Iranian Journal of Astronomy and Astrophysics
Vol. 10, No. 4, Autumn 2023, 327–334
©Available online at http://ijaa.du.ac.ir
DOI: 10.22128/ijaa.2023.736.1170
Online ISSN: 2383–403X

Iranian Journal of Astronomy and Astrophysics

Research Paper

Identification of Gamma-Ray Point Sources in Fermi-LAT Data with Minimum Spanning Tree Algorithm

Mehran Soor*[1] · Fatemeh Akhondi[2] · Hadi Hedayati[3]

[1] Department of Physics, K.N. Toosi University of Technology, Tehran, Iran;
*email: Mehran.Soor@email.kntu.ac.ir

[2] Department of Physics, K.N. Toosi University of Technology, Tehran, Iran;
email: f.akhondi@email.kntu.ac.ir

[3] Department of Physics, K.N. Toosi University of Technology, Tehran, Iran;
email: hedayati@kntu.ac.ir

Received: 28 September 2023; **Accepted:** 28 December 2023; **Published:** 30 December 2023

Abstract. Gamma rays are the most energetic photons in the electromagnetic spectrum, detected with ground-based and space-based detectors in different energy ranges from sources in our galaxy and beyond. Gamma-ray point sources can be identified by special clustering of these photons. The minimum spanning tree (MST) algorithm is a graph-based method in order to find clusters. In this paper, we use the MST algorithm for finding point sources in Fermi gamma-ray space telescope data which is sensitive to photons with energies of 20 MeV up to more than 300 GeV. To this end, we selected eight completely random ($10° \times 10°$) fields of Fermi gamma-ray sky and tested the algorithm on the 12-year Fermi-LAT sky (Pass 8) at energy ranges above 3 GeV and above 6 GeV and with different cluster selection criteria. The calculation of Precision and Recall for both fields shows that MST is a useful algorithm in order to identify the point.

Keywords: Astronomy data analysis, Clustering, Gamma-ray sources

1 Introduction

Gamma rays, the most energetic electromagnetic waves, have the shortest wavelength. Photons with energy above 100KeV are considered as gamma rays. High-energy gamma rays (between a few MeV upto 30 GeV) are detected by space-based detectors, and very high-energy gamma rays ($\gtrsim$30 GeV) by ground-based experiments. Exploring the universe through the gamma-ray window is very interesting because it gives us a view of its non-thermal radiation and allows us to detect extreme events on different astronomical scales. In our galaxy, sources such as pulsars, pulsar wind nebula, supernova remnants, etc., have been detected that emit gamma rays. Outside the Milky Way, gamma rays from galaxies with high star production rates and jets of super relativistic particles escaping from supermassive black holes at the center of some galaxies have also been observed [1].

The Fermi Gamma-Ray Space Telescope is a space-based telescope for observing gamma rays. The Large Area Telescope is one of the instruments of the Fermi telescope that detects

* *Corresponding author*

high-energy gamma rays in the range of 30 MeV to more than 300 GeV [2]. This telescope has a converter, a precise detector, and a calorimeter that converts the gamma-ray into a pair of electron and positron ($e^+ - e^- pair$) [3]. Gamma-ray sources are detected in Fermi telescope data by spatial clustering, and several methods, including minimum spanning tree, have been proposed for this purpose. The minimum spanning tree algorithm is a graph-based method for cluster detection that starts by selecting clusters from a specific tree that connects all points with the minimum possible total edge weight.

The method used in the Fermi source catalog is the maximum likelihood (ML) analysis method, in which we consider the sources identified by this method in the fourth Fermi catalog as a reference. Then, by applying the minimum spanning tree algorithm on Fermi gamma-ray telescope data, we identify gamma-ray point sources, match them with the Fermi sources, and measure the performance of this algorithm to identify gamma-ray point sources.

