC2992	0,576	0,491	0,3	0,074	0,007	0,515	0,157	0,499	0,587	0,604
NGC3031	0,931	0	0,308	0,011	0,001	0,641	0,163	0,446	0,567	0,597
NGC3227	0,644	0,462	0,245	0,074	0,005	0,563	0,186	0,449	0,623	0,55
NGC3256	0,121	0,779	0,591	0,089	0,024	0,54	0,173	0,437	0,591	0,632
NGC3511	0,171	0,25	1	0,065	0,323	0,501	0,212	0,457	0,638	0,519
NGC3556	0,162	0,414	0,866	0,112	0,022	0,509	0,164	0,416	0,569	0,713
NGC3628	0,171	0,322	0,936	0,412	0,251	0,446	0,162	0,423	0,57	0,747
NGC3786	0,587	0,351	0,411	0,042	0,007	0,507	0,151	0,409	0,54	0,774
NGC3982	0,283	0,312	0,81	0,04	0,018	0,444	0,224	0,48	0,696	0,45
NGC4088	0,154	0,533	0,769	0,113	0,024	0,533	0,15	0,413	0,506	0,786
NGC4388	0,571	0,519	0,28	0,202	0,004	0,393	0,181	0,527	0,715	0,484
NGC4501	0,798	0,031	0,441	0,116	0,001	0,412	0,175	0,572	0,718	0,424
NGC4579	0,967	0	0,265	-0,048	0,005	0,499	0,147	0,404	0,511	0,821
NGC4594	0,842	0	0,415	-0,022	0,001	0,415	0,171	0,578	0,737	0,405
NGC4676	0,148	0,386	0,906	0,18	0,266	0,503	0,155	0,434	0,543	0,736
NGC4945	0,192	0,386	0,853	0,853	0,016	0,461	0,155	0,459	0,552	0,723
NGC5033	0,403	0,161	0,802	0,069	0,017	0,407	0,153	0,475	0,558	0,736
NGC5135	0,315	0,415	0,68	0,131	0,016	0,372	0,152	0,502	0,558	0,725
NGC5194	0,337	0,208	0,839	0,023	0,018	0,409	0,168	0,579	0,74	0,409
NGC5929	0,24	0,636	0,575	0,097	0	0,403	0,162	0,583	0,739	0,42
NGC5953	0,023	0,684	0,793	0,073	0,021	0,4	0,163	0,566	0,71	0,467
NGC6810	0,305	0,459	0,653	0,07	0,019	0,304	0,136	0,505	0,501	0,845
NGC6890	0,544	0,414	0,407	0,026	0,007	0,423	0,174	0,601	0,765	0,339
NGC7130	0,318	0,747	0,381	0,128	0,012	0,414	0,179	0,602	0,773	0,327
NGC7252	0,105	0,375	0,967	0,035	0,264	0,311	0,14	0,557	0,576	0,702
NGC7469	0,474	0,496	0,417	0,051	0,012	0,315	0,153	0,652	0,655	0,485

Continua na próxima página...

Tabela 2.1 - Continuação

Galáxias	(1) r_AGN	(2) r_HII	(3) r_PDR	(4) Ssil	(5) EW6.2	(6) X0_30	(7) X31_150	(8) X151_225	(9) X226_330	(10) maior_300
NGC7496	0,281	0,765	0,411	0,054	0,016	0,304	0,143	0,679	0,605	0,521
NGC7582	0,485	0,301	0,578	0,234	0,011	0,33	0,141	0,721	0,584	0,475
NGC7603	1	0	0,225	-0,046	0,002	0,316	0,14	0,728	0,584	0,476
NGC7714	0,122	0,866	0,512	-0,014	0,282	0,306	0,138	0,737	0,579	0,482
SST172458.3+591545	0,814	0,161	0,307	0,338	0,004	0,281	0,119	0,763	0,538	0,536
SWIRE4_J105840.62+582124.7	0,158	0,525	0,77	0,048	0,022	0,257	0,204	0,624	0,505	0,603
SWIRE4_J105943.83+572524.9	0,457	0,517	0,421	0,518	0,008	0,261	0,157	0,768	0,559	0,457
UGC12138	0,631	0,507	0,22	0,037	0,004	0,252	0,11	0,809	0,562	0,491
UGC5101	0,495	0,415	0,464	0,42	0,008	0,271	0,106	0,821	0,537	0,496

Resultados e Discussões

Neste capítulo são expostos os resultados das análises de agrupamento, PCA e CA realizadas na amostra constituída de 148 galáxias dominadas por Starburst. As propriedades selecionadas para análise foram citadas no capítulo anterior e incluem as cinco classes de PAHs com tamanhos moleculares selecionadas através do número de átomos de carbono, EW do PAH em $6.2\mu\text{m}$ e silicatos, bem como as três principais fontes ionizantes, são elas: AGNs, regiões HII e regiões PDR.

3.1 *Análise de Componentes Principais (PCA)*

Inicialmente, foi definida uma matriz de dados por meio da correlação de Pearson, iniciando a análise estatística da amostra. A matriz de correlação indica de forma visual quais propriedades relacionam-se entre si, bem como o grau de relação entre elas. Se as variáveis apresentam valores positivos, ou seja, correlação positiva significa que o aumento de uma é acompanhado pelo aumento da outra. Pelo contrário, quando os valores são negativos, ou seja, correlação negativa significa que a diminuição de uma é acompanhada por aumento na outra e vice-versa.

Quando as variáveis não são relacionadas, a correlação será zero ou nula. Os valores entre -1 e +1 indicam a relação linear entre as variáveis, isto é, o quanto cada variável impacta no crescimento ou diminuição de outra. É necessário ressaltar que embora correlação expresse relação estatística, não implica necessariamente em causalidade, porém como tratamos de dados fisicamente relacionados é possível inferir importantes considerações a respeito da amostra.

A Figura 3.1 apresenta o resultado da exploração, onde nota-se correlação negativa entre

a ionização por AGN e as outras duas fontes ionizantes, HII e PDR. Por outro lado, nota-se uma correlação positiva entre a ionização por AGN e conjuntos moleculares pequeno e médio. O conjunto molecular muito pequeno possui correlação positiva apenas com as espécies pequenas. As espécies grandes possuem correlação positiva com as propriedades rHII e Ssil, enquanto as espécies muito grandes possuem correlação positiva apenas com rPDR e EW62. Também é possível perceber que a propriedade Ssil apresenta correlação positiva com rHII e moléculas grandes, onde, será investigada mais profundamente em trabalhos futuros.

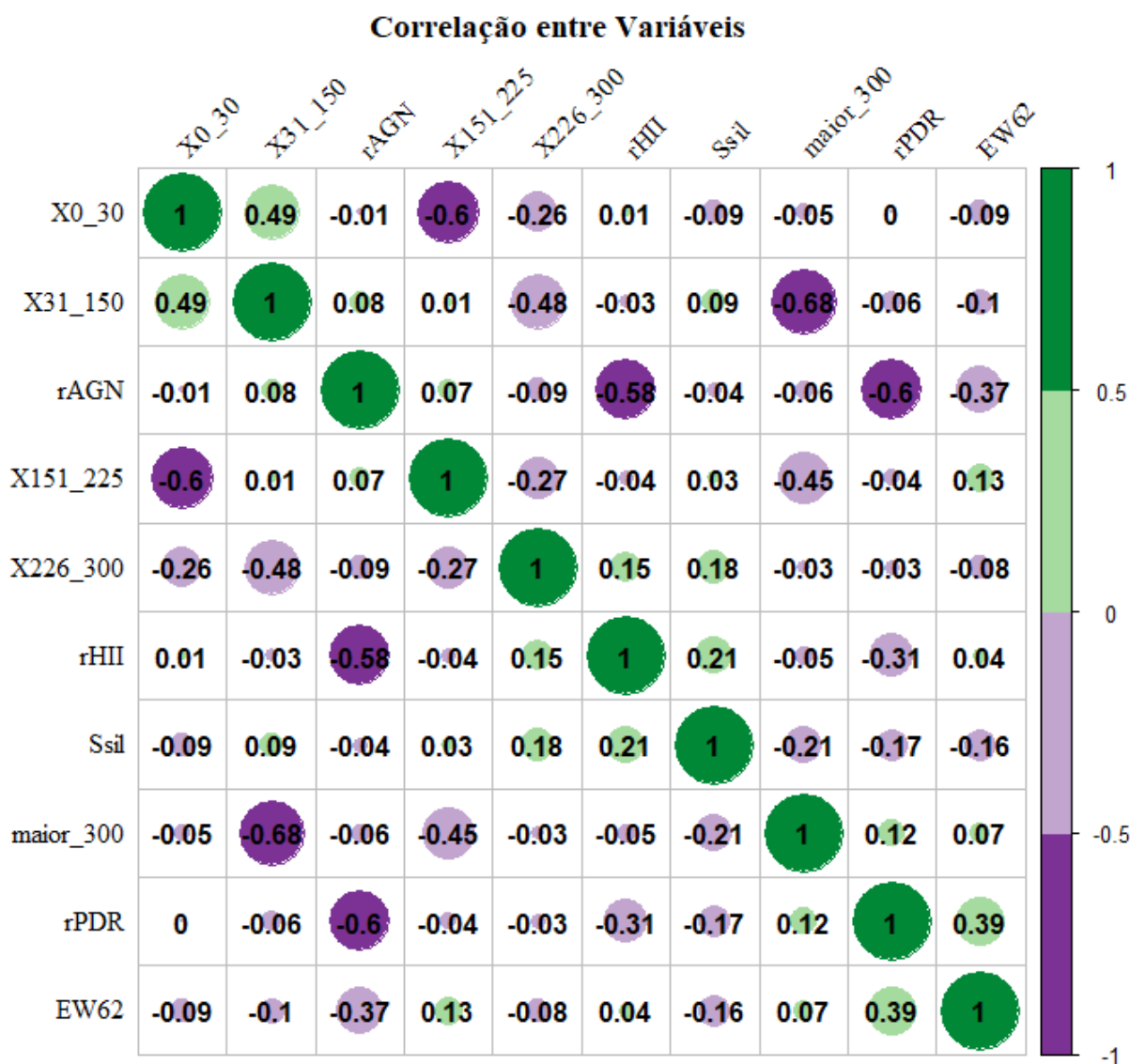


Figura 3.1: Matriz de Correlação de Pearson entre propriedades da amostra. Os rótulos indicam o grau de correlação, positivamente correlacionadas em verde e, em roxo, negativamente correlacionadas.

Posteriormente, foi realizada a Análise de Componentes Principais (PCA). Esse método

busca identificar padrões contidos na amostra original, bem como reduz dimensões das variáveis originais e com alta correlação, na menor quantidade possível, mas que explique a variância englobada nas originais. Estas novas variáveis são os PCs (*“principal components”*) ou dimensões, cujo autovalor médio indica o quanto cada dimensão contribui para as diferenças encontradas nos dados originais. Dessa forma, os primeiros componentes ou dimensões possuem autovalores grandes.

Na Figura 3.2, pode-se constatar que a contribuição dos 5 primeiros componentes principais somam o total de 84,6% da variância da amostra original. Peres-Neto et al. (2005) sugerem que o número de componentes principais selecionados para a análise, seja determinado pelo limite em que os demais PCs sejam pequenos, sem causar grandes diferenças na variância total. As três primeiras dimensões são responsáveis por 58,1% da variância total da amostra, assim é necessário entender a participação das 10 propriedades originais para a composição dos componentes principais.

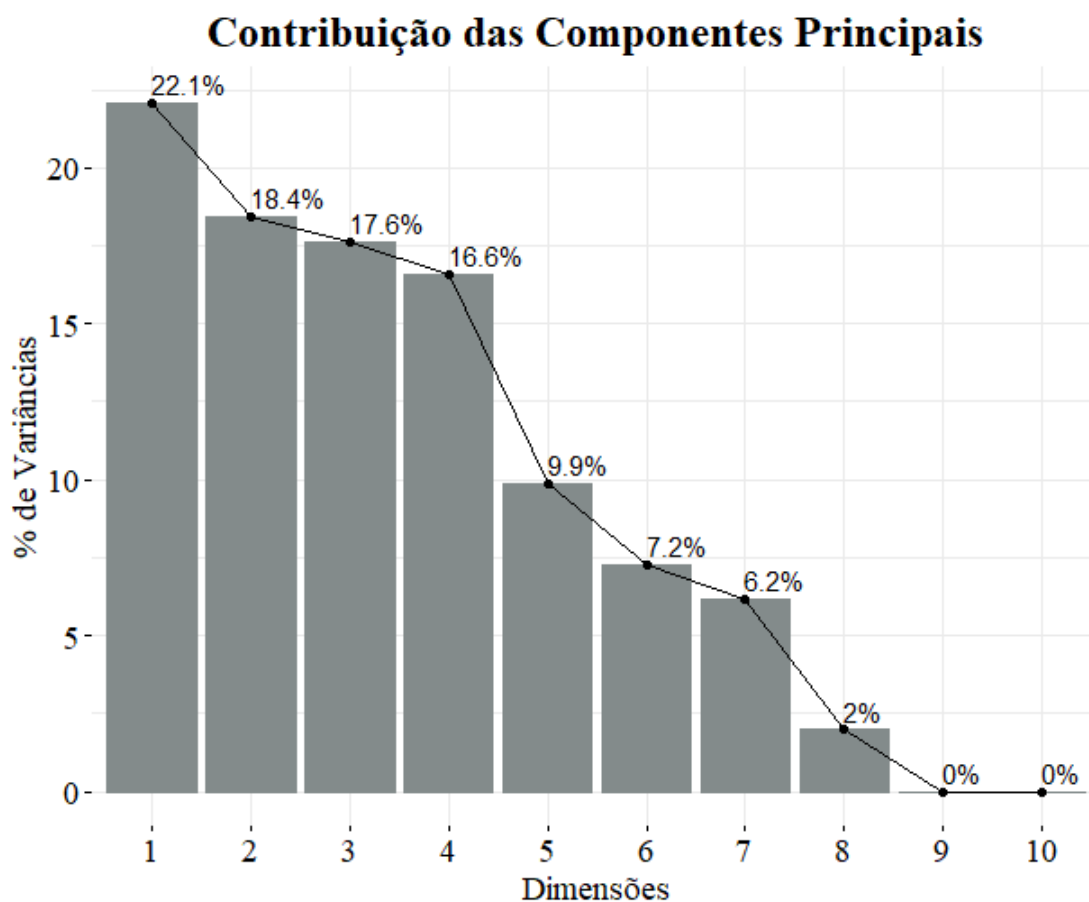


Figura 3.2: Porcentagem da contribuição de cada componente principal para a amostra. Cinco PCs concentram mais de 80% de toda a variabilidade encontrada na amostra.

A Figura 3.3 mostra a contribuição percentual de cada propriedade individualmente para cada uma das cinco dimensões. Pode-se verificar que a Dimensão 1 abrange os tamanhos moleculares, sendo que os conjuntos pequeno (X31_150) e muito grande (maior_300) respondem por mais de 46% dessa componente. Outra contribuição importante vem das variáveis relacionadas à ionização por AGN e PDR que correspondem a $\sim 16\%$ e 12% , respectivamente. A assinatura de silicato e HII são pouco representativas nesta dimensão, contando com menos de 2% da variância.

Contribuição das propriedades para as dimensões

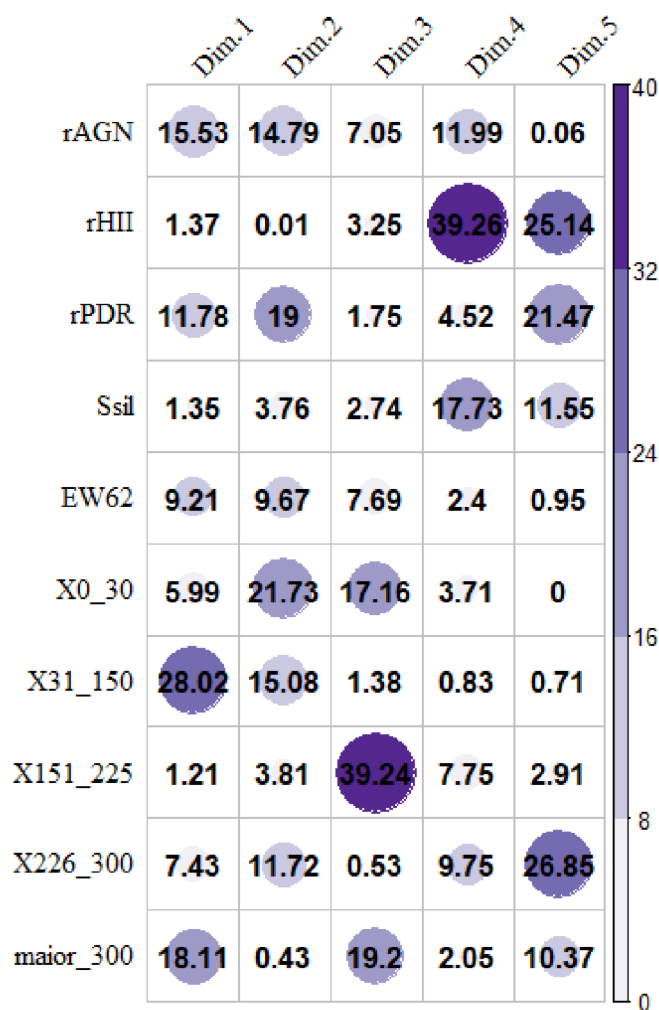


Figura 3.3: Porcentagem da contribuição das propriedades originais para a composição dos componentes principais, obtida com a PCA. O tamanho e a transparência do círculo roxo é proporcional ao valor desta contribuição.

