Mass segregation and cluster properties

Author: Enric Fernández Salgado, efernasa48@alumnes.ub.edu
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Gemma Busquet Rico, gemma.busquet@ub.edu

Abstract: The infrared dark cloud G14.225−0.506 hosts two protoclusters in its northern and southern hubs, making it an ideal target for investigating clustering and mass segregation in early stages of cluster formation. By analyzing multi-wavelength data from millimeter, centimeter, and infrared catalogs, we constructed Minimum Spanning Trees (MSTs) to study the spatial distribution of young stellar popullation. Key statistical parameters, such as the Q parameter and the mass segregation ratio ($\Lambda_{\rm MSR}$), were calculated to assess whether the clusters exhibit fractal or centrally concentrated structures. Our findings reveal subclustering in some of the sources and peak centering in others and a notable mass segregation in the northern hub, while the southern hub displays a more uniform distribution. These results enhance our understanding of the physical conditions, dynamics, and evolution of protoclusters.

Keywords: Infrared dark clouds, Mass segregation, Minimum spanning tree, Star formation, Cluster properties, Subclustering

I. INTRODUCTION

It is a generally observed phenomenon around the Universe that young stars develop in close groups. Stars are born commonly in dense gas clouds, being the main star generator the ones with masses higher than 10^3 $M_\odot$ and H_2 density above 10^4 cm^{-3} [1]. These clouds collapse and as the density, pressure and temperature rise stars begin to form. The reason why we observe that Young stellar objects (YSOs) are normally in large groups pretty close to each other is due to fragmentation of the cloud during its gravitational collapse because of inhomogeneities in its mass density [6]. This leads to the formation of star clusters formed by vast amounts of young stars in relatively close distances. These stars are mainly low-mass stars (< 0.5 $M_\odot$ as these kind conform the majority of the universe star population. The most massive clouds can also form protoclusters containing intermediate- and high-mass stars [11].

If we look at massive clouds with a mix of YSOs with different masses, it is reasonable to assume that the gravitational pull of the msot massive stars in the cluster can affect how the rest of stars distribute, putting the most massive ones in the center and leaving the lighter ones at the edges. This effect is known as mass segregation and it is generally observed in most clusters [3], even though there's room for debate whether this is a phenomenon that happens as a consequence of the most massive stars moving the smaller ones or if high-mass stars actually tend to form at the center of the cluster.

To observe this phenomenon in detail it is interesting to find some statistical properties of the cluster geometry. To start with, we it is necessary to obtain a graphical object known as the Minimum Spanning Tree (MST), which will be explained in detail later. The branches of this graph give crucial information about the distribution of the cluster members and it is necessary for the calculations of other important parameters (see Sec. II) related to the cluster's properties. . Below we explain in detail the main information that can be obtained from these different analysis techniques to characterize the cluster properties.

Mass segregation characterizes the phenomenon described earlier. The value of the mass segregation will be higher if the most massive stars are close to the cluster center and the low-mass stars are distributed throughout the whole cluster. This can give information about various properties of the cluster:

- Cluster's stellar dynamic: we can explain mass segregation as a consequence of gravitational exchanges between high-mass and low-mass stars. This would lead the cluster to a point in which, time after its formation we can see that the stellar population organizes to a lower energy system. In this "final" state (understanding final as the end of the process when the stars organize and segregate, not as the final state of the cluster as the YSOs will still keep evolving with time) we could say the cluster has hit its relaxation state, the time it takes to reach this point is called relaxation time.

- Initial conditions: mass segregation can also be the consequence of some initial conditions of the natal cloud that formed the cluster. The central parts of the cluster, which are closer to the center of mass could tend to augment in density faster than the edges, as the force of the gravitational collapse is more intense close to the center.

- Evolution of the cluster: dissipation of the cluster could also affect how its mass distributes, also the stars who are closer have a higher probability of collision. This last point is unlikely to be the cause of the phenomena as the stars formed by collisions of other stars are not the kind of sources we are studying in this case.

The Q parameter [10] characterizes whether the cluster stellar density is centrally peaked or presents subclustering (fractal distribution), which can be divided with subgroups each one with its own mass segregation phenomenon. This depends on the value of Q as we will explain in Sec. II.

