

Construção de um banco de dados de propriedades físicas de galáxias observadas no ultravioleta

André Luiz de Amorim

Orientador:

Prof. Dr. Roberto Cid Fernandes Jr.

Coorientador:

Prof. Dr. Abilio Mateus Jr.

•••

Universidade Federal de Santa Catarina
Centro de Ciências Físicas e Matemáticas
Curso de Pós-Graduação em Física

Dissertação de mestrado apresentada ao Curso de Pós-Graduação em Física da UFSC em preenchimento parcial dos requisitos para obtenção do título de Mestre em Física.

Florianópolis (SC) – 1 de março de 2012

A meus pais e avós.

Agradecimentos

Este trabalho só foi possível graças à insistência do meu orientador Prof. Roberto Cid, sem o qual eu provavelmente não teria voltado para concluir o mestrado. Agradeço profundamente a ele por acolher um filho pródigo do Grupo de Astrofísica. Agradeço ao meu coorientador Prof. Abilio Mateus, por me guiar na etapa final do meu trabalho. Agradeço também ao Prof. Antônio Kanaan, que foi meu orientador na iniciação científica e na minha primeira tentativa no mestrado, e cujas influências carrego até hoje.

Agradeço ao Time da Gaivota, por desenvolverem o STARLIGHT, alicerce onde este trabalho se apoia; à Dra. Natalia, a principal “bug catcher” das tabelas do STARLIGHT, pelas várias ideias e sugestões; ao William Schoenell, à Carol Deggerone e ao Eduardo Lacerda, pelo apoio e companheirismo; aos colegas do Laboratório de Astrofísica, e em especial aos membros do “Comitê de Insalubridade do Grupo de Astrofísica”.

Sou grato ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, pelo apoio financeiro; ao Curso de Pós-Graduação em Física, pela estrutura e instalações adequadas; aos professores e funcionários – especialmente ao Chefe de Expediente do PGFSC, Antônio Machado, sempre prestativo para resolver as questões burocráticas durante a minha passagem pelo curso; às equipes do *Sloan Digital Sky Survey*, *SDSS* e do *Galaxy Evolution Explorer*, *GALEX*, por proverem os dados e as ferramentas sem os quais este trabalho não seria possível.

Agradeço à minha família pelo carinho e dedicação, eu não seria nada sem eles. E por fim, à Marilize, por me incentivar e sempre estar do meu lado, me aturando nos momentos difíceis desta jornada.

Resumo

Megalevantamentos de dados astronômicos estão gerando um volume imenso de dados, e com isso surge a necessidade de desenvolver formas de armazenamento e técnicas de mineração de dados. Neste trabalho explora-se o uso de bancos de dados relacionais (RDBMS) para facilitar o acesso e o gerenciamento dos dados gerados por um levantamento de dados de grande porte. Em particular, descreve-se a importação para um banco de dados do catálogo de propriedades físicas de galáxias obtido através do código STARLIGHT. O STARLIGHT é um código de síntese espectral, que tem como entrada uma base de populações estelares simples e espectros de quase 1 000 000 de galáxias do *Sloan Digital Sky Survey* (SDSS), e gera como saída informações relacionadas ao histórico de formação estelar das galáxias. Baseada inteiramente no espectro óptico, esta análise serve de guia para interpretar dados em outras faixas espectrais para galáxias de diferentes tipos (como as que estão formando estrelas, as galáxias ativas de diferentes classes, as passivas e as aposentadas). O atual *data release* do *Galaxy Evolution Explorer* (GALEX GR6) inclui a identificação cruzada de seus objetos com os objetos do SDSS. Isso permite correlacionar as informações resultantes do STARLIGHT com dados de fotometria no ultravioleta. Uma amostra de galáxias limitada em *redshift* é analisada com base no diagrama de classificação [N II]/H α (WHAN), avaliando as suas propriedades físicas em relação às cores em ultravioleta.

Abstract

Astronomical mega-surveys are creating a large volume of data, creating new challenges regarding data management and archiving. Data mining techniques are needed to make use of this data deluge. The use of Relational Database Management Systems (RDBMS) for management and data access in large surveys is explored in this study. The STARLIGHT catalog is imported into a RDBMS, using *CasJobs* for data access. These data are produced feeding about 1 000 000 galaxy spectra from Sloan Digital Sky Survey (*SDSS*) and a set of simple stellar populations to STARLIGHT, yielding a catalog of physical properties of those galaxies. This analysis is based entirely on data from the optical spectrum, serving as a benchmark for studies in other wavelengths. The Galaxy Evolution Explorer (*GALEX*) latest data release (GR6) includes a crossmatch between GALEX and SDSS catalogs. Data from STARLIGHT are then matched to GALEX, adding ultraviolet photometry to the STARLIGHT catalog. As a case study, a redshift-limited sample is taken from this new catalog, galaxies are classified based on the [N II]/H α (WHAN) diagram, and the physical properties for each class are analyzed against their ultraviolet colours.

Sumário

1	Introdução	1
1.1	A nova era da astronomia	1
1.1.1	Megalevantamentos de dados	2
1.1.2	Mineração de dados	3
1.2	O catálogo de propriedades físicas do <i>STARLIGHT</i>	3
1.3	Este trabalho	6
1.3.1	Organização deste trabalho	6
2	O <i>Galaxy Evolution Explorer (GALEX)</i>	9
2.1	Objetivos do <i>GALEX</i>	9
2.2	Histórico do estudo do céu no UV	12
2.3	Resultados obtidos pelo <i>GALEX</i>	13
2.4	Data releases e banco de dados	14
3	Estendendo a base de dados <i>SDSS/STARLIGHT</i> para o ultravioleta	17
3.1	Banco de dados do <i>SDSS</i>	17
3.1.1	Migração de OODBMS para RDBMS	17
3.1.2	<i>SkyServer</i>	18
3.1.3	<i>CasJobs</i>	19

3.2	Banco de dados do <i>STARLIGHT</i>	20
3.2.1	Importação para o RDBMS	21
3.2.2	Estrutura do banco de dados	21
3.2.3	Amostra do <i>STARLIGHT</i>	22
3.3	Identificação cruzada entre <i>SDSS</i> e <i>GALEX</i>	24
3.4	Obtendo dados UV para a amostra <i>STARLIGHT</i>	25
3.4.1	Relação de <i>crossmatch</i> entre <i>SDSS</i> e <i>GALEX</i>	25
3.4.2	Dados UV	25
3.4.3	Correções aplicadas à fotometria UV	27
3.5	Definição da amostra <i>STARLIGHT+UV</i>	29
4	Análise da amostra <i>STARLIGHT+UV</i>	33
4.1	Diagrama cor–cor	33
4.2	Classificação das galáxias	36
4.3	Propriedades físicas no diagrama cor–cor	42
5	Conclusões e perspectivas	51
5.1	Este trabalho	51
5.2	Trabalhos futuros	53
A	<i>Queries SQL</i>	55
	Referências Bibliográficas	I

Lista de Figuras

1.1	Exemplos de ajuste de espectro com o STARLIGHT.	5
2.1	Curvas de transmissão dos filtros do GALEX.	10
2.2	Footprint dos surveys GALEX AIS, MIS e SDSS	12
2.3	Diagrama cor–magnitude em ultravioleta.	13
2.4	Tela do programaGalexView.	15
3.1	Telas do SkyServer.	18
3.2	Tela do CasJobs.	19
3.3	Esquema do banco de dados do STARLIGHT.	23
3.4	Magnitude M_r em função do <i>redshift</i>	30
3.5	Histogramas das medidas da amostra STARLIGHT+UV.	31
4.1	Frequência de galáxias no diagrama cor–cor.	34
4.2	Diagrama cor–cor para os diversos parâmetros STARLIGHT.	35
4.3	Diagrama de diagnóstico WHAN.	37
4.4	Cores UV no diagrama WHAN.	37
4.5	Cores UV no diagrama BPT.	38
4.6	Diagrama cor–cor UV de acordo com o tipo de galáxia.	39
4.7	Histogramas de cores para as classes de galáxias.	39

4.8	Histogramas de cores em função de $W_{H\alpha}$	40
4.9	Cor UV em função de $W_{H\alpha}$	41
4.10	Idade estelar média ponderada em fluxo no diagrama cor–cor.	43
4.11	Idade estelar média ponderada em massa no diagrama cor–cor.	44
4.12	Metalicidade estelar média ponderada em fluxo no diagrama cor–cor.	45
4.13	Metalicidade estelar média ponderada em massa no diagrama cor–cor.	46
4.14	Massa estelar das galáxias no diagrama cor–cor.	47
4.15	Atenuação por poeira no diagrama cor–cor.	48
5.1	Localização dos acessos ao <i>CasJobs</i> do STARLIGHT.	52
A.1	<i>Query</i> para atualizar os índices da amostra de galáxias do STARLIGHT.	55
A.2	<i>Query</i> listando identificações entre <i>SDSS</i> e <i>surveys</i> do <i>GALEX</i>	56
A.3	<i>Match</i> entre os objetos da amostra do STARLIGHT e <i>GALEX</i>	57
A.4	<i>Query</i> para listar as detecções por banda UV e <i>survey</i>	58
A.5	Extração da amostra STARLIGHT+UV.	59

Lista de Tabelas

2.1	<i>Surveys</i> realizados pelo <i>GALEX</i>	11
2.2	Campos observados em cada <i>General Release</i> do <i>GALEX</i>	15
3.1	Identificações entre <i>SDSS DR7</i> e <i>survey AIS</i> do <i>GALEX GR6</i>	24
3.2	Identificações entre <i>SDSS DR7</i> e <i>survey AIS</i> do <i>GALEX GR6</i>	25
3.3	Descrição dos campos da tabela \times <i>SDSSDR7</i>	26
3.4	Número de detecções por banda no catálogo UV do <i>STARLIGHT</i>	27
3.5	Propriedades físicas das galáxias utilizados na amostra.	29

Capítulo 1

Introdução

1.1 A nova era da astronomia

Os primeiros *surveys*¹ espectroscópicos da astronomia extragaláctica começaram a ser feitos de forma sistemática na década de 80 (Huchra 1988). Foram observadas cerca de 2400 galáxias pelo *Center for Astrophysics* em Harvard (Huchra et al. 1983) e aproximadamente 2000 galáxias pelo *Southern Redshift Survey* (da Costa et al. 1988). Estes *surveys* pioneiros, assim como alguns dos atuais, visavam primariamente medir *redshifts* para estudos da estrutura em larga escala do Universo.

Nos últimos anos, com o advento dos *mega-surveys*, está ocorrendo uma revolução na forma de fazer ciência na astronomia. Os diversos *surveys* em execução atualmente estão produzindo dados a uma taxa da ordem de petabytes por ano. Este talvez seja o primeiro campo da ciência onde as informações coletadas por máquinas tenham – nas próximas décadas – um volume maior do que os seres humanos são capazes de digerir. Uma espécie de *Singularidade Tecnológica*² da astrofísica, onde máquinas coletam, analisam e classificam dados. O papel do cientista num cenário como este ainda não está muito claro (Norris 2010).

¹Um *survey* astronômico é um levantamento de informações ou mapeamento de regiões do céu utilizando telescópios e detetores.

²O termo *Singularidade Tecnológica* se refere a um futuro hipotético onde uma inteligência superior à humana emerge através da tecnologia. Qualquer previsão após tal fato se torna muito difícil, algo similar a um horizonte de eventos, dada a dificuldade em entender uma inteligência superior à humana.

1.1.1 Megalevantamentos de dados

O *Sloan Digital Sky Survey* (*SDSS*; York et al. 2000) é referência quando falamos em *surveys* modernos. Em seus 8 anos de funcionamento, obteve imagens em 5 filtros (ver Figura 2.1) de um quarto do céu e espectros de um milhão de galáxias. O seu catálogo contém 4 terabytes de dados fotométricos e espectroscópicos, e 12 terabytes de imagens. O *SDSS* foi praticamente o primeiro *survey* a conseguir popularizar o acesso aos seus dados. Estes foram feitos públicos desde o princípio, iniciando uma “corrida do ouro” no seu vasto volume de dados. Esta filosofia é compartilhada hoje pela grande maioria dos *surveys* de grande porte.

Outros *surveys* pioneiros nesta era de mega-levantamentos são o *2dF Galaxy Redshift Survey* (2dFGRS; Colless 1999) e o *Two Micron All Sky Survey* (2MASS; Skrutskie et al. 2006). Embora estes *surveys* já tenham sido concluídos, os seus dados permanecem disponíveis publicamente. Outro exemplo é o arquivo do *Hubble Space Telescope*, onde, embora não sendo um *survey* estritamente falando, todas as observações são disponibilizadas após um ano.

Existem diversos *surveys* em operação atualmente. O *Wide-field Infrared Survey Explorer* (WISE) é um telescópio espacial da NASA³ que está mapeando o céu inteiro nas faixas de 3,4, 4,6, 12 e 22 μm do infravermelho (Wright et al. 2010). O *Visible and Infrared Survey Telescope for Astronomy* (VISTA) é um telescópio no Chile fazendo um *survey* do céu do hemisfério sul no infravermelho próximo (Born 2010). O *Kepler* é um telescópio espacial da NASA (Borucki et al. 2010), semelhante ao WISE, e está fazendo um *survey* de uma região da Via Láctea para descobrir a fração de estrelas com planetas similares à Terra na nossa Galáxia. Tratado com mais detalhes no capítulo 2, o *Galaxy Evolution Explorer* (*GALEX*) está mapeando o céu em ultravioleta.

O projeto J-PAS (*Javalambre Physics of the Accelerating Universe Astrophysical Survey*) é um *survey* que pretende imagear um quinto do céu em 56 cores (Benítez et al. 2009). Os filtros, de banda estreita, irão cobrir toda a região óptica, formando um espectro de baixa resolução para cada pixel do *survey*. Serão mais de 200 terabytes de dados brutos (imagens). O projeto é uma colaboração entre Espanha e Brasil, com mais de 70 pesquisadores e engenheiros envolvidos, incluindo integrantes do Grupo de Astrofísica da UFSC. O objetivo do *survey* é a exploração das causas da aceleração do universo, relacionadas à energia escura. Entretanto, uma quantidade considerável de ciência adicional poderá ser feita com base no espectro de uma região tão ampla do céu. Seu início está previsto para 2013.

O LSST (*Large Synoptic Survey Telescope*) mapeará metade do céu a cada mês, aproxima-

³NASA Explorer Mission - <http://explorers.gsfc.nasa.gov/missions.html>

damente, durante cerca de dez anos (Ivezic et al. 2008). Com este “filme”, em seis bandas espectrais similares às do *SDSS*, espera-se fazer um inventário do objetos do Sistema Solar e explorar os eventos transientes como Novas, núcleos ativos de galáxias, surtos de raios gama, para citar alguns exemplos. Este *survey* também pretende explorar a natureza da energia escura, embora, da mesma forma que o J-PAS, os dados possam ser aproveitados para diversos outros fins. As suas operações científicas têm início previsto para 2020. Serão mais de um petabyte em imagens brutas por ano, muito mais do que poderia ser revisado por humanos.

1.1.2 Mineração de dados

Tradicionalmente astrônomos armazenam seus dados em arquivos texto ou binários contendo um registro por linha, de uma forma tecnicamente conhecida como *flat file*. Buscas neste tipo de banco de dados são feitas examinando individualmente cada registro do arquivo. Com o volume de dados obtido, por exemplo, pelo *SDSS* ou pelo *GALEX* (aproximadamente 222 milhões de objetos, 34 mil campos), o uso de arquivos simples para armazenamento de dados se torna inviável. É preciso “profissionalizar” o gerenciamento de dados de um *survey* desta escala e aprender a trabalhar com volumes imensos de dados (Bell et al. 2006).

Técnicas de detecção de padrões em grandes conjuntos de dados, conhecidas como *Mineração de dados*, tiveram um interesse crescente nos últimos anos (Apte 2011). Estas técnicas têm naturalmente um caráter interdisciplinar, indo desde aplicações científicas até áreas como Economia (Spanos 2000) e Inteligência de Mercado⁴.

A equipe de arquivamento do *SDSS* desenvolveu uma ferramenta chamada *CasJobs* (ver Seção 3.1.3), que permite que múltiplos usuários possam fazer buscas e coleta de dados em seu banco de dados. Embora o *CasJobs* não seja, estritamente falando, uma ferramenta de mineração de dados, ele facilita, e muito, o acesso aos dados de um *survey*. O presente trabalho se baseia fortemente no uso desta ferramenta.