Em relação a Dimensão 2, é possível verificar que as maiores contribuições provêm dos conjuntos moleculares muito pequeno e pequeno, além das fontes de ionização por AGN (rAGN) e ionização por PDR (rPDR), somando mais de 70% da variância na componente.

Na sequência tem-se a contribuição dos tamanhos moleculares grande e da variável EW62, somando pouco mais de 21%. As variáveis relacionadas à ionização por HII, a assinatura de silicato e moléculas médias e muito grandes possuem pouca contribuição. As quatro propriedades quando somadas, representam 8% para esta dimensão.

Na Dimensão 3, a contribuição do conjunto molecular médio X151.225 responde por 39,24%. A segunda maior contribuição, 19,2% atribui-se ao conjunto das moléculas muito grande, enquanto moléculas muito pequenas possuem 17,16%. A Dimensão 4 concentra uma maior contribuição da propriedade rHII com 39,26%, enquanto a assinatura de silicato e rAGN contribuem, respectivamente, com 17,73% e 11,99% da variância na componente. Por fim, a Dimensão 5 possui 26,85% de contribuição proveniente do conjunto molecular grande. Enquanto as propriedades HII e PDR correspondem a, respectivamente, 25,14% e 21,47% da variabilidade desta componente.

A partir dessa análise, é possível deduzir que na Dimensão 1 predominam as propriedades relacionadas aos tamanhos das moléculas de PAHs. Na Dimensão 2, destacam-se o tamanho molecular muito pequeno e a propriedade rPDR, bem como moléculas pequenas e a ionização por AGN. Já a Dimensão 3 é dominada pelo conjunto moleculares médio e muito pequeno. A Dimensão 4 se destaca pela contribuição da propriedade rHII e Ssil. Enquanto a Dimensão 5, possui a maior parcela de moléculas muito grandes.

Os resultados da PCA mostram que os dados originais tiveram a dimensionalidade reduzida, onde, mais de 70% da variância do conjunto de 10 propriedades das galáxias dominadas por starburst sendo explicadas pelas 4 primeiras componentes, porém, como as dimensões são ortogonais foi mantido a variância da amostra e adequação da modelagem (Jolliffe, 2002).

Além disso, a PCA mostra que embora a amostra de galáxias seja dominada por starbursts a Dimensão 1 possui importante contribuição da fonte ionizante não térmica do AGN ($\sim 15\%$) e também por moléculas pequenas ($\sim 28\%$). Por outro lado, as dimensões 4 e 5 tem domínio das fontes de ionização proveniente de estrelas (rHII e rPDR) e moléculas grandes e muito grandes. Esses resultados em conjunto com a análise de correlação de Pearson nos levam a concluir que galáxias com contribuições altas de rPDR e rHII possuem moléculas grandes, muito grandes e Ssil.

Esta análise sugere que o tamanho molecular dos PAHs, bem como as fontes de ionização por PDR e AGN produzem maior variância na amostra. Em outras palavras,

embora seja composta por galáxias dominadas por starbursts, a luminosidade proveniente de AGNs influencia na variabilidade total por ser dessemelhante da maior parte da amostra. Maiores discussão sobre os resultados de PCAs podem ser encontrado na dissertação de Miranda (2020).

O intervalo espectral do MIR permite a investigação dos diversos processos que ocorrem nos objetos astrofísicos, como a formação estelar e a energia emitida pela acreção de matéria no SMBH (Hernán-Caballero e Hatziminaoglou, 2011). Portanto, é necessário entender a influência exercida pela fonte de excitação nas galáxias, bem como sua relação com as propriedades observadas, como tamanho molecular e emissão de silicatos.

3.2 *Análise de Agrupamento*

A análise de agrupamento hierárquico ou variância mínima de Ward decompõe um conjunto de dados e ao final do processo, reúne grupos que seguem determinado padrão de semelhança. Baseados na distância Euclidiana, objetos de grupos diferentes possuem maiores diferenças entre si, enquanto dentro dos mesmos as desigualdades são mínimas. Desse modo verifica semelhanças ou diferenças entre elementos de acordo com sua posição no dendrograma.

Nesta seção a amostra será analisada com o propósito de identificar a influência das fontes ionizantes, bem como a contribuição do tamanho das moléculas de PAHs para a luminosidade total das galáxias (ver Miranda, 2020). E, embora os dados contenham uma população dominada por starburst, possui significativas parcelas de AGNs, logo é importante derivar de que modo as características das fontes de ionização das galáxias influenciam no comportamento físico-químico das moléculas de PAHs.

O método hierárquico permite dendrogramas com diversos grupos, sendo necessário decidir o número adequado. Kassambara (2017a) apresenta o pacote NbClust como alternativa para a escolha do número de grupos sem interferência do pesquisador. A Figura 3.4 apresenta o resultado obtido com o código para a métrica Euclidiana e o método Ward. Dentre todos os 30 índices, dois propõe 0 grupos e um índice propõe apenas 1 grupo. Três sugerem 2 grupos, dezessete sugerem 3 grupos, um sugere 5 grupos e por fim 6 sugerem 6 grupos.

Em concordância com o método sugerido por Kassambara (2017a), foram utilizados 3

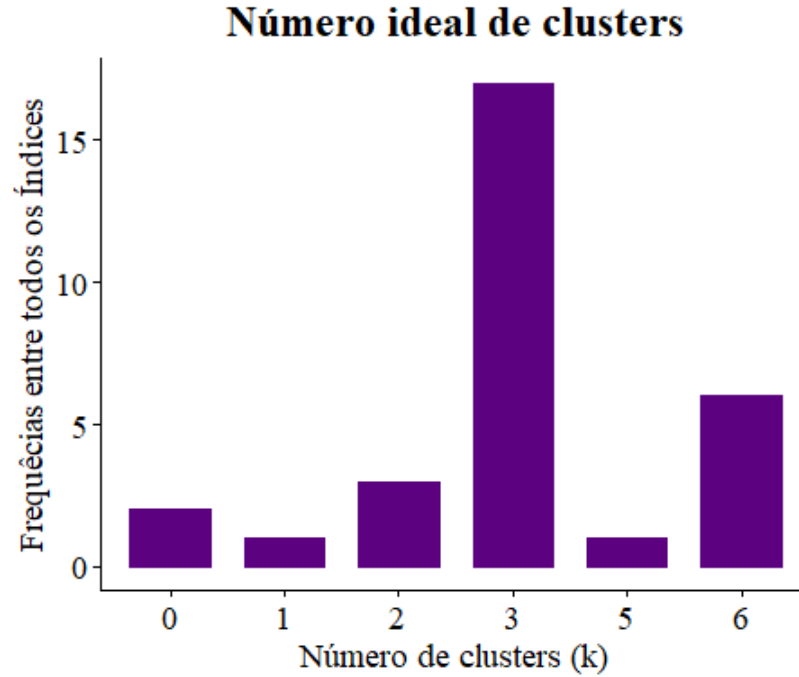


Figura 3.4: Histograma com o número ideal de clusters, o eixo horizontal mostra a quantidade de grupos sugeridas, indo de 0 a 6. No eixo vertical, quantos índices sugerem essa quantidade. De acordo com a regra da maioria, o melhor número de grupos é 3.

grupos para realizar o agrupamento hierárquico. Na Figura 3.5 encontra-se o resultado do procedimento, é possível verificar os três grupos definidos anteriormente, com acentuada homogeneidade interna e desigualdade entre eles. O Grupo 1, possui 114 galáxias e se encontra em rosa; o Grupo 2, se encontra na cor azul é formado por 14 galáxias e o Grupo 3, composto por 23 galáxias, está em verde.

A Tabela 3.1, relaciona os nomes de cada galáxia e a qual grupo pertencem, os quais serão discutidos na Seção 3.2.1.

Tabela 3.1 - Galáxias pertencentes aos grupos 1, 2 e 3.

Grupo 1	Grupo 2	Grupo 3
3C293	3C31	IRAS_02411+0353
AGN15	Arp220	IRAS_03521+0028
EIRS-13	CGCG381-051	IRAS_08201+2801
EIRS-2	E12-G21	IRAS_09039+0503
EIRS-41	EIRS-14	IRAS_09116+0334
GN26	EIRS-16	IRAS_09463+8141

IRAS_00091-0738	EIRS-21	IRAS_09539+0857
IRAS_00482-2721	EIRS-37	IRAS_10485-1447
IRAS_01166-0844	EIRS-55	IRAS_11095-0238
IRAS_01298-0744	EIRS-9	IRAS_11180+1623
IRAS_01355-1814	IRAS_00456-2904	IRAS_11582+3020
IRAS_01569-2939	IRAS_01199-2307	IRAS_12032+1707
IRAS_03250+1606	IRAS_01494-1845	IRAS_12112+0305
IRAS_10035+2740		IRAS_12359-0725
IRAS_10190+1322		IRAS_13469+5833
IRAS_10494+4424		IRAS_13509+0442
IRAS_11028+3130		IRAS_14060+2919
IRAS_11130-2659		IRAS_14252-1550
IRAS_11387+4116		IRAS_14348-1447
IRAS_11506+1331		IRAS_16300+1558
IRAS_13335-2612		IRAS_16333+4630
IRAS_13539+2920		IRAS_16487+5447
IRAS_15206+3342		
IRAS_15225+2350		
IRAS_16090-0139		
IRAS_16468+5200		
IRAS_16474+3430		
IRAS_17028+5817		
IRAS_17068+4027		
IRAS_20414-1651		
IRAS_21208-0519		
IRAS_21329-2346		
IRAS_22206-2715		
IRAS_22491-1808		
IRAS_23129+2548		
IRAS_23234+0946		

IRAS02021-2103
IRAS02480-3745
IRAS03209-0806
IRAS04074-2801
IRAS05020-2941
IRAS08474+1813
IRAS08591+5248
IRAS10594+3818
IRAS12447+3721
IRAS13106-0922
IRAS14121-0126
IRAS14197+0813
IRAS14202+2615
IRAS14485-2434
IRAS15043+5754
IRAS21477+0502
IRAS22088-1831
M-2-33-34
M-2-40-4
M-5-13-17
MIPS15928
MIPS8342
MIPS8521
Mrk1066
Mrk273
Mrk334
Mrk471
Mrk52
Mrk609
Mrk622

Mrk883
Mrk938
NGC1056
NGC1097
NGC1125
NGC1222
NGC1365
NGC1566
NGC2146
NGC2273
NGC2623
NGC2639
NGC2992
NGC3031
NGC3227
NGC3256
NGC3511
NGC3556
NGC3628
NGC3786
NGC3982
NGC4088
NGC4388
NGC4501
NGC4579
NGC4594
NGC4676
NGC4945
NGC5033
NGC5135

NGC5194
NGC5929
NGC5953
NGC6810
NGC6890
NGC7130
NGC7252
NGC7469
NGC7496
NGC7582
NGC7603
NGC7714
SST172458.3+591545
SWIRE4_J105840.62+582124.7
SWIRE4_J105943.83+572524.9
UGC12138
UGC5101

3.2.1 Detalhamento dos Grupos

Após utilizar o método Ward para especificar os grupos, passamos a compreender as propriedades das galáxias dos três grupos selecionados através desse método não-supervisionado e buscando características físicas e químicas da fonte de ionização das galáxias da amostra e suas respectivas classes de tamanhos de PAHs. Os dados que fornecem essas informações, como já foi citado são: parcelas de contribuição da fonte ionizante, r_{AGN} , r_{HII} e r_{PDR} ; largura equivalente da banda de PAH em $6.2 \mu m$; assinatura de silicato e os cinco tamanhos de moleculares de PAHs.

Na Tabela 3.2 encontram-se as medidas estatísticas dos Grupos 1, 2 e 3. É possível verificar que no Grupo 1 formado por 114 galáxias, a propriedade r_{PDR} é mais pronunciada (0,53) em relação aos demais agrupamentos. Ainda sobre o primeiro grupo, existem con-

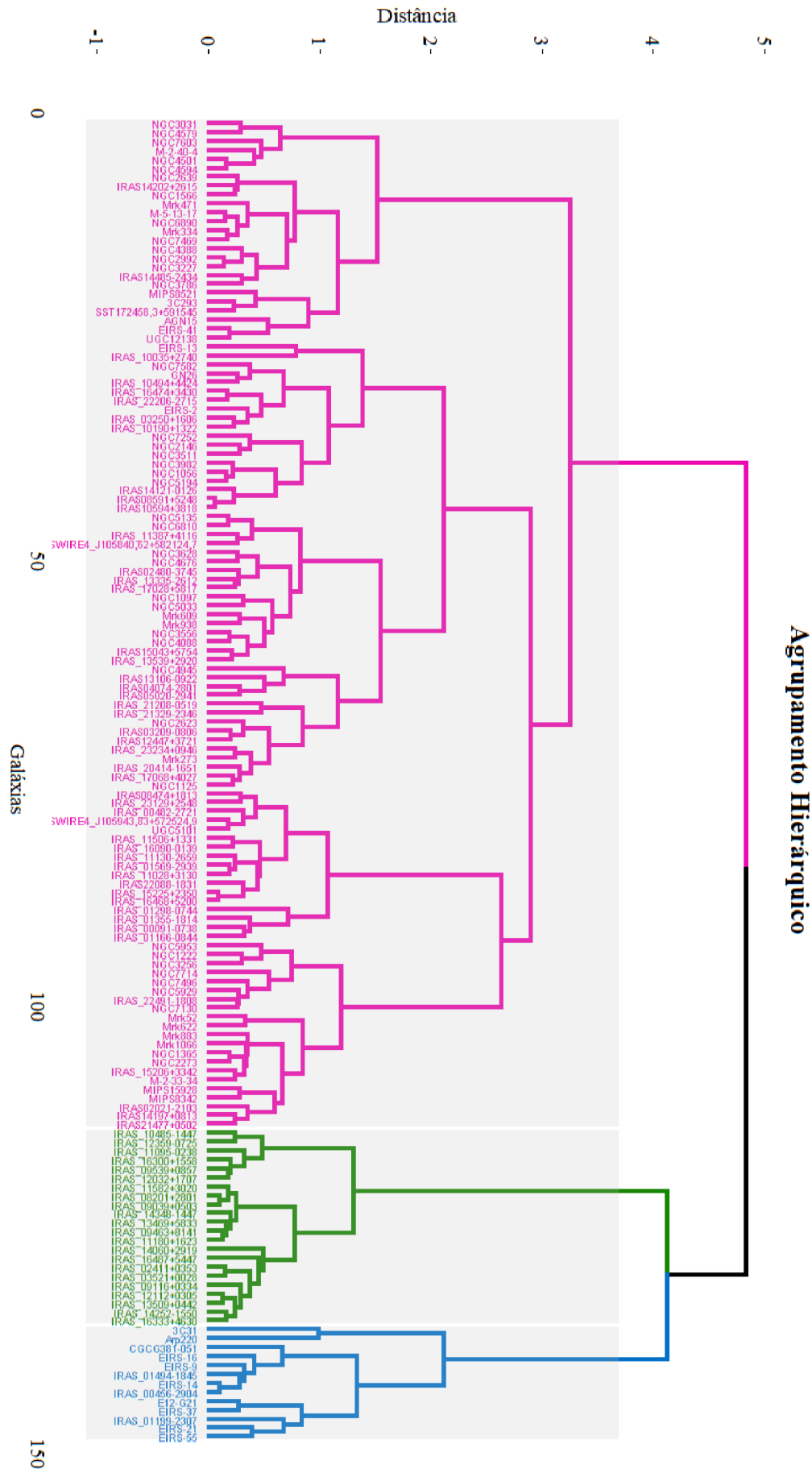


Figura 3.5: Dendrograma gerado pelo método Ward. Foi realizado um corte próxima a distância euclidiana 3, delimitando os três grupos: em rosa, o Grupo 1; em azul, o Grupo 2 e em verde o Grupo 3. O eixo horizontal apresenta a amostra total, colorida por grupos e nos rótulos são apresentados os nomes das galáxias.

tribuições de todas as classes moleculares, no entanto as médias estatísticas das moléculas grande e muito grande (X226_300, maior_300), são maiores do que nos outros grupos. Assim, a análise de agrupamentos destaca o primeiro grupo dominado por rPDR e moléculas muito grandes corroborando com os resultados da análise de correlação de Pearson.

