Clustering Properties of Intermediate and High-mass Young Stellar Objects[*]

Miguel Vioque[1,2] , Manuel Cavieres[3] , Michelangelo Pantaleoni González[4] , Álvaro Ribas[5] , René D. Oudmaijer[6] ,
Ignacio Mendigutía[4] , Lena Kilian[7] , Héctor Cánovas[8] , and Michael A. Kuhn[9]

[1] Joint ALMA Observatory, Alonso de Córdova 3107, Vitacura, Santiago 763-0355, Chile; miguel.vioque@alma.cl
[2] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[3] Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile
[4] Centro de Astrobiología (CSIC-INTA), Departamento de Astrofísica, ESA-ESAC Campus, E-28691 Madrid, Spain
[5] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
[6] School of Physics and Astronomy, Sir William Henry Bragg Building, University of Leeds, Leeds LS2 9JT, UK
[7] School of Earth and Environment, University of Leeds, Woodhouse Lane, LS2 9JT, Leeds, UK
[8] Telespazio UK for the European Space Agency (ESA), European Space Astronomy Centre (ESAC), Camino Bajo del Castillo s/n, E-28692 Villanueva de la Cañada, Madrid, Spain
[9] Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK

Abstract

We have selected 337 intermediate- and high-mass young stellar objects (YSOs; $1.5–20\,M_\odot$) well-characterized with spectroscopy. By means of the clustering algorithm HDBSCAN, we study their clustering and association properties in the Gaia DR3 catalog as a function of stellar mass. We find that the lower-mass YSOs ($1.5–4\,M_\odot$) have clustering rates of 55%–60% in Gaia astrometric space, a percentage similar to that found in the T Tauri regime. However, intermediate-mass YSOs in the range $4–10\,M_\odot$ show a decreasing clustering rate with stellar mass, down to 27%. We find tentative evidence suggesting that massive YSOs ($>10\,M_\odot$) often (yet not always) appear clustered. We put forward the idea that most massive YSOs form via a mechanism that demands many low-mass stars around them. However, intermediate-mass YSOs form in a classical core-collapse T Tauri way, yet they do not appear often in the clusters around massive YSOs. We also find that intermediate- and high-mass YSOs become less clustered with decreasing disk emission and accretion rate. This points toward an evolution with time. For those sources that appear clustered, no major correlation is found between their stellar properties and the cluster sizes, number of cluster members, cluster densities, or distance to cluster centers. In doing this analysis, we report the identification of 55 new clusters. We tabulated all of the derived cluster parameters for the considered intermediate- and high-mass YSOs.

Unified Astronomy Thesaurus concepts: Star formation (1569); Clustering (1908); Young star clusters (1833); Star clusters (1567); Young stellar objects (1834); Herbig Ae/Be stars (723); Massive stars (732); T Tauri stars (1681); Emission line stars (460); Protoplanetary disks (1300)

Supporting material: machine-readable table

1. Introduction

A significant fraction of the young stellar object (YSO) population in the Galaxy is spatially clustered to some degree (e.g., Kuhn et al. 2014; Zari et al. 2018) and appears associated with star-forming regions and molecular clouds (Carpenter 2000; Lada & Lada 2003; Grasha et al. 2019; Krumholz et al. 2019; Kuhn et al. 2021a). This is expected from a theory of turbulent collapse driving the hierarchical evolution of molecular clouds (e.g., Fujii & Portegies Zwart 2016; Barnes et al. 2019; Grudić et al. 2021, 2022). Often, YSOs can be found in comoving groups of stars (e.g., Prisinzano et al. 2022). Some of these groups can be fitted with precision by a single isochrone, thus hinting at a common origin (e.g., Castro-Ginard et al. 2022). Nevertheless, evidence points against a simple linear star formation history in many of these groups, which are often gravitationally unbound and hierarchically distributed

(Kuhn et al. 2014, 2019; Rodríguez et al. 2019; Guszejnov et al. 2022; Mendigutía et al. 2022; Sun et al. 2022).

