

# Project - Interest Trends in Astrophysics

Parker Levesque  
parker.levesque@mail.utoronto.ca

May, 2022

## 1 Summary

The goal of this project is to investigate the growth of interest in low surface brightness galaxies. This will be done using metrics from the astrophysics data system (ads), retrieved using the ads python package.

The first part of this project will be to characterize ‘interest’. There are three main metrics that can be used in order to assign a certain level of interest when it comes to topics or subject matter in the field of astronomy: number of citations, number of papers published, and number of reads. It is difficult to narrow down interest levels and compare them between two topics, however, it is straight foreword to say that the more people are citing, reading and publishing papers related to a topic, the more the scientific community is interested as a whole.

To continue, we will reduce this data in order to discover, analyse and compare trends. For relatively new topics in astronomy, such as low surface brightness galaxies (LSBG), it can be predicted that the interest will be low and spike in the recent years. All the while more dated topics may show interesting dynamics influenced by later discoveries, causing a resurgence in interest.

## 2 Using the ADS Package

**Getting an API key and installing the package:** Creating an account with the [ADS](#) will provide a personal API key needed to use the python package. Saving the key to `~/.ads/dev_key` and running `pip install ads` in the command line will set everything up.

**Searching the Database:** The important function used to search the ADS database is the **Search-Query()** function from the ads package. Many arguments can be used to specify a query, telling the function to gather a list of papers that have been registered containing these parameters. The relevant arguments for this project are:

- **q** - Query the registered papers using keywords in substructures of the text. (e.g. title, abstract, body, year.)
- **sort** - Sorting the returned list of identified papers using a specified metric (e.g. citation count or read count)
- **fl** - Limiting the attributes fetched by the query. This helps control the rate limit usage when creating a request to the API. (e.g. title, abstract, body, citation count, read count)

Information about the papers can then be used in python through the attributes. A simple example:

```
import ads

papers_2020 = ads.SearchQuery(q='title:_galaxy_year:_2020',
                              sort='read_count', fl=['title', 'read_count'])

for paper in papers_2020:
    print(paper.title, paper.read_count)
```

## 3 The ads\_query\_pkg Package

This package, created for this project, contains four functions designed to illustrate the overall interest in a certain topic, as well as the evolution of the interest over a specified period of time.

### 3.1 cite\_count() and read\_count()

The **cite\_count** function will arrange the all-time most cited papers, given a keyword, from most to least cited. It returns the citation counts arranged in that same order. The keyword can be specified to search in the title of the paper, the abstract, or both. The **read\_count()** function will do the same but using the read count of the papers instead.

This property will show an overall representation of the interest of a given topic; getting an idea of how many greatly influential papers that have been published, and the degree to which they have influenced the scientific community.

### 3.2 top\_cite() and num\_papers()

These two functions take a keyword input and a time frame (in years). The time frame is divided into a number of time periods (decades by default), and the functions will return the total number of citations or reads in each of the time periods respectively. Again, the keyword can be specified to search in the title of the paper, the abstract, or both.

This information will show when the general interest in the topic was sparked and how it has evolved over the desired time frame. This is especially useful when comparing two underlying subjects within the umbrella of a general topic. (e.g. black-holes vs worm-holes, under the topic of general relativity).

## 4 Analysing Trends and Comparing Interest

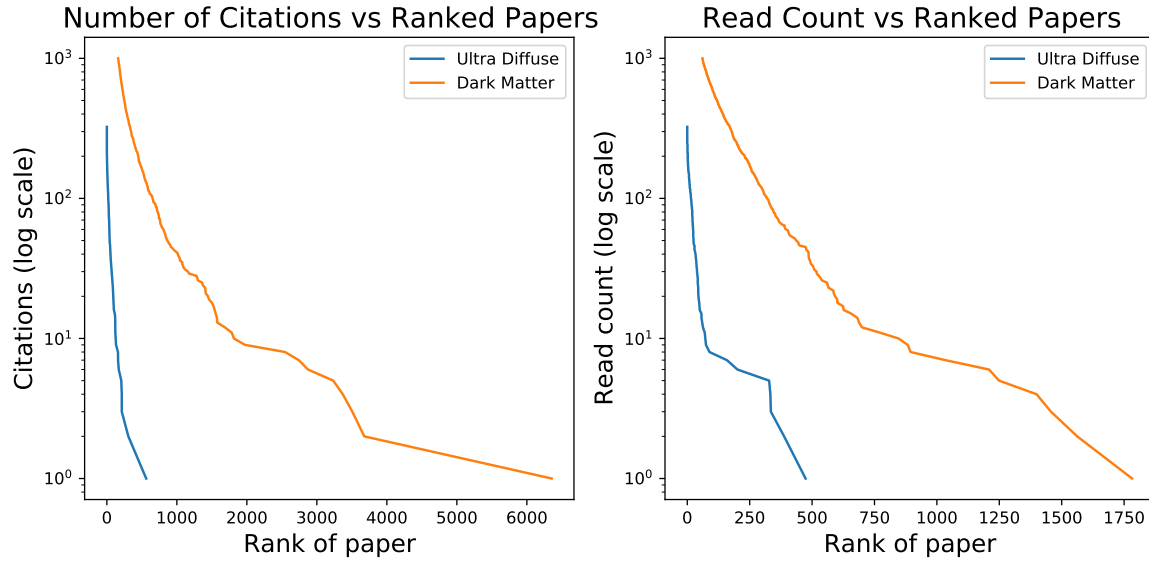
For this analysis, we will be comparing growth in interest in different subject areas related to low surface brightness galaxies.