1.2 O catálogo de propriedades físicas do STARLIGHT

O STARLIGHT é um código de síntese espectral desenvolvido por Cid Fernandes et al. (2005). Dado um espectro de galáxia, o programa retorna as frações de massa e luz correspondentes às populações estelares componentes desta galáxia. O STARLIGHT expressa o espectro desta

⁴Ferramentas como OLAP (*On-line Analytical Processing*) são usadas por empresas para avaliação de mercado e tomada de decisão, agindo sobre um grande volume de dados.

galáxia como uma combinação linear de espectros de populações estelares simples⁵ (SSP, do inglês *Simple Stellar Population*) com diferentes idades e metalicidades. Matematicamente isto é equivalente a encontrar as componentes do vetor espectro da galáxia numa base de espectros de SSP, ou seja,

$$F_\lambda = \sum_{i=1}^{N_*} x_i F_\lambda^*(t_i, Z_i) g_\lambda(A_{V_i}).$$

Nesta equação, F_λ é o fluxo em cada comprimento de onda. Os N_* espectros de base $F_\lambda^*(t_i, Z_i)$, com t_i e Z_i representando respectivamente a idade e a metalicidade do elemento de base, são somados com pesos x_i . O conjunto $\{x_i, i = 1, 2, \dots, N_*\}$ é chamado *vetor de população* para a galáxia considerada. O termo g_λ corrige o espectro pelo efeito de atenuação interestelar. Os espectros da base provêm de modelos de síntese evolutiva de populações estelares. A maior parte dos resultados publicados do STARLIGHT usa os modelos de Bruzual & Charlot (2003, BC03), do quais foram extraídas SSPs cobrindo uma ampla gama de idades e metalicidades. São ao todo 150 SSP diferentes (25 idades e 6 metalicidades) utilizadas na base. O problema torna-se então um ajuste num espaço de parâmetros bastante grande. A Figura 1.1 mostra o ajuste feito para uma galáxia do SDSS.

Analisando o vetor de população das galáxias é possível obter algumas de suas propriedades físicas. É possível também extrair as medidas das linhas de emissão – não modeladas no ajuste – com bastante precisão. A técnica foi aplicada aos espectros de galáxias do SDSS e o resultado da síntese gerou um catálogo de propriedades físicas e linhas de emissão de quase um milhão de galáxias. Disponibilizado em <http://www.starlight.ufsc.br/>, este banco de dados já foi utilizado por dezenas de pesquisadores em todo o mundo, resultando em vários artigos (Bian et al. 2006; Liang et al. 2007; Peebles et al. 2009; Lara-López et al. 2009, 2010, para citar alguns) e teses (Mateus 2006; Gomes 2009; Asari 2010).

⁵Uma SSP consiste num conjunto de estrelas formadas ao mesmo tempo com a mesma metalicidade.

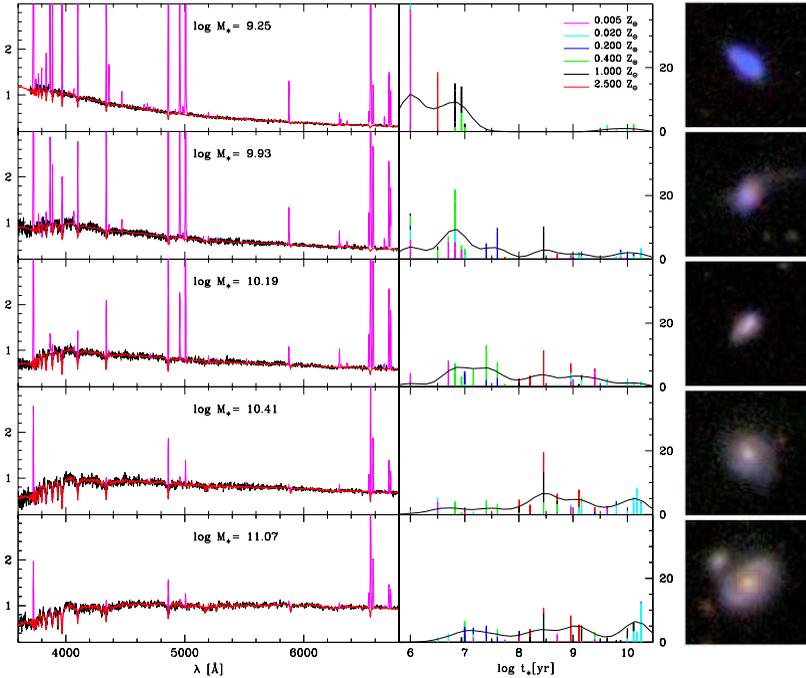


Figura 1.1: Exemplos de ajuste de espectros de galáxias com o STARLIGHT (Asari et al. 2007). À esquerda, espectros observados (preto), espectros modelados (vermelho), com regiões mascaradas em magenta. No meio, a fração da luz associada à cada uma das 25 idades das SSPs usadas na síntese, com a curva representando a versão suavizada do histórico de formação estelar. À direita, imagens do SDSS correspondentes às galáxias.

1.3 Este trabalho

Historicamente a região ultravioleta (UV) do espectro eletromagnético tem sido pouco estudada na astronomia. Não por falta de interesse dos astrônomos, mas pelo simples fato de que observações feitas de dentro da atmosfera são impossibilitadas pela camada de ozônio. É preciso ir para o espaço. Foram lançados diversos satélites com o intuito de estudar o céu em UV (ver Seção 2.2). O *GALEX* foi o primeiro telescópio espacial a fazer um *survey* do céu inteiro em UV. A missão tem como objetivo estudar a evolução da taxa de formação estelar em galáxias (Martín et al. 2005).

Embora este e outros grandes *surveys* tenham um grande valor individualmente, esta é somente a ponta do *iceberg*. Cruzando informações de *surveys* em diversos comprimentos de onda pode-se ter uma nova percepção da relação entre os processos físicos subjacentes. Este trabalho utiliza dados obtidos dos *surveys* *SDSS* e *GALEX*, nas bandas espectrais óptica e UV, respectivamente. Estes dois *surveys* foram indexados (ver Seção 3.3) de forma a facilitar a identificação entre as suas detecções. A técnica desenvolvida por eles pode ser aplicada em virtualmente qualquer outro *survey* astronômico.

As galáxias utilizadas na síntese do *STARLIGHT* foram obtidas do catálogo do *SDSS*. Neste trabalho foi feita a identificação destas galáxias no catálogo do *GALEX*, obtendo-se assim as suas medidas de fotometria UV. Este conjunto de dados formou uma amostra de galáxias com as suas propriedades físicas, cores ópticas e cores UV. As galáxias desta amostra são classificadas utilizando o diagrama de diagnóstico *WHAN*, conforme Cid Fernandes et al. (2011). Este método de classificação utiliza as linhas de emissão $H\alpha$ e $[N II] \lambda 6584$ para separar as galáxias nas classes de formação estelar, núcleo ativo, “aposentadas” e passivas. As cores UV das galáxias de cada classe são analisadas em busca de correlações e tendências.

1.3.1 Organização deste trabalho

No capítulo seguinte, a missão *GALEX* é tratada em mais detalhes. As características do *survey* são discutidas e são apresentados alguns resultados importantes. Os produtos gerados pelos *surveys* do *GALEX* e o acesso a estes dados são descritos no final do capítulo.

O terceiro capítulo introduz o uso de bancos de dados em astronomia. Em seguida é tratado o problema genérico de identificação cruzada (*crossmatch*) de objetos em catálogos de *surveys* astronômicos. Os bancos de dados do *SDSS* e do *STARLIGHT* são apresentados, e o processo de *crossmatch* é feito entre os catálogos do *SDSS* e do *GALEX*. Finalmente, define-se

uma amostra com magnitudes absolutas (ópticas e UV) e propriedades físicas das galáxias.

No quarto capítulo é feita a análise da amostra obtida no capítulo anterior. As galáxias são classificadas através do diagrama WHAN, introduzido por Cid Fernandes et al. (2010, 2011), e as propriedades físicas destas galáxias (obtidas do STARLIGHT) são analisadas de acordo com as suas cores ópticas e UV.

No quinto e último capítulo são apresentadas as conclusões e perspectivas deste trabalho.

As *queries* SQL utilizadas neste trabalho estão listadas no apêndice A.

Capítulo 2

O *Galaxy Evolution Explorer* (GALEX)

2.1 Objetivos do GALEX

O *Galaxy Evolution Explorer* (GALEX) é um telescópio espacial de pequeno porte (espelho primário com 50 cm) da NASA¹, lançado em 28 de abril de 2003 para conduzir um *survey* de todo o céu numa faixa espectral do UV, entre 1350 e 2750 Å. O objetivo principal do GALEX é estudar a evolução da taxa de formação estelar em galáxias (Martin et al. 2005). Os dados coletados pela missão são publicados em *Data Releases* periódicos, denominados *General Releases*. Este trabalho foi realizado sobre os dados do sexto *General Release*, GR6.

A missão consiste em uma série de *surveys* fotométricos e espectroscópicos (ver tabela 2.1). Destes, os principais *surveys* são o *All Sky Survey* (AIS), cobrindo todo o céu numa profundidade menor, e o *Medium Imaging Survey* (MIS), cobrindo uma área limitada porém mais profunda, que foram utilizados neste trabalho. O imageamento é feito em duas bandas espectrais: ultravioleta distante (*far ultraviolet*, FUV), de 1350 a 1750 Å, e ultravioleta próximo (*near ultraviolet*, NUV), de 1750 a 2750 Å. As curvas de transmissão dos filtros utilizados nessas bandas podem ser vistos na Figura 2.1. A espectroscopia é feita inserindo-se no caminho ótico um grisma, que consiste num prisma combinado com uma rede de difração. Obtém-se deste modo um espectro de baixa resolução para cada objeto na imagem, conforme descrito por Morrissey et al. (2007).

Os *surveys* do GALEX foram planejados de forma a se valer de outros *surveys* já existentes

¹NASA *Small Explorer* (SMEX) - <http://explorers.gsfc.nasa.gov/missions.html>

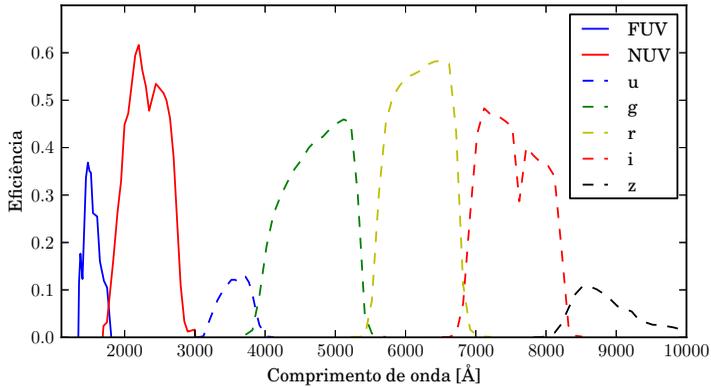


Figura 2.1: Curvas de transmissão dos filtros do *GALEX*, medidas em laboratório, representadas por linhas sólidas (Morrissey et al. 2005). Os comprimentos de onda efetivos dos filtros são $\lambda_{FUV} = 1528\text{\AA}$ e $\lambda_{NUV} = 2271\text{\AA}$. Para comparação é mostrada a curva eficiência para os filtros do *SDSS* (linhas tracejadas). Dados retirados do *website* mantido por Peter Capak: <http://www.astro.caltech.edu/~capak/cosmos/filters/>

Tabela 2.1: *Surveys* realizados pelo GALEX. O CAI consiste em observações de anãs brancas para calibração. A cobertura do céu é dada em graus quadrados. No caso do NGS, a magnitude limite é dada em unidades de densidade superficial de magnitude [arcsec⁻²]. Informações retiradas de Martin et al. (2005).

<i>Survey</i>	Cobertura [graus ²]	Mag. AB limite
<i>Calibration Imaging (CAI)</i>	-	-
<i>All-sky Imaging Survey (AIS)</i>	26 000	20.5
<i>Medium Imaging Survey (MIS)</i>	1 000	23
<i>Deep Imaging Survey (DIS)</i>	80	25
<i>Nearby Galaxy Survey (NGS)</i>	80	27.5
<i>Wide Field Spectroscopic Survey (WSS)</i>	80	20
<i>Medium-deep Spectroscopic Survey (MSS)</i>	8	21.5–23
<i>Deep Spectroscopic Survey (DSS)</i>	2	23–24

em outros comprimentos de onda. A Figura 2.2 mostra a sobreposição da área observada² pelos *surveys* AIS e MIS do GALEX e do *Sloan Digital Sky Survey (SDSS)*. O objetivo primário da missão do GALEX é calibrar a taxa de formação estelar no universo local e determinar o histórico cosmológico de formação estelar entre os *redshifts* $0 < z < 2$ (Martin et al. 2005). A comparação com dados de *surveys* em outros comprimentos de onda tem um papel fundamental no cumprimento deste objetivo.

²*Footprint*, no linguajar astronômico.

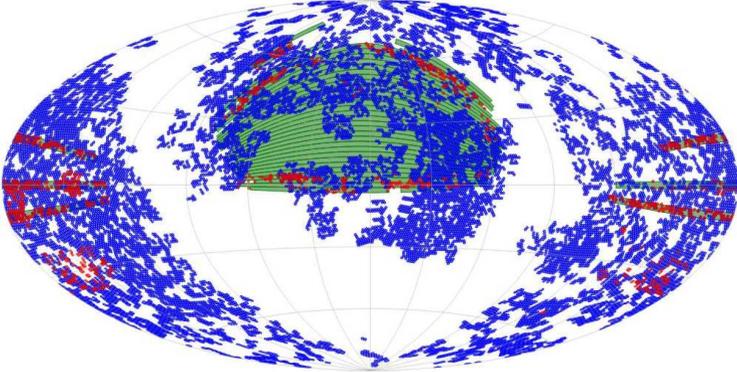


Figura 2.2: Footprint dos surveys GALEX GR2+3 AIS (azul), MIS (vermelho) e SDSS DR6 (verde), de Budavári et al. (2009).

2.2 Histórico do estudo do céu no UV

Devido à alta absorção da luz na banda UV causada pela camada de ozônio, observações do céu na banda UV precisam ser feitas fora da atmosfera terrestre. Não é de se estranhar, portanto, que o trabalho nesta faixa espectral tenha progredido menos do que na faixa do óptico e do infravermelho.

O primeiro trabalho sistemático de observação em UV foi feito pelo *Orbiting Astronomical Observatory 2* (Code et al. 1970), obtendo fotometria e espectroscopia de estrelas brilhantes, aglomerados globulares e galáxias próximas. Durante as décadas de 1970 e 1980, este e outros satélites como o TD-1 (Boksenberg et al. 1973), o *Astronomical Netherlands Satellite* (van Duinen et al. 1975) e o *International Ultraviolet Explorer* (Kondo & Wamsteker 1987) – o primeiro satélite a utilizar um detetor de imageamento UV – forneceram os dados fundamentais para os modelos de síntese de população estelar de galáxias. Surveys de campo amplo foram feitos por uma câmera lunar erguida por astronautas da *Apollo 16* (Carruthers 1973), a bordo do *Skylab* (Henize et al. 1975) e pelo instrumento *FAUST* a bordo do *Spacelab* (Bowyer et al. 1993). Muitas imagens UV também foram obtidas pelo *Ultraviolet Imaging Telescope* em duas missões em ônibus espacial (Stecher et al. 1997).

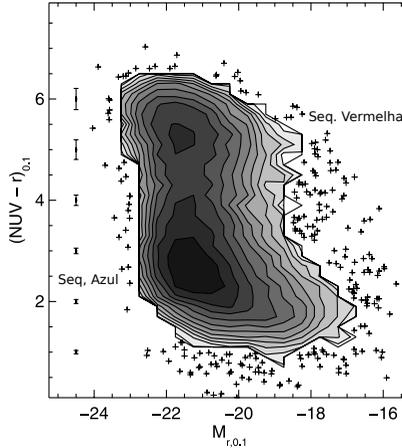


Figura 2.3: Diagrama cor-magnitude em ultravioleta. (Wyder et al. 2007, figura 7).

2.3 Resultados obtidos pelo GALEX

O *GALEX* fez o primeiro *survey* do céu inteiro em UV. As regiões próximas ao plano da Galáxia foram evitadas para não danificar os detectores. Pode-se ter uma ideia do sucesso desta missão considerando a grande quantidade de artigos publicados³. Abaixo segue um resumo dos resultados mais notáveis.

Wyder et al. (2007) analisaram a distribuição de galáxias em função da cor UV e da magnitude absoluta no universo local. Esta distribuição é conhecida como *Diagrama Cor-Magnitude* (CMD, na sigla em inglês para *Color-Magnitude Diagram*). Os autores usam *redshifts* e fotometria óptica obtidas do *SDSS* junto com fotometria UV do *survey* MIS do *GALEX*. A amostra do *SDSS* é correlacionada com a do *GALEX* procurando o objeto do *GALEX* mais próximo de cada objeto *SDSS* até um limite de 4 segundos de arco.

O diagrama cor-magnitude ($NUV - r$ contra M_r) elaborado por Wyder et al. mostra a separação das galáxias nas sequências azul e vermelha (Figura 2.3). Esta distribuição bimodal é um resultado bem conhecido na astronomia (Baldry et al. 2004). Porém, diferente do diagrama cor-magnitude para a faixa espectral do óptico, a distribuição de cores em UV não

³Há uma lista com as mais de 500 publicações relacionadas ao projeto do *GALEX* em <http://www.galex.caltech.edu/researcher/publications.html>

pode ser ajustada somente pela soma de duas gaussianas, há um excesso de objetos nas cores intermediárias entre os picos azul e vermelho. A boa separação entre as sequências é atribuída a uma maior sensibilidade à formação estelar recente.