O Grupo 2 contém 14 galáxias e é o menor dos agrupamentos. A propriedade rAGN apresenta uma média maior em relação aos demais, no entanto as componentes referentes à ionização estelar possuem médias significativas. Para Hernán-Caballero e Hatziminaoglou (2011), a presença dos dois tipos de fonte ionizante pode ser explicada pela abertura do instrumento utilizado (IRS) captar uma grande parcela da luminosidade da galáxia, bem como a existência de objetos com significativa contribuição vinda de AGNs coexistindo com as atividades estelares do starburst (Hernan-Caballero et al., 2009).

As propriedades referentes às moléculas de PAHs pequenas, muito pequenas e médias possuem as maiores médias nesse grupo. Os conjuntos moleculares grande e muito grande não possuem contribuição nesse grupo, assim, corroborando com os resultados na análise de correlação de Pearson, onde, galáxias dominadas por AGNs estão correlacionadas com moléculas muito pequenas e pequenas.

Por último o Grupo 3, que apresenta-se como um subconjunto que possui galáxias cujas fonte de ionização térmica e não-térmica coexistem. Esse grupo é o segundo maior, composto por 43 galáxias e as componentes estelares (rHII e rPDR) apresentam mais de 0,5 de média, enquanto, a fonte rAGN possui 0,36 de média. Os maiores valores de contribuição provém dos tamanhos moleculares médio, grande e muito pequeno. Além disso, esse cluster possui a maior média para propriedade Ssil em relação aos demais grupos.

Os resultados obtidos pela análise de agrupamento são corroborados pelos trabalhos de Hernán-Caballero e Hatziminaoglou (2011) e Hernan-Caballero et al. (2009). O primeiro classifica a amostra utilizada como dominada por formação estelar, mas com significativa contribuição vinda do AGN, enquanto o segundo aponta para a existência de AGNs com consideráveis taxas de formação estelar. É possível verificar que os três grupos possuem contribuições de todos os tamanhos moleculares, no entanto as classes grande e muito grande possuem maiores valores no Grupo 1. No Grupo 2, as moléculas médias e pequenas se destacam e no Grupo 3, surge a maior contribuição de moléculas grandes em relação aos outros clusters. Na seção seguinte, será tratada a Análise de Correspondência (CA) executada na amostra.

Tabela 3.2 - Medidas estatísticas do Grupos 1, 2 e 3.

	Grupo 1 - Rosa		Grupo 2 - Azul		Grupo 3 - Verde	
Propriedade	Média	Desvio	Média	Desvio	Média	Desvio
r_AGN	0,38	0,20	0,45	0,23	0,36	0,07
r_HII	0,49	0,23	0,44	0,20	0,56	0,12
r_PDR	0,53	0,21	0,50	0,17	0,50	0,13
Ssil	0,24	0,22	0,25	0,25	0,45	0,19
EW6.2	0,06	0,14	0,06	0,10	0,03	0,06
X0.30	0,45	0,13	0,59	0,17	0,47	0,08
X31.150	0,16	0,03	0,58	0,18	0,31	0,02
X151.225	0,52	0,16	0,77	0,33	0,54	0,02
X226.300	0,62	0,08	0,00	0,00	0,91	0,04
maior_300	0,58	0,15	0,00	0,00	0,00	0,00

3.3 Análise de Correspondência (CA)

Indicada para tratamento de grandes quantidades de dados, a Análise de Correspondência permite visualizar relações existentes entre linhas e colunas da matriz de dados. De forma análoga a PCA, também é utilizada para reduzir as dimensões da amostra original, porém, essa técnica estatística multivariada não-supervisionada é complementar a PCA pois ela encontra correspondência entre as variáveis de uma matriz de dados enquanto as PCAs encontram discriminação desses dados. Portanto, essa técnica permite ainda, especificar de forma visual quais variáveis estão relacionadas e de quais formas estas relações acontecem.

Em prosseguimento ao estudo, foi realizada a Análise de Correspondência na amostra de galáxias dominadas por starburst. A Figura 3.6 apresenta os resultados dessa técnica no gráfico de variâncias explicadas por dimensões, onde o tracejado azul indica o autovalor médio. A Dimensão 1 compreende cerca de 27,8% da variância total observada na amostra, enquanto a Dimensão 2 representa 21.6%. A Dimensão 3 descreve 16,7% de variância, enquanto a Dimensão 4 e a Dimensão 5, respectivamente 10,2% e 8,9%, correspondem variância total da amostra.

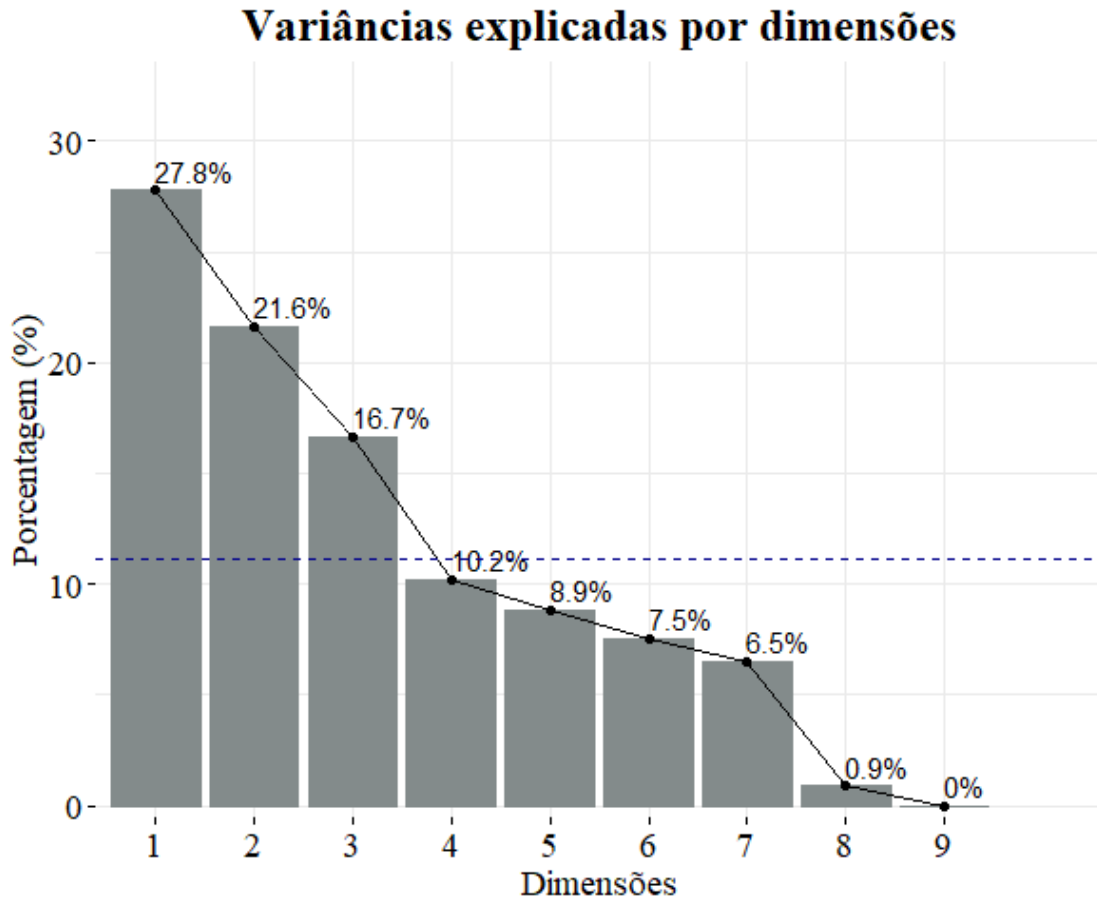


Figura 3.6: Porcentagem da variância para cada dimensão, obtida com a CA. O tracejado azul indica o autovalor médio, mostrando que as três primeiras dimensões concentram a dessemelhança de toda amostra.

Se os dados fossem distribuídos arbitrariamente, os autovalores, ou seja a variância retida em cada dimensão, seriam idênticos. Bendixen (1995) sugere que os eixos mantidos na análise, sejam escolhidos a partir do autovalor médio da contribuição, obtido pela expressão: $100/(N_{colunas} - 1)$. Como esta amostra é composta por 10 colunas de propriedades das galáxias, o autovalor médio relevante é considerado 11,11% da variabilidade, como indicado na imagem anterior. Neste sentido, as três primeiras dimensões que somam 66,1%, explicam de forma satisfatória as similaridades encontrada na amostra.

Uma vez determinadas as dimensões com maior impacto na variabilidade da amostra, é necessário verificar quais variáveis possuem maior relevância para as dimensões. A contribuição da propriedade é dada em porcentagem, sendo que aquelas com valores elevados são mais significativas para a dimensão correspondente. A ferramenta estatística utilizada nesta seção fornece a contribuição das propriedades e também das galáxias, para descrever a variabilidade encontrada na amostra. Isso acontece pela capacidade da CA verificar

relações entre variáveis de colunas e os indivíduos das linhas.

A Figura 3.7 mostra as dez colunas que descrevem características da amostra e suas respectivas contribuições para cada dimensão. Nota-se a expressiva contribuição da largura equivalente da banda em 6.2 (EW62) para a Dimensão 1, que detém 59,05% da variabilidade. Outra variável interessante é a contribuição da assinatura de silicato, que apresenta 13,39% de variância. As demais colunas quando somadas contribuem com $\sim 27\%$ para o total desta dimensão.

Contribuição das propriedades para cada dimensão

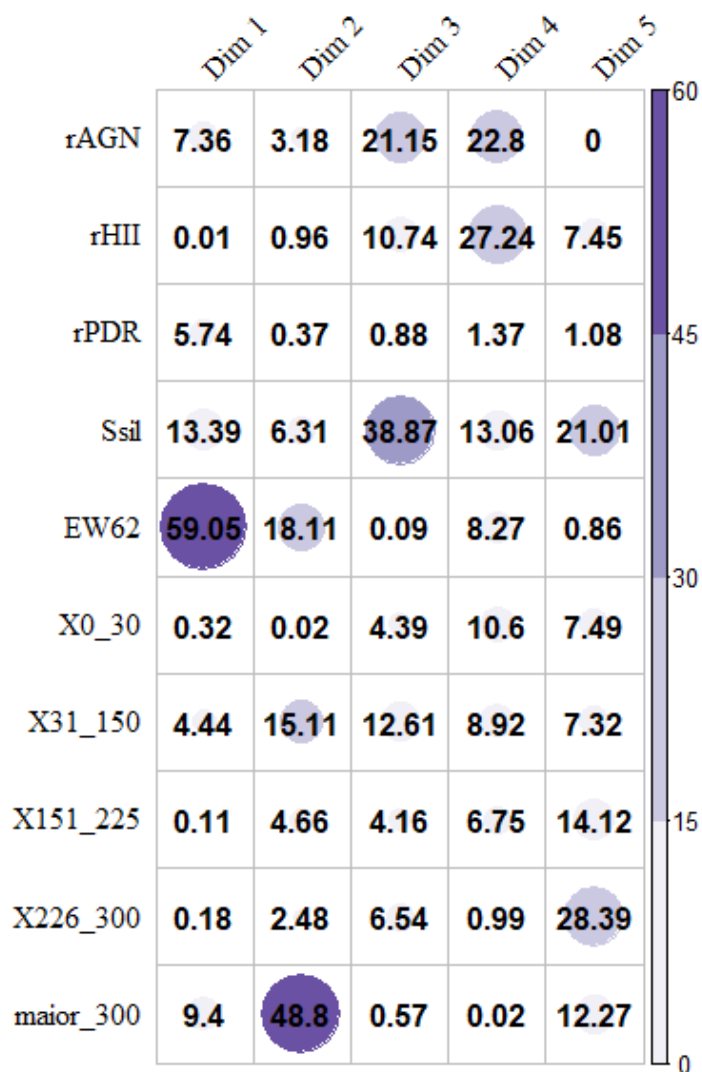


Figura 3.7: Porcentagem da contribuição das propriedades originais para cada dimensão. O tamanho e a transparência do círculo roxo é proporcional ao valor desta contribuição.

As propriedades mais relevantes para a Dimensão 2 são o tamanho molecular muito grande, a largura equivalente EW62 e tamanho molecular médio que contribuem com,

48,8%, 18,11% e 15,11%, respectivamente. Percebe-se que os menores percentuais de contribuição, vêm das propriedades X0_30 e rPDR, 0,02% e 0,37%, respectivamente.

A Dimensão 3 possui as maiores contribuições provenientes da assinatura de silicato e da fonte de ionização por AGN, 38,87% e 21,15%. A propriedade X31_150 retém 12,61% da variância dessa componente, enquanto 10,74% de contribuição provém da componente rHII. Com limitada contribuição para a Dimensão 3, 0,09% vêm da largura equivalente de PAH em 6.2 μm .

Conforme mencionado anteriormente, as duas últimas dimensões não possuem valor médio relevante para a análise de correspondência. Porém, vale ressaltar a expressiva parcela das fontes ionizantes rAGN e rHII para a Dimensão 4, respectivamente, 27,24% e 22,8%. Bem como moléculas muito pequenas ($\sim 11\%$), pequenas ($\sim 9\%$) e Ssil (13%). Na Dimensão 5 duas propriedades contribuem significativamente e ambas estão relacionadas ao material molecular, são elas PAHs grandes (X226_330 $\sim 28\%$) e assinatura de silicato (Ssil $\sim 21\%$). As maiores contribuições de PAHs grandes e Ssil aconteceram nas Dimensão 5 e 3.

A delimitação das contribuições permitiu formular conclusões sobre a análise realizada. A propriedade mais importante para a Dimensão 1, EW62, também é utilizada como indicador de formação estelar, enfatizando a composição principal da amostra, que são galáxias starbursts. Bem como a reduzida contribuição de galáxias com núcleo ativo, representada pela presença da componente rAGN. A Dimensão 2, possui a maior parcela de contribuição vinda da moléculas muito grandes, enquanto a segunda maior contribuição está associada largura equivalente das bandas de PAHs.

Encontradas as propriedades mais importantes para a variância das dimensões, foram determinados os indivíduos com maior contribuição para as mesmas. Na Figura 3.8 encontra-se o resultado para a Dimensão 1, onde o tracejado vermelho representa o autovalor médio e indica quais galáxias são mais relevantes. É possível perceber que as galáxia EIRS-13 e IRAS_10035+2740, contribuem com $\sim 20\%$ e $\sim 6\%$ da variância na primeira componente, respectivamente.