A. The target: G14.225−0.506

We will study the Infrared Dark Cloud (IRDC) G14.225−0.506 (from now on G14.225), also known as M17 SWex, which is found inside a vast and massive molecular cloud located in the southwest of the HII region called M17 [4]. It has an extension of 77×15 pc and a total mass superior than 10^5 $M_\odot$. Its distance is approximately 1488–1574 pc as obtained in recent observations by using *Gaia* DR2 parallax measurements, but we will assume a distance of $1.6^{+0.3}_{-0.1}$ kpc, similarly to the recent work of [4]. This cloud contains a rich population of protostars and YSOs that concentrate in two main region, the North hub and the South hub, each of them harboring a deeply embedded cluster (i.e, protocluster) in the early stages of its formation.
In order to investigate the cluster properties in G14.225 we used the source catalog obtained in previous works. Specifically, the ALMA millimeter data [2], the VLA centimeter data [4], and the infrared (IR) catalog from *Spitzer* [13].

By using these variety of wavelengths we can observe sources at different evolutionary stages, from Class 0 (protostars) to Class III (pre-main sequence stars). These different kinds of YSOs have different temperatures, which influences the wavelength of the radiation they emit. This way we know their evolutionary stage: centimeter and millimeter sources tend to correspond to protostellar objects (Class 0 and I) which have a cold core that emits at low frequencies, whereas IR sources mainly correspond to Class II and II.
properties of the Infrared Dark Cloud (IRDC) G14.225, focusing on the spatial distribution and mass segregation of YSOs during the early stages of cluster formation. We will employ a combination of statistical tools and graph theory techniques, utilizing catalogs of YSOs identified across various wavelength regimes.

II. METHODS AND RESULTS

To obtain information about the spatial distribution of the cluster we will use a graph theory concept known as a Minimum Spanning Tree (MST), which is defined as a graph when the connections in the nodes follow these conditions:

- All nodes must be connected, there must be a path between any two random nodes.
- There cannot be closed paths or cycles inside the graph.
- Defining the "weight" of every possible branch as a concrete value, the graph must be generated optimizing the total sum of the weight to be the minimum while still respecting the other two rules. In our case the "weight" will be the euclidean distance between the two nodes that are connected by each branch.

There are different algorithms to generate an MST, we will use a code with the python library *mistree* [8], which uses the *Kruskal* [7] algorithm.

The Kruskal algorithm follows these simple steps:

- Sort all the possible branches by weight, in ascendant order.
- Include de first branch in the graph.
- Check if the graph has any closed path using Union-Find algorithm.
- If the graph doesn't have any cycle keep the branch.
- Repeat until $N_{\text{branch}} = N_{\text{node}} - 1$

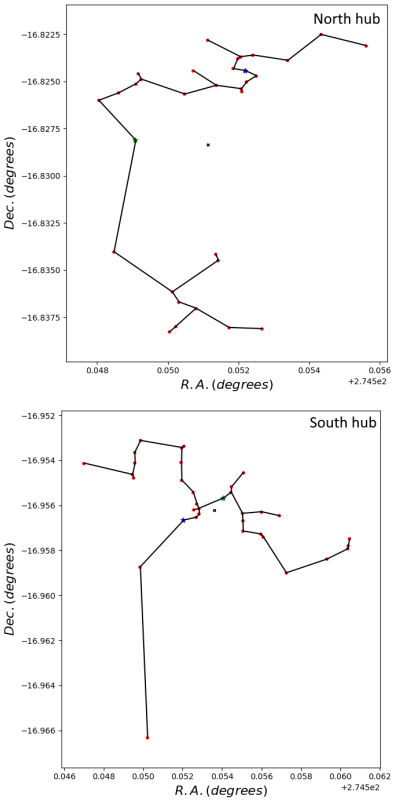

FIG. 1: MST for the ALMA millimeter continuum sources in G14.225-North (top) and G14.225-South (bottom). The most massive star is marked in blue and the second one in green, black mark represents the center of the cluster, obtained as the mean position of the sources.

Figure 1 shows the MST for both the North and South hubs of the millimeter sources detected in G14.225. Table 1 presents the results of the MST analysis for both hubs, which have very similar values, though some parameters are notably different. Mainly, the deviation and maximum distance. This is visible in Fig. 1 as the South hub has its lowest star quite separated from the rest, making both the deviation and maximum distance to increase compared to the North hub.

Measure	**North Hub**	**South Hub**
Median Distance	2.485	2.274
Mean Distance	3.794	3.713
Deviation	3.956	4.868
Max Distance	21.386	27.393
Min Distance	0.444	0.461

TABLE I: Values for the median, mean, maximum and minimum projected distance (in units of arcseconds)±0.001 of the branches of the MST of both the North and South hub.