Martin et al. (2007) investigaram as propriedades das galáxias entre as sequências vermelha e azul para a mesma amostra citada acima. As galáxias nesta região intermediária são preferencialmente galáxias com núcleo ativo (*Active Galactic Nucleus*, AGN). Os autores estimam o fluxo de massa de galáxias indo da sequência azul para a vermelha.

Ainda para a mesma amostra, Schiminovich et al. (2007) investigaram a correlação entre a morfologia das galáxias e a sua posição no CMD. A função de luminosidade UV do universo local é medida – pela primeira vez, segundo os autores – com relação aos parâmetros estruturais e à inclinação das galáxias.

A missão do *GALEX* se encerrou em 31 de dezembro de 2011. Dados coletados após o GR6, como as observações no mesmo campo utilizado na missão *Kepler*, observações de M31 e da Nuvem de Magalhães, entre outros, serão liberados num último *data release*, GR7. Os dados obtidos pelo *GALEX* permanecerão disponíveis publicamente no *Multi-Mission archive at the Space Telescope Science Institute* (MAST).

2.4 Data releases e banco de dados

A liberação dos dados do *GALEX* é feita anualmente em *General Releases* (GR). Os dados consistem basicamente em imagens e catálogos, divididos em campos (*tiles*) com área de aproximadamente 1,2 graus quadrados. Devido ao modo como o *GALEX* faz as observações, um determinado objeto pode estar presente em mais de um campo. A tabela 2.2 mostra o número cumulativo de campos observados por *survey* em cada GR⁴. Observações de pesquisadores convidados (*Guest Investigators*, GI) foram selecionadas de forma a complementar os *surveys*.

Para facilitar o acesso aos dados do *GALEX*, o MAST desenvolveu uma ferramenta chamada *GalexView* (Figura 2.4), utilizando tecnologia *Adobe Flex*⁵. Desta forma o *GalexView* pode ser acessado através de seu *website*⁶ em qualquer *web browser* que tenha suporte ao

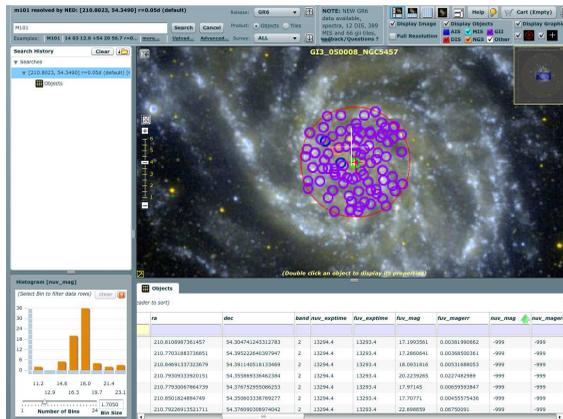
⁴Informações retiradas do *website* do GR6: <http://galex.stsci.edu/GR6/>

⁵*Adobe Flex* é um *framework* de código aberto que permite desenvolver aplicações para *web browsers*. Ver <http://www.adobe.com/products/flex.html>.

⁶*GalexView*: <http://galex.stsci.edu/GalexView/>

Tabela 2.2: Campos observados em cada *General Release* do *GALEX*.

Release	AIS	DIS	MIS	NGS	GI	CAI	Espectros	Total
GR1	3 074	14	112	52	-	-	7	3 259
GR2/GR3	15 721	165	1 017	296	288	20	41	17 548
GR4/GR5	28 269	292	2 161	458	788	38	174	32 180
GR6	28 889	338	3 479	480	1 314	51	-	34 551

**Figura 2.4:** Tela do programa *GalexView*, com a visualização da galáxia M101.

*Adobe Flash Player*⁷.

Através do *GalexView* é possível fazer buscas, visualizar e obter imagens e catálogos dos campos do *GALEX*. As buscas podem ser feitas de forma bastante versátil, tanto pelo nome do objeto quanto pelas coordenadas do céu. O formato de entrada é flexível o suficiente para evitar os problemas causados por idiosincrasias na notação de coordenadas (por exemplo, tanto “14h03m12.6s +54d20m56.7s” quanto “14 03 12.6 54 20 56.7” ou “210.83 54.35” apontam para a mesma região). A sua interface permite filtrar o conteúdo retornado pelas buscas, separando por *surveys*. Há também uma ferramenta de histograma, permitindo filtrar pelos valores

⁷*Adobe Flash Player* é uma extensão multiplataforma para *web browsers* que provê capacidade de visualização de conteúdo *flash* gerado tanto pelos seus editores proprietários quanto por ferramentas de terceiros. Ver <http://www.adobe.com/products/flashplayer/>.

das colunas dos catálogos. Os objetos selecionados na busca aparecem marcados na visualização da imagem. Utilizando um sistema do tipo “carrinho de compras”, pode-se selecionar campos e objetos de interesse, para, ao final do uso do sistema, baixar toda a seleção de uma vez.

Tanto o *GalexView* quanto outras ferramentas de busca do MAST, como o *GALEXSearch Form* e o *GALEXTilelist*, são construídos sobre um *banco de dados relacional* acessado através da linguagem *SQL* (Chamberlin & Boyce 1974). Muito comum na indústria, bancos de dados relacionais dispõem em geral de uma vasta gama de ferramentas para gerenciamento dos dados. Uma de suas grandes vantagens é o uso de índices⁸ para agilizar o acesso a dados. Embora a tecnologia exista desde a década de 1970 (Codd 1970), até uma década atrás suas vantagens eram praticamente negligenciadas na astronomia.

Bancos de dados relacionais e ferramentas para gerenciamento e acesso a dados serão tratados com mais detalhes no capítulo 3.

⁸Um índice numa tabela de banco de dados é uma estrutura que copia partes da tabela numa determinada ordem, de forma a aumentar a velocidade de acesso aos dados ao custo de espaço de armazenamento.

Capítulo 3

Estendendo a base de dados *SDSS/STARLIGHT* para o ultravioleta

3.1 Banco de dados do *SDSS*

Um dos maiores responsáveis pela promoção do uso de bancos de dados relacionais na astronomia é o projeto *Sloan Digital Sky Survey (SDSS)*. Inicialmente o *SDSS* utilizou um sistema de gerenciamento de banco de dados orientado a objetos (*Object Oriented Database Management System, OODBMS*; Maier et al. 1986). Após pouco mais de um ano a abordagem se mostrou inadequada: entre os principais problemas, uma linguagem de *query* inadequada e *performance* fraca. O motivo, segundo Thakar et al. (2004), foi a incapacidade da empresa desenvolvedora do OODBMS em prover novas funcionalidades requisitadas pelo projeto e correção de *bugs*, bem como em acompanhar o crescimento da performance do *hardware*.

3.1.1 Migração de OODBMS para RDBMS

Todo o banco de dados do *SDSS* foi migrado para um sistema de gerenciamento de banco de dados relacional (*Relational Database Management System, RDBMS*; Codd 1970) num esforço guiado por Thakar et al.. RDBMS pode ser considerado o padrão da indústria, praticamente todas as linguagens de programação têm bibliotecas de interface às implementações de RDBMS comerciais mais comuns (Oracle, IBM, Microsoft e Postgres). Há uma diversidade de ferramentas para desenvolvimento e gerenciamento de RDBMS. E talvez o maior benefício de todos, o acesso aos dados é feito utilizando uma linguagem padronizada: *Simple*

Query Language, ou simplesmente SQL (Chamberlin & Boyce 1974). A migração dos dados do SDSS para um RDBMS comercial implicou num aumento significativo da performance do acesso aos dados, e resultou no desenvolvimento do *SkyServer*¹. O servidor de banco de dados escolhido pelo SDSS foi o *Microsoft SQL Server*.

A comparação entre OODBMS e RDBMS no caso particular do SDSS não implica necessariamente na superioridade do segundo em relação ao primeiro. Tanto a abordagem orientada a objetos quanto a abordagem relacional têm suas vantagens e desvantagens. O estudo de caso do SDSS é apenas uma evidência anedótica em favor do uso de bancos de dados relacionais. Dado que o SDSS obteve êxito com o uso de um RDBMS, é esperado que outros *surveys* astronômicos com grandes volumes de dados sigam o seu exemplo.

3.1.2 SkyServer

O *SkyServer* é um *website* (Figura 3.1) que provê acesso aos dados armazenados no banco de dados do SDSS (Szalay et al. 2002). O acesso mais simples pode ser feito através de um atlas de locais famosos (*famous places*), que mostra imagens coloridas de objetos celestes conhecidos. Há formulários para buscas mais sérias, gerando coleções de imagens, espectros e tabelas de dados. No *SkyServer* é possível fazer buscas avançadas utilizando SQL, embora haja limites de tempo de execução e de quantidade de objetos retornados. Esta limitação é contornada através do sistema *CasJobs*, que é tratado na próxima seção.

É importante ressaltar que é possível (de fato, a equipe do SDSS encoraja) criar *mirrors*² do *SkyServer*. Tanto o banco de dados do SDSS quanto o código fonte do *SkyServer* estão disponíveis no próprio *website* do *SkyServer*. Temos um clone do banco de dados do *Data Release 8* do SDSS no servidor *CasJobs* do STARLIGHT³.

3.1.3 CasJobs

O *Catalog Archive Server Jobs* (*CasJobs*) é um serviço online desenvolvido pela equipe do SDSS para expandir a capacidade do *SkyServer* (Li & Thakar 2008). Nele o usuário pode executar consultas SQL no banco de dados do SDSS da mesma forma que no *SkyServer*. Porém, além de consultas rápidas, é possível agendar a execução de consultas mais longas. O *CasJobs*

¹SDSS SkyServer: <http://skyserver.sdss.org/>

²*Mirror*: Espelho, em inglês. Clone de um *website*.

³*CasJobs* do STARLIGHT: <http://casjobs.starlight.ufsc.br/casjobs/>

The screenshot shows two main panels of the SkyServer interface. The left panel, titled 'SQL Search', contains a warning: 'Please note: To be fair to other users, queries run from SkyServer search tools are restricted in how long they can run and how much output they return, by timeout and row limits. Please see the Query limits help page. To run a query that is not restricted by a timeout or number of rows returned, please use the CasJobs batch query service.' Below this is a 'Query' input field and a 'Submit' button. The right panel, titled 'Explore', shows details for SDSS J020212.33-000602.2, including its name 'GALAXY J020212.33-000602.2', coordinates, and a table of photometric data. The table has columns for 'Filter', 'Flux', 'Flux Error', 'Flux Limit', 'Flux Limit Error', 'Flux Limit Lower Bound', 'Flux Limit Upper Bound', 'Flux Lower Bound', 'Flux Upper Bound', 'Flux Lower Bound Error', 'Flux Upper Bound Error', 'Flux Lower Bound Lower Bound', 'Flux Lower Bound Upper Bound', 'Flux Upper Bound Lower Bound', 'Flux Upper Bound Upper Bound'. The table shows data for filters 'i', 'z', 'g', 'r', 'u'. Below the table is a plot of the galaxy's light profile, showing flux density versus position.

Figura 3.1: Telas do SkyServer. À esquerda, formulário para submeter uma *query* SQL. À direita, ferramenta *Explore* mostrando a galáxia NGC 799.

The screenshot shows the 'CasJobs' interface. At the top, there's a navigation bar with 'Help', 'Tools', 'Query', 'History', 'MyDB', 'Import', 'Groups', 'Output', 'Profile', 'Queues', 'SkyServer', and 'Logout'. Below this is a 'Context' section with 'Table (optional)' and 'Task Name'. The main area shows a SQL query:


```
select top 10
  speckobjid, s.z as redshift,
  s.fibmag_g as mag_g,
  s.fibmag_r - s.fibmag_g as color,
  'a href=http://cas.sdss.org/dr7/en/tools/chart/mavi.asp?ra=' +
  cast(s.ra as varchar(15)) +
  '&dec=' + cast(s.dec as varchar(10)) +
  '&size=' +
  '&img_sco=http://casjobs.sdss.org/ImgOutoutDR7/getImage.asp?ra=' +
  cast(s.ra as varchar(15)) +
  '&dec=' + cast(s.dec as varchar(10)) +
  '&size=40xwidth=40&height=60&out=1/'> '
  as pic
from speckphotoall s
```

 Below the query is a table with columns 'speckobjid', 'redshift', 'mag_g', 'color', and 'pic'. The table shows several rows of data, with the 'pic' column containing small thumbnail images of galaxies. The first row is:

speckobjid	redshift	mag_g	color	pic
7509397271163692	0.0213136	17.40857	0.8991566	
7509397271583126	0.213991	20.50339	1.270136	
75093972720025600	0.65181	20.11052	0.1381607	
7509397272424904	0.12666	20.25174	1.388601	

 At the bottom, there's a 'Contact' section and a 'Query Results' button.

Figura 3.2: Tela do CasJobs. Resultado da *query* buscando o *redshift*, a magnitude na banda *g*, a cor *g* – *r* e uma amostra da imagem de objetos com espectroscopia.

gerencia estas consultas agendadas numa fila de execução, de modo a não sobrecarregar a rede ou os servidores de banco de dados.

Cada usuário possui seu próprio banco de dados, chamado *MyDB*. Pode-se importar tabelas para o *MyDB* para utilizar em *queries* correlacionando com os dados presentes no *CasJobs*. O *MyDB* serve como armazenamento de tabelas do usuário, e há mecanismos para exportar estas tabelas para arquivos nos formatos FITS, CSV, XML e VOTable. Estes arquivos podem ser lidos por programas de análise de dados como o *TopCat*⁴, ou mesmo importados para outros bancos de dados.

É possível utilizar o *CasJobs* para acessar virtualmente qualquer banco de dados. No momento, o Grupo de Astrofísica da UFSC possui um servidor *CasJobs* com bancos de dados do STARLIGHT, SDSS DR8, GalaxyZoo (Lintott et al. 2008) e uma amostra do GALEX e um catálogo de *redshifts* fotométricos (O'Mill et al. 2011). O *CasJobs* também foi adotado por outros projetos como o GALEX, *Kepler*⁵ e o *Palomar Quest*⁶ (Djorgovski et al. 2008).

A Figura 3.2 mostra uma tela típica de uma sessão no *CasJobs*.

3.2 Banco de dados do STARLIGHT

O STARLIGHT é um código de síntese espectral (Cid Fernandes et al. 2005). O programa é executado uma vez para cada galáxia do SDSS, recebendo seu espectro como um arquivo texto. Ele usa uma biblioteca de espectros SSPs com diferentes idades e metalicidades como uma base do espaço de espectros galácticos possíveis. De forma simplificada, o que o STARLIGHT faz é encontrar as frações de massa e luz correspondentes a cada elemento da base, ou seja, cada SSP. Analisando as relações entre as componentes determinadas pela síntese, o programa determina diversas propriedades físicas da galáxia, como massa estelar total, idade estelar média, metalicidade estelar média, atenuação causada por poeira e dispersão de velocidades, para citar apenas algumas. Quase um milhão de espectros foram analisados, e o resultado da síntese foi armazenado em arquivos texto.

Somente os componentes estelares do espectro são obtidos desta forma. Subtraindo o espectro sintetizado de luz estelar é possível medir as linhas de emissão do espectro. Esta é uma etapa de pós-processamento, que gera um catálogo complementar de linhas de emissão.

⁴*TopCat* é um visualizador gráfico interativo e editor de dados tabulares usado em astronomia. Ver <http://www.star.bris.ac.uk/~mbt/topcat/>.

⁵*Kepler CasJobs*: <http://mastweb.stsci.edu/kplrscasjobs/>

⁶*Palomar Quest CasJobs*: <http://webvoy.cacr.caltech.edu/CasJobs/>

3.2.1 Importação para o RDBMS

Os arquivos da síntese gerados pelo STARLIGHT ocupam quase trezentos gigabytes. Mesmo o catálogo de propriedades físicas das galáxias, sozinho, ocupa mais de um gigabyte. Embora seja um volume razoavelmente grande de dados, é possível trabalhar com esta quantidade de dados num computador atual⁷. A transferência de arquivos com tamanho da ordem de gigabytes pela *internet* também é lugar-comum atualmente. Poder-se-ia argumentar que distribuir os dados neste formato seja a forma mais adequada.

Entretanto, deve-se admitir que uma das maiores razões para o sucesso do *CasJobs* em prover acesso aos dados do *SDSS* não é o tamanho da base de dados, e sim a facilidade com que o usuário pode acessar os dados e filtrar somente o que lhe for conveniente. Além disso, manter a base de dados num local central permite que sejam feitas correções e revisões, o que implicaria normalmente numa nova transferência caso cada usuário tivesse a sua cópia local.

