Milky Way Bulge Photometry Visualization

Siva Harshini Dev Bonam
Rutgers University
Piscataway, NJ, USA
harshini.bonam@rutgers.edu

Kanya Kreprasertkul
Rutgers University
Piscataway, NJ, USA
kk1003@scarletmail.rutgers.edu

Yifan Liao
Rutgers University
Piscataway, NJ, USA
yl1463@scarletmail.rutgers.edu

***Abstract*—** We established a multilevel interactive map to demonstrate over 600 million stars inside the milky-way bulge provided by the European Southern Observatory (ESO). [1] On our map, selected information is shown which is regarded to its spatial position, brightness, and spectrum. We are able to process these huge amounts of data in a relatively short time using distributed computation in Spark Streaming ecosystem and by applying graph waves theory [2] , which is an algorithm that separates graphs into sub-graphs efficiently introduced by James Abello and Daniel Nakhimovich.

***Keywords*—** Milky way, graph waves, graph building, graph cities, PSF Photometry, astronomical statistics, right ascension, declination, band magnitude J and K, galactic distance, heat map, star features distribution, histogram, MVT Architecture, PySpark, Django framework.

I. Project Description

Nowadays, with the development of modern industries, human beings are using more and more metal materials. Someday we will eventually use up the worldâs resources on the earth. In a quest to solve problems like these, most of us escape into gazing at the sky, turning to the boundless universe. However, as we know, there are countably infinite planets. Traversing through all those trillions of celestial bodies will take more than a personâs entire life. So, our spectacular interstellar journey may start with research on the brightness and spectrum to scoop the secrets of the stars at the center of the galaxy, starting with an analysis of the huge data (approximately 100 GB). [3]

The purpose of this project is to establish an interactive map to demonstrate the information of stars. For example, the total number of stars, minimum brightness, maximum brightness and average brightness of the each tile. With the help of this map, astronomers are able to grow up with a better idea about which planet may have certain metal according to the spectrum and brightness shown in. Also, graph waves theory was applied to our data, which is an algorithm used to separate graphs into sub-graphs efficiently.

A. Stage1 - The requirements gathering stage.

Our dataset is based on a catalog released by the European Southern Observatory (ESO) with a total size of about 100 GB. [1] It contains over 600 million records which represent a star respectively. In this project, 7 columns will be mainly focused on, i.e. source ID, 4 columns of its spatial position and 2 columns of its brightness. We performed exploratory data analysis with Python on Jupyter Notebook using several types of plot such as heat map and histogram. It is downloaded through the download script provided by the ESO[1] through the scientific portal data request. This huge dataset is downloaded and stored in the LCSR facility at Rutgers University.

This application is designed for: The knowledge about the dataset is provided in ESO [4]. We learnt more about the ESO VVV Experiment[5] and data released in the phase 3 of this experiement [6] [7] [8] [9].

This application is designed for:

- **Technical Non-CS User:** The users we designed for the map are Technical Non-CS Users, to be more specific, they are astronomers. Astronomers spend their whole life to observe outside the scope of Earth, trying to discover something about the universe that would be considered astronomical progress. However, we had to agree, human's power is limited. Even if they work 365 days a year and field 1 star every minute. They can only scan 500 thousand stars. As we mentioned before, our data set has over 600 million records, which means a professional astronomer needs to work 1200 years without any rest. That is obviously impossible. However, The world is developing, the era is making progress all the time.
  Scenario 1: We can provide an interactively map in front of them where they can filter the statistics first, then choose the planet they are interested in by visualization of its position in space, brightness and spectrum.
  - This smart way can save them a lot of time. The second they save from scanning all, the more time they can find targeted stars. At some point, we will eventually hit our goal.

- Project Timeline and Division of Labor.