Next, we start obtaining the Q parameter, mass segregation and distribution function.

A. Q parameter

The Q parameter is defined as:

$$Q = \frac{\overline{m}}{\overline{s}} \tag{1}$$

where $\overline{m}$ is the normalized mean distance of the branches of the graph, which is obtained as:

$$\overline{m} = \frac{\sum L_i \cdot (N_{\text{node}} - 1)}{\sqrt{N_{\text{node}} \cdot A}} \tag{2}$$

With L_i being the weight (distance) of every branch of every node of the MST, N_{node} is the total amount of nodes (sources) and A the projected area of the cluster, obtained as the area of a disk of radium $\text{R}_{G14.225} = 35.909''$ for the North hub and $38.437''$ for the South hub, which corresponds to the maximum distance between a source and the average position of the nodes. And $\overline{s}$ is the normalized mean distance between all the points:

$$\overline{s} = \frac{2 \cdot \sum s_i}{N_{Nodes} \cdot (N_{Nodes} - 1) \cdot R_{G14.225}} \tag{3}$$

The value of the Q parameter is a powerful tool to differentiate whether the cluster has a stellar density centrally peaked with a smooth radial star distribution gradient ($Q > 0.8$) or if it has a fractal geometry, which indicate subclustering inside it in various scales ($Q < 0.8$). The results we got were Q_N=0.475 and Q_S=0.685, for the North and South hub, respectively. This means that both protoclusters present a subclustering behavior, which can be visually observed when comparing the center of the cluster in the MST (marked with a black cross) with the zones where we see the highest number of stars it makes clear that there is no peak centered behavior.

B. Mass segregation

The next element to analyze in the graph is the mass segregation, which we determine using a parameter defined as the mass segregation ratio:

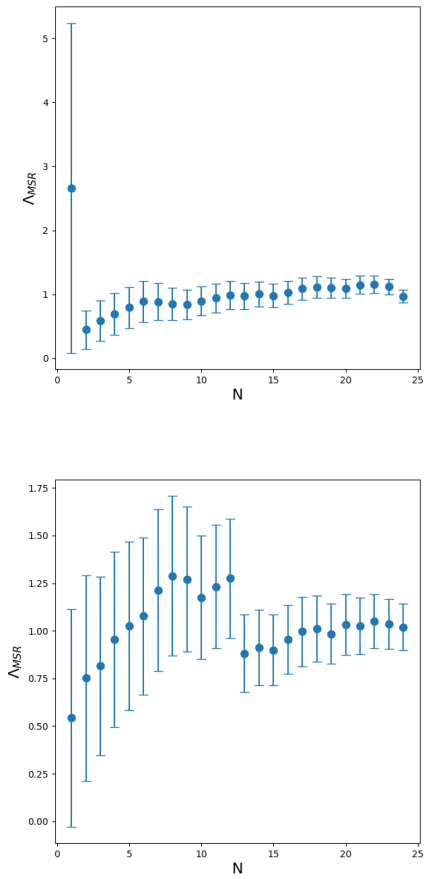

FIG. 2: Mass segregation ratio for the millimeter sources detected with ALMA for the North hub (top) and the South hub (bottom) as a function of the number of sources N.

$$\Lambda_{\text{MSR}} = \frac{L_{\text{random}}}{L_{\text{massive}}} \pm \frac{\sigma_{\text{random}}}{L_{\text{massive}}} \tag{4}$$

where L_{random} is defined as the mean weight of the sum of all the branches connected to N different nodes chosen randomly. To be sure we aren't picking non representative nodes by luck we will repeat this process 3000 times and then we calculate L_{random} as its average value,

σ_{random} is the standard deviation of this value in the 3000 iterations, and L_{massive} is the sum of the weight of the branches of the N most massive sources in the cluster. If the value of the mass segregation ratio is higher than unity that means the distance of the average N sources with their neighbors is higher than the distance of the most massive ones, which is a clear indicator than massive stars are closer to their neighbors indicating the phenomena of the mass segregation. Also it is logical to think that as we increase N the ratio will progressively get closer to 1, as we will be picking less massive stars, and at the limit $N = N_{\text{nodes}}$ then $\Lambda_{MSR} = 1$.