O *CasJobs* requer um servidor rodando *Windows Server* com *Internet Information Services* (IIS) e *Microsoft SQL Server* (MSSQL). A instalação do *CasJobs* está documentada no *website* do *SkyServer* (ver Seção 3.1.2). Com um servidor *CasJobs*, o trabalho consiste em importar os dados em arquivos texto para um banco de dados no MSSQL. A ferramenta principal para a manipulação dos bancos de dados no MSSQL é o *Microsoft SQL Server Management Studio*. Nele há um assistente para importação de dados baseado no *SQL Server Integrated Services* (SSIS). A importação das tabelas de propriedades físicas do STARLIGHT foi feita rapidamente através desta ferramenta. Entretanto, foi necessário normalizar⁸ a tabela de linhas de emissão para facilitar a inclusão de novas linhas no futuro.

3.2.2 Estrutura do banco de dados

O esquema do banco de dados do STARLIGHT pode ser visto na Figura 3.3. Os dados observacionais obtidos do *SDSS* estão na tabela *observational_params*. Estes dados, junto com os parâmetros de entrada da síntese (tabela *synthesis_params*), foram usados como parâmetros para a obtenção das propriedades físicas das galáxias, armazenadas na tabela *synthesis_results*. Estas tabelas estão ligadas através dos campos *SpecObjID* e *SynID*.

⁷Quando esta dissertação foi escrita, era bastante comum comprar um computador pessoal novo com 4 gigabytes de memória RAM ou mais.

⁸A normalização consiste em decompor uma tabela em tabelas menores (com menos campos) de forma que elas fiquem melhor estruturadas. No caso das linhas de emissão, a tabela passa de “todas as linhas de um dado objeto num único registro” para “uma linha para cada registro”. No primeiro caso, adicionar um novo tipo de linha de emissão implicaria em mudar a estrutura da tabela, o que é evitado utilizando a segunda abordagem.

As linhas de emissão estão armazenadas na tabela `el_fit`, ligada à `synthesis_results` através do campo `SynID`. A definição das linhas (comprimento de onda e faixas contínuo usadas na medida) está contida na tabela `cfg_el_fit`. As tabelas referentes ao *GALEX* são explicadas na Seção 3.4.1.

3.2.3 Amostra do STARLIGHT

A amostra de galáxias do STARLIGHT contém 926246 espectros do *SDSS* DR7 (Abazajian et al. 2009). A identificação de cada espectro é feita através de um tripleto: a data juliana média da observação (MJD, *Mean Julian Date*), a identificação da placa de suporte das fibras ópticas (Plate) e a identificação da fibra utilizada para a obtenção do espectro (FiberID). Este tripleto (MJD, Plate, FiberID) identifica unicamente um espectro. Porém, é mais conveniente (e eficiente) ter um identificador único (uma chave primária⁹) para os registros num banco de dados. No caso do *SDSS*, a tabela de espectros (`SpecObjAll`) tem um identificador chamado `SpecObjID`.

Além de espectros, o banco de dados do *SDSS* contém fotometria de 1/4 do céu. Cada objeto do catálogo de fotometria tem um identificador único, `ObjID`. Existe uma coluna na tabela de espectros chamada `BestObjID`, que aponta para o registro de fotometria (tabela `PhotoObjAll`) mais provável para cada espectro. É importante salientar que nem todo espectro tem um `BestObjID` definido.

A tabela de índices da amostra de galáxias do STARLIGHT contém inicialmente os tripletos [MJD, Plate, FiberID]. Esta tabela é importada para o ambiente `CasJobs` do *SDSS* DR7¹⁰. Através da execução da *query* da Figura A.1 a tabela tem os valores de `SpecObjID` e `BestObjID` preenchida. Entre os objetos na amostra do STARLIGHT, 622 objetos não possuem um valor de `BestObjID` definido, ou seja, não foi possível encontrar a sua contrapartida fotométrica.

⁹Chave primária é um conjunto de um ou mais campos tais que a combinação de todos os campos da chave não se repete.

¹⁰*CasJobs* *SDSS* DR7 - <http://casjobs.sdss.org/CasJobs/>

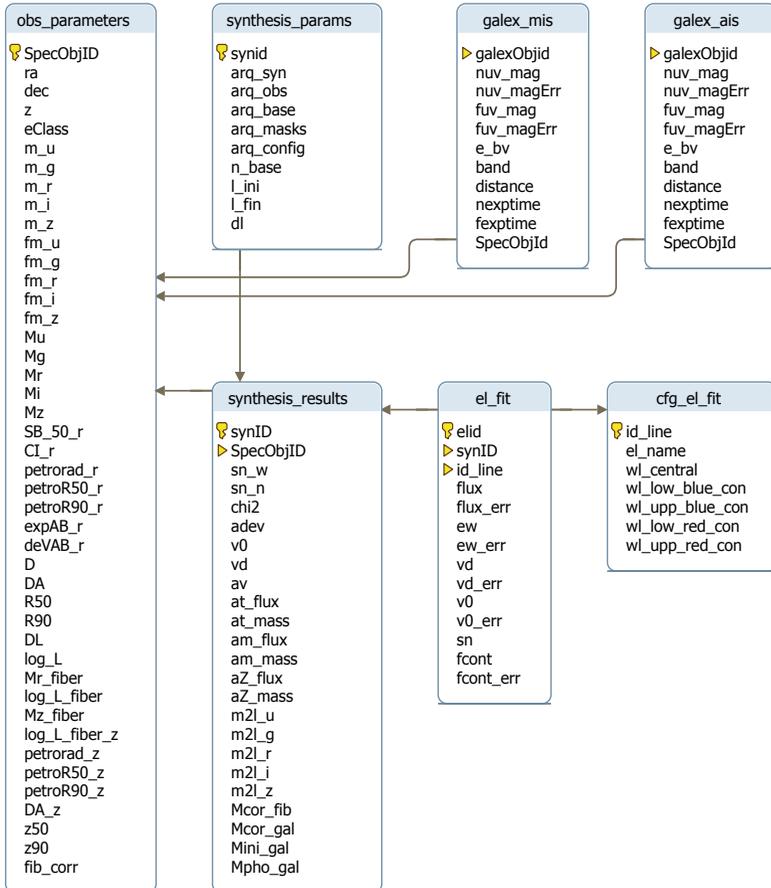


Figura 3.3: Esquema do banco de dados do STARLIGHT. As linhas indicam chaves estrangeiras, relacionando dois campos (marcados com uma chave) de duas tabelas diferentes. Os triângulos indicam campos indexados. A descrição completa dos campos pode ser encontrada no site do STARLIGHT através do [link](http://casjobs.starlight.ufsc.br/casjobs/field_list.html) http://casjobs.starlight.ufsc.br/casjobs/field_list.html.

3.3 Identificação cruzada entre SDSS e GALEX

A identificação cruzada (*crossmatch*) de objetos em *surveys* diferentes é um problema razoavelmente complicado. A cobertura do céu de cada *survey* em geral não é a mesma. Por outro lado, os objetos presentes em um *survey* podem não ter sido detectados no outro. A probabilidade de duas fontes em catálogos diferentes corresponderem a um mesmo objeto pode ser calculada como função da separação entre elas e a precisão astrométrica das medidas (Budavári & Szalay 2008).

Budavári et al. (2009) aplicam este método probabilístico ao SDSS e ao GALEX. O *crossmatch* espacial é feito dentro de um RDMS (MSSQL, o mesmo usado no *CasJobs*), utilizando técnicas avançadas de indexação (Kunszt et al. 2000, *Hierarchic Triangular Mesh*). A tabela resultante é uma relação “muitos para muitos”, onde a maioria dos objetos GALEX tem apenas um objeto SDSS associado, mas outras associações podem ocorrer. Um exemplo onde pode ocorrer uma associação “um para muitos” é o caso onde existe uma fonte fraca em UV (presente no SDSS mas não detectada pelo GALEX) próxima a uma fonte presente tanto no UV quanto no óptico. O algoritmo irá apontar estes dois objetos no SDSS como candidatos a serem a contrapartida óptica do objeto detectado no GALEX. O caso inverso implicaria numa associação “muitos para um”. Nas tabelas 3.1 e 3.2 há a quantidade de identificações para cada tipo de associação, referentes aos *surveys* AIS e MIS, respectivamente. Os valores foram determinados para o *crossmatch* entre SDSS DR7 e GALEX GR6, disponível no *CasJobs* do GALEX¹¹). O artigo citado acima mostra a mesma tabela, com os resultados para dados do SDSS DR6 e GALEX GR3. A técnica utilizada por Budavári et al. agora faz parte do *pipeline* do GALEX. A distribuição do *CasJobs* inclui as ferramentas necessárias para fazer o *crossmatch* espacial entre bancos de dados.

¹¹*CasJobs* GALEX: <http://galex.stsci.edu/casjobs/>

Tabela 3.1: Número de identificações entre *SDSS DR7* e *survey AIS* do *GALEX GR6*, por associação. A lista foi gerada pela *query* mostrada na Figura A.2.

<i>GALEX</i>	<i>SDSS</i>		
	1	2	Muitos
1	15 267 818	9 150 919	4 623 197
2	4 524 337	2 504 786	1 162 463
Muitos	770 645	426 691	184 680

Tabela 3.2: Número de identificações entre *SDSS DR7* e *survey MIS* do *GALEX GR6*, por associação. A lista foi gerada pela *query* mostrada na Figura A.2.

<i>GALEX</i>	<i>SDSS</i>		
	1	2	Muitos
1	8 201 735	5 923 551	3 775 187
2	2 120 174	1 580 701	984 067
Muitos	276 447	234 016	150 894

3.4 Obtendo dados UV para a amostra STARLIGHT

3.4.1 Relação de *crossmatch* entre *SDSS* e *GALEX*

Como comentado na Seção 3.3, no banco de dados do *GALEX* há uma tabela de *crossmatch* entre os objetos do *GALEX* e os seus correspondentes ópticos no catálogo do *SDSS*, chamada *xSDSSDR7*. A descrição completa dos campos pode ser vista na tabela 3.3. Dado que a identificação não é necessariamente unívoca, existem informações extras nesta tabela a fim de facilitar a seleção dos melhores candidatos: *DistanceRank* e *MultipleMatchCount*.

A identificação cruzada também é feita na direção oposta. Dado um objeto do *SDSS*, foram encontrados os objetos do *GALEX* candidatos. Para um par [*objID*, *SDSSobjID*], há também os campos *ReverseDistanceRank* e *ReverseMultipleMatchCount*.

Tabela 3.3: Descrição dos campos da tabela xSDSSDR7.

Campo	Descrição
ObjID	Identificador único de objeto do <i>GALEX</i> .
SDSSObjID	Identificador único do <i>SDSS</i> .
Distance	Separação angular em segundos de arco.
DistanceRank	Um número inteiro, onde o valor 1 indica que o objeto do <i>GALEX</i> é o mais próximo do objeto <i>SDSS</i> , o valor 2 indica que ele é o segundo mais próximo, etc.
ReverseDistanceRank	Um número inteiro, onde o valor 1 indica que o objeto do <i>SDSS</i> é o mais próximo do objeto <i>GALEX</i> , o valor 2 indica que ele é o segundo mais próximo, etc.
MultipleMatchCount	Um número inteiro indicando quantos objetos <i>SDSS</i> foram encontrados para o objeto <i>GALEX</i> dentro do raio de busca (5 segundos de arco).
ReverseMultipleMatchCount	Um número inteiro indicando quantos objetos <i>GALEX</i> foram encontrados para o objeto <i>SDSS</i> dentro do raio de busca (5 segundos de arco).

3.4.2 Dados UV

A amostra do STARLIGHT descrita na Seção 3.2.3 contém o identificador do catálogo de fotometria do *SDSS*. Este identificador é o mesmo utilizado na tabela xSDSSDR7. A *query* da Figura A.3 preenche a tabela chamada `galex_ais` com todos os objetos da amostra do STARLIGHT, junto com suas respectivas magnitudes FUV (`FUV_mag`) e NUV (`NUV_mag`), o erro na medida das magnitudes (`FUV_magErr` e `NUV_magErr`), o tempo de exposição em cada filtro (`fexptime` e `nexptime`), o excesso de cor $E(B-V)$ (`e_bv`, ver Seção 3.4.3) e a distância entre a detecção do objeto no *GALEX* e no *SDSS* (`distance`). Esta tabela é esparsamente populada, com um registro para cada objeto do STARLIGHT, e os dados UV preenchidos somente para os objetos com identificação positiva. Caso o objeto não tenha um correspondente *GALEX*, os valores serão nulos¹². De forma similar, os dados UV para o *survey* MIS foram armazenados na tabela `galex_mis`. Estas tabelas podem ser vistas no contexto do banco de dados do STARLIGHT na Seção 3.2.2.

¹²Numa tabela, quando um campo de um registro não possui valor definido, seu valor é dito “nulo”. Em SQL, a palavra-chave que representa um valor nulo é “*null*”.

Foram escolhidas somente identificações cruzadas do tipo um-para-um, ou seja, registros com `multipleMatchCount` e `reverseMultipleMatchCount` iguais a 1. Em situações diferentes desta, caso houvesse mais de um candidato para uma mesma detecção, seria necessária uma análise mais detalhada. Desta forma, optou-se por uma confiabilidade maior nos dados utilizando apenas identificações unívocas, em troca de uma amostra menor.

Outro artefato importante é que somente um objeto *GALEX* foi escolhido para cada objeto do catálogo do STARLIGHT, independente do *survey* utilizado. Se existe um correspondente AIS e outro MIS para o mesmo objeto da amostra, apenas o mais próximo é considerado. Nos campos onde os dois surveys se sobrepõem, há uma maior quantidade de objetos no MIS do que no AIS, devido ao maior tempo de exposição no MIS. Assim, nesses campos, os objetos do MIS têm maior chance de estarem mais próximos ao objeto do *SDSS*. Isto também não é um problema muito grave para o AIS, já que a sua cobertura do céu é muito maior do que a do MIS. A fração exata de objetos perdidos em cada caso ainda precisa ser determinada. Esta correção será feita futuramente numa revisão do catálogo.

No total foram obtidas 173 218 detecções no AIS e 41 274 no MIS. Nem todos os objetos tiveram detecção simultânea nas bandas FUV e NUV. A tabela 3.4 lista a quantidade de objetos que possuem valor definido de acordo com cada banda UV. Alguns objetos podem ser muito fracos em FUV. Isto combinado com a baixa eficiência do filtro (Figura 2.1) e do detetor podem fazer com que a contagem de fótons deste objeto seja muito baixa, ficando abaixo do nível limite de sinal-ruído do *survey*.

Tabela 3.4: Número de detecções por banda no catálogo UV do STARLIGHT. Nas linhas FUV e NUV, é listada a quantidade de objetos com detecção para os *surveys* AIS e MIS nas bandas FUV e NUV, respectivamente. A linha FUV+NUV lista a quantidade de objetos com detecção simultânea em FUV e NUV. A Figura A.4 contém a *query* que gera esta lista.

Banda	AIS	MIS
FUV	107 902	13 685
NUV	168 126	38 828
FUV+NUV	102 811	11 239

3.4.3 Correções aplicadas à fotometria UV

O catálogo do *GALEX* fornece a fotometria NUV e FUV dos objetos em magnitude aparente, sem qualquer correção a não ser por fatores instrumentais. Para que as galáxias possam ser analisadas de forma adequada, algumas correções precisam ser feitas.

A luz proveniente de outras galáxias sofre atenuação causada por poeira dentro da nossa Galáxia. Esta atenuação depende da direção de onde vem a luz, e a sua influência depende do comprimento de onda. O efeito final, em geral, é o avermelhamento do espectro. O excesso de cor $E(B - V)$ representa quantitativamente este avermelhamento.

O catálogo do *GALEX* provê os valores de excesso de cor $E(B - V)$ das galáxias com base nos mapas de atenuação interestelar de Schlegel et al. (1998). A correção é feita usando o modelo de atenuação CCM (Cardelli, Clayton, & Mathis 1989, equações 4a e 4b). Para um comprimento de onda λ qualquer, a atenuação absoluta (em magnitudes) é dada por

$$A_\lambda = R_V E(B - V) a(x) + E(B - V) b(x),$$

$$x = \frac{10000 \text{ \AA}}{\lambda}.$$

No caso das bandas FUV ($\lambda_{eff} = 2271 \text{ \AA}$) e NUV ($\lambda_{eff} = 1528 \text{ \AA}$),

$$x_{FUV} = 6,54$$

$$x_{NUV} = 4,40$$

com os coeficientes $a(x)$ e $b(x)$ dados por

$$a(x) = 1,752 - 0,326x - 0,104/[(x - 4,67)^2 + 0,341] + F_a(x)$$

$$b(x) = -3,090 + 1,825x + 1,206/[(x - 4,62)^2 + 0,263] + F_b(x)$$

$$F_a(x_{FUV}) = F_b(x_{FUV}) = 0$$

$$F_a(x_{NUV}) = -0,04473(x_{NUV} - 5,9)^2 - 0,09779(x_{NUV} - 5,9)^3$$

$$F_b(x_{NUV}) = 0,2130(x_{NUV} - 5,9)^2 + 0,1207(x_{NUV} - 5,9)^3.$$

Usando $R_V = 3,1$, a atenuação nas bandas FUV e NUV em função do excesso de cor $E(B - V)$

é dada por

$$A_{FUV} = 8,15 E(B - V)$$

$$A_{NUV} = 9,17 E(B - V).$$

Esta correção é então aplicada às magnitudes do catálogo. Vale notar que, para um mesmo objeto, $A_{FUV} < A_{NUV}$. Ou seja, ao contrário do que ocorre para outros filtros (por exemplo, B e V), a atenuação causa um azulamento do espectro na faixa do UV observado pelo *GALEX*. Isto ocorre devido à banda de 2100 Å na curva de atenuação da nossa Galáxia.