Despite this clustering nature, all-sky systematic surveys have reported many YSOs in relative isolation. For the low-mass regime ($M < 1.5\,M_\odot$), ~40%–50% of YSOs have been found relatively isolated (Gutermuth et al. 2009; Schmeja et al. 2009; Gouliermis et al. 2012; Winston et al. 2020; Kuhn et al. 2021b). Even within massive star-forming regions, ~20% of the YSOs appear unclustered (Kuhn et al. 2015). This number is consistent with theoretical expectations (e.g., Kruijssen 2012; Krumholz & McKee 2020). These isolated YSOs are often the result of dispersed short-lived clusters, ejection from nearby clusters, isolated star formation, or a combination of all of the previous phenomena.

A large-scale analysis of the clustering properties of intermediate- and high-mass YSOs ($M > 1.5\,M_\odot$) is still missing. At this mass regime, the isolated nature of some forming stars is of particular interest. These sources evolve much faster than their lower-mass counterparts (e.g., Bressan et al. 2012), and thus there is significantly less time for them to be scattered in the field. In addition, they require larger reservoirs of material for their formation, which suggests a stronger connection with clustered environments. Indeed, more-massive YSOs tend to appear more clustered (e.g.,

---

[*] Table 1 is also available online at the CDS in the VizieR archive service.

Hillenbrand 1995; Testi et al. 1999), and larger clusters have more-massive YSOs (e.g., Kirk & Myers 2012; Ksoll et al. 2021). However, many intermediate- and high-mass YSOs have been found relatively isolated, outside star-forming regions, and not belonging to any stellar overdensity (Sitko et al. 2008; Bressert et al. 2012; Law et al. 2022; Fedriani et al. 2023; Kuhn et al. 2023). Studies of OB stars also point in the direction that not all massive stars form in clusters (de Wit et al. 2005; Ward et al. 2020; Wright et al. 2022). This places important constraints on massive star formation theories, particularly in the case of the more-massive stars ($M >$ $10\,M_\odot$). Relatively isolated massive YSOs are often indicative that a monolithic core collapse has happened (e.g., McKee & Tan 2003), whereas the absence of a significant population of isolated massive YSOs would be supportive of a competitive-accretion scenario dominating massive star formation (e.g., Wang et al. 2010).

In addition, environmental and clustering properties can significantly affect YSOs, particularly when in the proximity of massive stars (Lodato & Manara 2023). High-density environments can affect their protoplanetary disks (e.g., Panić et al. 2021; Concha-Ramírez et al. 2022; Mendigutía et al. 2022; Winter & Haworth 2022), stellar accretion (e.g., Winter et al. 2020a), and planet forming mechanisms (e.g., Kruijssen et al. 2020; Winter et al. 2020b; Chevance et al. 2021; Longmore et al. 2021). In addition, larger stellar densities increase the chances of flybys, which can greatly affect star and planet formation (see Cuello et al. 2023).

The advent of the Gaia mission (Gaia Collaboration et al. 2016) has brought a revolution in the study of stellar clusters and associations (e.g., Cantat-Gaudin et al. 2020). In this paper, we make use of its Data Release 3 astrometry (Lindegren et al. 2021b) and photometry (Riello et al. 2021), in combination with a new large sample of well-characterized sources (Vioque et al. 2022), to present the first systematic study of the clustering properties of intermediate- and high-mass YSOs. In Section 2, we present the sample of 337 intermediate- and high-mass YSOs we consider in this study. In Section 3, we describe how we have identified clusters and associations of Gaia DR3 sources containing these YSOs using the machine-learning HDBSCAN algorithm. We analyze these clusters and the isolated YSOs in Section 4, including a comparison with published catalogs of open clusters. We discuss our findings in Section 5, and we conclude in Section 6.

2. Sample

We compiled 337 well-characterized intermediate- and high-mass YSOs ($M > 1.5\,M_\odot$) with derived stellar parameters from spectra and Gaia DR3 five-parameter astrometrical detections (α, δ, ϖ, $\mu_{\alpha*}$, μ_δ; i.e., R.A., decl., parallax, proper-motion R. A., and proper-motion decl.). We selected 222 of these sources from the catalogs of historically considered and well-studied intermediate- and high-mass YSOs of Vioque et al. (2018) and Guzmán-Díaz et al. (2021). To these, we added 115 of the newly identified intermediate- and high-mass YSOs of Vioque et al. (2022). The latter were selected from the Vioque et al. (2020) catalog, and hence are not biased toward any preferred location in the Galaxy (although the Vioque et al. 2020 catalog is limited to the Galactic plane, $-5 < b < 5$).