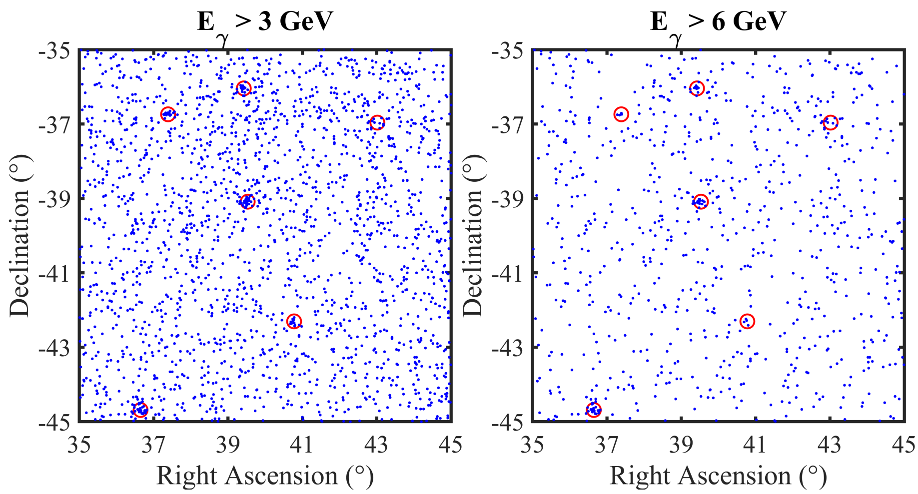

Figure 1: Scatter plot of photons for one of eight random fields. red circles mark the location of the point sources based on the LAT 12-year Source Catalog (4FGL-DR3). (Left side shows photons with energy range higher than 3 GeV and right side shows photons with energy range higher than 6 GeV)

2 Data selection: Femi-LAT

We selected eight completely random 10°×10° fields and tested the algorithm on the 10-year Fermi-LAT sky (Pass 8) from 1 January 2012 to 1 January 2022 at two energy ranges: above 3Gev and 6Gev. All data are available on the LAT data server. It contains photons above 3GeV and is filtered by applying the standard selection by Fermitools 2.2.0 and fermipy 1.2.0, source class events, evclsss = 128, and both front and back converting, evtype = 3, with up to a maximum zenith angle of 90°. Sources of the field were selected from LAT 12-year Source Catalog (4FGL-DR3) [4].

The result will be eight random fields with the Right ascension and Declination of the arrival directions of photons (see Appendix A). Figure 1 shows the random field in two different energy ranges, where the sources of the fourth Fermi catalog are also specified.

3 Minimum Spanning Tree (MST)

Equatorial coordinates (Ra and Dec) of photons form a set of two-dimensional points. This set of points is assumed to be the points of a graph. Now, the minimum spanning tree algorithm can be used for this set of two-dimensional points, where a point is randomly selected and connected to the nearest point in its neighborhood by an edge. After drawing N-1 edges, the graph is complete (N is the total number of points). Once the tree is complete, edges larger than a certain value are removed. This parameter is called Λ_c which is defined as a function of the average length of all edges (Λ_m). After applying this condition and cutting several edges, among the remaining clusters, the clusters whose number of points is less than the specified value (N_c) are removed. A Minimum Spanning Tree based method was first proposed for the analysis of COS-B satellite data. [5,6] developed this method and used it for Fermi-LAT gamma-ray data. In their recent paper, [7], they applied this algorithm to Fermi-LAT data to detect the supernova remnants (SNRs) in the large Magellanic Cloud galaxy. To verify the correctness of the MST algorithm that we wrote, we applied it to a part of same data and with the same parameters (Λ_c, N_c) that were in [7]. As can be seen from Figure 2, our algorithm found every predetermined source.

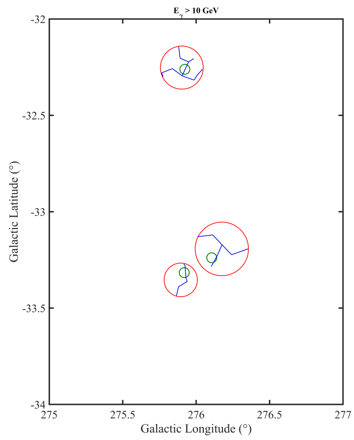

Figure 2: MST algorithm applied to the data in [7], green circles are the sources found by [7] and red circles shows sources that we found.