In Fig. 2 we can see the mass segregation ratio of both hubs depending on the number of stars we pick as the "massive ones". We can see that the North hub has a clear segregated point at the single one most massive star, but as we start adding more stars it drastically falls to 1. Graphically this can be seen in the MST as the most massive point is well connected with low separation with its neighbors, but de second one is quite far away from them. On the other hand, the South hub doesn't show signs of mass segregation as the mass segregation ratio is really close to 1. The first values are below the unity and the next ones start to increase to even surpass it just to fall again at value of 1 as expected.

C. Distribution function

The next step is aiming to get the distribution function, which is defined as the probability to find a pair of nodes depending on the projected separation between them. The normalized expression of the probability to find a pair in a section between s and $s + \Delta$s is:

$$\rho(s) \cdot \Delta s = \frac{2 \cdot N_i}{N_{\text{nodes}} \cdot (N_{\text{nodes}} - 1)} \tag{5}$$

As we mentioned previously N_{nodes} is the total number of sources, N_i is the amount of pairs that have a projected separation on the range of $[s, s+\Delta s]$ and Δs is the width of the range we are measuring. We used Δs $= \frac{2 \cdot R_{G14.2}}{20} = 3.5909''$ for the North hub and $3.8437''$ for the South hub which were calculated assuming the maximum possible distance between two sources is the diameter of the projected area obtained earlier, and using 20 segments to see decently the behavior of the distribution. It is also important to understand that N_i isn't related to the branches of the MST, here any possible branch between any pair of sources is taken into account. The reason why we divide by the normalization factor $\frac{N_{nodes} \cdot (N_{nodes}-1)}{2}$ is because it is the total amount of connections between all nodes, dividing by two to avoid counting the same connection twice.

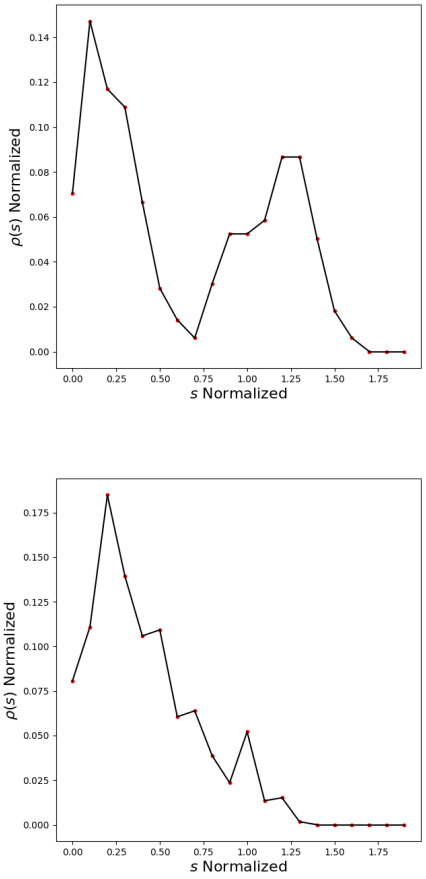

FIG. 3: Normalized distribution function for the millimeter sources detected with ALMA for the North hub (top) and the South hub (bottom), s is normalized to go from 0 to 2 dividing it by the cluster radius .

In Fig. 3 we can observe the distribution functions for the North hub and the South hub of the millimeter sources. It is clear that the North hub has two clear peaks, the first one at a distance of approximately 0.03 pc, and the second one is located at 0.3 pc. On the other hand, the South hub has a main peak at a distance of 0.06 pc and then it smoothly fades, with a small secondary peak at 0.3 pc. These distances are calculated as:

$$D = \Delta s \cdot s_{\text{norm}} \cdot 10 \cdot 1600 \tag{6}$$

where Δs must be in radians. We multiply s by ten to get how many times we have to multiply Δs and we multiply it by the radial distance of the cluster which is 1600 pc.

D. MST and cluster properties at other wavelengths

We generated the MST, Q parameter and distribution function for the North and South hub of the centimeter

and IR sources. Mass segregation cannot be studied at these wavelengths because the catalogs do not include this information. For the IR data, we note that we did not separate the cloud into North and South hubs. Instead we divided it in sub-catalogs according to their evolutionary stage: Stage 0/I (protostars) and Stage II/III (pre-main sequence stars). The MST and the distribution functions are presented in the appendix. The most notable differences are that the distribution function for the IR has a wide peak at low separation and then it decays very smoothly to zero for Stage 0/I as well as for Stage II/III. This effect is due to the fact that the IR catalog has a vast amount of sources, far greater than the other two and its distribution is more homogeneous. The centimeter sources present a first peak and then some secondary peaks as the value gets lower, specially at the North hub.