Na Figura 3.9 se encontram as galáxias mais relevantes para a Dimensão 2, indicadas pelo tracejado vermelho. Com maior contribuição destacam-se as duas primeiras galáxias EIRS-13 e EIRS-9 com $\sim 6\%$ e 5% , respectivamente. Na sequência as galáxias CGCG381-051 com pouco mais de 5% , Arp220 com $\sim 4\%$ e NGC4579 com $\sim 3\%$. Por fim, a Figura

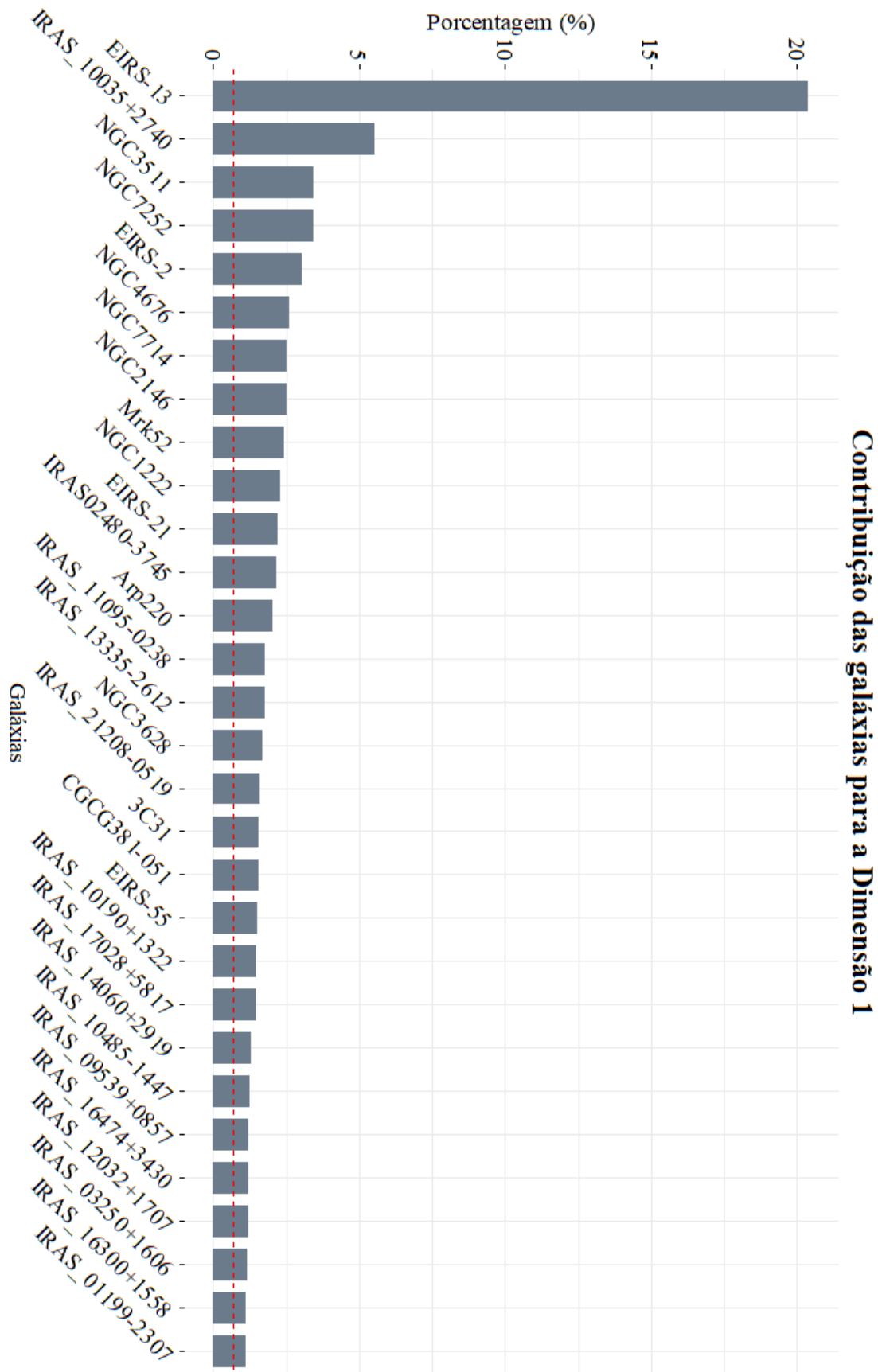


Figura 3.8: Percentil da contribuição das galáxias para a Dimensão 1, o autovalor médio é indicado pelo tracejado vermelho.

3.10 apresenta as galáxias que possuem maior relevância para a Dimensão 3, indicadas pelo tracejado do autovalor médio. O primeiro indivíduo com maior contribuição para esta componente é galáxia 3C31, correspondendo a $\sim 8\%$. A segunda e terceira galáxia EIRS-37 e E12-G21, representam ~ 7 e 6% .

Após verificar as propriedades e galáxias com maiores contribuições, buscou-se inferir como estas contribuem para a variância das dimensões principais. De acordo com Kassambara (2017b), o gráfico padrão da CA é o biplot simétrico, onde variáveis e dados são representados no mesmo gráfico e o grau de correlação é indicado pela distância entre dois pontos e pela localização no eixo do diagrama. O biplot foi proposto por Gabriel (1971), como forma complementar de analisar as dimensões principais dos dados pois é capaz de demonstrar as posições das propriedades dos dados nos quatro quadrantes. Assim, de fácil compreensão permite derivar quais informações possuem propriedades positivamente ou negativamente correlacionadas.

Sendo assim, este método é amplamente utilizado em análises estatísticas multivariadas, pois mostra de forma simultânea como variáveis se relacionam nas dimensões principais, bem como permite uma visão global do comportamento da amostra (Kassambara, 2017b). Portanto, de modo geral, variáveis positivamente correlacionadas (propriedades e/ou galáxias) são agrupadas e as negativamente correlacionadas, se localizam em quadrantes opostos. A distância da origem determina a relevância dessa variável, quanto mais longe da origem, maior a contribuição para a respectiva dimensão.

A Figura 3.11 mostra um diagrama biplot com as propriedades da Dimensão 1 versus Dimensão 2, onde, as propriedades são representadas por triângulos pretos e as galáxias, coloridas de acordo com a contribuição para a composição dessas dimensões. Pode-se verificar que EIRS-13 apresenta forte correlação com EW62, enquanto Mrk609 apresentam forte correlação com maior_300. O diagrama também mostra que a propriedade EW62 está anti-correlacionada às contribuições do núcleo ativo, o que é consistente com os resultado de Hernán-Caballero e Hatziminaoglou (2011) que utilizam essa propriedade como indicador direto de formação estelar. Esse resultado também corrobora com os resultados obtidos por Sales et al. (2010), onde a EW dos PAHs diminuem ao aumentar a contribuição do AGNs.

Além disso, é possível perceber no diagrama com a Dimensão 2 (ver Figura 3.11) que as propriedades de moléculas grandes (X226_300) e muito grandes (maior_300) estão

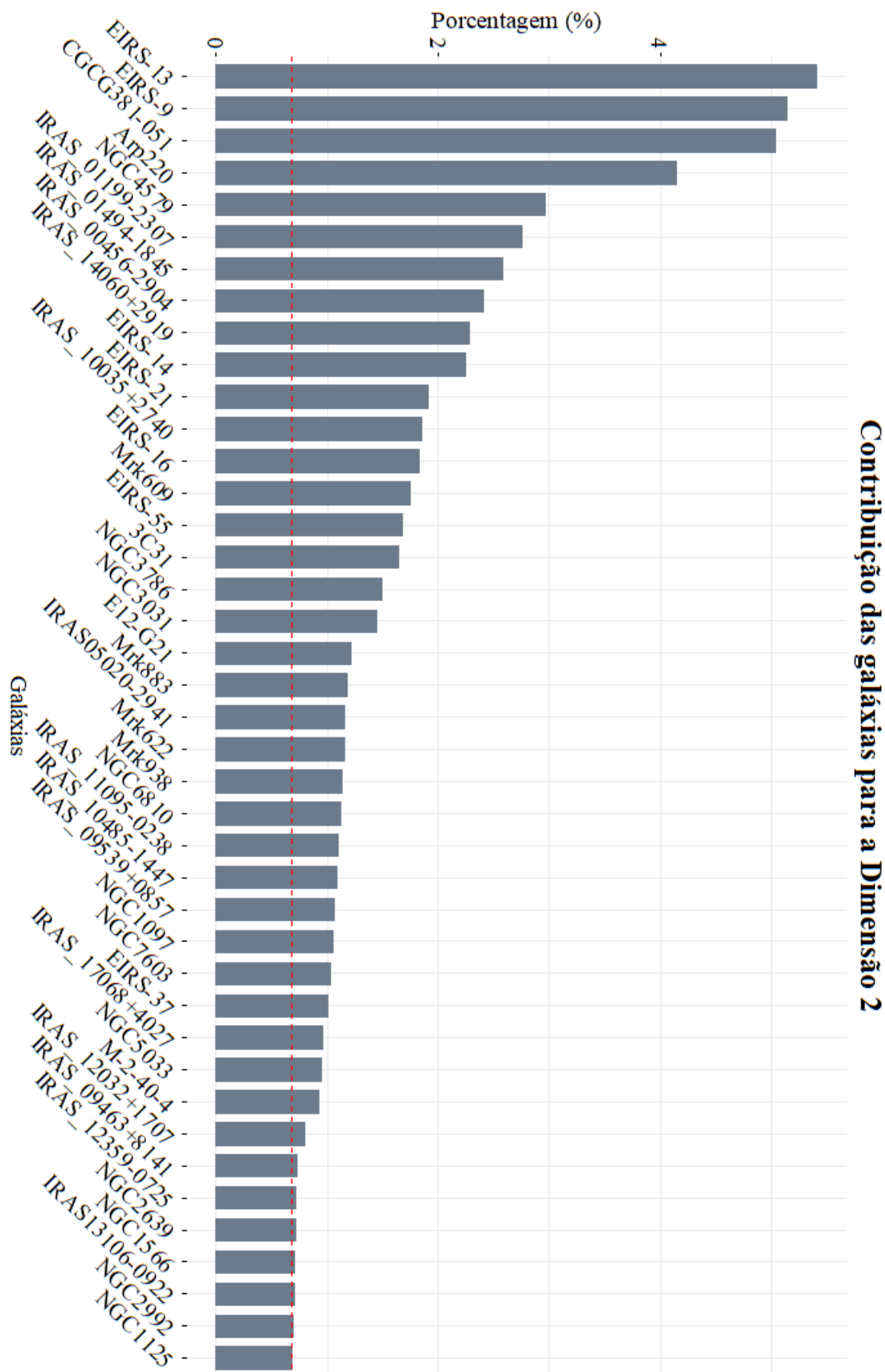


Figura 3.9: Percentil da contribuição das galáxias para a Dimensão 2, o autovalor médio é indicado pelo tracejado vermelho.

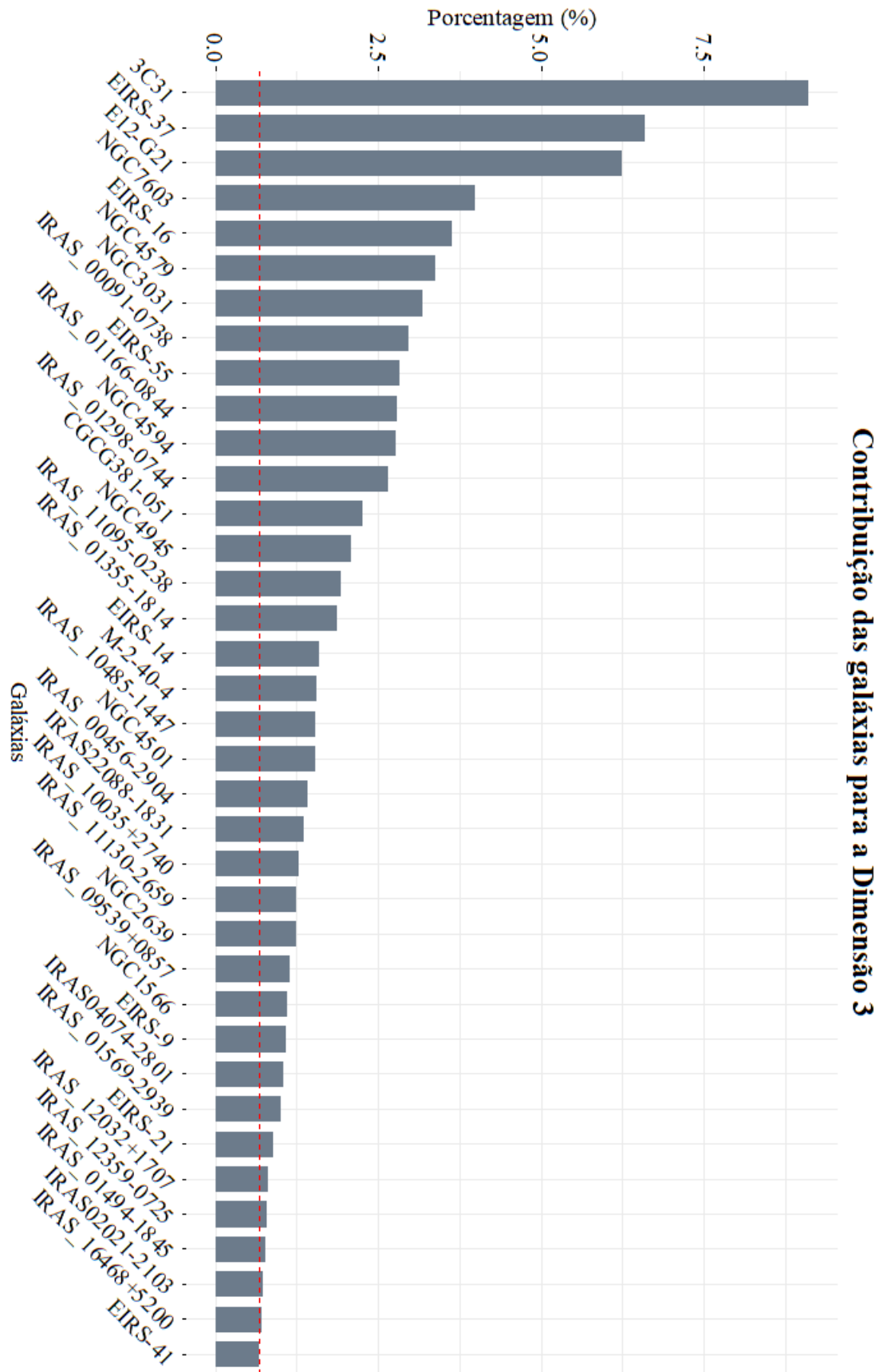


Figura 3.10: Percentil da contribuição das galáxias para a Dimensão 3, o autovvalor médio é indicado pelo tracejado vermelho.

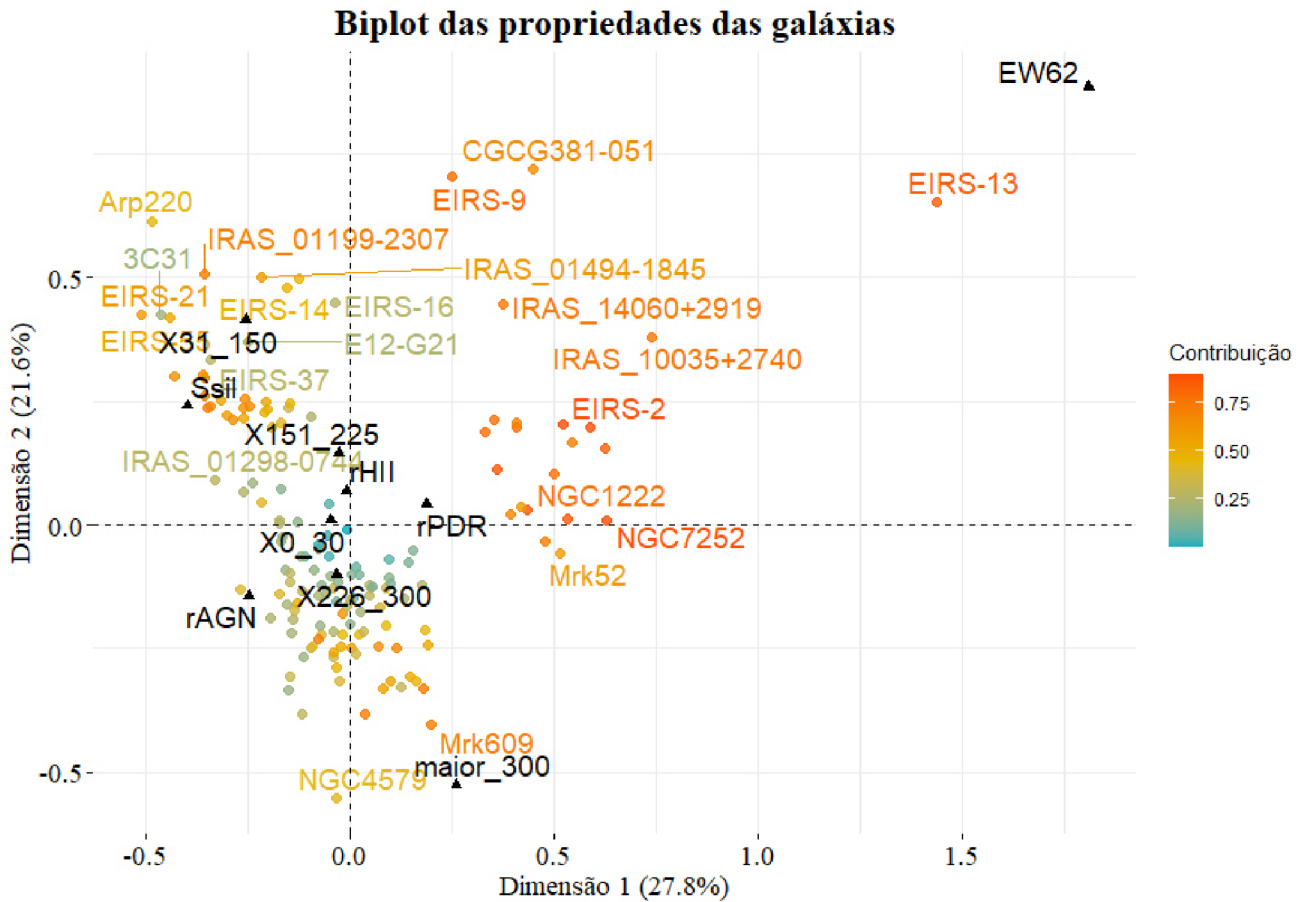


Figura 3.11: Biplot das propriedades das galáxias das Dimensões 1 e 2. As propriedades são representadas por triângulos pretos e as galáxias, coloridas de acordo com a contribuição para a composição das dimensões

no mesmo quadrante que rAGN. Enquanto, o quadrante oposto a essas propriedades encontram-se moléculas médias, pequenas, Ssil e rHII. Esses resultados corroboram os trabalhos de Sales (2012), Sales et al. (2010), Sales et al. (2013) e Ruschel-Dutra et al. (2014), onde se demonstrou com outras metodologias que em galáxias com AGNs preferencialmente encontramos moléculas de PAHs maiores que em galáxias dominadas por formação estelar.