In this paper, to apply a minimum spanning tree on Fermi's data based on previous researches, value for parameter Λ_c is considered $0.7\Lambda_m$ [5,6,8,9]. For the N_c parameter, the values 3, 4, and 5 were taken to determine the best value for this parameter. Figure 3 and Figure 4 shows the implementation of the minimum spanning tree for one of the random regions. To check the algorithm's performance, we have used Precision and Recall; Precision means the exactitude of the algorithm to find real sources and rejecting false sources, and Recall means how many real sources the algorithm has correctly identified [10]. After running the algorithm on all eight fields and calculating Precision and Recall for all areas,

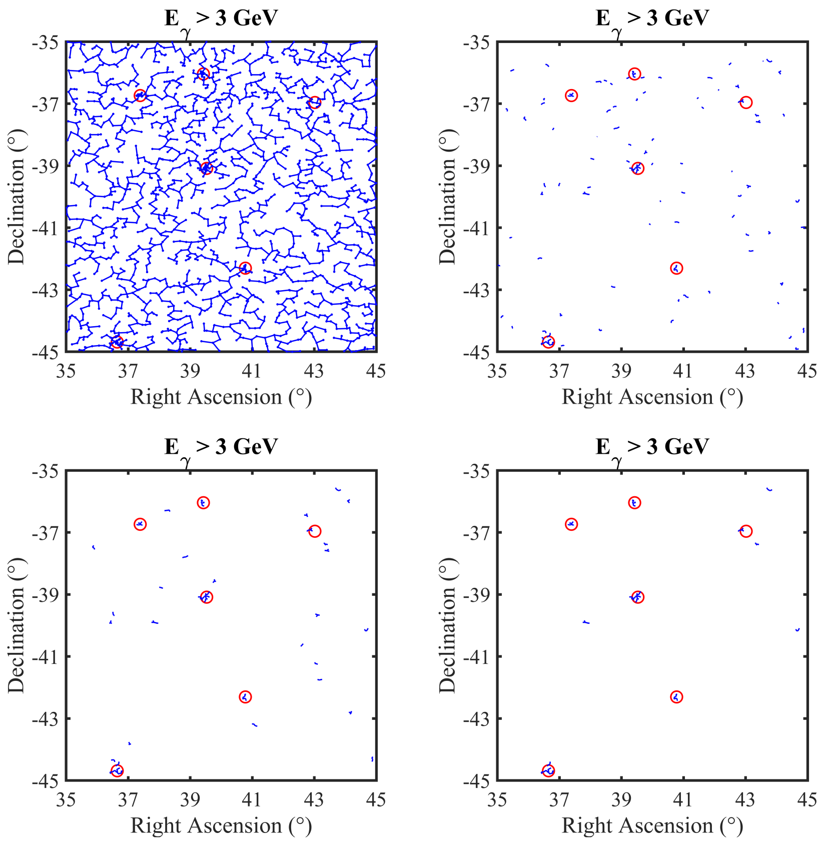

Figure 3: The upper left figure shows the application of the algorithm on field of Figure 1 for energy range above 3 GeV. The upper right figure shows the MST after applying the separation and elimination with parameter $N_c = 3$. The lower figures show $N_c = 4$ and $N_c = 5$, respectively, from left to right.

Figure 4 was obtained (see Appendix A).