	mm N	**mm S**	**cm N**	**cm S**	**IR 0/1**	**IR 2/3**
Median	2.485	2.274	45.401	20.950	68.903	58.314
Mean	3.794	3.713	45.477	36.239	87.421	76.442
Deviation	3.956	4.868	45.477	33.926	70.576	57.606
Min	0.444	0.461	4.045	2.816	2.538	3.158
Max	21.386	27.393	136.148	128.358	411.582	320.261
Q	0.475	0.685	0.875	0.932	0.709	0.722
$\mathbf{R}_{G14.225}$	35.909	38.437	249.510	247.257	2368.990	2380.273

TABLE II: Table of statistical data of the distances (all in arcseconds $\pm$ 0.001), and Q parameters for the different sources and hubs.

Table 2 shows the values of the statistical data that we collected from the catalogs at different wavelengths. We can see that milimeter and mid-IR sources tend to subclustering and centimeter are more peak centered.

III. CONCLUSIONS

Our aim in this project was to analyse the structure of the distribution of the stellar popullation in the IRDC G14.225, analyse its cluster properties and to see if they were notable differences between YSO's of different class as we got our sources from different wavelengths form a variety of cathalogs for the milimeter [2], centimeter[4], and mid-IR[12].

Overall we can see that young stellar objects in the IRDC G14.225 tend to form in fractals, we observe clustering behavior as we have to separate North and South hub, and we even still see clustering behavior inside this hubs. On the other hand it is not sure to affirm that mass segregation is a phenomenon that occurs in G14.225 at least not at the millimeter sources. Also there is no tangible difference between the North and South hub of the millimeter and centimeter sources in regard of the data of table 2 and The mid-IR doesn't show many differences between the stage 0/I and II/III sources.

We have observed that the field of the IR and cm sources does not contain the field of the millimeter ALMA sources, this doesn't allow us to take a common field thus making comparison hard in some aspects. Even though this already shows segregation in the G14.225 by age. In a future it would be interesting to try to merge all the sources in one catalog to observe the phenomenon of the mass segregation at a larger scale. It is also worth noting that we lack some information about the actual geometry of the cluster, even though we know the IRDC G14.225 is at a distance of $\sim$ 1600 pc, we don't know the 3D distribution, even though all this projects relies on observed distances these are only projected two dimensional angular data, two stars that appear to be close in our projection could be quite far away in the radial axis.

Acknowledgments

Many thanks to my advisor Gemma Busquet Rico, for guiding me trough the process and helping me clear my doubts in the parts of the work that weren't familiar for me, as well for providing me with a very complete bibliography to get the necessary information. As well I would like to thank all the support from my parents, my roommates and my partner Anna for helping me organize my time and to take pauses when needed.

[1] Busquet et al. "Unveiling a cluster of protostellar disks around the massive protostar GGD 27 MM1"(2019)
[2] Brichs "Identification of the protosetllar popullation in the massive cloud IRDC G14.225" (2021)
[3] Cartwright & Whitworth (2004)
[4] Díaz-Márquez et al. "Radio survey of the stellar population in the infrared dark cloud G14.225-0.506" (2024)
[5] Jeans et al. "The Stability of a Spherical Nebula" (1902)
[6] Kruskal algorithm
[7] Mistree
[8] Krumholz et al. (2019)
[9] Povich et al. "Radio survey of the stellar population in the infrared dark cloud G14.225-0.506" (2016)
[10] Povich & Whitney (2010)
[11] Sadaghiani, M. et al. "Physical properties of the star-forming clusters in NGC 6334. A study of the continuum dust emission ith ALMA". A&A, **635**, A2 (2020)
[12] Zinnecker et al. (1993)
[13] Zucker et al. (2020)

Mass segregation and cluster properties

Author: Enric Fernández Salgado, efernasa48@alumnes.ub.edu
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Gemma Busquet Rico, gemma.busquet@ub.edu

Resum: El núvol fosc infraroig G14.225 alberga dos protoclústers als seus nuclis nord i sud, fet que el converteix en un objectiu ideal per investigar l'agrupament i la segregació de masses en les primeres etapes de formació estel·lar. Analitzant dades de múltiples longituds d'ona procedents de catàlegs de fonts mil·limètriques, centimètriques i d'infraroig mitjà, hem construït arbres d'expansió mínima (MSTs) per estudiar la distribució espacial dels objectes estel·lars joves (YSOs). Es van calcular paràmetres estadístics clau, com el paràmetre Q i el coeficient de segregació de masses ($\Lambda_{\rm MSR}$), per avaluar si els cúmuls presenten estructures fractals o concentrades al centre. Els nostres resultats revelen subagrupaments ($Q < 0.8$) en algunes fonts i una concentració central en altres, així com una notable segregació de masses al nucli nord, mentre que el nucli sud mostra una distribució més uniforme. Aquests resultats milloren la comprensió sobre les condicions físiques, la dinàmica i l'evolució dels protoclústers.