A Figura 3.12 mostra um diagrama biplot com as propriedades da Dimensão 2 versus Dimensão 3, onde, as propriedades são representadas por triângulos pretos e as galáxias por pontos coloridas de acordo com a contribuição para a composição dessas dimensões. O gráfico indica que as galáxias NGC4579, NGC1566 e NGC2639 possuem intensa correlação com rAGN. Percebe-se que a Dimensão 3 também é caracterizada pela correlação negativa entre os componentes de origem não térmica e estelar, uma vez que se encontram em quadrantes opostos.

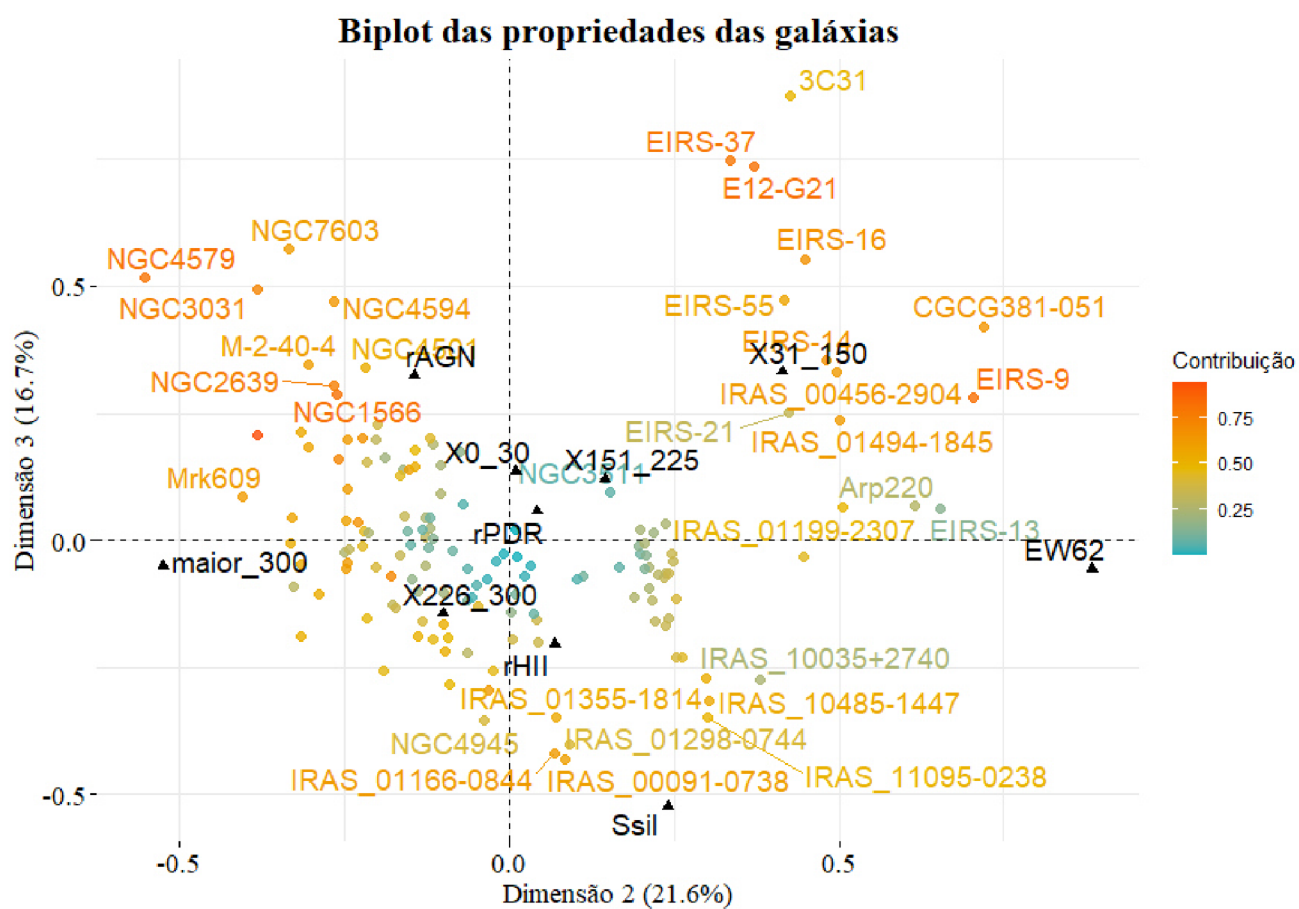


Figura 3.12: Biplot das propriedades das galáxias das Dimensões 2 e 3. As propriedades são representadas por triângulos pretos e as galáxias, coloridas de acordo com a contribuição para a composição das dimensões.

Conclusões

No decorrer deste trabalho investigamos, as propriedades físico-químicas do gás e poeira de uma amostra composta por 148 galáxias dominadas por starburst, observadas pelo Telescópio Espacial Spitzer. O banco de dados consiste em galáxias do catálogo ATLAS, cujas propriedades relacionadas às fontes de ionização, EW de PAHs em $6.2\mu\text{m}$ e assinatura de silicato foram derivadas e coletadas por Hernán-Caballero e Hatziminaoglou (2011). Também foram usadas propriedades referentes aos tamanhos moleculares, extraídas da base de dados NASA Ames PAH IR e derivadas por Espinoza (2020).

Foram realizados diversos métodos de análise estatística multivariada na amostra, a saber: análise de componentes principais (PCA); agrupamento hierárquico pelo método de Ward e análise de correspondência (CA). No primeiro momento deste estudo, foi definida uma matriz de correlação utilizando o coeficiente de Pearson, para verificar associação entre os dados da amostra. Foi possível inferir através de estudo estatístico com algoritmo de modelagem não-supervisionado as seguintes informações:

- A componente da fonte ionizante rAGN possui correlação negativa com as fontes de ionização rHII e rPDR. Também está correlacionada negativamente com a classe das moléculas muito pequenas e grandes e com as propriedades EW62, Ssil. Essa componente apresenta correlação positiva apenas com moléculas pequenas e médias.
- A propriedade rHII é negativamente correlacionada com moléculas pequenas, médias e grandes. Bem como com as fontes de ionização por AGN e PDR. Ainda, essa propriedade possui correlação positiva com moléculas muito pequenas e grandes. De mesmo modo está positivamente correlacionada às propriedades Ssil e EW62.
- A componente rPDR está positivamente correlacionada apenas com a classe molecu-

lar muito grande e com EW62. Enquanto está negativamente correlacionada com as fontes ionizantes AGN e HII, com a assinatura de silicato e com moléculas pequenas, médias e grandes.

- A propriedade da largura equivalente da banda em $6.2 \mu\text{m}$ possui correlação positiva com as componentes estelares (rPDR e HII) e com moléculas médias e muito grandes. Bem como possui correlação negativa com a propriedade rAGN, com a assinatura de silicato e com os conjuntos moleculares muito pequeno, pequeno e grande.
- Por fim, a propriedade relacionada à assinatura de silicato, que está positivamente correlacionada aos conjuntos moleculares pequeno, médio, grande e com a componente rHII. Enquanto está negativamente correlacionada aos conjuntos moleculares muito pequeno e muito grande. Da mesma forma possui correlação negativa às propriedades rAGN, rPDR e EW62.

As análises realizadas pela PCA mostrou que as três primeiras dimensões concentram mais de 58% da variância encontrada na amostra. Onde a Dimensão 1 engloba as classes moleculares pequena (X31_150), muito grande (maior_300), rAGN e rPDR que correspondem $\sim 74\%$ da variância nessa componente. A Dimensão 2 concentra $\sim 56\%$ da contribuição proveniente das propriedades X0_30, rPDR e X31_150. Enquanto na Dimensão 3, $\sim 75\%$ da contribuição tem origem nos conjuntos moleculares médio, muito grande e muito pequeno.

Além disso, a técnica de PCA reduziu a amostra em dimensões totalmente ortogonais e agrupou os dados de acordo com suas variâncias. Nesse sentido, este método é bastante satisfatório, pois condensou em poucas componentes a estrutura da amostra, composta principalmente de galáxias starbursts, mas com relevante parcela de ionização por AGNs. Além da significativa contribuição do conjunto de tamanhos moleculares. Além disso, por se tratar de uma análise discriminante, é possível verificar quais propriedades mais contribuem para a variação dos dados encontrados na amostra.

Na segunda parte do trabalho, o pacote NbClust foi utilizado para indicar o número adequado de grupos, utilizando o método de agrupamento hierárquico de Ward e a métrica Euclidiana. O algoritmo sugeriu como apropriado o número 3, produzindo assim os seguintes clusters, que são: Grupo 1, contendo 114 galáxias; Grupo 2 com 14 galáxias e Grupo 3, composto por 23 galáxias. Após delimitarmos o número de grupos, foi realizada a média estatística e o desvio padrão, para indicar o quanto essa média flutua, nas propriedades de

cada um. A partir dessas análises, foi possível inferir as seguintes conclusões:

- O Grupo 1 mostra acentuada contribuição da componente rPDR e a assinatura de silicato possui a menor média em relação aos demais grupos. Este grupo possui contribuição de todos as classes de tamanhos de PAHs, no entanto moléculas grandes e muito grandes retém maiores parcelas neste cluster.
- No Grupo 2 são observadas propriedades típicas de núcleos ativos, como dominância da componente AGN e baixa contribuição das componentes estelares. A classe dos PAHs médios possui predominância nesse grupo.
- O Grupo 3 apresenta grande contribuição das regiões HII e de AGNs. Nota-se alta contribuição da assinatura de silicato, em relação aos clusters anteriores e prevalência de espécies moleculares grandes.

Nesse sentido, o agrupamento hierárquico atingiu os resultados esperados, reconhecendo os padrões existentes na amostra e identificando as características que esperava-se que fossem alcançadas. Como por exemplo, a indicação do número de clusters, classificados de acordo com a fonte de ionização dominante e com parcelas dos tamanhos moleculares. Além de corroborar com outros métodos, como o estudo desenvolvido por Hernán-Caballero e Hatziminaoglou (2011).

A última etapa do trabalho foi submeter a amostra à Análise de Correspondência (CA). Esse método indica que as três primeiras dimensões representam mais de 66% da variabilidade amostral. A descrição das propriedades mostra que a largura equivalente EW62, que é relacionada às regiões de formação estelar (Hernán-Caballero e Hatziminaoglou, 2011), contribui com $\sim 59\%$ da Dimensão 1. O conjunto molecular muito grande e EW62 contribuem com $\sim 48\%$ e $\sim 18\%$, respectivamente, para a Dimensão 2. Enquanto a Dimensão 3 possui as maiores parcelas vindas das propriedades Ssil e rAGN, $\sim 38\%$ e $\sim 21\%$ respectivamente.

O método CA mostra reduzida contribuição dos tamanhos moleculares na primeira dimensão, indicando que essas características têm pouco efeito na variabilidade da amostra. Além de indicar quais são as propriedades mais relevantes para a amostra, a CA também especifica quais galáxias possuem maior contribuição. Além de indicar as propriedade correlacionadas a estas galáxias nas respectivas dimensões, por meio do gráfico biplot, .

Na Dimensão 1 foram extraídas as seguintes galáxias: EIRS-13 e IRAS_10035+2740 com $\sim 20\%$ e $\sim 5\%$ de contribuição, respectivamente. O diagrama biplot aponta que essas galáxias estão relacionadas com a propriedade EW62, dominante nesta componente. Na Dimensão 2, as galáxias com maior contribuição são EIRS-13, EIRS-9 e CGCG381-051, respectivamente $\sim 6\%$, $\sim 5\%$ e $\sim 5\%$. Com pouco mais de 8% a galáxia 3C31 possui a maior contribuição para a Dimensão 3, assim como EIRS-37 e E12-G21, com $\sim 6\%$ e $\sim 7\%$, respectivamente.

Diferentemente da PCA, que é uma análise discriminante e busca distinguir objetos, a CA é uma análise descritiva que tem o intuito de esquematizar tendências ou semelhanças nos dados. Pode-se observar que as fontes ionizantes rHII, rPDR e rAGN, bem como a largura equivalente em $6,2 \mu\text{m}$, têm efeito relevante na caracterização da amostra. Este fato é consistente com os autores e mostra a confiabilidade da técnica, pois não se sabia a capacidade do algoritmo de identificar o comportamento da amostra.

De modo geral, as 148 galáxias utilizadas neste estudo puderam ser satisfatoriamente compreendidas com poucos elementos. Através de algoritmos não-supervisionados foi possível validar resultados obtidos anteriormente por Sales et al. (2010, 2013) utilizando técnicas complementares. Portanto, o grande diferencial desse trabalho é fazer uso de estatística multivariada que possibilita o tratamento de grandes volumes de dados, por meio da modelagem não-supervisionada.

Dessa forma, a pesquisa realizada se mostrou uma poderosa ferramenta na análise de características astroquímicas de galáxias, sendo alicerçado em técnicas e referências anteriores. Além disso, é importante ressaltar que os algoritmos de modelagem não-supervisionadas, além de validar os resultados publicados anteriormente eles também se mostraram ferramentas importante para a nova era da astroquímica onde será necessário trabalhar com um grande volume de dados como os que serão disponibilizados pelos grandes telescópios, ex. “*James Web Space Telescope, Legacy Survey of Space and Time, Extremely Large Telescope*” e etc.

Referências Bibliográficas

- Alonso-Herrero A., Ramos Almeida C., Esquej P., Roche P. F., Hernán-Caballero A., Hönig S. F., González-Martín O., Aretxaga I., Mason R. E., Packham C., Levenson N. A., Rodríguez Espinosa J. M., Siebenmorgen R., Pereira-Santaella M., Díaz-Santos T., Colina L., Alvarez C., M. T. C., Nuclear 11.3 μ m PAH emission in local active galactic nuclei, MNRAS, 2014, vol. 443, p. 2766
- Antonucci R. R. J., Unified Models for Active Galactic Nuclei and Quasars, ARA&A, 1993, vol. 31, p. 473
- Antonucci R. R. J., Miller J. S., Spectropolarimetry and the nature of NGC 1068, ApJ, 1985, vol. 291, p. 621
- Baldwin J. A., Phillips M. M., Terlevich R., Classification Parameters for the Emission-line Spectra of Extragalactic Objects, PASP, 1981, vol. 93, p. 5
- Beckmann V., Shrader C. R., Active Galactic Nuclei. Weinheim: Wiley-VCH, 2012, 250 p.
- Bendixen M. T., Compositional Perceptual Mapping Using Chi-squared Trees Analysis and Correspondence Analysis., Journal of Marketing Management, 1995, vol. 11, p. 571
- Boersma C., Bauschlicher C. W., Ricca A., Mattioda A. L., Cami J., Peeters E., Sánchez de Armas F., Puerta Saborido G., Hudgins D. M., Allamandola L. J., The NASA Ames PAH IR Spectroscopic Database Version 2.00: Updated Content, Website and On/Offline Tools, ApJS, 2014, vol. 211, p. 8