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{2}$$

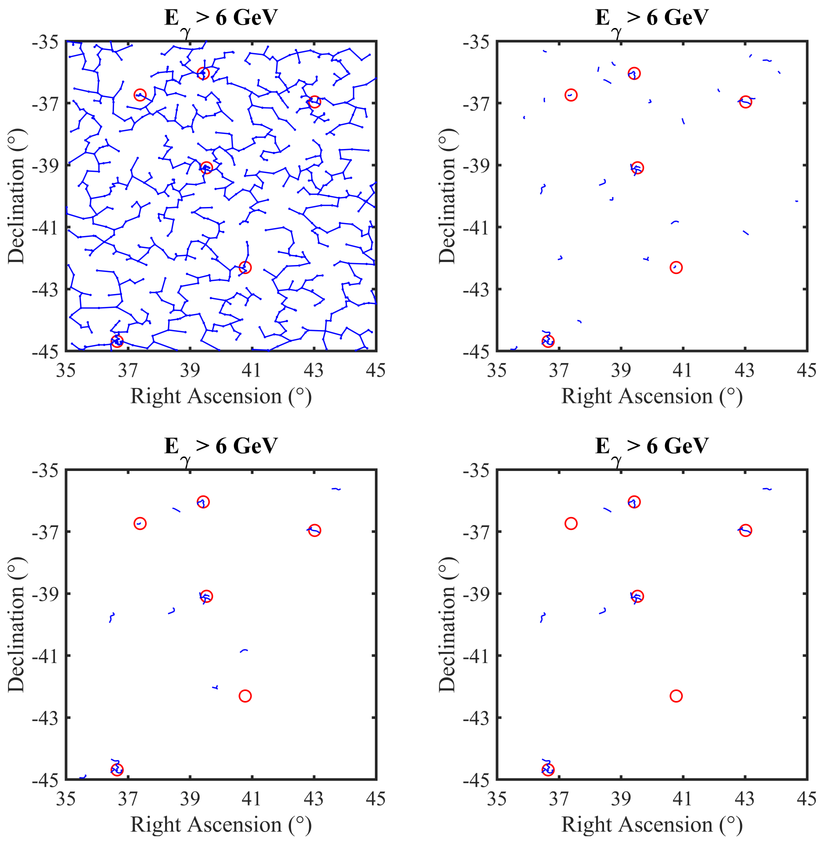

Figure 4: The upper left figure shows the application of the algorithm on field of Figure 1 for energy range above 6 GeV. The upper right figure shows the MST after applying the separation and elimination with parameter $N_c = 3$. The lower figures show $N_c = 4$ and $N_c = 5$, respectively, from left to right.

4 Discussion

According to the examples, a large number of possible sources are proposed by applying $N_c = 3$, so it is expected that the minimum spanning tree is a suitable method for finding the location of existing sources and even suggesting new sources in that area. Precision and Recall graphs also confirm this statement for both areas.

The higher the $N_c = 3$ value, the higher the Precision (Figure 5) and the lower the recall. Several ways exist to increase this method's precision and identify the location of point sources: considering different N_c and checking its results with $N_c = 3$. In this article, $N_c = 5$ has been investigated, and it was seen that it reduces the existing errors to a great extent (Figure 5). Also, using several other clustering methods, such as DBSCAN and -means, and applying them to the results obtained from MST to optimize the results, increase the probability of sources, and identify new gamma rays sources. After detecting the location of the sources, the next step in our research is to identify their nature.