Paraules clau: Núvols de pols foscos infrarojos, segregació de masa, arbres d'expansió mínima, formació estelar, propietats de cúmul, subcúmuls.

Objectius de Desenvolupament Sostenible (ODSs o SDGs)

1. Fi de la es desigualtats	10. Reducció de les desigualtats
2. Fam zero	11. Ciutats i comunitats sostenibles
3. Salut i benestar	12. Consum i producció responsables
4. Educació de qualitat	13. Acció climàtica
5. Igualtat de gènere	14. Vida submarina
6. Aigua neta i sanejament	15. Vida terrestre
7. Energia neta i sostenible	16. Pau, justícia i institucions sòlides
8. Treball digne i creixement econòmic	17. Aliança pels objectius
9. Indústria, innovació, infraestructures	

IV. APPENDIX

All codes used in this project are publicly uploaded at Github.

A. Plots of the MST at other wavelengths

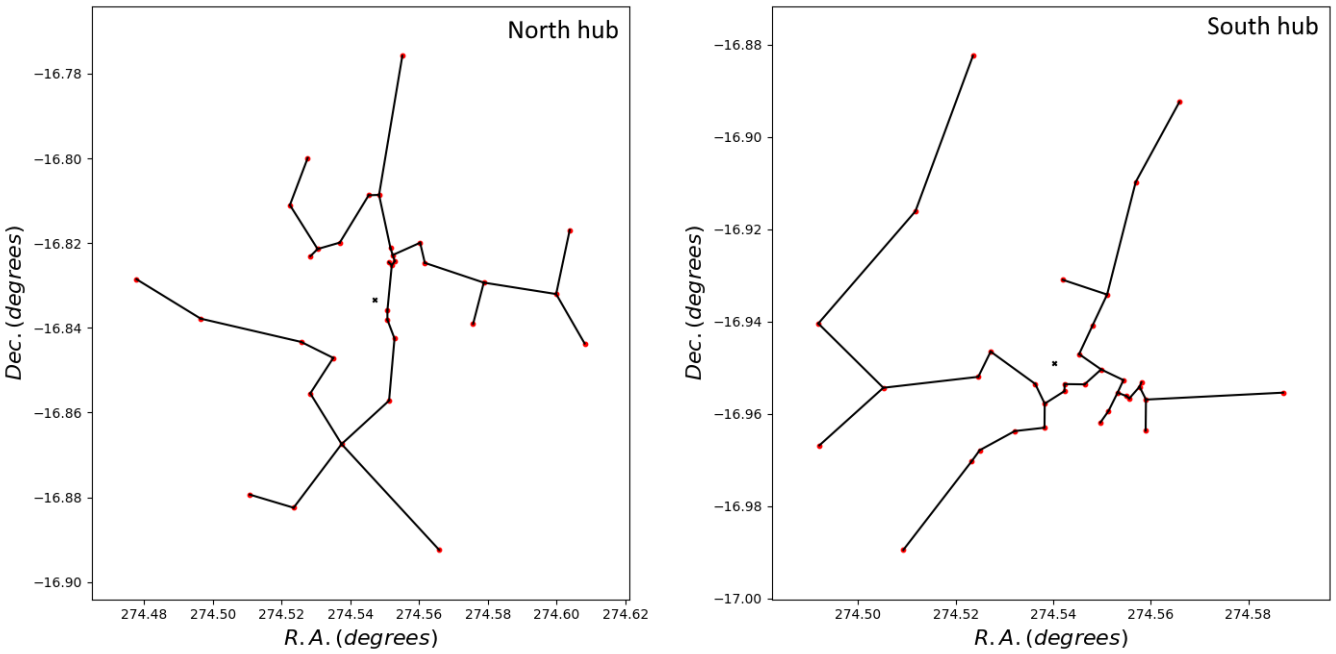

FIG. 4: MST for the centimeter sources in G14.225-North (left) and G14.225-South (right). Black mark represents the center of the cluster, obtained as the mean position of the sources.