Quando se observa, através de um determinado filtro, galáxias em *redshifts* diferentes, é preciso levar em conta que a “janela” observada do espectro depende do *redshift*. Pensando de outro modo, é como se o filtro não fosse o mesmo para todas as galáxias. Para obter medidas comparáveis, neste caso, é preciso compensar o efeito do *redshift*; este procedimento é chamado de correção $k(z)$. Assim, as magnitudes aparente NUV e FUV das galáxias da amostra foram transformadas em magnitudes absolutas utilizando a correção $k(z)$ através do código `KCORRECT (v4_2)` de Blanton & Roweis (2007).

3.5 Definição da amostra STARLIGHT+UV

A *query* mostrada na Figura A.5 monta a amostra a ser utilizada no capítulo seguinte. As colunas selecionadas são as magnitudes *ugriz* do *SDSS*, a magnitude NUV do AIS do *GALEX* e algumas propriedades físicas das galáxias (tabela 3.5). Foram selecionadas também a largura equivalente e o fluxo das linhas de emissão $H\alpha$, $H\beta$, $[O III] \lambda 5007$ e $[N II] \lambda 6584$.

Tabela 3.5: Propriedades físicas das galáxias, obtidas do STARLIGHT.

Coluna	Descrição
<code>mcor_gal</code>	Logaritmo da massa estelar [M_{\odot}]
<code>at_flux</code>	Média do logaritmo da idade estelar [anos] ponderada em fluxo
<code>at_mass</code>	Média do logaritmo da idade estelar [anos] ponderada em massa
<code>am_flux</code>	Metalicidade estelar média [Z_{\odot}] ponderada em fluxo
<code>am_mass</code>	Metalicidade estelar média [Z_{\odot}] ponderada em massa
<code>AV</code>	Atenuação causada por poeira [magnitude]

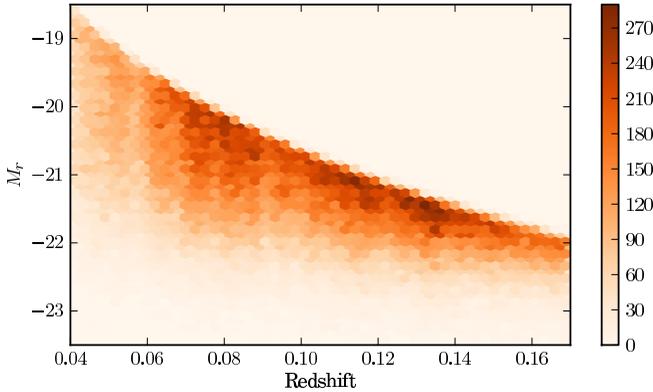


Figura 3.4: Magnitude M_r em função do *redshift* para a amostra STARLIGHT+UV. A cor dos bins hexagonais indica a densidade de pontos.

É preciso fazer algumas considerações com respeito ao *redshift* das galáxias que pertencem à amostra. O *redshift* não deve ser inferior a 0,04 para evitar efeitos de abertura da fibra¹³. Já em *redshifts* superiores a 0,17, a linha [N II] $\lambda 6584$ fica deslocada para além do limite vermelho dos espectros do SDSS, escapando da detecção. Assim, as galáxias da amostra devem ter $0,04 < z < 0,17$. Este foi o mesmo critério usado por Cid Fernandes et al. (2011).

Foram selecionadas somente galáxias com magnitude aparente $r < 17,77$, que é o limite nominal da amostra principal de galáxias do SDSS (*SDSSMain Galaxy Sample*; Strauss et al. 2002). Obtém-se assim uma amostra limitada em fluxo. A Figura 3.4 mostra a distribuição da magnitude absoluta M_r das galáxias da amostra em função do *redshift*. As propriedades físicas e observacionais das galáxias estão distribuídas conforme os histogramas da Figura 3.5. Estas propriedades serão exploradas com mais detalhes na Seção 4.3.

Resumindo, a amostra STARLIGHT+UV contém os objetos do AIS com fotometria NUV definida, limitada em *redshifts* entre 0,04 e 0,17, e com magnitude aparente r menor que 17,77, num total de 130 362 galáxias.

¹³Os espectros obtidos pelo SDSS foram feitos através de fibras ópticas. Cada fibra coleta a luz de uma região de 3 segundos de arco de diâmetro. Em *redshifts* próximos, uma fração considerável da luz das galáxias (em geral das partes externas) cai fora da fibra. Nestes casos, o espectro obtido não é o espectro integrado da galáxia, mas sim o espectro das regiões centrais.

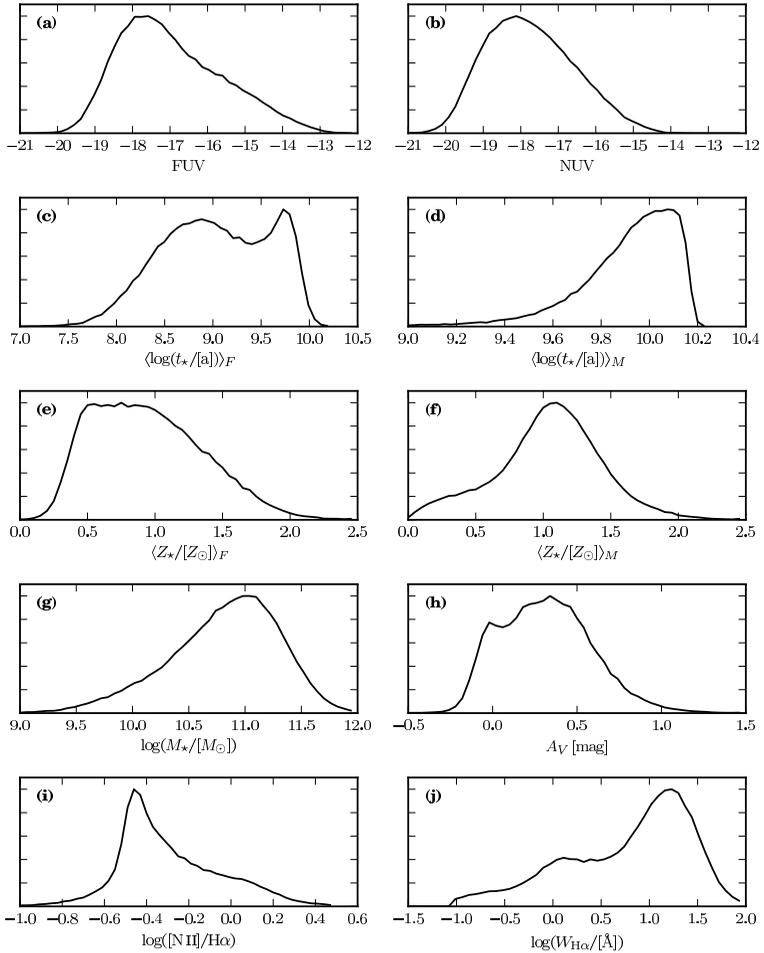


Figura 3.5: Histogramas das medidas da amostra *STARLIGHT+UV*.

Capítulo 4

Análise da amostra *STARLIGHT+UV*

4.1 Diagrama cor–cor

De posse de uma amostra de galáxias com uma informação adicional (as magnitudes em ultravioleta), é natural tentar ver como estas novas medidas se relacionam às medidas conhecidas. Chilingarian & Zolotukhin (2011) mostram que, num gráfico tridimensional das cores $NUV-r$ e $g-r$ contra a magnitude z , a distribuição de galáxias pode ser aproximada por uma superfície polinomial de baixa ordem. Os autores mostram que há uma forte correlação entre a cor $NUV-r$ e a morfologia da galáxia. Eles também estudam o histórico de formação estelar (SFH) no diagrama $NUV-r$ contra $g-r$ (diagrama cor–cor), mas a exploração é um tanto superficial.

O diagrama cor–cor para a amostra *STARLIGHT+UV* é mostrado na Figura 4.1. A cor das caixas hexagonais representa o logaritmo do número de galáxias. Através das curvas de nível é possível ver a bimodalidade da distribuição, separando as galáxias nas sequências vermelha e azul. A Figura 4.2 mostra as propriedades físicas das galáxias obtidas através do *STARLIGHT* no diagrama cor–cor. Os eixos são os mesmos da Figura 4.1, com a cor dos pontos indicando o valor de cada parâmetro. Chilingarian & Zolotukhin chegam numa relação semelhante à do painel (a) para a idade estelar das galáxias, porém através de outros meios. Estas propriedades físicas são vistas com mais detalhes na Seção 4.3.

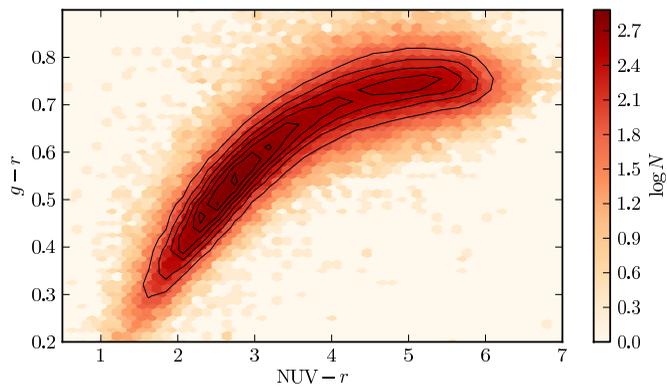


Figura 4.1: Frequência de galáxias no diagrama cor-cor ($NUV - r$ versus $g - r$) da amostra STARLIGHT+UV.

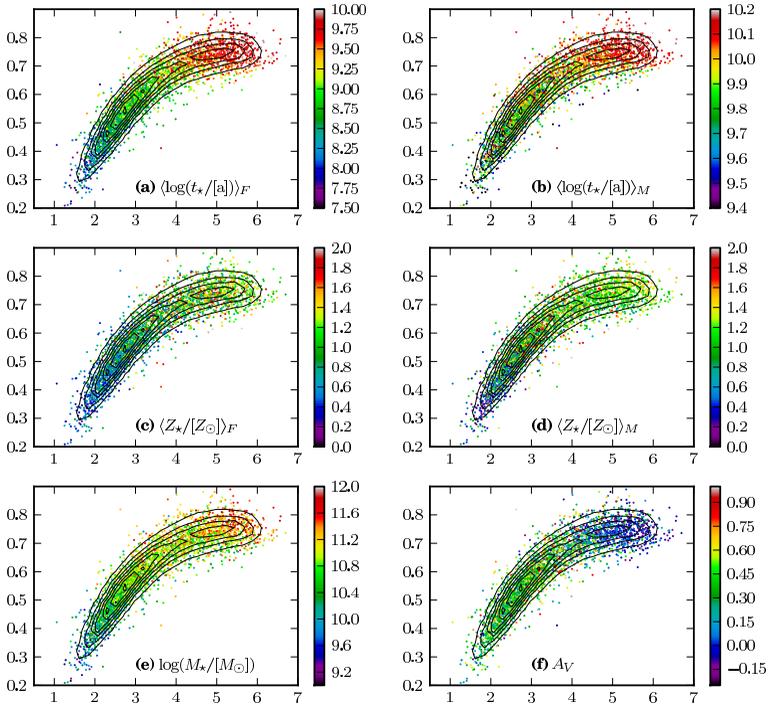


Figura 4.2: Propriedades físicas das galáxias em função de cor UV e cor óptica. Os contornos das figuras representam os níveis referentes à Figura 4.1, e os eixos horizontal e vertical são os mesmos utilizados naquela figura. As cores dos pontos nos painéis correspondem a: **(a)** Média do logaritmo da idade das SSP componentes da galáxia, ponderada pelo fluxo. **(b)** O mesmo que a anterior, mas ponderada pela massa. **(c)** Metalicidade média das SSP componentes da galáxia, ponderada pelo fluxo. **(d)** O mesmo que a anterior, ponderada pela massa. **(e)** Logaritmo da massa estelar da galáxia, em massas solares. **(f)** Atenuação por poeira na galáxia, na banda V.

4.2 Classificação das galáxias

Nesta seção são discutidas formas de classificação de galáxias e a forma como a cor UV das galáxias está relacionada às classes. São utilizadas as linhas de emissão $H\alpha$, $H\beta$, $[N II] \lambda 6584$ e $[O III] \lambda 5007$ (daqui em diante chamados apenas de $[N II]$ e $[O III]$).

A classificação feita a seguir divide as galáxias em dois grupos principais: as galáxias com linha de emissão (ELG) e as galáxias passivas (PG). As ELG ainda podem ser divididas dependendo do processo físico por trás das linhas de emissão. Essencialmente, os agentes ionizantes responsáveis pelas linhas observadas em galáxias são estrelas jovens, núcleos ativos e objetos do tipo HOLMES (*Hot Low-Mass Evolved Stars*). Cid Fernandes et al. (2011) elabora um procedimento simples para separar as galáxias em classes, de acordo com o agente ionizante que domina a emissão de linhas na galáxia. A distinção é feita de acordo com a largura equivalente da linha $H\alpha$ ($W_{H\alpha}$) e a razão entre o fluxo de linhas $[N II]/H\alpha$, num diagrama conhecido como WHAN. Este diagrama relaciona duas quantidades físicas diferentes: $W_{H\alpha}$ mede a quantidade de fótons ionizantes em relação à massa estelar, e $[N II]/H\alpha$ mede a abundância de nitrogênio, o estado de ionização e a temperatura do gás. Assim, as galáxias são separadas em classes neste diagrama conforme os seguintes critérios:

- (a) Galáxias com formação estelar (SF): $\log([N II]/H\alpha) < -0,4$ e $W_{H\alpha} > 3 \text{ \AA}$,
- (b) Galáxias com núcleo ativo forte (sAGN): $\log([N II]/H\alpha) > -0,4$ e $W_{H\alpha} > 6 \text{ \AA}$,
- (c) Galáxias com núcleo ativo fraco (wAGN): $\log([N II]/H\alpha) > -0,4$ e $3 \text{ \AA} > W_{H\alpha} > 6 \text{ \AA}$,
- (d) Galáxias “aposentadas” (RG): $0,5 \text{ \AA} < W_{H\alpha} < 3 \text{ \AA}$,
- (e) Galáxias passivas (PG): $W_{H\alpha} < 0,5 \text{ \AA}$ e $W_{[N II]} < 0,5 \text{ \AA}$.

Neste esquema, as SF têm linhas de emissão devido a estrelas jovens e massivas, as sAGN e wAGN têm linhas devido a núcleos ativos, as RG têm linhas devido a HOLMES e as PG não têm emissão mensurável.

O diagrama WHAN para a amostra STARLIGHT+UV pode ser visto na Figura 4.3. A cor dos pontos para cada classe é a mesma utilizada por Cid Fernandes et al. (2011). São 42 694 galáxias com formação estelar, 32 193 galáxias com núcleo ativo forte, 11 972 galáxias com núcleo ativo fraco, 24 385 galáxias aposentadas e 13 425 galáxias passivas. O mesmo diagrama, porém com a cor dos pontos representando a cor NUV – r das galáxias (Figura 4.4), mostra que a cor UV das galáxias está relacionada à sua classe.

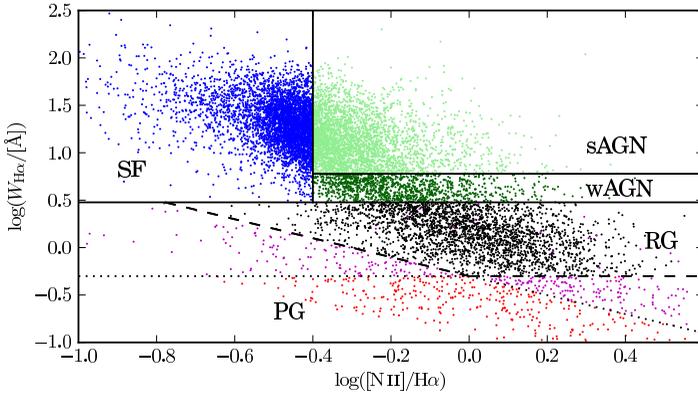


Figura 4.3: Diagrama de diagnóstico WHAN. As linhas tracejadas separam as galáxias em classes. **Azul:** galáxias com formação estelar (SF). **Verde claro:** galáxias com núcleo ativo forte (sAGN). **Verde forte:** galáxias com núcleo ativo fraco (wAGN). **Preto:** galáxias aposentadas (RG). **Vermelho:** galáxias passivas (PG). **Magenta:** Galáxias que não se encaixam em nenhuma destas classes.