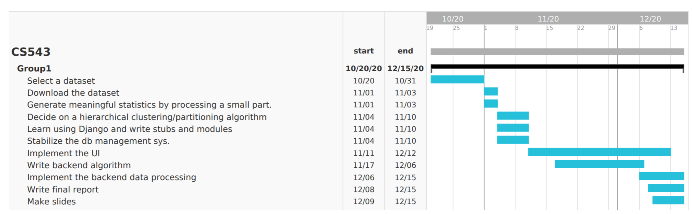

Fig. 1. 8 weeks Gantt chart that was followed to complete the project

Kanya Kreprasertkul: UI and exploratory data analysis and statistical analysis
Harshini Bonam: UI and background data processing
Yifan Liao: UI and reporting

B. Stage2 - The design stage.

- Short Textual Project Description: In our web application, we mainly provide two types of information to a user. The first one is the interactive statistical graphs which give detailed statistics of our dataset. The second one is the interactive graph cities, which also display graph buildings and graph drawings.
- Each star in each tile represents vertices in a graph. To create edge in a graph, the distance between the stars was calculated with the following equation, where right ascension (RA or α) and declination (DEC or δ) are the angular positions. [10]

$$Distance = \sqrt{[(\alpha_1 - \alpha_2)(cos\frac{\delta_1 + \delta_2}{2})]^2 + (\delta_1 - \delta_2)^2} \quad (1)$$

Then, we connected edge for the stars if their angular distance is less than 0.0001 radians.

- Data Flow Diagram:

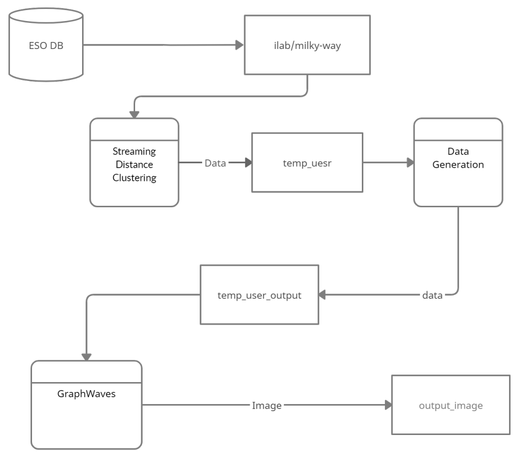

Fig. 2. Data flow chart

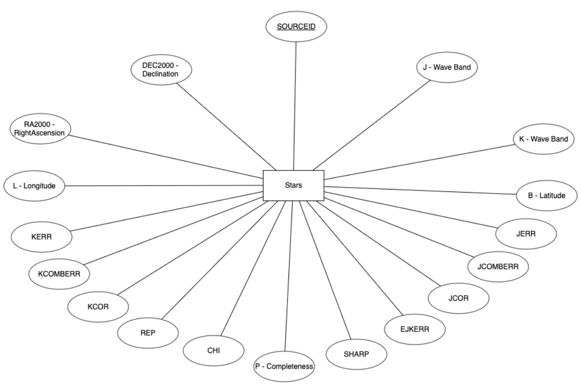

Fig. 3. The ER Diagram.

C. Stage3 - The implementation stage.

- This web application primarily uses client-server architecture with MVT (Model, View, Template) design pattern.
- Raw dataset had multiple csv files with over three million records. All the data is stored in MySQL database (Model). MySQL workbench has been used for importing csv data files into the respective tables. The only hierarchies that were present in the raw data are World and Country. Continent hierarchy was added on top of country level by means of aggregation. Indexing has been used to achieve faster data retrievals in the case of tables having more than a million records.
- View in the design pattern is implemented as Python functions that serve the calls from client. Templates are the webpages that are viewed by the client. HTML, CSS and JavaScript is used for coding the template.
- Django web framework with python language is used for the project. Django automatically connects to the MySQL database whenever application starts running.

The steps involved in the implementation are as follows:

- Downloaded the dataset as 196 FITS files from the download script provided by the ESO.
- Read the data in batches of 10000 using Spark Streaming feature.
- Performed the exploratory data analysis on this data to understand the distribution of number of star per tile, the max and min distances, brightness and spectral wavelengths, etc.
- Generate 61 Million clusters for 612 Million stars for the input data using Streaming K-Means algorithm provided by Spark ecosystem.
- Generate graph cities for each cluster by adopting graph waves algorithm[2], where each node in the graph is a cluster center star and each edge between a pair of stars represents distance less then 0.0001 radians.