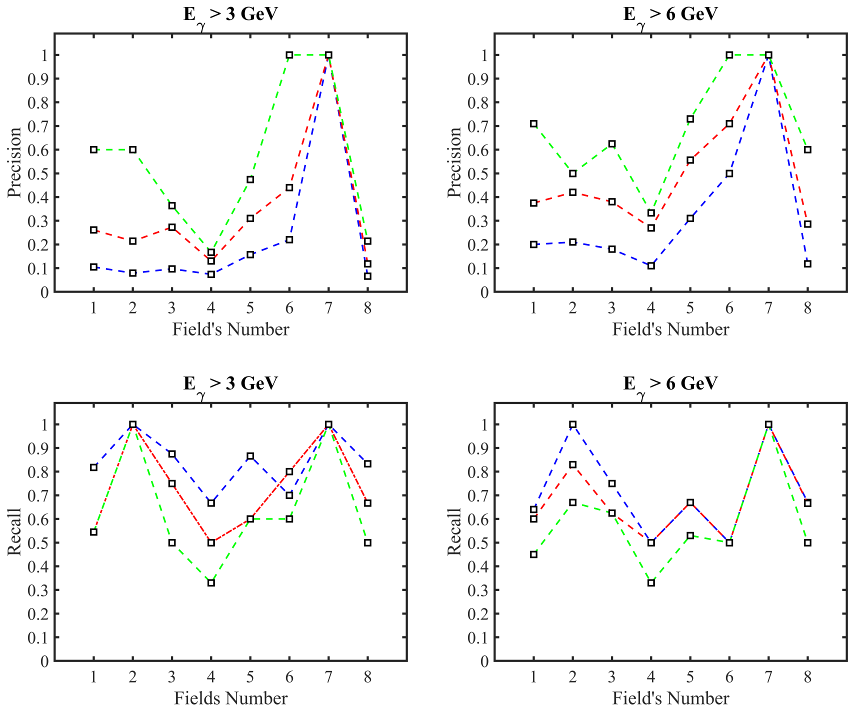

Figure 5: The figure shows the calculation of Precision and Recall for all eight investigated fields. The blue dashed line shows calculations assuming parameter $N_c = 3$, the red dashed line shows calculations considering parameter $N_c = 4$, and the green dashed line shows calculations considering parameter $N_c = 5$ to all fields.

Authors' Contributions

All authors have the same contribution.

Data Availability

No data available.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

The authors have diligently addressed ethical concerns, such as informed consent, plagiarism, data fabrication, misconduct, falsification, double publication, redundancy, submission, and other related matters.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or nonprofit sectors.

References

[1] De Angelis, A., & Mallamaci, M. 2018, EPJ Plus, 133.

[2] Albert, A., & et al. 2020, ApJ, 903, L14.

[3] Atwood, W. B., Abdo, A. A., Ackermann, M., & et al. 2009, ApJ, 697, 1071.

[4] Abdollahi, S., Acero, F., Baldini, L., & et al. 2022, ApJS, 260, 53.

[5] Campana, R., Bernieri, E., Massaro, E., & et al. 2013, Astrophys Space Sci., 347, 169.

[6] Campana, R., Massaro, E., Gasparrini, D., & et al. 2008, MNRAS, 383, 3.

[7] Campana, R., Massaro, E., Bocchino, F., Miceli, M., & et al. 2022, MNRAS, 515, 1676.

[8] Campana, R., & Massaro, E. 2021, A&A, 652, A6.

[9] Campana, R., Massaro, E., & Bernieri, E. 2018, A&A, 619, A23.

[10] Fränti, P., & Mariescu-Istodor, R. 2023, Pattern Recogn. Lett., 167, 115.

A Appendix: Additional Tables

Table 1: More information on the details of each field. N_γ shows the number of photons.

NO. Field	Equatorial Coordinates (°) (Ra, Dec)	N_γ (E_γ >3 GeV)	N_γ (E_γ >6 GeV)	NO. Sources
1	(30, -30)	2148	881	11
2	(40, -40)	2128	856	6
3	(50, -50)	1994	784	8
4	(13, -72)	1198	442	6
5	(20, -20)	2765	1078	15
6	(160, 55)	2998	1357	10
7	(38, 90)	194	60	0
8	(60, -60)	1711	696	6

Table 2: Details of results for sample field that mentioned in paper.

Fields	N_c	TP	FP	FN	Precision	Recall
Field 2 ($E_\gamma > 3$)	3	6	70	0	0.0789	1
	4	6	22	0	0.2142	1
	5	6	4	0	0.6	1
Field 2 ($E_\gamma > 6$)	3	6	22	0	0.2142	1
	4	5	7	1	0.4167	0.8333
	5	4	4	2	0.5	0.6667