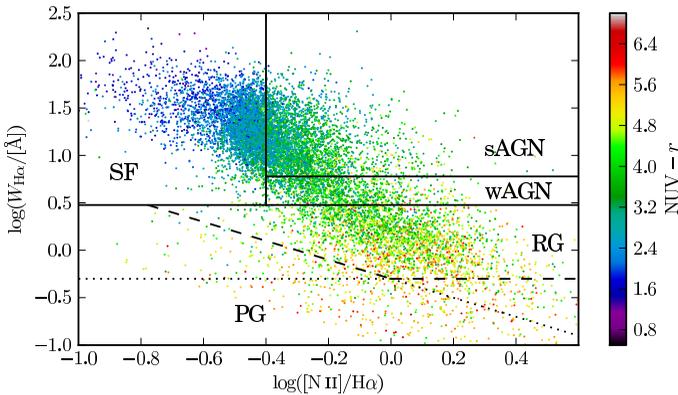


Figura 4.4: Diagrama WHAN semelhante ao da Figura 4.3. A cor dos pontos representa $\text{NUV} - r$. Pode-se notar que a cor UV das galáxias é diferente para cada classe.

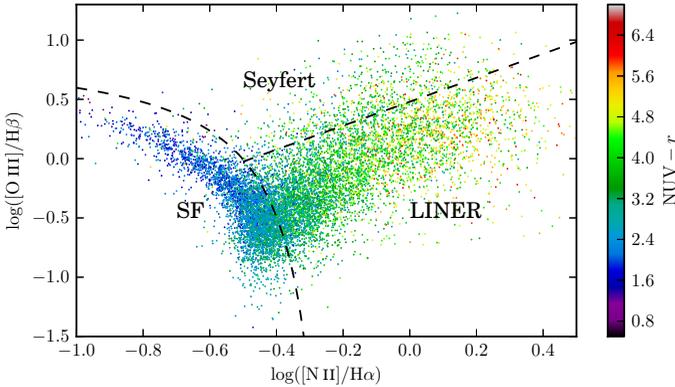


Figura 4.5: Diagrama BPT, também usado para classificar galáxias. A cor dos pontos representa $NUV - r$. As linhas tracejadas separam as galáxias nas classes *Seyfert*, *LINER* e formação estelar, conforme Cid Fernandes et al. (2010, linhas S06 e K06 da tabela 1). As classes possuem cores UV consistentes com a classificação proveniente do diagrama WHAN.

Outra forma de classificar as galáxias é através das razões entre linhas de emissão $[N II]/H\alpha$ e $[O III]/H\beta$. Este diagrama, conhecido como BPT (Baldwin, Phillips, & Terlevich 1981) é bastante utilizado na astronomia extragaláctica. Em Cid Fernandes et al. (2010), discute-se os detalhes da classificação das galáxias utilizando o diagrama BPT. As linhas tracejadas separam as galáxias nas classes *Seyfert* (que corresponde à classe sAGN na classificação pelo WHAN), *LINER*¹ (correspondente às wAGN e aposentadas) e galáxias com formação estelar (SF). A cor UV das galáxias da amostra no diagrama BPT (Figura 4.5) é consistente com as cores para o diagrama WHAN.

As classes de galáxias ocupam regiões distintas do diagrama cor–cor na Figura 4.6. A cor dos pontos representa a classe das galáxias utilizando o mesmo código de cores da Figura 4.3. Embora não esteja muito claro para as RG e PG, as classes formam uma sequência neste diagrama. Isto pode ser visto mais facilmente na Figura 4.7. Esta figura é de certo modo uma versão resumida da Figura 4.6. Mesmo havendo uma sobreposição considerável entre as classes, a sequência está bem definida.

¹*Low-Ionization Nuclear Emission-Line Region*: região nuclear com linhas de emissão de baixa ionização.

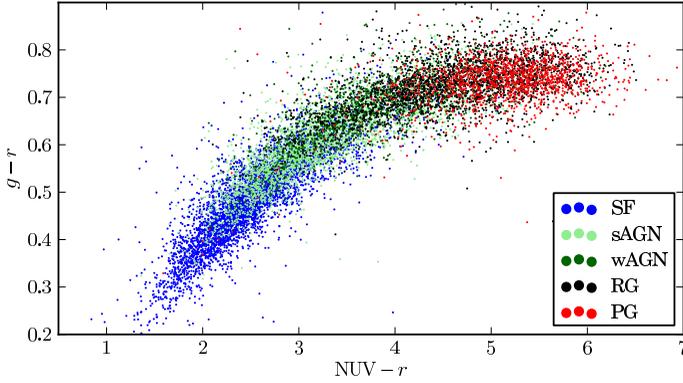


Figura 4.6: Classes de galáxias no diagrama cor-cor UV. As cores referentes às classes de galáxia são as mesmas do diagrama WHAN (Figura 4.3). É possível notar uma separação entre as classes, embora haja uma sobreposição severa entre elas.

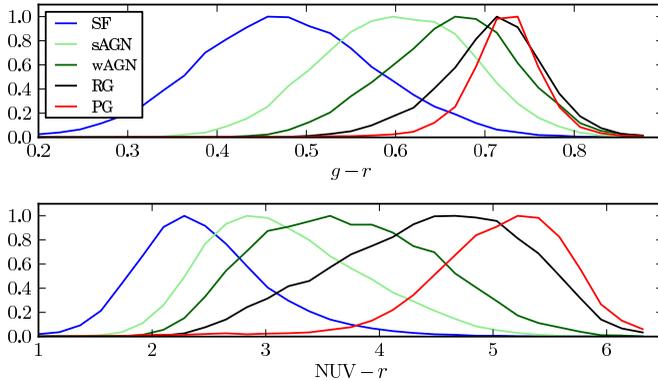


Figura 4.7: Histogramas normalizados das cores óptica ($g-r$) e ultravioleta ($NUV-r$) para as classes de galáxias. A cor das linhas representa a classe conforme a Figura 4.3. Em ultravioleta aparece uma separação entre as classes de galáxias aposentadas (RG) e passivas (PG). Note que os histogramas seguem o agrupamento das classes na Figura 4.6.

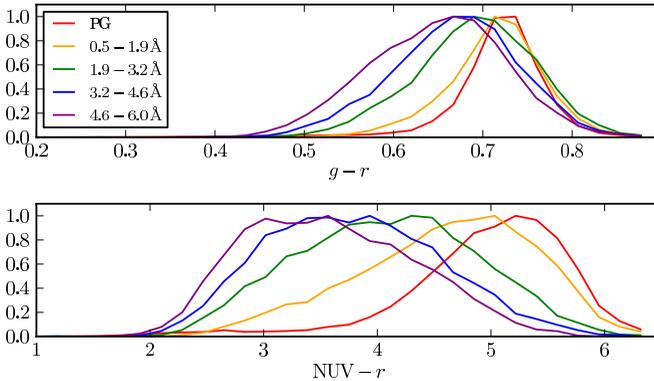


Figura 4.8: Histogramas normalizados das cores óptica ($g-r$) e ultravioleta ($NUV-r$) para várias faixas de $W_{H\alpha}$. Para comparação, é mostrado (em vermelho) também o histograma para as PG. Foram selecionadas galáxias com $\log([N II]/H\alpha) < -0,4$ e $0,5 \text{ \AA} < W_{H\alpha} < 6 \text{ \AA}$, ou seja, todas as galáxias wAGN e RG.

É interessante notar que a separação entre as classes é mais pronunciada em UV. As classes RG e PG mal se distinguem no óptico, enquanto que em UV as RG se mostram consistentemente mais azuis². Através de uma inspeção visual pode-se inferir, ainda que qualitativamente, que as distribuições em UV são significativamente assimétricas em comparação com as distribuições no óptico. Isto pode ser indício de uma contaminação entre as classes, evidenciada agora pelos novos dados em UV.

O mesmo pode ser dito sobre as distribuições de galáxias RG e wAGN, cuja diferença no óptico é magnificada em $NUV-r$, embora, como no caso das RG e PG, persista uma grande superposição. Isto não é surpreendente, já que as classes são definidas de modo completamente independente das cores ópticas e UV. Além disso, qualquer linha divisória entre uma classe e outra é inevitavelmente difusa, devido a erros observacionais e ao próprio fato de que os fenômenos que representam duas classes podem coexistir. Nada impede, por exemplo, que uma galáxia seja aposentada, no sentido de não formar mais estrelas, e possua um núcleo ativo fraco alimentando-se de tão pouco gás que não seja suficiente para formar estrelas.

Para apreciar com um pouco mais de detalhe a diferença em UV entre as galáxias das

²O conceito de “mais azul” não está relacionado à percepção de cor, mas sim ao fato de que um objeto emite mais luz em comprimentos de onda menores.

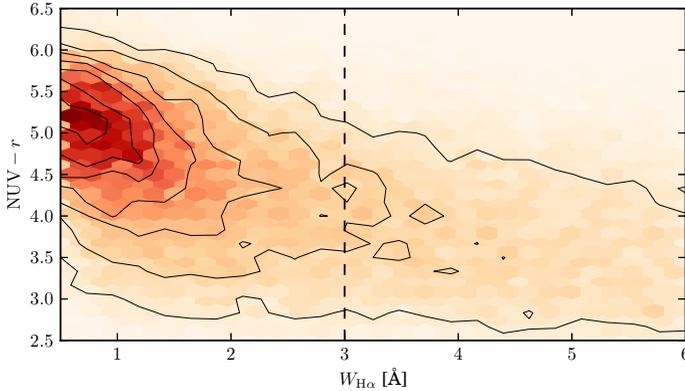


Figura 4.9: Frequência de galáxias em função da cor UV e de $W_{H\alpha}$ para as mesmas galáxias da figura 4.8, ou seja, $\log([N II]/H\alpha) < -0,4$ e $0,5 \text{ \AA} < W_{H\alpha} < 6 \text{ \AA}$. A linha tracejada indica a fronteira entre RG ($W_{H\alpha} < 3 \text{ \AA}$) e wAGN ($W_{H\alpha} > 3 \text{ \AA}$)

classes wAGN, RG e PG, a Figura 4.8 repete os histogramas da Figura 4.7, mas desta vez limitada a objetos com $\log([N II]/H\alpha) > -0,4$ e $W_{H\alpha} < 6 \text{ \AA}$, excluindo, portanto, galáxias SF e sAGN. As diferentes cores representam faixas em $W_{H\alpha}$, com uma quantidade de subdivisões suficiente para mostrar a tendência sem comprometer a clareza. O que se observa na figura é uma progressão gradual e unívoca na direção de cores mais vermelhas a medida em que $W_{H\alpha}$ diminui. As galáxias RG mais azuis são aquelas que têm os maiores valores de $W_{H\alpha}$ permitidos para esta classe ($W_{H\alpha} < 3 \text{ \AA}$). De modo similar, as galáxias wAGN que têm $W_{H\alpha}$ próximo à fronteira entre wAGN e RG são aqueles com cores mais vermelhas. Esses resultados também podem ser vistos na Figura 4.9, onde as mesmas galáxias são apresentadas em um diagrama $NUV - r$ contra $W_{H\alpha}$.

Estas figuras corroboram a suspeita levantada acima de que a superposição em $NUV - r$ entre classes espectrais definidas com base em $W_{H\alpha}$ é, em grande medida, devida ao conflito entre uma descrição discreta e uma realidade contínua. Neste caso limites rígidos de classificação têm apenas um sentido estatístico. Ao mesmo tempo, na medida em que revelam uma correlação entre $W_{H\alpha}$ e $NUV - r$, as Figuras 4.8 e 4.9 reforçam o uso de $W_{H\alpha}$ como uma propriedade observacional útil para classificação. Como salientado por Cid Fernandes et al. (2011), o uso de uma largura equivalente vai contra a história da classificação espectral baseada em linhas de emissão (baseada exclusivamente em razões de fluxos), mas é a única maneira de

separar AGNs verdadeiros de falsos.

4.3 Propriedades físicas no diagrama cor–cor

Os agentes por trás das linhas de emissão das galáxias são fundamentalmente diferentes (estrelas jovens, núcleos ativos e HOLMES) para cada classe. Não é de se estranhar, portanto, que as propriedades físicas das galáxias estejam relacionadas à sua classificação.

A idade estelar média para cada galáxia ponderada em fluxo e em massa (média sobre as suas SSP componentes), é mostrada nas Figuras 4.10 e 4.11. O painel (a) destas figuras mostra histogramas normalizados para cada classe, semelhantes aos obtidos por Cid Fernandes et al. (2011, figura 10). Os painéis de (b) até (f) mostram o diagrama cor–cor, com eixos idênticos aos da Figura 4.1, para as galáxias de cada classe. Como é esperado, as galáxias SF têm estrelas em média mais jovens do que as demais, seguidas das sAGN, wAGN, RG e PG, que consistem basicamente de populações de estrelas velhas. Em geral, as galáxias na região azul do diagrama cor–cor têm predominantemente populações jovens, enquanto galáxias vermelhas têm populações estelares velhas.

A metalicidade estelar, ponderada em fluxo, das PG, RG e wAGN é em média superior à metalicidade solar e são virtualmente indistinguíveis, conforme os painéis (d), (e) e (f) da Figura 4.12. As SF e sAGN, respectivamente nos painéis (b) e (c) da mesma figura, por outro lado têm metalicidades sub-solares. No diagrama cor–cor, as galáxias azuis têm em geral metalicidade estelar mais baixa, embora não haja uma sequência tão clara quanto para a idade estelar. A Figura 4.13 mostra a mesma relação para a metalicidade estelar ponderada em massa, porém neste caso as sAGN têm metalicidade mais próxima das wAGN.

A massa estelar das SF é consideravelmente menor do que as outras classes (Figura 4.14). Estas têm praticamente a mesma distribuição de massa estelar, com as sAGN ligeiramente menos massivas em média. Como visto, as SF são em geral mais azuis, a distribuição de massa estelar no diagrama cor–cor mostra uma certa tendência para massas menores no extremo azul.

A atenuação por poeira na banda V no diagrama cor–cor pode ser vista para cada classe de galáxias na Figura 4.15. A atenuação também forma uma sequência: PG, RG, wAGN e sAGN, em ordem crescente de atenuação, com as SF tendo atenuação similar às sAGN. A cor também está correlacionada com a atenuação, com galáxias azuis tendo atenuação maior do que as vermelhas. Uma coisa que chama a atenção é o fato de haver galáxias com atenuação negativa. Isto pode indicar que a galáxia tem populações mais vermelhas do que as existentes

Média do logaritmo da idade estelar [a] ponderada em fluxo

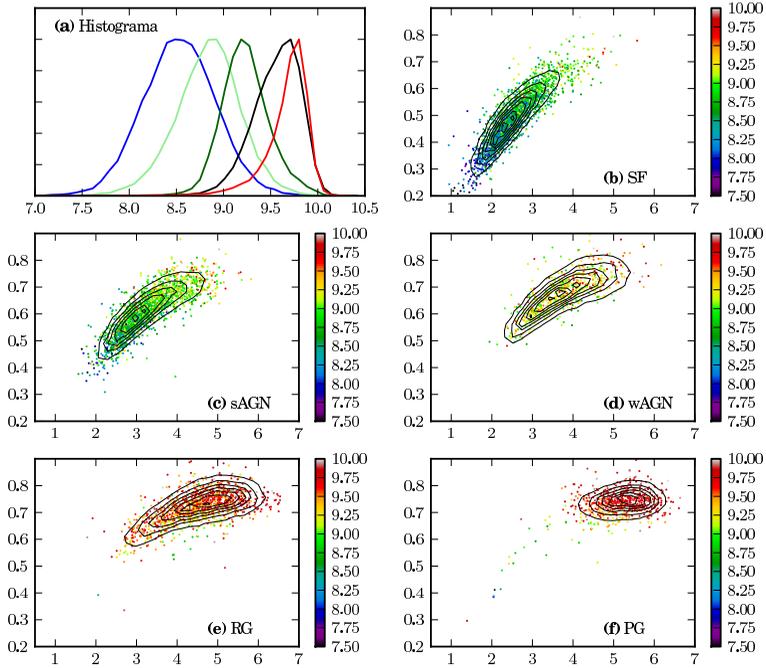


Figura 4.10: Média do logaritmo da idade estelar das galáxias, ponderada pelo fluxo, em função de $NUV - r$ e $g - r$. O painel (a) mostra o histograma normalizado das idades para cada classe, utilizando o mesmo código de cores da Figura 4.3. Os painéis de (b) a (f) são diagramas cor–cor com eixos iguais aos da Figura 4.1 código de cor indicando a idade, contendo somente as galáxias com formação estelar (SF), AGN fortes (sAGN), AGN fracas (wAGN), galáxias aposentadas (RG) e galáxias passivas (PG), respectivamente. Os contornos indicam a densidade de galáxias.