- Canelo C. M., Friaça A. C. S., Sales D. A., Pastoriza M. G., Ruschel-Dutra D., Variations in the 6.2 μm Emission Profile in Starburst-dominated Galaxies: A Signature of Polycyclic Aromatic Nitrogen Heterocycles (PANHs)?, *MNRAS*, 2018, vol. 475, p. 3746
- Charrad M., Ghazzali N., Boiteau V., Niknafs A., NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set, *J. Stat. Softw.*, 2014, vol. 61, p. 1
- De Becker M., Astrochemistry: the Issue of Molecular Complexity in Astrophysical Environments, *Bull. Soc. R. Sci. Liege*, 2013, vol. 82, p. 33
- Ehrenfreund P., Rasmussen S., Cleaves J., Chen L., Experimentally Tracing the Key Steps in the Origin of Life: The Aromatic World, *Astrobiology*, 2006, vol. 6, p. 490
- Espinoza J., Técnicas de análise espectral aplicadas para identificação de PAHs em Galáxias Starbursts: Transformada Rápida de Fourier e o Algoritmo SAM, Rio Grande: Universidade Federal do Rio Grande, 2020, Dissertação de Mestrado, 100 p.
- Everitt B. S., Landau S., Leese M., Cluster Analysis. 4 ed. London: Arnold, 2001
- Fath E. A., The Spectra of Some Spiral Nebulae and Globular Star Clusters, *LicOB*, 1909, vol. 149, p. 71
- Gabriel K. R., The Biplot Graphic Display of Matrices with Application to Principal Component Analysis, *Biometrika*, 1971, vol. 58, p. 453
- Gallagher J. S., Hunter D. A., Structure and Evolution of Irregular Galaxies, *ARA&A*, 1984, vol. 22, p. 37
- Galvão N. D., Marin H. F., Técnica de mineração de dados: uma revisão da literatura, *Acta Paul Enferm*, 2009, vol. 22, p. 686
- Gastão B. L. N., Astronomia Extragaláctica e Cosmologia: Notas de Aula. São Paulo: Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG), 2020, 268 p
- Hair J. F., Black W. C., Babin B. J., Anderson R. E., Tatham R. L., Análise multivariada de dados. 6^a ed. Porto Alegre: Bookman, 2005, 688 p.
- Hernan-Caballero A., Pérez-Fournon I., Hatziminaoglou E., Afonso-Luis A., Rowan-Robinson M., Rigopoulou D., Farrah D., Lonsdale C. J., Babbedge T., Clements D.,

- Serjeant S., Pozzi F., Vaccari M., Montenegro-Montes F. M., Valtchanov I., González-Solares E., Oliver S., Shupe D., Gruppioni C., Vila-Vilaró B., Lari C., La Franca F., Mid-infrared spectroscopy of infrared-luminous galaxies at $z \sim 0.5-3$, *MNRAS*, 2009, vol. 395, p. 1695
- Hernán-Caballero A., Hatziminaoglou E., An atlas of mid-infrared spectra of star-forming and active galaxies, *MNRAS*, 2011, vol. 414, p. 500
- Hubble E. P., *The Realm of the Nebulae*. New Haven: Yale University Press, 1936, 207
- Joblin C., Léger A., Martin P., Contribution of Polycyclic Aromatic Hydrocarbon Molecules to the Interstellar Extinction Curve, *ApJ*, 1992, vol. 393, p. 79
- Jolliffe I. T., *Principal Component Analysis*. 2nd ed. New York: Springer-Verlag, 2002
- Kassambara A., *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA, 2017a
- Kassambara A., *Practical Guide to Principal Component Methods in R: PCA, (M)CA, FAMD, MFA, factoextra*. STHDA, 2017b
- Kepler S. O., Saraiva M. F. O., *Astronomia e Astrofísica*. 3^a. São Paulo: Livraria da Física, 2017, 640 p
- Khachikian E. Y., Weedman D. W., An Atlas of Seyfert Galaxies, *ApJ*, 1974, vol. 192, p. 581
- Leitherer C., Leão J. R. S., Heckman T. M., Lennon D. J., Pettini M., Robert C., Ultraviolet Line Spectra Of Metal-poor Star-forming Galaxies, *ApJ*, 2001, vol. 550, p. 724
- Li A., *Interaction of Nanoparticles with Radiation*, ASP Conference Series, Astrophysics of Dust, 2004, vol. 309, p. 417
- Magdis G. E., Rigopoulou D., Helou G., Farrah D., Hurley P., Alonso-Herrero A., Bock J., Burgarella D., Chapman S., Charmandaris V., Mid- to far-infrared properties of star-forming galaxies and active galactic nuclei, *A&A*, 2013, vol. 136

- McGuire B. A., 2021 Census of Interstellar, Circumstellar, Extragalactic, Protoplanetary Disk, and Exoplanetary Molecules, arXiv e-prints, 2021
- Miranda B. M. A., Propriedades dos hidrocarbonetos aromáticos policíclicos em galáxias Starbursts: uma análise estatística multivariada, Rio Grande: Universidade Federal do Rio Grande, 2020, Dissertação de Mestrado, 104 p.
- Miranda B. M. A., Sales D. A., Amaral J. T. S., Mariotti C. B., Vicenti J. R. M., Estudo da Emissão do Benzeno em Galáxias Usando a Teoria do Funcional de Densidade, *Rev. Mundi Eng., Tecnol. Gest.*, 2019, vol. 4, p. 136
- O'Dowd M. J., Schiminovich D., Johnson B. D., Treyer M. A., Martin C. D., Wyder T. K., Charlot S., Heckman T. M., Martins L. P., Seibert M., van der Hulst J. M., Polycyclic Aromatic Hydrocarbons in Galaxies at $z \sim 0.1$: the Effect of Star Formation and Active Galactic Nuclei, *ApJ*, 2009, vol. 705, p. 885
- Pastoriza M. G., Studies of galaxies with peculiar nuclei, *Astrophys. Space Sci.*, 1975, vol. 33, p. 173
- Pearson K., LIII. On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine and Journal of Science*, 1901, vol. 2, p. 559
- Peres-Neto P., Jackson D. A., Somers K. M., How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited, *Computational Statistics & Data Analysis*, 2005, vol. 49, p. 974
- Peterson B. M., An Introduction To Active Galactic Nuclei. Cambridge: Cambridge University Press, 1997, 238 p.
- Rodríguez-Ardila A., Viegas S. M., Pastoriza M. G., Prato L., Near-Infrared Coronal Lines in Narrow-Line Seyfert 1 Galaxies, *ApJ*, 2002, vol. 579, p. 214
- Rosa R. C., Sales D. A., Canelo C. M., Miranda B. M. A., Análise Computacional do Hidrocarboneto Aromático Policíclico Antraceno e sua Aplicação na Astroquímica, *Cad.Astro*, 2021, vol. 2, p. 132

- Ruschel-Dutra D., Pastoriza M. G., Riffel R., Sales D. A., Winge C., A mid-IR comparative analysis of the Seyfert galaxies NGC 7213 and NGC 1386, *MNRAS*, 2014, vol. 438, p. 3434
- Sales D. A., Propriedades Físicas dos Silicatos e Hidrocarbonetos Aromáticos Policíclicos Presentes na Região Nuclear das Galáxias Seyferts e Starburst, Universidade Federal do Rio Grande do Sul, 2012, Tese de Doutorado, 166 p.
- Sales D. A., Pastoriza M. G., Riffel R. A., Polycyclic aromatic hydrocarbon and emission line ratios in active galactic nuclei and starburst galaxies, *ApJ*, 2010, vol. 725, p. 605
- Sales D. A., Pastoriza M. G., Riffel R. A., Winge C., Polycyclic aromatic hydrocarbon in the central region of the Seyfert 2 galaxy NGC 1808, *MNRAS*, 2013, vol. 429, p. 2634
- Schneider P., *Extragalactic Astronomy and Cosmology: An Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, 630 p.
- Seyfert C. K., Nuclear Emission on Spiral Nebulae, *ApJ*, 1943, vol. 97, p. 28
- Shannon M. J., Boersma C., Organic Molecules in Space: Insights from the NASA Ames Molecular Database in the era of the James Webb Space Telescope, *Proceedings of the 17th Python in Science Conference*, 2018, pp 99–105
- Silva-Ribeiro A., Krabbe A. C., Canelo C. M., Monteiro A., Sales D. A., Hernandez-Jimenez J. A., Andrade D. P. P., Polycyclic aromatic hydrocarbons and the ionized gas in galaxies with active nuclei, *MNRAS*, 2022, vol. 509, p. 327
- Smith J. D. T., Draine B. T., Dale D. A., Moustakas J., Kennicutt R. C., Helou G., Armus L., Roussel H., Sheth K., Bendo G. J., The Mid-Infrared Spectrum of Star-forming Galaxies: Global Properties of Polycyclic Aromatic Hydrocarbon Emission, *ApJ*, 2007, vol. 656, p. 770
- Spoon H. W. W., Marshall J. A., Houck J. R., Elitzur M., Hao L., Armus L., Brandl B. R., Charmandaris V., Mid-Infrared Galaxy Classification Based on Silicate Obscuration and PAH Equivalent Width, *ApJ*, 2007, vol. 654, p. 49
- Sérsic J. L., Pastoriza M. G., Properties of Galaxies with Peculiar Nuclei, *PASP*, 1967, vol. 79, p. 152

- Tielens A. G. G. M., *The Physics and chemistry of the interestelar medium*. Cambridge: Cambridge University Press, 2005
- Tielens A. G. G. M., *Interstellar Polycyclic Aromatic Hydrocarbon Molecules*, *ARA&A*, 2008, vol. 46, p. 289
- Tielens A. G. G. M., *The Molecular Universe*, *RMP*, 2013, vol. 85, p. 1021
- Tielens A. G. G. M., Hollenbach D., *Photodissociation regions. I - Basic model. II - A model for the Orion photodissociation region*, *ApJ*, 1985, vol. 291, p. 722
- Voit M. G., *Destruction and Survival of Polycyclic Aromatic Hydrocarbons In Active Galaxies*, *MNRAS*, 1992, vol. 258, p. 841
- Ward J. H., *Hierarchical Grouping to Optimize an Objective Function*, *JASA*, 1963, vol. 58, p. 236
- Woltjer L., *Emission nuclei in Galaxies*, *ApJ*, 1959, vol. 130, p. 38
- Zier C., Biermann P., *Binary black holes and tori in AGN II. Can stellar winds constitute a dusty torus*, *A&A*, 2002, vol. 396, p. 91

Apêndice

Apêndice A

Artigo Completo publicado na Revista CADERNOS
DE ASTRONOMIA

A.1 Rosa, Sales, Canelo & Matoso, v. 2, n. 2, p. 132, (2021)

Análise computacional do hidrocarboneto aromático policíclico antraceno e sua aplicação na astroquímica

Rosiane C. da Rosa¹, Dinalva A. Sales¹, Carla M. Canelo² e Brenda M. A. Miranda¹

¹Universidade Federal do Rio Grande - FURG

²Universidade de São Paulo

Resumo

Este trabalho objetiva analisar o espectro de emissão na região entre 5 e 15 μ m do antraceno, uma espécie química de relevância astroquímica, obtido por modelagem computacional. E também comparar resultados obtidos com conjuntos de bases diferentes. A classe molecular escolhida possui características que permitem sua presença abundante em ambientes hostis do espaço, despertando interesse acerca de suas propriedades físico-químicas. A estrutura molecular foi desenhada na plataforma Gabedit e os cálculos da abordagem quântica realizados pelo software ORCA. Foram utilizadas duas funções de base, 6-31G* e 6-31G**, para comparação entre resultados. Para a análise dos métodos e inferências de semelhanças e diferenças, foi utilizado o NASA Ames PAH IR Spectroscopic Database (PAHdb), como banco de dados de referência. Os resultados obtidos apresentam um avermelhamento nas bandas vibracionais, mesmo utilizando a função descrita pelos autores do PAHdb, expondo a direta relação entre complexidade e eficácia dos métodos comparados. Ainda assim, os resultados obtidos foram satisfatórios. Uma vez que as bandas de emissão desta molécula, em comparação com dados espectroscópicos, permitem inferir a presença da mesma em galáxias ativas, como Seyfert 1, Seyfert 2 e Starburst.

Abstract

This work aims to analyze the emission spectrum in the region between 5 and 15 μ m of anthracene, a chemical species of astrochemical relevance, obtained by computational modeling. And also to compare the results obtained with different sets of bases. The chosen molecular class has characteristics that allow its abundant presence in hostile space environments, arousing interest in its physicochemical properties. The molecular structure was designed on the Gabedit platform and the calculations of the quantum approach were performed by the ORCA software. Two base functions, 6-31G* and 6-31G**, were used to compare the results. For the analysis of methods and inferences of similarities and differences, the NASA Ames PAH IR Spectroscopic Database (PAHdb) was used as a reference database. The obtained results show a reddening in the vibrational bands, even using the function described by the authors of the PAHdb, exposing the direct relationship between complexity and effectiveness of the compared methods. Nevertheless, the obtained results were satisfactory. Since the emission bands of this molecule, in comparison with spectroscopic data, allow us to infer its presence in active galaxies, such as Seyfert 1, Seyfert 2 and Starburst.

Palavras-chave: astroquímica, PAH, modelagem computacional, DFT.

Keywords: astrochemistry, PAH, computational modeling, DFT.

DOI: [10.47456/Cad.Astro.v2n2.35751](https://doi.org/10.47456/Cad.Astro.v2n2.35751)

1 Introdução

As últimas décadas apresentaram pesquisas promissoras sobre a física e a química em ambientes astrofísicos. Isso se deve à grande quantidade de dados observacionais disponíveis, além do avanço nas técnicas computacionais. Decor-

rido mais de 80 anos que as primeiras moléculas simples com dois átomos foram identificadas no espaço, outros espécimes complexos também foram descobertos [1].

Todos esses exemplares contribuem para o ciclo de vida no meio interestelar (ISM), onde explo-

sões de supernova ou estrelas no ramo assintótico das gigantes (AGB) liberam elementos sintetizados em seu núcleo, que mais tarde evoluíram para estruturas como galáxias, estrelas e planetas [2]. Dentre as diversas classes moleculares, verificou-se que vários objetos emitiam bandas espectrais até então não identificadas e que mais tarde foram atribuídas aos hidrocarbonetos aromáticos policíclicos (PAHs) [1].

Aproximadamente 20% ou mais do carbono existente no meio interestelar encontra-se na forma de PAHs, e grande parte da emissão no infravermelho médio (MIR), conhecida por AIBs (do inglês *Aromatic Infrared Bands*), deriva dessa classe molecular [3]. Além disso, PAHs podem ter alguns de seus átomos de carbono substituídos por nitrogênio ou oxigênio, dando origem a compostos complexos que desempenham importante participação na origem da vida.

Neste artigo será analisado o espectro de emissão no infravermelho (IR) do PAH antraceno, composto por três anéis aromáticos. Esta espécie foi escolhida por representar um PAH simples, o que facilita a análise de sua estrutura e processos numéricos posteriores. Além disso, pesquisas recentes inferem à presença de sua versão catiônica na Nuvem Molecular Perseus [4].

O espectro foi simulado utilizando o Gabedit¹, uma interface gráfica utilizada em pacotes de química computacional, e ORCA², um software disponível para acadêmicos e que apresenta métodos modernos de estrutura eletrônica. Dentre nossos objetivos estão entender algumas das propriedades que permitem a existência dessas moléculas em ambientes tão inóspitos, mas que podem dar origem aos tijolos da vida no Universo.

Uma vez que a região do IR fornece informações fundamentais a respeito de ambientes astrofísicos como a temperatura e a taxa de formação estelar [5], a compreensão da química desses compostos pode colaborar com o desenvolvimento de tecnologias voltadas para o meio ambiente, visto que em ambientes terrestres os PAHs são altamente poluentes.

2 Materiais e métodos

Neste trabalho foram utilizados dados obtidos por meio da modelagem computacional. A es-

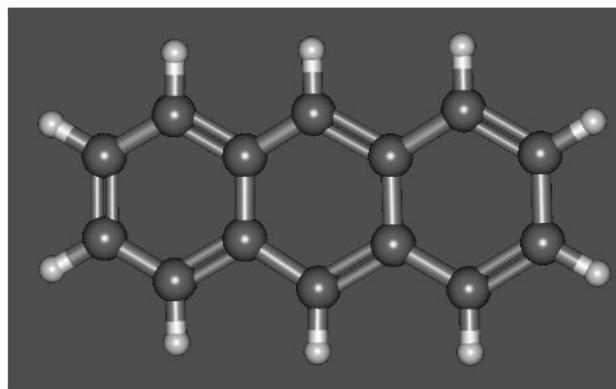


Figura 1: Molécula de Antraceno simulada na interface gráfica do software Gabedit. Fonte: Gabedit

colha deste método se deu pela necessidade de se detalhar comportamentos moleculares específicos e, tendo em vista que a resolução de métodos analíticos para um sistema multi-eletrônico seria inviável, os softwares de química computacional permitem que cálculos mais complexos sejam realizados.

A estrutura inicial da molécula foi desenhada no Gabedit. Este programa é uma interface gráfica de usuário para pacotes de química computacional. Exibe diversos resultados de cálculos e suporta grande parte dos formatos de arquivos moleculares. Também gera arquivos de entrada e permite desenhar e visualizar estruturas em três dimensões, bem como os espectros após a realização dos cálculos.

A realização dos cálculos e aproximações quânticas foi através do programa ORCA. Com fácil e eficiente uso, esta ferramenta permite aplicar códigos de química computacional, especificamente voltados para propriedades espectroscópicas de sistemas. Apresenta diversos métodos quânticos, dentre semi-empíricos e *ab initio*, bem como métodos híbridos que combinam parcelas dos métodos anteriores e que será o foco deste trabalho.