D. Stage4 - User interface.

The User Interface were developed with Django framework. Meanwhile Interactive statistical graphs were built using AnyChart.js. We examined our hierarchical clustering and partitioning algorithm on our dataset with Python on Jupyter Notebook, and we determined interesting statistics we should display. Several essential libraries were used as listed below.

- astropy
- astroquery
- numpy
- pandas
- matplotlib

There are 2 types of brightness (J magnitude and K magnitude) contained in our dataset. Therefore, we show both of them. We display histogram for the brightness distribution of each tile, and provide bar chart and map for the following statistics.

- Minimum brightness of each tile
- Maximum brightness of each tile
- Average brightness of each tile
- Total number of stars in each tile

Our map is created with heat map, so several colors are used to represent different values and we can see that there are pattern in the average brightness as shown below.

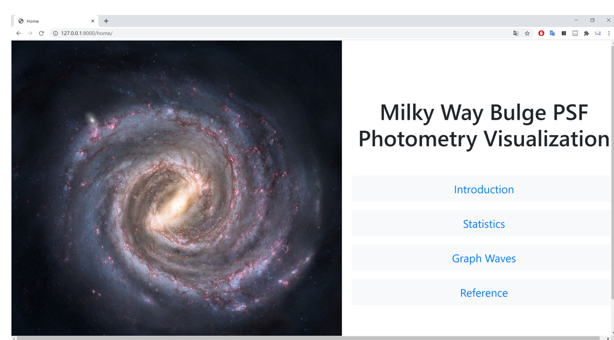

Fig. 4. Interactive map of the average K magnitude.

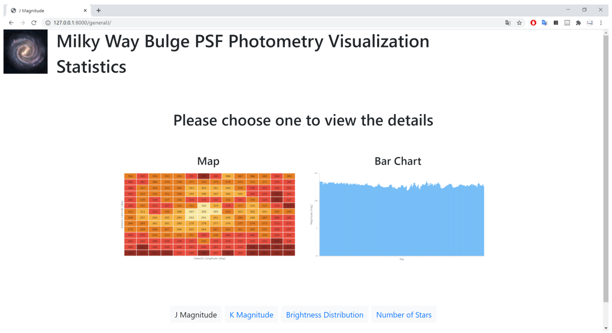

Fig. 5. Interactive map of the average K magnitude.

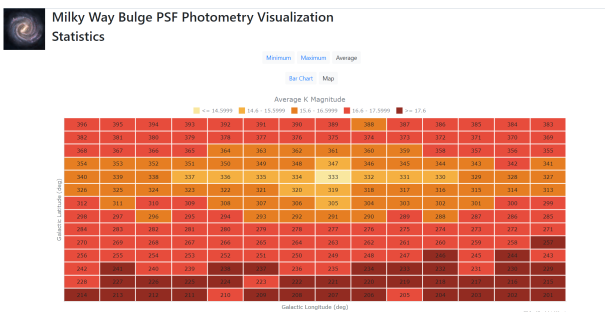

Fig. 6. Interactive map of the average K magnitude.

https://www.overleaf.com/project/5fd67d6d82d7e66037625e37

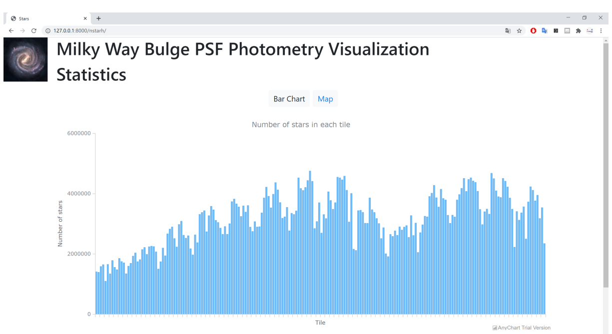

Fig. 7. Interactive map of the average K magnitude.

Graph waves theory was applied to our dataset using PySpark on Jupyter Notebook. In this feature, we display graph buildings and graph drawings for each tile. Plotly.js was used to constructed the interactive graph buildings. A user can see the result by giving a tile number. The example of graph buildings and graph drawings are shown below.