Média do logaritmo da idade estelar [a] ponderada em massa

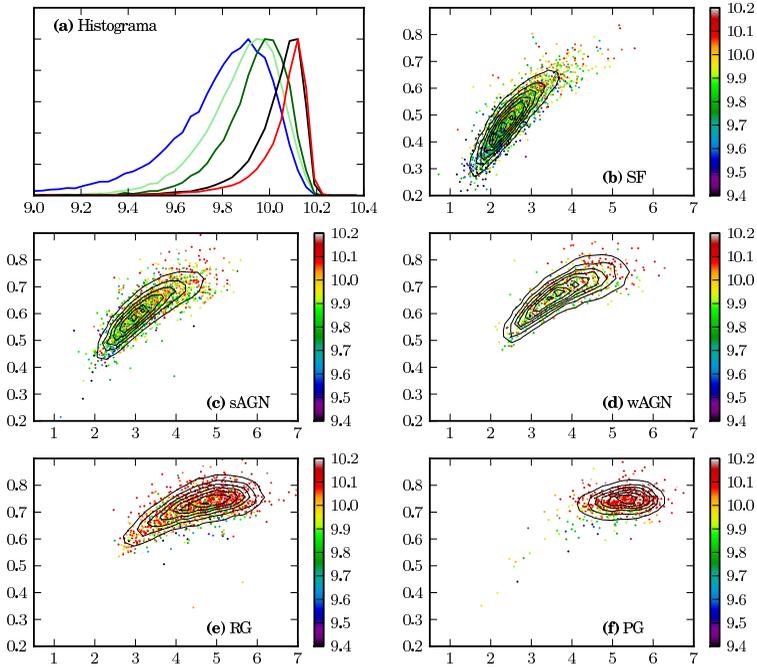


Figura 4.11: O mesmo que a Figura 4.10, para a média do logaritmo da idade estelar das galáxias, ponderada pela massa. Note que a escala de idades não é a mesma.

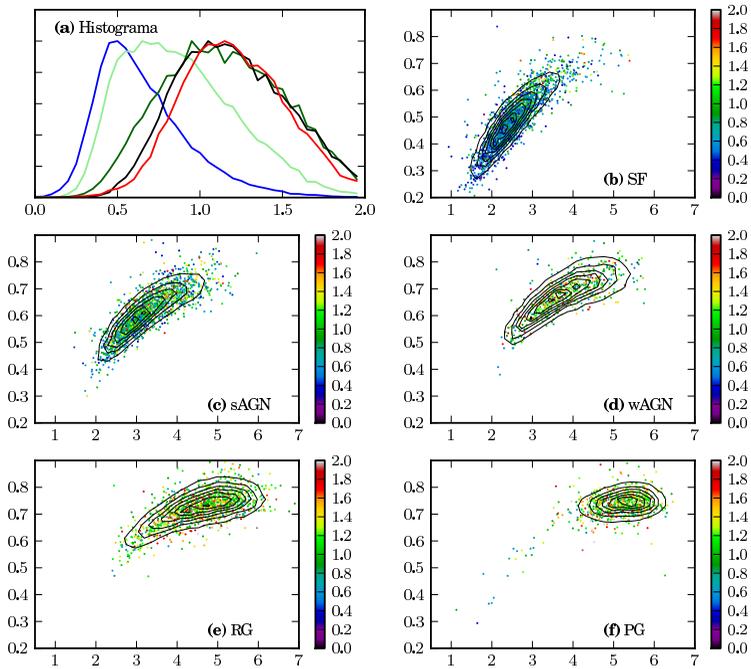
Metallicidade estelar média $[Z_{\odot}]$ ponderada em fluxo

Figura 4.12: O mesmo que a Figura 4.10, para a metalicidade estelar média das galáxias, ponderada pelo fluxo.

Metallicidade estelar média $[Z_{\odot}]$ ponderada em massa

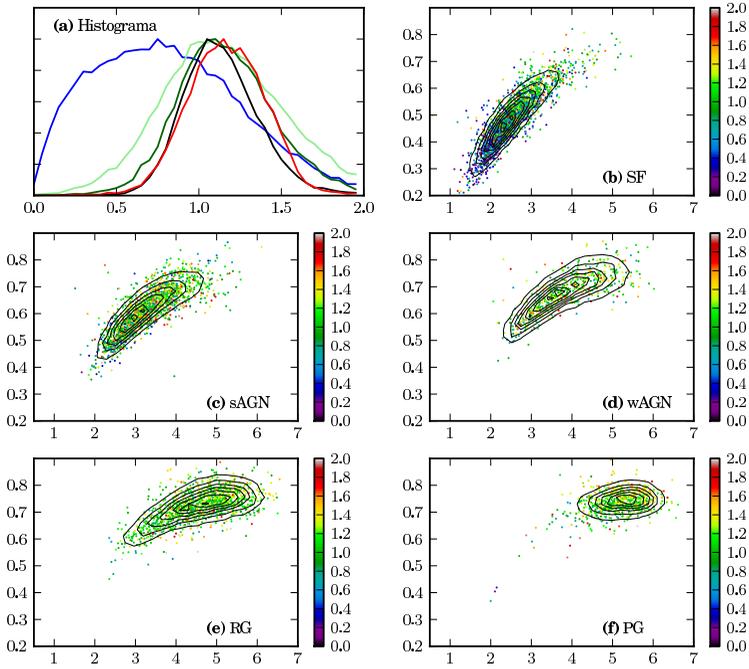


Figura 4.13: O mesmo que a Figura 4.10, para a metalicidade estelar média das galáxias, ponderada pela massa.

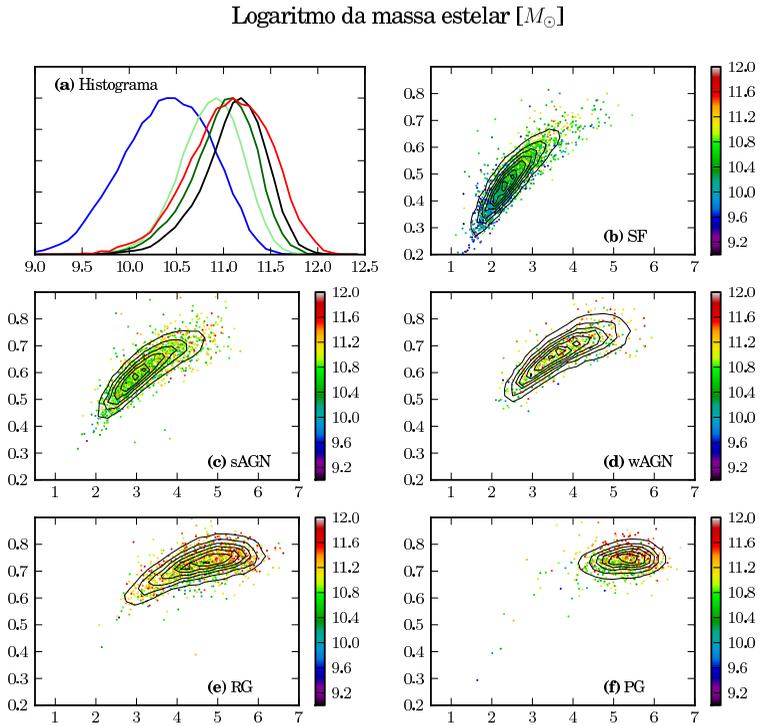


Figura 4.14: O mesmo que a Figura 4.10, para o logaritmo da massa estelar das galáxias em massas solares.

Extinção por poeira [magnitude]

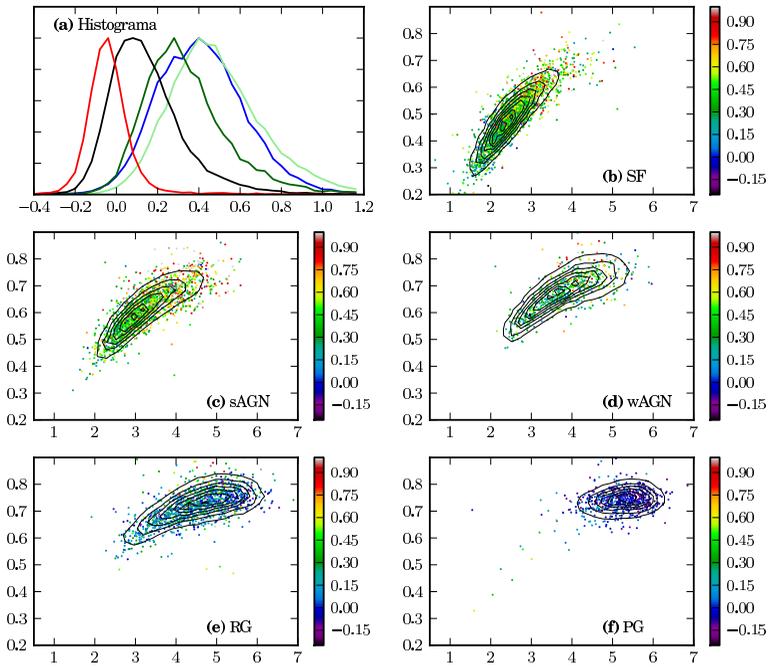


Figura 4.15: O mesmo que a Figura 4.10, para a atenuação por poeira na banda V das galáxias.

na base de SSP. Como o espectro da galáxia é mais vermelho do que todos os espectros da base, o excesso de vermelho acaba sendo compensado através de um “azulamento” do espectro, que é equivalente a um avermelhamento A_V negativo. Também, galáxias com A_V próximo a zero e com um espectro muito ruidoso podem ter um ajuste melhor desta forma. Este resultado só ocorre porque A_V é um parâmetro livre na síntese do STARLIGHT.

Em geral pode-se notar que a cor UV das galáxias está relacionada de alguma forma com as suas propriedades físicas. Para algumas propriedades, como a idade estelar média, esta relação é bastante pronunciada. Os novos dados observacionais, isto é, a magnitude NUV das galáxias, estão de acordo com o que seria esperado com base numa classificação feita por linhas de emissão, e também com a síntese feita pelo STARLIGHT, ambos provenientes dos espectros óticos do SDSS. A concordância destas informações vindas de comprimentos de onda diferentes (e *surveys* diferentes) fortalece a validade tanto do esquema de classificação quanto da síntese.

Capítulo 5

Conclusões e perspectivas

5.1 Este trabalho

Este estudo teve como objetivo principal explorar o uso de bancos de dados em astronomia. Com a previsão de um dilúvio de dados, é de extrema importância que sejam desenvolvidas técnicas e ferramentas adequadas para absorver este volume de dados.

A importação do catálogo do STARLIGHT para um banco de dados relacional foi uma tarefa relativamente simples. O uso de um banco de dados relacional, neste caso *MS SQL Server*, facilitou muito o gerenciamento do catálogo. As diversas ferramentas disponíveis permitem que se manipule os dados de uma forma eficiente. Em especial, grande parte do presente trabalho foi feito nos ambientes *CasJobs*, tanto do STARLIGHT quanto do *SDSS* e do *GALEX*. O catálogo tem sido usado intensamente por colaboradores do STARLIGHT em São Paulo e na Espanha, além de ter diversos usuários espalhados pelo mundo (Figura 5.1).

Por si só, o uso dos dados provenientes de um único *survey* de grande porte tem o potencial de gerar resultados até então inimagináveis. A identificação cruzada de objetos em *surveys* diferentes, por sua vez, promete ampliar, e muito, este espaço de possibilidades. O esforço de Budavári et al. (2009), utilizando as técnicas de indexação espacial descritas por Kunszt et al. (2001), na identificação cruzada de objetos entre *SDSS* e *GALEX* é de suma importância neste sentido: a maioria dos *surveys* atuais provê (ou planeja) formas de integração com o *SDSS* e outros *surveys*. Esta informação cruzada foi utilizada para obter os dados fotométricos em UV dos objetos do catálogo do STARLIGHT. Assim, o banco de dados agora contém as propriedades físicas de quase um milhão de galáxias, os seus correspondentes parâmetros observacionais provenientes do *SDSS*, e as magnitudes UV absolutas e corrigidas (poeira e correção $k(z)$) para



Figura 5.1: Localização dos acessos ao *CasJobs* do STARLIGHT. O número de acessos é referente ao mês de novembro de 2011.

aproximadamente 170 mil galáxias, obtidas pelo *GALEX*.

Como um estudo de caso, foi montada uma amostra de galáxias do catálogo do STARLIGHT, com as suas propriedades físicas, algumas linhas de emissão e fotometria óptica e UV (chamada amostra STARLIGHT+UV). Dados os critérios de seleção da amostra, a extração da amostra resume-se a executar uma *query* no *CasJobs* e obter o seu resultado em um arquivo no formato desejado, FITS por exemplo. É claro, este processo pode ser repetido caso os critérios de seleção mudem (o que foi feito muitas vezes neste estudo, diga-se de passagem). Ainda assim, a ideia é abandonar a cultura de ter todo o conjunto de dados do catálogo presentes no computador do usuário. Com os catálogos imensos que estão sendo planejados (J-PAS e LSST, por exemplo), é bastante provável que este seja o caminho mais eficiente a se seguir.

No capítulo 4 mostrou-se que as classes de galáxias da amostra STARLIGHT+UV, segundo os critérios de classificação descritos por Cid Fernandes et al. (2011), tendem a ocupar regiões distintas do diagrama cor-cor. Pôde-se notar, nos diagramas cor-cor para as propriedades físicas das galáxias, que a distribuição de RG tem um espalhamento considerável na direção das AGN e SF. Com um exercício simples, foi possível verificar que este espalhamento é devido não somente à incerteza nas medidas, mas também por se utilizar uma classificação discreta sobre fenômenos que são de natureza contínua.

Tecnicamente, as PG e as RG deveriam ser essencialmente as mesmas, com as RG apresentando linhas de emissão fracas. Em geral as propriedades físicas destas duas classes apresen-

tam uma sobreposição considerável, como pode ser visto nos diagramas cor–cor (Figuras 4.10 até 4.15), e são distintas das duas classes AGN. Ainda assim, é preciso ter em mente de que parece haver diferenças sistemáticas entre as PG e as RG. É preciso estudar estas diferenças mais a fundo.

5.2 Trabalhos futuros

O estudo da amostra *STARLIGHT*+UV presente neste trabalho foi feito de forma superficial. A intenção foi apenas demonstrar uma forma de trabalhar com dados provenientes de diversos catálogos. As observações feitas no capítulo 4 são em geral de caráter qualitativo. Entretanto, os dados sugerem que existe uma contaminação entre as classes de galáxias. Um provável candidato é o erro na medida de $W_{H\alpha}$ na amostra. Isto causa uma contaminação de galáxias SF e wAGN nas RG. É preciso determinar se o erro nas medidas é, sozinho, capaz de explicar esta contaminação, ou se ela é inerente à classificação simplificada adotada.

A síntese feita pelo *STARLIGHT* levou em conta apenas a região óptica do espectro, mas os modelos de SSP utilizados possuem dados que se estendem até além do UV. É possível calcular a magnitude UV esperada para uma galáxia dados os seus componentes determinados na síntese. Porém, para poder comparar os dados previstos pela síntese com os observados, é preciso garantir que os dados são compatíveis. A fotometria do *GALEX* é integrada sobre toda a galáxia. Já os espectros do *SDSS*, foram obtidos utilizando uma fibra com aproximadamente 3 segundos de arco. Cada fibra não captura, necessariamente, toda a luz das galáxias. Em geral a fração da luz de uma galáxia capturada pela fibra varia com o tamanho da galáxia e o seu *redshift*. Assim, por exemplo, regiões de formação estelar, com emissão em UV significativa, que estão na periferia de uma galáxia relativamente próxima podem ter sido excluídas. Para fazer uma comparação adequada é necessário selecionar uma amostra de galáxias mais distantes, tal que uma fração considerável da sua luz caia dentro da fibra. Esta comparação pode ajudar a determinar a acurácia dos modelos de SSP em UV, além de testar a lei de atenuação usada na síntese¹. Outra possibilidade é usar a fotometria UV desta amostra de galáxias como pontos extras do espectro de entrada do *STARLIGHT*. Desta forma espera-se que o ajuste possa ser mais sensível à formação estelar recente, e possa separar uma formação estelar fraca de objetos velhos e quentes, como anãs brancas, por exemplo. A versão mais recente do *STARLIGHT* já está preparada para trabalhar com dados fotométricos em conjunto com espectrais.

¹As leis de atenuação Cardelli, Clayton, & Mathis (1989, CCM) e Calzetti, Kinney, & Storchi-Bergmann (1994) têm previsões diferentes em UV devido à presença ou não da corcova em 2175 Å.

Embora a seleção da amostra tenha sido feita no banco de dados, uma boa parte da análise foi feita num computador pessoal. Isto custou diversas iterações de extração e diagnóstico da amostra. Alguns procedimentos como a classificação das galáxias e histogramas de propriedades físicas poderia ser feita diretamente no servidor. O próximo passo, com relação ao banco de dados do STARLIGHT, é criar ferramentas de diagnóstico que rodem no servidor. A ideia é que estas ferramentas possam ser facilmente estendidas e abstraíam a forma como as informações estão armazenadas no banco de dados.