### Case Study: Dark Matter and Ultra-Diffuse Galaxies

Due to the observed rotation curves and mass-to-light ratios of LSBGs, we know that they are dark matter dominated, making them great objects to further investigate the properties of dark matter.

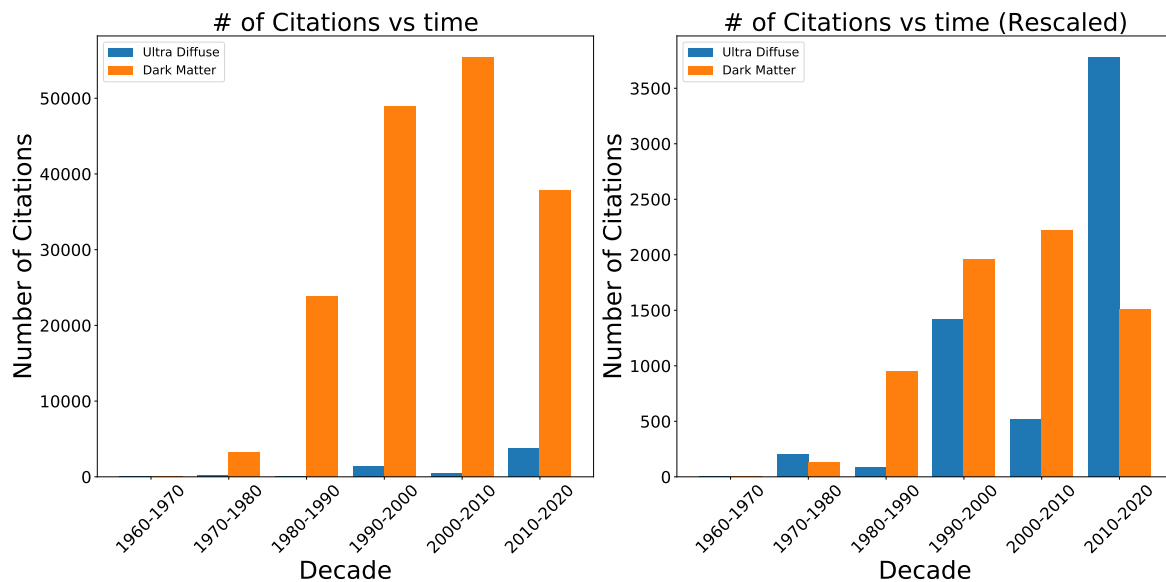
The term 'Ultra-Diffuse galaxy', coined by Dr Abraham, is a galaxy that is on the extreme end of the low-surface-brightness spectrum. Some are known to contain mostly dark matter, while others are known to contain no dark matter at all.

Let's analyse the growth in interest related to these two subject areas, which are both connected to LSBGs in their own ways.



**Figure 4.1:** (a) Left: The top cited papers of all time with the specified keyword appearing in their title, arranged from most cited to least cited. (b) Right: The top read papers of all time with the specified keyword appearing in their title, arranged from most read to least read. The number of citations is plotted against this in log scale in order to accentuate the characteristics of the curves.

As shown in Figure 4.1, we can immediately see that dark matter has been a term used amongst scientist for more time than ultra-diffuse galaxies (UDGs), allowing for it to gain appearances in more papers over the year, as well as gaining higher read and citation counts within that pool of papers. Having more time as a popular subject area, will heighten your curve and will cause it to sport a less steep gradient in log space, since less popular papers will still have a chance to gain some level of traction.



**Figure 4.2:** (a) Left: The number of citations from papers with the keywords 'Dark Matter' and 'Ultra Diffuse Galaxies' in the title, organized in decades from the 1960s to the present day. (b) Right: Same as the Left but with the Dark Matter data re-scaled by  $\frac{1}{25}$ .

As seen in Figure 4.2, the trends can be more easily compared when re-scaled to a similar size (plot (b)). From plot (a), we can see that UDGs started to rise in interest a couple decades after dark matter. Since UDGs are difficult to measure and are sometimes comprised of mostly dark matter, it makes sense that the study of dark matter rises in popularity earlier on. From plot (b), We get a sense of how UDGs have completely sky-rocketed in interest over the last decade, while dark matter has decreased slightly. However, this decrease could be an artifact of our query: more recent papers are possibly more likely to contain specific terms related to dark matter in the title, while it being mentioned in the abstract of the body of the paper.

## 5 Conclusion

The **ads.query.pkg** package currently only contains four functions used in order to get a general idea of the trends in interest for any topic in astronomy. However, the functions defined in the package are easily modifiable to suit the specific needs of a user, by simply adding parameters into the ads functions themselves. This creates a large amount of degrees of freedom when it comes to personalizing a function in the **ads.query.pkg** package, and opens up the floor for others to propose different or better ways to analyse and compare interest.

An improvement that should be implemented into these functions appears when checking the papers that are gathered by the **ads.SearchQuery()** function: a compound keyword is not searched as a single object, but is recognized as two key words being searched independently. This leads to inconveniences related to spelling conventions, e.g. using a hyphen or a space in ‘Ultra(-/ )Diffuse’.