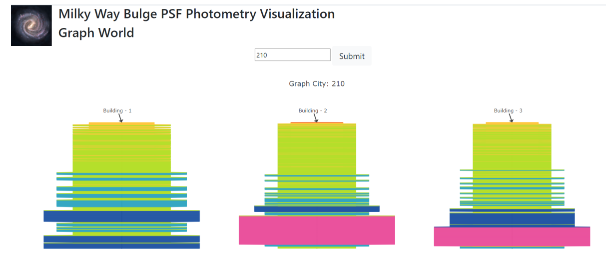

Fig. 8. A Graph buildings of tile 210.

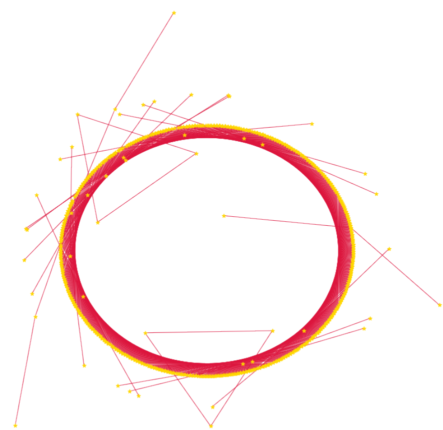

Fig. 9. A graph drawing of tile 201.

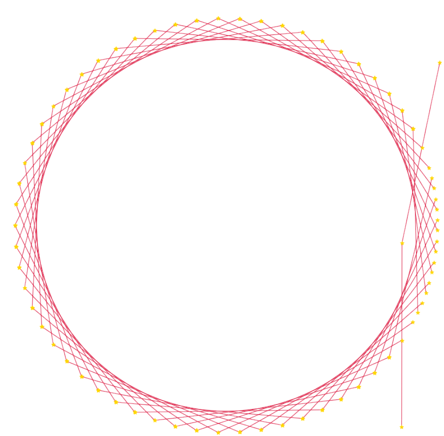

Fig. 10. A graph drawing of tile 201.

II. Project Highlights.

Our system has yielded a multitude of interesting and useful facts about Milky Way Bulge PSF Photometry. They have been mentioned as follow.

1) Our dataset is based on a catalog released by the European Southern Observatory (ESO) with a total size of about 100 GB.
2) Graph waves theory was applied to our data, which is an algorithm that separates graphs into sub-graphs efficiently.
3) The dataset is too large to be processed in a normal way, so we used stream processing to solve this problem.
4) By examining map of the average brightness, we found out that the average brightness of the stars in the middle of the Milky Way are lower than the outside.

III. Future work.

- An advanced search feature to search for the information of stars in a given astronomical position can be developed.
- A compare feature between selected stars in certain tile can be developed.
- Data analysis tools can be applied to get more insightful information.

References

[1] E. Valenti. (2019) Milky way bulge psf photometry. [Online]. Available: https://www.eso.org/rm/api/v1/public/releaseDescriptions/140

[2] J. Abello and D. Nakhimovich. (2020) Graph waves. [Online]. Available: http://ceur-ws.org/Vol-2578/BigVis7.pdf

[3] E. Valenti. (2019) Milky way bulge psf photometry. [Online]. Available: https://www.eso.org/qi/catalog/show/272

[4] Eso archive science portal. [Online]. Available: http://archive.eso.org/scienceportal/home

[5] Eso vvv exprtiment. [Online]. Available: https://arxiv.org/pdf/1902.01695.pdf

[6] The vvv phase 3 data release vircam. [Online]. Available: https://www.eso.org/sci/observing/phase3/data$_{r}eleases/vvv_{d}r3.pdf$

[7] Vista variables in the via lactea (vvv): The public eso near-ir variability survey of the milky way. [Online]. Available: https://arxiv.org/pdf/0912.1056.pdf

[8] J band astronomy (infrared). [Online]. Available: https://en.wikipedia.org/wiki/J$_{b}and_{(}infrared)$

[9] [Online]. Available: https://en.wikipedia.org/wiki/K$_{b}and_{(}infrared)$

[10] M. Richmond. Some basic astrometry. [Online]. Available: http://spiff.rit.edu/classes/phys373/lectures/astrom/astrom.html