Ambos os programas são gratuitos, disponíveis para uso não-comercial. Finalizada a estrutura como mostra a Figura 1, é possível que ligações e ângulos estejam diferentes dos valores de equilíbrio, sendo necessário o uso de um campo de força para ajustar a geometria molecular, de acordo com a energia potencial estabelecida para os mesmos.

A Tabela 1 fornece os comandos utilizados para este estudo. A opção frequência é apresentada como tipo de trabalho, uma vez que o modo de vibração molecular corresponde ao espectro IR da

¹<http://gabedit.sourceforge.net/>

²<https://orcaforum.kofo.mpg.de/app.php/dlext/>

substância.

O cálculo de energia foi feito no estado neutro, uma vez que grande parte da contribuição da emissão em galáxias se dá por espécies neutras e também devido a otimização do tempo computacional. A escolha de uma função de base permite maior precisão no cálculo da estrutura eletrônica. As funções tipo Gaussianas (GTFs) do conjunto de bases de Pople foram escolhidas, uma vez que a literatura aponta vantagens em relação a outros conjuntos [6].

Ainda, é um dos conjuntos de bases utilizado pelo NASA Ames PAH IR Spectroscopic Database (PAHdb) [7–9], que será usado para fins de comparação com os resultados obtidos. Como método de trabalho foi escolhido um dos mais utilizados em química quântica, a Teoria do Funcional da Densidade, mais especificamente o funcional híbrido B3LYP. Este funcional combina dois ou mais funcionais, permitindo uma sensível melhora nas soluções apresentadas. Determinados os procedimentos de entrada para a modelagem, a ferramenta iniciará o processo.

2.1 Teoria do funcional da densidade

A mecânica quântica facilitou uma compreensão mais profunda da estrutura da matéria. Por meio da equação que descreve a função de onda do elétron é possível obter toda a informação necessária para determinar o estado de um sistema, seja átomo, molécula ou sólido. Porém, poucos sistemas físicos possuem solução analítica, e mesmo a solução numérica aproximada pode ser computacionalmente inviável, sendo necessárias aproximações que possibilitem o uso em sistemas moleculares.

Uma vez que a função de onda depende de 3N coordenadas (x, y e z de cada elétron), só é possível a resolução para o átomo de hidrogênio, que possui um único elétron. Para contornar esse problema, é necessário o uso de aproximações baseadas na densidade eletrônica ρ , já que essa variável não depende do número de elétrons e sim das coordenadas x, y e z .

O uso da densidade eletrônica foi proposto em 1927 por Thomas e Fermi. E, apesar deste modelo possuir limitações nas previsões de sistemas reais, precedeu o surgimento da moderna teoria usada até hoje. Por considerar uma distribuição homogênea de elétrons e negligenciar o termo de troca na interação entre elétrons, o método foi

aprimorado por Paul Dirac em 1930, ao adicionar um termo de troca [10].

O desenvolvimento formal da teoria do funcional da densidade (DFT) por Pierre Hohenberg e Walter Kohn ocorreu em 1964, utilizando a densidade eletrônica ρ como variável básica para calcular a equação de Schrödinger e propondo soluções para o impasse no modelo Thomas-Fermi por meio de dois teoremas [11].

O primeiro teorema afirma que o hamiltoniano e todas as propriedades resultantes podem ser obtidas a partir da densidade eletrônica. O segundo teorema mostra que o funcional da densidade também pode ser obtido pelo método variacional. Neste processo, o hamiltoniano correspondente pode ser determinado de acordo com a densidade aproximada eletrônica, enquanto a energia pode ser obtida através da equação de Schrödinger. As funções da densidade eletrônica que resultam no menor valor de energia serão as mais corretas.

No ano seguinte, Walter Kohn e Lu Jeu Sham demonstraram um conjunto de equações nas quais o sistema de partículas interagentes é substituído por um de partículas não-interagentes, submetidas a um potencial e reproduzindo as condições do sistema interagente. Essa mudança considera o fato de que em ambos, a densidade eletrônica será sempre a mesma. A equação geral de Kohn-Sham para a energia eletrônica do estado fundamental se torna

$$E[\rho] = \int V_{ext}(r)\rho(r)dr + F[\rho], \quad (1)$$

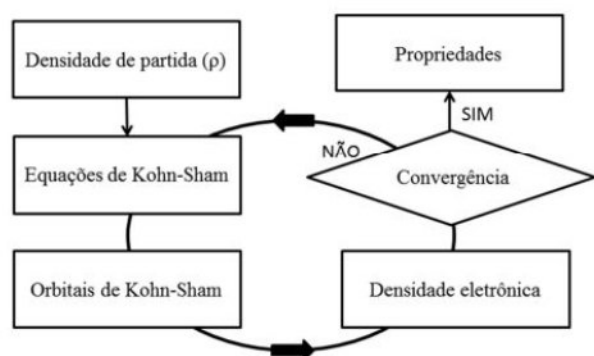
sendo V_{ext} o potencial externo e $F[\rho]$ o funcional da densidade independente do potencial externo. Conhecer a densidade permite o cálculo do potencial externo, do hamiltoniano e da função de onda, mas ainda é necessária uma função de onda para que ρ obtenha bons resultados. E para chegar na energia eletrônica de diversas contribuições, surge a proposta de utilizar funcionais aproximados,

$$E[\rho] = T[\rho] + V_{Ne}[\rho] + V_{ee}[\rho] + E^{XC}[\rho], \quad (2)$$

onde $T[\rho]$ é a energia cinética dos elétrons não-interagente, $V_{Ne}[\rho]$ o potencial externo, $V_{ee}[\rho]$ o potencial de interação elétron-elétron e $E^{XC}[\rho]$ a energia de troca e correlação. O último termo pode ser separado em um termo de troca $E^X[\rho]$ e um termo de correlação $E^C[\rho]$. O segundo

Tabela 1: Parâmetros de entrada para o cálculo realizado pelo ORCA.

ORCA input	
	Especificação de Trabalho
Tipo de Trabalho	Frequências
Tipo de Método	Funcionais Híbridos
Método	B3LYP
Estado excitado	Não
Bases	6-31G* / 6-31G**

**Figura 2:** Modelo do ciclo de cálculos do DFT. Fonte: [12]

teorema de Hohenberg-Kohn permite o cálculo de um conjunto de orbitais ϕ^{KS} , que minimizam $E[\rho]$, utilizando o operador de um elétron de Kohn-Sham $\hat{h}_i^{KS}$, equacionado como

$$\hat{h}_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_k^{\text{núcleos}} v_i \frac{Z_k}{|r_i - r_k|} + \int \frac{\rho(r')}{|r_j - r'|} dr' + V^{XC}, \quad (3)$$

em que $V^{XC} = \delta E^{XC}/\delta \rho$ é o potencial de troca e correlação, acessível caso se tenha valores da energia de troca e correlação E^{XC} . Após a resolução das equações de Kohn-Sham, os orbitais são calculados e tem-se uma nova densidade eletrônica [12], em um ciclo que é refeito tantas vezes for necessário até que os critérios iniciais sejam satisfeitos, como mostra a Figura 2.

2.2 Aproximações dos potenciais e funcionais híbridos

O potencial de troca e correlação E^{XC} não possui solução analítica e a utilização de aproximações é necessária. Dentre as mais conhecidas, estão a Aproximação da Densidade Local (LDA -

Local Density Approximation) e a Aproximação do Gradiente Generalizado (GGA - Generalized Gradient Approximation). Ambas as aproximações têm alta eficácia e baixo custo computacional, porém podem não ser adequadas em outras utilizações mais complexas. Neste caso, a literatura aponta para os *funcionais híbridos*, desenvolvidos a partir da combinação de funcionais para a energia de troca com um funcional para a energia de correlação de diferentes métodos.

Este trabalho utiliza o funcional híbrido B3LYP, também utilizado pelo NASA Ames PAH IR Spectroscopic Database (PAHdb) na modelagem do antraceno disponível na biblioteca [13,14]. Além disso, entre 2015-2020 o B3LYP aparece em aproximadamente 4.697 ocorrências dos 5.926 funcionais híbridos da categoria Físico-Química do Science Citation Index Expanded - SCIE, aproximadamente 80% [15]. O aspecto híbrido desse funcional é comprovado na energia de troca e correlação E_{xc} , descrito pela seguinte equação,

$$E_{xc} = (1 - a_0)E_x^{LDA} + a_0E_x^{HF} + a_xE_x^{B88} + a_cE_c^{LYP88} + (1 - a_c)E_c^{VWN80}, \quad (4)$$

onde os três primeiros termos são as aproximações LDA, HF e Becke 1988 (B88) para interação de troca, os dois últimos termos são aproximações Lee-Yang-Parr 1988 (LYP88) e Vosko-Wilks-Nusair 1980 (VWN80) para correlação eletrônica, e os coeficientes $a_0 = 0,2$, $a_x = 0,72$ e $a_c = 0,81$ são parâmetros definidos experimentalmente. Devido à combinação de diferentes parâmetros diferentes de cálculo, o método B3LYP permite a obtenção de resultados satisfatórios e que dependem das propriedades desejadas do objeto de estudo. Também existe uma variação desse funcional, o CAM-B3LYP, que possui resultados superiores aos do funcional B3LYP, porém

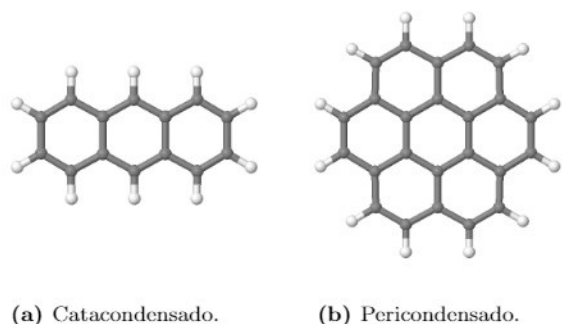


Figura 3: Exemplos de estruturas moleculares. Fonte: NASA Ames PAH IR Spectroscopic Database (PAHdb).

com um alto custo computacional. Por isso, não será abordado aqui, mas deixado como possibilidade de um novo estudo utilizando este método.

2.3 Hidrocarbonetos aromáticos policíclicos (PAHs)

Os PAHs são compostos orgânicos arranjados em anéis aromáticos, apresentam estruturas hexagonais ligadas à sua estabilidade e são classificados em pericondensados, catacondensados e aqueles com uma ligação simples entre carbonos. Como visto na Figura 3, enquanto os catacondensados são lineares, os pericondensados são núcleos compactos de carbono, o que garante mais estabilidade à molécula.

Vários objetos astronômicos como regiões HII e estrelas ricas em carbono possuem bandas de emissão atribuídas aos PAHs, bem como nebulosas planetárias e galáxias de núcleo ativo (AGNs). Isso infere que são materiais resistentes e abundantes no ISM [2]. A intensa radiação ultravioleta (UV) dessas regiões atinge essas moléculas, que ficam altamente excitadas, vibram e emitem no IR, durante o processo de desexcitação [16]. A Figura 4 apresenta espectros de pequenos PAHs, em que é possível perceber a semelhança nas posições de banda, uma vez que são construídos de uma mesma unidade básica.

Perto de 50% da luminosidade no MIR pode ser associada às bandas de emissão mais intensa dessas moléculas. Estas variam de acordo com o tamanho, a geometria molecular e são características dos modos de excitação vibracional, *bending* (dobramento) e *stretching* (estiramento) [17]. As características da emissão dos PAHs são bem conhecidas na literatura [16], no entanto este trabalho se dedica ao intervalo entre 5-15 μm , apre-

Tabela 2: Características de emissão IR interestelar e PAHs. Fonte: Adaptado de [16].

Bandas de emissão e modo de excitação	
5.2 μm	C-H flexão e C-C alongamento
5.65 μm	C-H flexão e C-C alongamento
6.2 μm	C-C alongamento
7.6 μm	C-C alongamento e C-H flexão
7.8 μm	C-C alongamento e C-H flexão
8.6 μm	C-H flexão
11.0 μm	C-H flexão
11.2 μm	C-H flexão
12.7 μm	C-H flexão
13.6 μm	C-H flexão
14.2 μm	C-H flexão

sentado na Tabela 2.

2.4 Base de dados do NASA Ames PAH IR

O PAHdb reúne espectros de simulações experimentais e computacionais de PAHs, na faixa espectral de 2,5 - 4000 μm . Contém dados de PAHs puros (carbono e hidrogênio) e espécies com nitrogênio, oxigênio ou magnésio, bem como em seus estados neutros e ionizados. Disponibiliza estruturas, fórmulas químicas, conversão de vibração, referências da literatura e outros atributos, quando disponíveis. Todas essas informações podem ser baixadas no módulo de tabela XML ou ASCII de uma única molécula ou um *cluster* molecular.

Além disso, fornece várias interfaces de programação de aplicativos (API, *Interface de Programação de Aplicativos*, em inglês), como AmesPAHdbIDLSuite, AmesPAHdbPythonSuite e uma versão simplificada do conjunto de ferramentas pyPAHdb [18]. Atualmente, contém 4.233 espécies teóricas na versão 3.20 e 84 espécies experimentais na versão 3.00. Além disso, será de grande importância nas futuras observações do James Webb Space Telescope (JWST).

Esta biblioteca online e de livre acesso será utilizada para compararmos os resultados obtidos com espectros disponíveis da mesma molécula. Também será possível inferir qual função de base melhor se adequa ao encontrado na literatura. Os

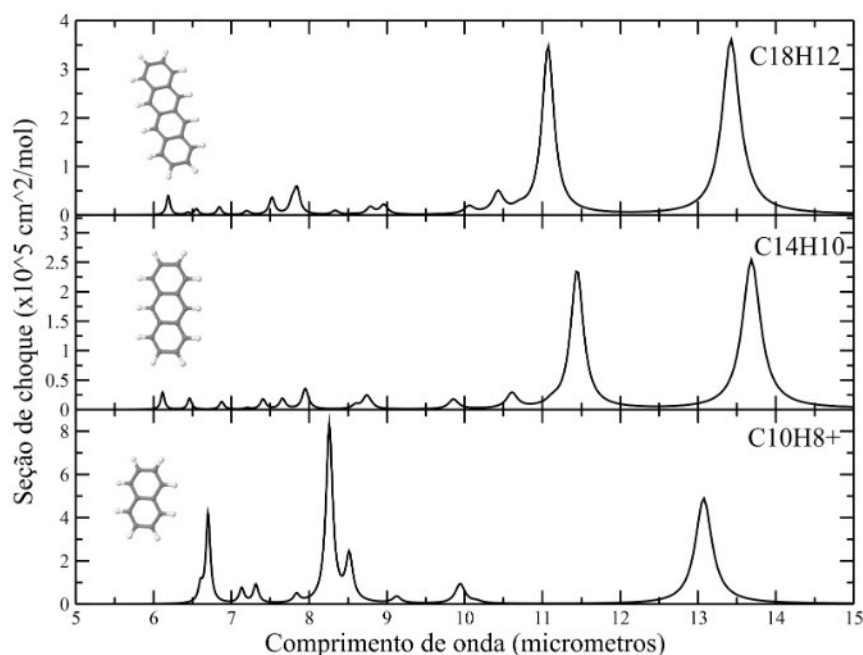


Figura 4: Espectro IR de PAHs. Fonte: NASA Ames PAH IR Spectroscopic Database (PAHdb).

resultados também serão comparados com dados espectroscópicos de galáxias, a fim de verificar a presença do PAH antraceno nessas regiões. Uma vez que encontra-se na literatura o uso desse tipo de modelagem para tal fim [19].

3 Resultados e discussões

A Figura 5 apresenta os espectros no IR do PAH antraceno, obtidos com as diferentes funções de base. É possível perceber os picos de emissão em regiões próximas às bandas de emissão associadas aos PAHs interestelares. Notamos leve diferença entre as posições e também na intensidade de bandas. A comparação entre os dados disponíveis na biblioteca de referência permite notar um avermelhamento nas bandas vibracionais, mesmo utilizando a função de base 6-31G*, descrita pelos autores do PAHdb [7–9].

A Figura 6 mostra os resultados obtidos pela nossa modelagem do antraceno e da modelagem realizada pelo PAHdb da mesma molécula. Podemos perceber que a modelagem realizada neste trabalho possui um desvio para o vermelho nas bandas de emissão, especialmente nos comprimentos de onda de 11,9 e 14,1 μm .

Uma possibilidade para essa ocorrência é o fato de existirem estados de excitação muito baixos na molécula em questão, o que em nosso trabalho foi

desconsiderado. Ainda, a aproximação realizada pelo B3LYP desconsidera a anarmonicidade, ou seja, reduz um sistema de 24 átomos, o antraceno em questão, a um oscilador harmônico.