Esta forma de trabalho deve se tornar a regra nos próximos anos. Pode-se pensar que será possível armazenar dezenas de petabytes obtidas pelo LSST no computador do pesquisador, dado o crescimento que vem ocorrendo na densidade de armazenamento. Entretanto, é preciso cautela. Way (2011) compara o crescimento da densidade de armazenamento com a largura de banda de entrada/saída (E/S) e a velocidade de processamento. Embora uma *query* em geral escale linearmente com o tamanho de um banco de dados, as operações científicas comumente aplicadas aos dados incluem classificação, cálculo de distância e densidade, por exemplo, e são geralmente de complexidade $O(N^2)$. Ou seja, mesmo que seja possível armazenar todo o conteúdo de um *survey* deste porte, é bastante provável que uma abordagem tradicional, acessando os dados de forma sequencial, esbarre na limitação de E/S e poder computacional. Tudo indica que a aposta mais segura seja manter os dados em centros de processamento de dados (*data centers*) poderosos, possivelmente espalhados geograficamente para melhorar a velocidade do acesso.

Apêndice A

Queries SQL

Queries SQL utilizadas no texto. As palavras-chave da linguagem estão em negrito e maiúsculas. Os servidores *CasJobs* referenciados na legenda das figuras estão listados abaixo.

STARLIGHT: <http://casjobs.starlight.ufsc.br/casjobs/>
SDSS: <http://casjobs.sdss.org/CasJobs/>
GALEX: <http://galex.stsci.edu/casjobs/>

```
UPDATE sample
  SET SpecObjID=so.SpecObjID, ObjID=so.BestObjID
FROM sample s2 INNER JOIN DR7..SpecObjAll so
  ON so.MJD=s2.MJD
  AND so.Plate=s2.Plate
  AND so.FiberID=s2.FiberID
```

Figura A.1: Atualização dos índices da amostra de galáxias do STARLIGHT. A *query* foi executada no *CasJobs* do SDSS para obter SpecObjID e BestObjID dado o tripuleto [MJD, Plate, FiberID].

```

SELECT
  'AIS' AS survey, '1' AS mg, '1' AS ms, COUNT(*) AS N
FROM xsdssdr7 x
INNER JOIN photoobjall p ON x.objid = p.objid
INNER JOIN photoextract pe ON p.photoextractid = pe.photoextractid
WHERE x.multipleMatchCount = 1 AND x.reverseMultipleMatchCount = 1
      AND pe.mpstype='AIS'

UNION SELECT
  'AIS' AS survey, '1' AS mg, '2' AS ms, COUNT(*) AS N
FROM xsdssdr7 x
INNER JOIN photoobjall p ON x.objid = p.objid
INNER JOIN photoextract pe ON p.photoextractid = pe.photoextractid
WHERE x.multipleMatchCount = 1 AND x.reverseMultipleMatchCount = 2
      AND pe.mpstype='AIS'

UNION SELECT
  'AIS' AS survey, 1 AS mg, '+' AS ms, COUNT(*) AS N
FROM xsdssdr7 x
INNER JOIN photoobjall p ON x.objid = p.objid
INNER JOIN photoextract pe ON p.photoextractid = pe.photoextractid
WHERE x.multipleMatchCount = 1 AND x.reverseMultipleMatchCount > 2
      AND pe.mpstype='AIS'

UNION SELECT
  ...

```

Figura A.2: Lista das identificações mútuas entre *SDSS DR7* e os *surveys* do *GALEX*. São contados quantos objetos de um dado *survey* têm identificação direta e reversa com apenas 1, 2 ou mais de objetos. Os campos *mg* e *ms* representam respectivamente o número de candidatos para o *GALEX* e para o *SDSS*. Ver seção 3.4.1 para mais detalhes sobre a identificação cruzada. Apenas uma parte da *query* foi incluída, o restante é similar ao código presente. *Query* executada no *CasJobs* do *GALEX*.

```
SELECT INTO mydb..galex_ais
    s.objid AS sdssobjid, x.objid AS galexobjid,
    s.mjd, s.plate, s.fiberid,
    g.fuv_mag, fuv_magErr,
    g.nuv_mag, g.nuv_magErr,
    g.e_bv,
    g.band,
    x.distance,
    pe.fexptime,
    pe.nexptime
FROM mydb..sample s
LEFT JOIN xSDSSDR7 x
    ON s.objid = x.sdssobjid
    AND x.distanceRank=1
    AND x.reverseDistanceRank=1
    AND x.multipleMatchCount=1
    AND x.reverseMultipleMatchCount=1
LEFT JOIN photoobjall g
    ON g.objid = x.objid
LEFT JOIN photoextract e
    ON e.photoextractid=g.photoextractid
WHERE e.mpstype='AIS'
```

Figura A.3: Query para o match entre os objetos da amostra do STARLIGHT e GALEX AIS. A mesma query foi usada para o MIS, trocando apenas o nome da tabela para galex_mis e modificando a última linha para e.mpstype='MIS'. Query executada no CasJobs do GALEX.

```

SELECT
    'AIS' AS survey, 'FUV' AS band, COUNT(*) AS N
FROM galex_ais
WHERE galexobjid <> 0
    AND fuv_mag <> -999
UNION SELECT
    'AIS' AS survey, 'NUV' AS band, COUNT(*) AS N
FROM galex_ais
WHERE galexobjid <> 0
    AND nuv_mag <> -999
UNION SELECT
    'AIS' AS survey, 'FUV+NUV' AS band, COUNT(*) AS N
FROM galex_ais
WHERE galexobjid <> 0
    AND fuv_mag <> -999
    AND nuv_mag <> -999

UNION SELECT
    'MIS' AS survey, 'FUV' AS band, COUNT(*) AS N
FROM galex_mis
WHERE galexobjid <> 0
    AND fuv_mag <> -999
UNION SELECT
    'MIS' AS survey, 'NUV' AS band, COUNT(*) AS N
FROM galex_mis
WHERE galexobjid <> 0
    AND uv_mag <> -999
UNION SELECT
    'MIS' AS survey, 'FUV+NUV' AS band, COUNT(*) AS N
FROM galex_mis
WHERE galexobjid <> 0
    AND fuv_mag <> -999
    AND nuv_mag <> -999

```

Figura A.4: Lista contendo quantidade de objetos no catálogo STARLIGHT +UV com detecções *GALEX* nas bandas FUV, NUV e em ambas. Valores de -999 indicam que o a coluna está indefinida. Valores de *galexObjID* iguais a zero indicam que o objeto STARLIGHT não tem correspondente *GALEX*. *Query* executada no *CasJobs* do STARLIGHT.

```

SELECT INTO MyDB..galex_ais_elines_z
    s.specobjid,
    g.nuv_mag AS NUV,
    o.Mu, o.Mg, o.Mr, o.Mi, o.Mz,
    o.m_u, o.m_g, o.m_r, o.m_i, o.m_z,
    s.mcor_gal, s.at_flux, s.at_mass, s.am_flux,
    s.am_mass, s.AV,
    e.oiii_5007_flux, e.oiii_5007_flux_err,
    e.oiii_5007_ew, e.oiii_5007_ew_err, e.oiii_5007_sn,
    e.nii_6584_flux, e.nii_6584_flux_err,
    e.nii_6584_ew, e.nii_6584_ew_err, e.nii_6584_sn,
    e.halpha_flux, e.halpha_flux_err,
    e.halpha_ew, e.halpha_ew_err, e.halpha_sn,
    e.hbeta_flux, e.hbeta_flux_err,
    e.hbeta_ew, e.hbeta_ew_err, e.hbeta_sn,
    o.z AS redshift
FROM galex_ais g
INNER JOIN synthesis_results s ON
    s.specobjid = g.specobjid
INNER JOIN el_fit_all e ON
    s.synid = e.synid
INNER JOIN obs_parameters o ON
    o.specobjid = s.specobjid
WHERE
    g.galexobjid <> 0
    AND g.nuv_mag <> -999
    AND o.z > 0.04
    AND o.z < 0.17
    AND o.m_r < 17.77

```

Figura A.5: Extração da amostra STARLIGHT+UV. Os limites em *redshift* ($0,04 < z < 0,17$) e magnitude ($r < 17,77$) são explicados na seção 3.5. O critério “*g.nuv_mag* <> -999” remove os objetos que não tiveram identificação em NUV. *Query* executada no *CasJobs* do STARLIGHT.

Referências Bibliográficas

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F. et al. 2009, *ApJS*, 182, 543
- Apte, C., ed. 2011, *KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA: ACM), 618112
- Asari, N. V. 2010, PhD thesis, Curso de Pós-Graduação em Física da Universidade Federal de Santa Catarina & École Doctorale d'Astronomie et Astrophysique d'île-de-France, Observatoire de Paris
- Asari, N. V., Cid Fernandes, R., Stasińska, G., Torres-Papaqui, J. P., Mateus, A., Sodré, L., Schoenell, W., & Gomes, J. M. 2007, *MNRAS*, 381, 263
- Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, *ApJ*, 600, 681
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Bell, G., Gray, J., & Szalay, A. 2006, *Computer*, 39, 110
- Benítez, N., Gaztañaga, E., Miquel, R., Castander, F., Moles, M., Crocce, M., Fernández-Soto, A., Fosalba, P. et al. 2009, *ApJ*, 691, 241
- Bian, W., Gu, Q., Zhao, Y., Chao, L., & Cui, Q. 2006, *MNRAS*, 372, 876
- Blanton, M. R. & Roweis, S. 2007, *AJ*, 133, 734
- Boksenberg, A., Evans, R. G., Fowler, R. G., Gardner, I. S. K., Houziaux, L., Humphries, C. M., Jamar, C., Macau, D. et al. 1973, *MNRAS*, 163, 291
- Born, A. J. 2010, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7738*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J. et al. 2010, *Science*, 327, 977
- Bowyer, S., Sasseen, T. P., Lampton, M., & Wu, X. 1993, *ApJ*, 415, 875

- Bruzual, G. & Charlot, S. 2003, *MNRAS*, 344, 1000
- Budavári, T., Heinis, S., Szalay, A. S., Nieto-Santisteban, M., Gupchup, J., Shiao, B., Smith, M., Chang, R. et al. 2009, *ApJ*, 694, 1281
- Budavári, T. & Szalay, A. S. 2008, *ApJ*, 679, 301
- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *ApJ*, 429, 582
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Carruthers, G. R. 1973, *Appl. Opt.*, 12, 2501
- Chamberlin, D. D. & Boyce, R. F. 1974, in *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control, SIGFIDET '74* (New York, NY, USA: ACM), 249–264
- Chilingarian, I. & Zolotukhin, I. 2011, *ArXiv e-prints*
- Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, *MNRAS*, 358, 363
- Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, *MNRAS*, 413, 1687
- Cid Fernandes, R., Stasińska, G., Vale Asari, N., Mateus, A., Schlickmann, M. S., Schoenell, W., & Schoenell. 2010, in *IAU Symposium, Vol. 267, IAU Symposium*, 65–72
- Codd, E. F. 1970, *Commun. ACM*, 13, 377
- Code, A. D., Houck, T. E., McNall, J. F., Bless, R. C., & Lillie, C. F. 1970, *ApJ*, 161, 377
- Colless, M. 1999, in *Large-Scale Structure in the Universe*, ed. G. Efstathiou & et al., 105
- da Costa, L. N., Pellegrini, P. S., Sargent, W. L. W., Tonry, J., Davis, M., Meiksin, A., Latham, D. W., Menzies, J. W. et al. 1988, *ApJ*, 327, 544
- Djorgovski, S. G., Baltay, C., Mahabal, A. A., Drake, A. J., Williams, R., Rabinowitz, D., Graham, M. J., Donalek, C. et al. 2008, *Astronomische Nachrichten*, 329, 263
- Gomes, J. M. 2009, PhD thesis, Curso de Pós-Graduação em Física da Universidade Federal de Santa Catarina
- Henize, K. G., Wray, J. D., Parsons, S. B., Benedict, G. F., Bruhweiler, F. C., Rybski, P. M., & Ocallaghan, F. G. 1975, *ApJ*, 199, L119
- Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, *ApJS*, 52, 89
- Huchra, J. P. 1988, in *Astronomical Society of the Pacific Conference Series, Vol. 5, The Minnesota lectures on Clusters of Galaxies and Large-Scale Structure*, ed. J. M. Dickey, 41–70

- Ivezic, Z., Tyson, J. A., Acosta, E., Allsman, R., Anderson, S. F., Andrew, J., Angel, R., Axelrod, T. et al. 2008, ArXiv e-prints
- Kondo, Y. & Wamsteker, W. 1987, Exploring the universe with the IUE satellite, *Astrophysics and space science library* (D. Reidel)
- Kunszt, P., Szalay, A. S., & R., T. A. 2001, in *Mining the sky: proceedings of the MPA/ESO/MPE Workshop, held at Garching, Germany, July 31-August 4, 2000*, ed. A. Bandy & M. Zaroubi, S. and Bartelmann, ESO astrophysics symposia (Springer), 631–637
- Kunszt, P. Z., Szalay, A. S., Csabai, I., & Thakar, A. R. 2000, in *Astronomical Society of the Pacific Conference Series, Vol. 216, Astronomical Data Analysis Software and Systems IX*, ed. N. Manset, C. Veillet, & D. Crabtree, 141–+
- Lara-López, M. A., Bongiovanni, A., Cepa, J., Pérez García, A. M., Sánchez-Portal, M., Castañeda, H. O., Fernández Lorenzo, M., & Pović, M. 2010, *A&A*, 519, A31
- Lara-López, M. A., Cepa, J., Bongiovanni, A., Pérez García, A. M., Castañeda, H., Fernández Lorenzo, M., Pović, M., & Sánchez-Portal, M. 2009, *A&A*, 505, 529
- Li, N. & Thakar, A. R. 2008, *Computing in Science Engineering*, 10, 18
- Liang, Y. C., Hammer, F., Yin, S. Y., Flores, H., Rodrigues, M., & Yang, Y. B. 2007, *A&A*, 473, 411
- Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C. et al. 2008, *MNRAS*, 389, 1179
- Maier, D., Stein, J., Otis, A., & Purdy, A. 1986, *SIGPLAN Not.*, 21, 472
- Martin, D. C., Fanson, J., Schiminovich, D., Morrissey, P., Friedman, P. G., Barlow, T. A., Conrow, T., Grange, R. et al. 2005, *ApJ*, 619, L1
- Martin, D. C., Wyder, T. K., Schiminovich, D., Barlow, T. A., Forster, K., Friedman, P. G., Morrissey, P., Neff, S. G. et al. 2007, *ApJS*, 173, 342
- Mateus, A. 2006, PhD thesis, Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo
- Morrissey, P., Conrow, T., Barlow, T. A., Small, T., Seibert, M., Wyder, T. K., Budavári, T., Arnouts, S. et al. 2007, *ApJS*, 173, 682
- Morrissey, P., Schiminovich, D., Barlow, T. A., Martin, D. C., Blakkolb, B., Conrow, T., Cooke, B., Erickson, K. et al. 2005, *ApJ*, 619, L7
- Norris, R. P. 2010, ArXiv e-prints
- O’Mill, A. L., Duplancic, F., García Lambas, D., & Sodr e, Jr., L. 2011, *MNRAS*, 413, 1395

- Peebles, M. S., Pogge, R. W., & Stanek, K. Z. 2009, *ApJ*, 695, 259
- Schiminovich, D., Wyder, T. K., Martin, D. C., Johnson, B. D., Salim, S., Seibert, M., Treyer, M. A., Budavári, T. et al. 2007, *ApJS*, 173, 315
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R. et al. 2006, *AJ*, 131, 1163
- Spanos, A. 2000, *Journal of Economic Methodology*, 7, 231
- Stecher, T. P., Cornett, R. H., Greason, M. R., Landsman, W. B., Hill, J. K., Hill, R. S., Bohlin, R. C., Chen, P. C. et al. 1997, *PASP*, 109, 584
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., Narayanan, V. K., Annis, J., Bernardi, M., Blanton, M., Burles, S. et al. 2002, *AJ*, 124, 1810
- Szalay, A. S., Gray, J., Thakar, A. R., Kunszt, P. Z., Malik, T., Raddick, J., Stoughton, C., & vandenBerg, J. 2002, in *Proceedings of the 2002 ACM SIGMOD international conference on Management of data, SIGMOD '02* (New York, NY, USA: ACM), 570–581
- Thakar, A. R., Szalay, A. S., Kunszt, P. Z., & Gray, J. 2004, eprint arXiv:cs/0403020
- van Duinen, R. J., Aalders, J. W. G., Wesselius, P. R., Wildeman, K. J., Wu, C. C., Luinge, W., & Snel, D. 1975, *A&A*, 39, 159
- Way, M. J. 2011, *ArXiv e-prints*
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D. et al. 2010, *AJ*, 140, 1868
- Wyder, T. K., Martin, D. C., Schiminovich, D., Seibert, M., Budavári, T., Treyer, M. A., Barlow, T. A., Forster, K. et al. 2007, *ApJS*, 173, 293
- York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R. et al. 2000, *AJ*, 120, 1579