B. Distribution function

Region	**Peak 1**	**Peak 2**	**Peak 3**	**Peak 4**
mm North	0.03	0.3	X	X
mm South	0.06	0.3	X	X
cm North	0.8	1.5	1.9	3.1
cm South	0.2	1.2	X	X
IR 0/I	9.2	X	X	X
IR II/III	9.2	X	X	X

TABLE III: Peak distances (in parsecs) for each source, hub or stage.

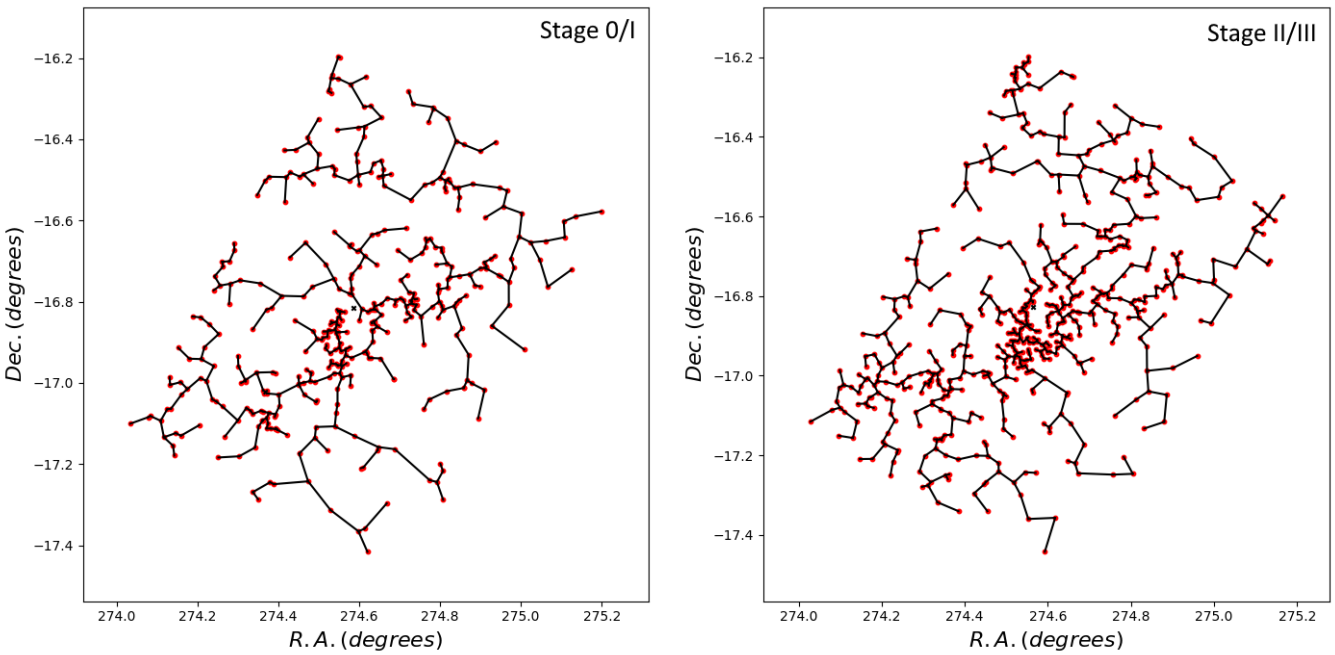

FIG. 5: MST for the IR sources in G14.225 for SED stage 0/I (left) and SED stage II/III (right). Black mark represents the center of the cluster, obtained as the mean position of the sources.