A Figura 7 mostra o espectro PAHdb contraposto ao resultado obtido com a função de base 6-31G**. É possível perceber uma maior semelhança entre ambos, com bandas de emissão mais próximas, em 11,7 μm e 13,9 μm . Utilizar uma função mais complexa contribuiu para um melhor resultado.

Os métodos de cálculos numéricos realizam aproximações que permitem a resolução para sistemas complexos, como moléculas de PAH. Embora a espécie escolhida seja pequena e simples, comparada à complexidade das emissões interestelares, é necessária a inclusão de funções que descrevam propriedades moleculares e mantenham semelhança com dados experimentais já conhecidos.

Para isso, constatou-se a necessidade da inclusão de funções que permitam maior aproximação dos resultados [6]. Essas funções são classificadas em dois tipos principais: funções de polarização e funções difusas. As funções de polarização descrevem distorções na nuvem eletrônica. Já as funções difusas são indicadas para sistemas iônicos, pois descrevem a expansão da nuvem eletrônica desses sistemas.

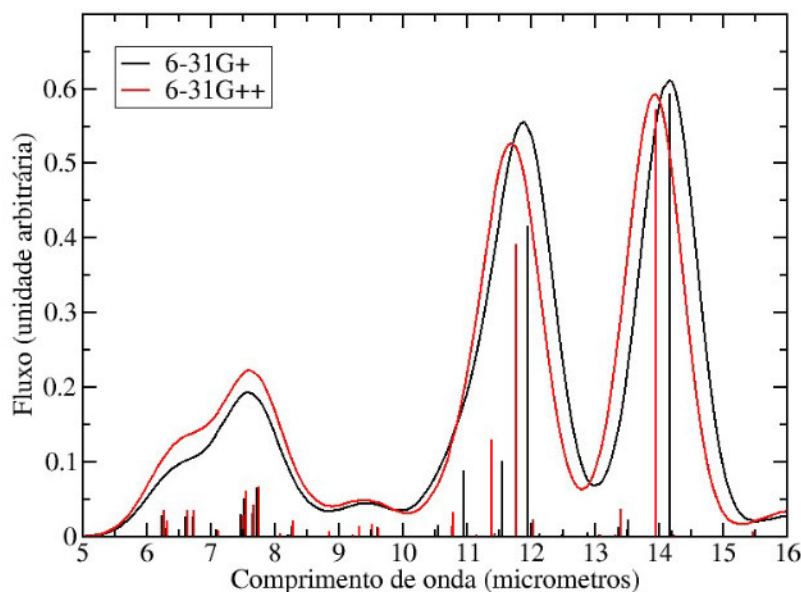


Figura 5: Espectros obtidos com as função de base 6-31G* e 6-31G**, utilizando FWHM = 5. Fonte: Gabedit

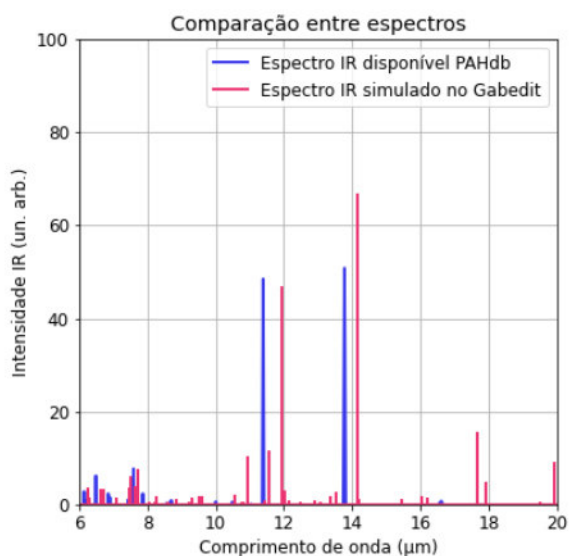


Figura 6: Comparação entre espectro PAHdb e simulado com a função de base 6-31G*.

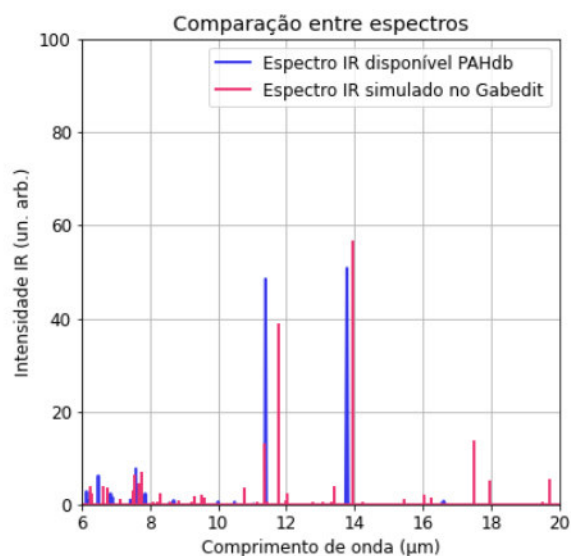


Figura 7: Comparação entre espectro PAHdb e simulado com a função de base 6-31G**.

Os dois conjuntos de base utilizados neste trabalho foram sugeridos por Pople e colaboradores [6], e os símbolos (* e **) indicam a inclusão de funções de polarização somente para átomos diferentes de hidrogênio ou a inclusão em todos os átomos do sistema, respectivamente. Esse fato justifica os resultados obtidos, uma vez que a espécie escolhida possui 10 átomos de hidrogênio que foram desconsiderados na primeira simula-

ção. Já o segundo conjunto, 6-31G**, aplica funções de polarização em todos os átomos do sistema, contribuindo para um melhor resultado.

A Figura 8 apresenta os resultados obtidos em paralelo com dados espectroscópicos de galáxias do tipo Starburst, Seyfert 1 e Seyfert 2, disponíveis na amostra do Projeto Spitzer/ATLAS [5]. Percebe-se que as emissões dos espectros simulados podem prever as bandas de PAH presentes

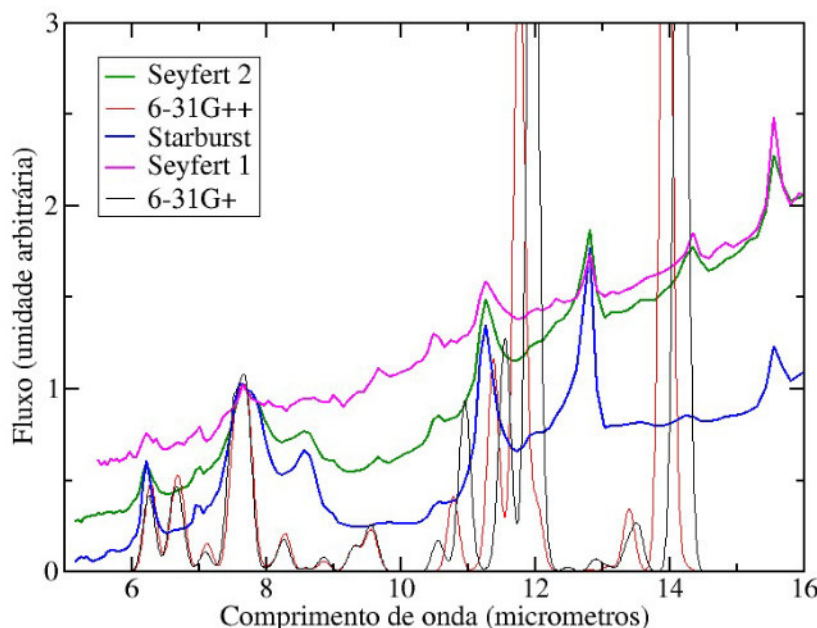


Figura 8: Análise dos resultados com dados espectroscópicos de galáxias Seyfert 1, Seyfert 2 e Starburst.

nessas galáxias, em comprimentos de onda mais azuis entre 6 e 10 μm . Essa característica em galáxias ativas permite estimar, ainda que indiretamente, a taxa de formação estelar [5].

É necessário citar que o funcional híbrido utilizado combina dois métodos de cálculo – uma parcela mais adequada para moléculas grandes (DFT), mas com resultados pouco confiáveis; e uma parcela complementar, indicada para moléculas pequenas (*ab initio*), mas que fornece mais precisão. Isso garante um resultado mais próximo ao experimental. O uso de diferentes funções de base altera importantes propriedades moleculares, o que possibilita analisar qual é mais adequada ao sistema proposto, sendo, neste caso, o PAH antraceno.

Os resultados obtidos através do uso da DFT, com dois conjuntos de base demonstraram-se satisfatórios. Uma vez que a modelagem computacional da molécula antraceno identificou contribuições em comprimentos de onda mais azuis no intervalo de 6 a 10 microns, as mesmas observadas em galáxias ativas. Nesse sentido podemos concluir que essa molécula pode estar presente em objetos extragalácticos. Além disso mudanças nos parâmetros de cálculo mostrou que ferramentas computacionais possibilitam a resolução de sistemas complexos como os PAHs.

4 Conclusões

A astroquímica vem se expandido nas últimas décadas devido à grande quantidade de dados disponíveis e ao avanço nos métodos de cálculo. Dentre as diversas classes moleculares existentes no ISM, os PAHs possuem grande importância para se entender a química de ambientes astrofísicos. Os PAHs são responsáveis por grande parte da luminosidade no MIR, pois eles aquecem ao absorver um fóton de radiação UV e, durante o processo de desexcitação, emitem em IR. Devido a sua resistência, compõem o material orgânico predominante do espaço e podem combinar-se com nitrogênio, tornando-se PANHs (heterocíclicos policíclicos aromáticos nitrogenados), que podem exercer importante papel na origem da vida no Universo.

A simulação realizada permitiu identificar as bandas de emissão predominantes do antraceno, bem como inferir as características nas frequências vibracionais encontradas ao se utilizar funções de base distintas, o que mostra a relação existente entre a complexidade e a eficiência do método. Nossos resultados foram bastante satisfatórios, uma vez que foi possível obter informações semelhantes ao espectro de referência. Além da identificação de bandas de emissão em intervalos observados em galáxias ativas.

Ainda, métodos computacionais permitem

aproximar os estudantes de processos complexos, como o espectro do PAH antraceno, sem a necessidade de supercomputadores. Como perspectiva de continuação para este trabalho, pretendemos simular moléculas maiores do que o disponível na literatura, uma vez que em galáxias ativas podemos encontrar espécies de até 1000 átomos de carbono [17], enquanto o maior PAH na amostra do PAHdb possui em torno de 300 átomos de carbono.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. D. A. Sales agradece ao CNPq e FAPERGS.

Sobre as autoras

Rosiane Carneiro da Rosa (rosianecarneiroda-rosa@outlook.com) é Graduada em Física pelo Instituto Federal de Educação, Ciência e Tecnologia Farroupilha - Campus São Borja (2019). Mestranda em Física na Universidade Federal do Rio Grande e possui foco de suas pesquisas na área de astroquímica, atuando principalmente nas pesquisas em moléculas complexas e simulações computacionais.

Dinalva Aires de Sales (dinalvaires@gmail.com) é Graduada em Matemática pela Universidade do Vale do Paraíba (2005), Mestre em Física e Astronomia pela Universidade do Vale do Paraíba (2007) e Doutora em Física pela Universidade Federal do Rio Grande do Sul (2012). Foi NASA Postdoctoral Fellow no Rochester Institute of Technology (2012/2013) e pós-doutoranda da DOCFIX-FAPERGS/CAPES na Universidade Federal do Rio Grande do Sul (2013/2016). Tem experiência na área de Astrofísica, atuando principalmente nos seguintes temas: núcleo ativo de galáxia, galáxias em interação, moléculas complexas, espectroscopia e imageamento. Atualmente é professora adjunta do Instituto de Matemática, Estatística e Física (IMEF) da Universidade Federal do Rio Grande (IMEF-FURG) e Coordenadora do Programa de Pós-Graduação em Física (PPG-Física) do IMEF-FURG.

Carla Martinez Canelo (camcanelo@gmail.com) é Graduada em Física pela Universidade de São Paulo (2013). Também é Mestra (2016) e Doutora (2020) em Astronomia pela mesma instituição. Possui experiência na área de Astronomia, com ênfase em Astrobiologia e Astroquímica, e em divulgação científica.

Brenda Matoso Abreu Miranda (brendamatoso@live.com) é Graduada em Física pelo Instituto Federal de Educação, Ciência e Tecnologia Farroupilha - Campus São Borja (2017). Especialista em Docência do Ensino Superior pela Universidade Anhanguera (2019) e Mestra em Física pela Universidade Federal do Rio Grande - FURG (2020).

Referências

- [1] M. D. Becker, *Astrochemistry: the issue of molecular complexity in astrophysical environments*, Bulletin de la Société Royale des Sciences de Liège **82**, 33 (2013). [ArXiv: 1305.6243](https://arxiv.org/abs/1305.6243).
- [2] A. G. G. M. Tielens, *The molecular universe*, *Rev. Mod. Phys.* **85**, 1021 (2013).
- [3] C. M. Canelo, *O mundo aromático – dos PAHs no meio interestelar às condições bióticas*, Dissertação de Mestrado, IAG - Universidade de São Paulo, São Paulo (2016). Disponível em <https://www.iag.usp.br/pos/>, acesso em jul. 2021.
- [4] S. Iglesias-Groth, A. Manchado et al., *A search for interstellar anthracene towards the Perseus anomalous microwave emission region*, *Monthly Notices of the Royal Astronomical Society* **407**(4), 2157 (2010).
- [5] A. Hernan-Caballero e E. Hatziminaoglou, *An atlas of mid-infrared spectra of star-forming and active galaxies*, *Mon. Not. Roy. Astron. Soc.* **414**, 500 (2011). [ArXiv: 1101.4794](https://arxiv.org/abs/1101.4794).
- [6] N. H. Morgon e R. Custódio, *Funções de base: o ajuste variacional*, *Revista Chemkeys* (2), 1 (2018).
- [7] J. Bauschlicher, Charles W., A. Ricca et al., *The NASA Ames PAH IR Spectroscopic Database: Computational version 3.00 with up-*

- dated content and the introduction of multiple scaling factors*, *The Astrophysical Journal Supplement Series* **234**(2), 32 (2018).
- [8] C. Boersma, C. W. Bauschlicher et al., *The NASA Ames PAH IR Spectroscopic Database version 2.00: Updated content, web site, and on(off)line tools*, *The Astrophysical Journal Supplement Series* **211**(1), 8 (2014).
- [9] A. L. Mattioda, D. M. Hudgins et al., *The NASA Ames PAH IR Spectroscopic Database: The laboratory spectra*, *The Astrophysical Journal Supplement Series* **251**(2), 22 (2020).
- [10] P. A. M. Dirac e R. H. Fowler, *Quantum mechanics of many-electron systems*, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **123**(792), 714 (1929).
- [11] P. Hohenberg e W. Kohn, *Inhomogeneous Electron Gas*, *Phys. Rev.* **136**, B864 (1964).
- [12] R. T. Santiago, *Novas parametrizações de funcionais híbridos para uso em cálculos relativísticos*, Dissertação de Mestrado, Universidade de São Paulo, São Paulo (2014). Disponível em <https://doi.org/10.11606/D.75.2014.tde-27012015-093530>, acesso em jul. 2021.
- [13] A. Mattioda, C. Bauschlicher et al., *Infrared spectroscopy of matrix-isolated neutral polycyclic aromatic nitrogen heterocycles: The acridine series*, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **181**, 286 (2017).
- [14] S. Langhoff, *Theoretical infrared spectra for polycyclic aromatic hydrocarbon neutrals, cations and anions*, *The Journal of Physical Chemistry* **100**(8), 2819 (1996).
- [15] *Science citation index expanded - scie*. Disponível em <https://www.webofscience.com>, acesso em jul. 2021.
- [16] A. G. G. M. Tielens, *The Physics and Chemistry of the Interstellar Medium* (Cambridge University Press, 2005).
- [17] D. A. Sales, M. G. Pastoriza et al., *Polycyclic aromatic hydrocarbon in the central region of the Seyfert 2 galaxy NGC1808*, *Mon. Not. Roy. Astron. Soc.* **429**(3), 2634 (2013). [ArXiv:1212.1357](https://arxiv.org/abs/1212.1357).
- [18] Matthew J. Shannon e Christiaan Boersma, *Organic molecules in space: Insights from the NASA Ames Molecular Database in the era of the James Webb Space Telescope*, in *Proceedings of the 17th Python in Science Conference*, editado por Fatih Akici, David Lippa et al. (2018), 99 – 105.
- [19] B. M. A. Miranda, D. A. Sales et al., *Estudo da emissão do benzeno em galáxias usando a teoria do funcional de densidade*, *Rev. Mundi Eng. Tecnol. Gest.* **4**, 136 (2019).