We obtained two sets of distances from Gaia DR3 parallaxes for this sample of 337 sources, and compare them against each other. The first set concerns the geometric distances of Bailer-

Jones et al. (2021). For the second set, we used a Bayesian approach with a prior for massive stars (see Pantaleoni González et al. 2021). Both distance sets are in great agreement, only showing significant differences for five (2%) of the sources. Indeed, 98% of the sample has $\varpi/\sigma(\varpi) > 3$ and 84% of the sample has ruwe < 2 (see Lindegren et al. 2021b for Gaia goodness-of-fit parameters). We conclude that our distances are accurate within error bars and independent of the underlying distance prior. Because of this accuracy, from now on we only consider the Bailer-Jones et al. (2021) distances. The main reason for this is that the other distance set assumes a massive star ($>10\,M_\odot$) prior, and all five discordant sources are known to be below this threshold.

For accuracy, only effective temperatures derived from spectra are considered (in order of precision, we use those listed in Fairlamb et al. 2015; Vioque et al. 2018, 2022; Wichittanakom et al. 2020, and references therein). We adjust the compiled luminosities from previous works to the Gaia DR3 Bailer-Jones et al. (2021) geometric distances. We then use these luminosities and effective temperatures to re-derive stellar masses homogeneously for the whole sample. For this, we use the PARSEC 1.2S pre-main-sequence tracks (Bressan et al. 2012). A Hertzsprung–Russell (HR) diagram of the sample is shown in Figure 1. The result is 317 sources in the sample with stellar mass determinations. We did not derive stellar masses for the 20 remaining sources, whose effective temperatures and luminosities place them to the left of the theoretical pre-main sequence in the HR diagram (Figure 1; this is often caused by unresolved binarity).

2.1. Sample Completeness and Biases

The intermediate- to high-mass YSO regime comprises different historical types of sources. Between $1.5 < M < 8\,M_\odot$, YSOs are categorized into the Herbig Ae/Be group (e.g., Vioque et al. 2018), at the latest stages of pre-main-sequence evolution, and their cooler predecessors the intermediate-mass T Tauris (IMTTs; e.g., Valegård et al. 2021). At $M > 8$ to $10\,M_\odot$, YSOs are generally referred to as massive YSOs (MYSOs; e.g., Koumpia et al. 2021; Marcos-Arenal et al. 2021). In this work, we use the term "intermediate- and high-mass" when referring to any YSO with $M > 1.5\,M_\odot$ and optically bright enough to be detected by Gaia (as opposed to the T Tauri regime containing optically bright low-mass YSOs). If we refer to particular stellar mass ranges, we explicitly mention it in the text. We note that most of the sources in the sample fit within the classical Herbig Ae/Be regime, as it is the one more accessible with Gaia (Figure 1; see Brittain et al. 2023).