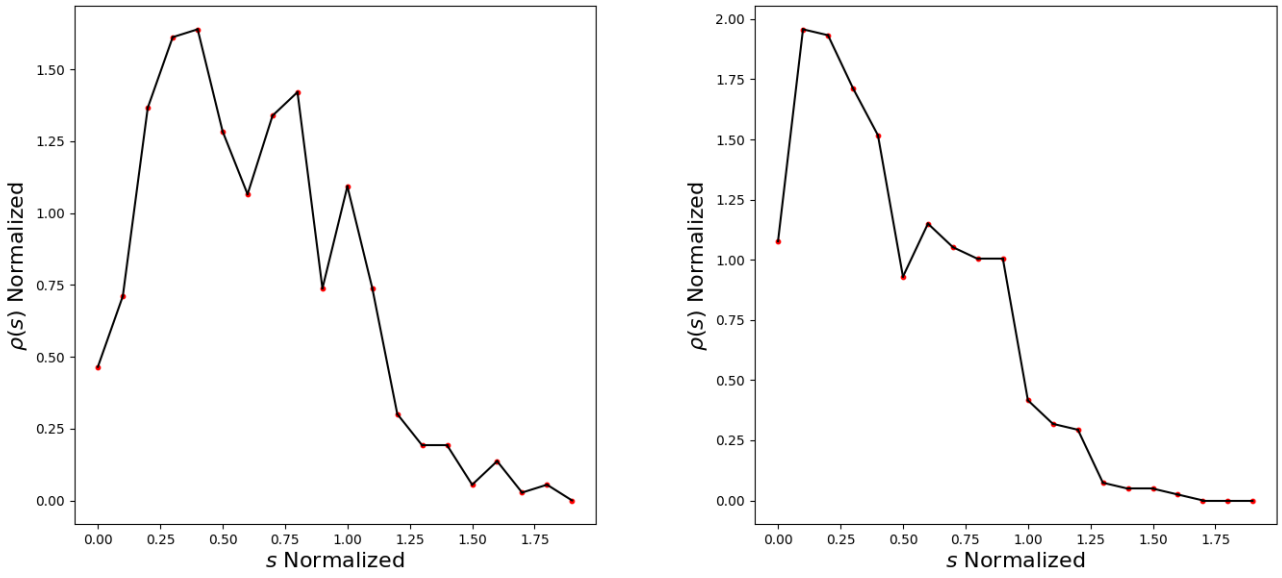

FIG. 6: Normalized distribution function for the centimeter sources detected with ALMA for the North hub (left) and the South hub (right), s is normalized to go from 0 to 2 dividing it by the cluster radius a

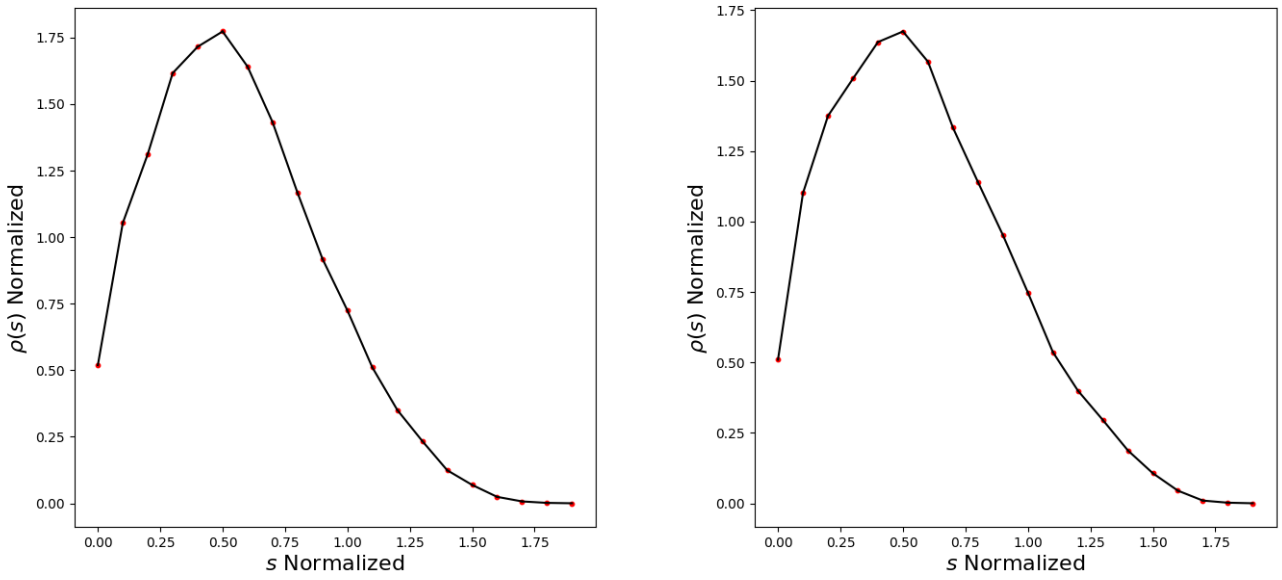

FIG. 7: Normalized distribution function for the mid-IR sources detected with ALMA for SED Stage 0/I (left) and SED stage II/III(right), s is normalized to go from 0 to 2 dividing it by the cluster radius and ρ(s) is normalized so that the total area of the probability has a value of 1.