The sample of 337 stars considered in this work contains most of the known intermediate- and high-mass YSOs, which are well characterized. We note that there are thousands of other proposed intermediate- and high-mass YSO candidates (e.g., Robitaille et al. 2008; Vioque et al. 2020; Creevey et al. 2023), but these lack spectroscopical confirmation of their nature and stellar parameters. We note that accurate stellar parameters and extinctions are necessary to derive stellar masses. For this reason, and the larger uncertainty on their true nature, we chose to leave these candidates out of this study. We thus expect the contamination of our sample to be very low. We note that the Gaia DR3 Apsis from DPAC was highly inefficient when identifying intermediate- and high-mass YSOs (mainly because they are a very small fraction of the Gaia
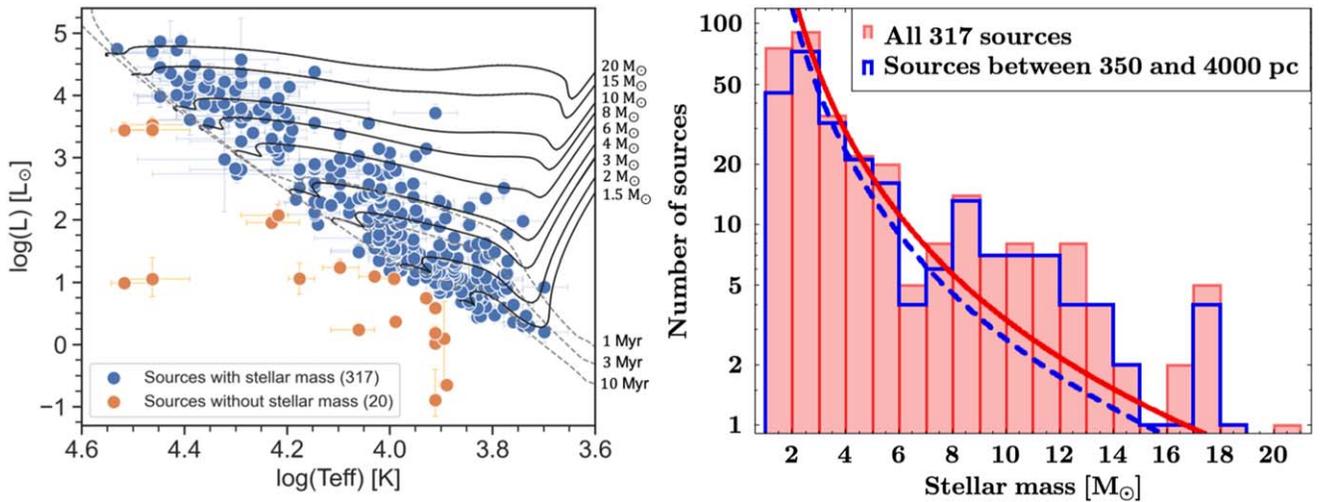
Figure 1. Left: Hertzsprung–Russell (HR) diagram of the considered sample of 337 intermediate- and high-mass YSOs. PARSEC 1.2S PMS tracks and isochrones corresponding to 1.5, 2, 3, 4, 6, 8, 10, 15, and 20 $M_\odot$ and 1, 3, and 10 Myr are presented (Bressan et al. 2012 and Marigo et al. 2017). Right: histogram of the number of YSOs considered in this work as a function of stellar mass. The lines indicate a Salpeter (1955) initial mass function (IMF) adjusted to the 2–3 $M_\odot$ bin.

Universe DPAC addresses; see Ae/Be stars in Table 2 of Fouesneau et al. 2023). Hence, no reliable sources can be included from Gaia DR3 Apsis' catalogs of young stars. There are a few other spectroscopically characterized intermediate-mass YSOs in the literature, which we have not considered to keep the sample as homogenous as possible (this includes some or all of the sources from Nuñez et al. 2021; Shridharan et al. 2021; Valegård et al. 2021; Zhang et al. 2022; Iglesias et al. 2023; Kuhn et al. 2023). We estimate that, within the distance range considered in the analysis and discussion of this work (350–4000 pc; see Section 3.2 and Figure 2), our sample contains at least 85% of the known and spectroscopically characterized intermediate- and high-mass YSOs.

We have not considered the YSOs that are very embedded in their parental material, and are not optically visible due to large extinctions (i.e., those who do not appear in Gaia, which has a faint limit of $G \lesssim 21$ mag). This implies we are mainly considering Class II and Class III YSOs in this study. Figure 1 shows an HR diagram and the stellar mass distribution of the sample. Looking at Figure 1, we see that our sample is covering well the range of 1.5–20 $M_\odot$. In principle, we would expect the above incompleteness to mainly affect the higher-mass regime, due to the faster evolution of more-massive stars (e.g., Figure 1 isochrones show that above 4 $M_\odot$ all sources are younger than 1 Myr). However, fitting an initial mass function (IMF; Figure 1) to the 2–3 $M_\odot$ bin shows that the more-massive population of YSOs is overrepresented in our sample ($M > 8\,M_\odot$). This is probably due to the fact that, although optically visible $M > 8\,M_\odot$ YSOs are rarer in their mass range, they can be observed at much larger distances. Therefore, our selected sample of sources can be considered as representative in mass of the evolved (Class II, Class III) Galactic population of intermediate- and high-mass YSOs.

In the YSO massive regime ($M > 8\,M_\odot$), we have estimated that only ∼20% of the total population appears in Gaia (comparing with the rms survey; Lumsden et al. 2013). However, due to the fast evolution of these massive sources, it is uncertain whether Gaia is tracing the more evolved population or the population that shredded its envelope through some unknown dynamical process. In any case, the dynamical cluster dissipation timescales are much larger than the massive
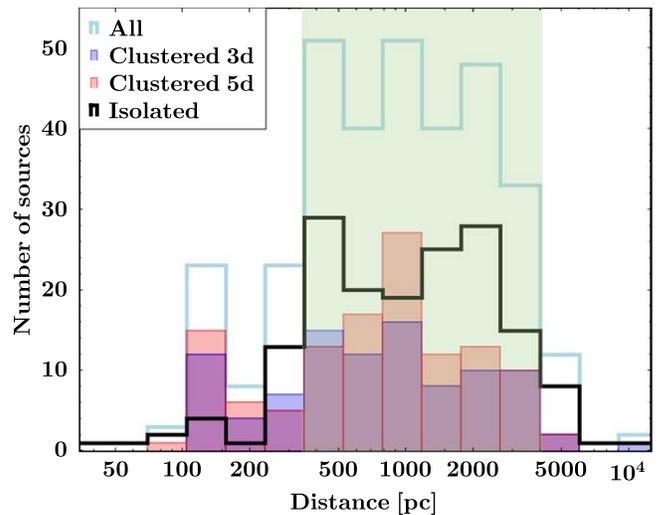

Figure 2. Histogram of the 337 YSOs considered in this work as a function of distance (light blue contour). Black contours are isolated stars. Blue and red bars indicate stars clustered in 3D or 5D space, respectively (we note 49 sources appear clustered in both). The green area shows the region within which the clustering algorithm HDBSCAN is most sensitive, and where we limit our analysis. Distances are Bailer-Jones et al. (2021) geometric distances.

star formation timescales (Farias & Tan 2023; Hao et al. 2023). Hence, this sample bias caused by the optical limitation of Gaia is likely to have a limited impact on our analysis of the clustering properties of intermediate- and high-mass YSOs. A more detailed and technical analysis of the biases and completeness of the methodology is presented in Appendix A.

In conclusion, the sample of 337 sources gathered in this work is not volume complete. However, it contains the majority of known and well-characterized intermediate- and high-mass YSOs with accurate Gaia data, and it is representative in mass of the optically visible ($G \lesssim 21$ mag) YSO population ranging 1.5–20 $M_\odot$. More in-depth information of the properties of this sample can be found in Fairlamb et al. (2015), Vioque et al. (2018), Arun et al. (2019), Wichittanakom et al. (2020), Guzmán-Díaz et al. (2021), Grant et al. (2022), and Vioque et al. (2022), among other works. A complete

review of this population of objects, with references, is provided by Brittain et al. (2023).

3. HDBSCAN Clustering Methodology

In this Section, we use the unsupervised machine-learning algorithm HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise; Campello et al. 2013; McInnes et al. 2017) to identify clusters and associations in the Gaia catalog containing the intermediate- and high-mass YSOs of the sample described in Section 2. Hunt & Reffert (2021) concluded that HDBSCAN is the most sensitive and effective algorithm for recovering open clusters in Gaia data. Cánovas et al. (2019) got to the same conclusion identifying members of the young (<5 Myr) Rho Oph star-forming region. We refer the reader to the aforementioned references for a detailed explanation on how HDBSCAN selects clusters. In the following subsections we describe our use of the Python implementation of HDBSCAN.[10]

3.1. HDBSCAN Methodology

HDBSCAN has three hyperparameters of importance (i.e., algorithm related variables that need to be decided by the user). The hyperparameter "cluster_selection_method" defines how clusters are selected from the cluster tree. We use the "leaf" method in this work, as recommended by Hunt & Reffert (2021) for identifying clusters in Gaia data. An additional advantage of "leaf" for this work is that it favors the selection of all of the clusters present in a field, down to the smaller ones, without excluding the possibility of detecting large clusters. The two other hyperparameters to consider are "min_cluster_size" and "min_samples." The first one defines the smallest sample size we can consider a "cluster." The second one can be understood as a quantifier of how conservative HDBSCAN is in selecting clusters. In this work, we set these two parameters to be equal unless stated otherwise, as it is customary in most HDBSCAN applications where there is no previous knowledge of the type and number of clusters present in the search field.

In order to select the field of stars to search for clusters, we queried Gaia DR3 for the HEALPix pixel level 5 (Hierarchical Equal Area isoLatitude Pixelization; see Górski et al. 2005), which contains each of the massive YSOs, and the eight HEALPix pixels around it. Each searched area, hence, has $\sim$31 deg^2 and side length approximately 5°. According to Hunt & Reffert (2021), this angular size is sufficient for detecting clusters at the wide range of distances considered. The only quality requirement is that sources must have a five-parameter Gaia DR3 astrometric solution (α, δ, ϖ, $\mu_{\alpha*}$, and μ_{δ}; faint limit $G \sim 21$ mag; Lindegren et al. 2021b). No other quality constraint to the astrometry was applied. The effects of this, and the selection biases induced by demanding a Gaia DR3 five-parameter astrometric solution are proved to be very minor in Appendix A.1. For 82 sources, we used a smaller HEALPix pixel, as the fields were very crowded (level 6, $\sim$8 deg^2, and side length of 2.5°) and HDBSCAN was failing to identify similar scale-size associations. This smaller field size does not bias the cluster identification (see Appendix A.2 for a more detailed analysis).

As a preprocessing step, we applied the zero-point bias correction to ϖ as described in Lindegren et al. (2021a). In

addition to this, the five dimensions of each field (α, δ, ϖ, $\mu_{\alpha*}$, and μ_{δ}) were re-scaled to have a median of zero and a unit interquartile range. This re-scaling process ensures that each parameter has an equal weight for HDBSCAN (Hunt & Reffert 2021).

We then applied the HDBSCAN algorithm to all Gaia DR3 stars in these fields using different combinations of parameters and hyperparameters. First, we looked for clusters in the 5D astrometric space (α, δ, ϖ, $\mu_{\alpha*}$, and μ_{δ}). We did this twice, using "min_cluster_size" and "min_samples" equaling [10, 10] and [30, 30]. Then, we repeated the search for clusters in these fields in a 3D astrometric space (α, δ, ϖ; "3D physical space"). In this latter case, we also ran the code twice, for "min_cluster_size" and "min_samples" equaling [10, 10] and [30, 30]. We thus ran HDBSCAN four times for each field containing each intermediate- and high-mass YSOs. This is done to cover the different combinations of possible cluster types (in physical space and proper motions or only in physical space) and minimum cluster sizes (10 and 30). An exploration of other HDBSCAN hyperparameters is presented in Appendix B. This exploration shows that the hyperparameters chosen in this work are appropriate for identifying clusters and associations with HDBSCAN.

We note that circumstellar extinction is often significant in YSOs. This extinction varies from source to source and can only be characterized with spectroscopic data. This has prevented us from using the color–magnitude diagram as a tool to identify or evaluate clusters.

3.2. Results of HDBSCAN

The methodology of Section 3.1 was applied homogeneously to all 337 intermediate- and high-mass YSOs in the sample. HDBSCAN assigns a normalized probability to every source in each field of belonging to a cluster or association. A probability of zero means the star is not in a cluster, while nonzero probabilities indicate different degrees of association to a cluster. To avoid biasing our results toward YSOs at cluster centers (which typically receive higher HDBSCAN probabilities), we considered as clustered the sources with a nonzero probability in either of the two configurations of "min_cluster_size" and "min_samples" (higher probability thresholds are considered in Section 4 to support the analysis). As a result, we obtain 121/337 sources that appear as clustered in 5D (physical space and proper motions), and 97/337 stars that appear as clustered in 3D (physical space only). There is not a large overlap between both groups, with 49 sources appearing clustered in both parameter spaces. This is not surprising, as all dimensions are treated equally by Section 3.1 methodology, and thus the proper motions make a significant difference in the clusters that can be identified or rejected (see Section 3.3 for a comparison between 3D and 5D clusters). All clusters have an HDBSCAN probability larger than 0.5, and 70% of the detected clusters have an associated HDBSCAN probability larger than 0.95. There are 168 sources (50% of the sample) that do not appear clustered in any HDBSCAN configuration, and hence they can be considered as isolated.

These results of HDBSCAN are shown in Figure 2 as a function of distance. From Hunt & Reffert (2021), we know the field angular size we are using is inefficient for detecting real associations below 350 pc. Figure 2 confirms this. Likewise, it is evident from Figure 2 that 4000 pc is likely a threshold for the HDBSCAN cluster identification sensitivity. Therefore, we

---

[10] https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html
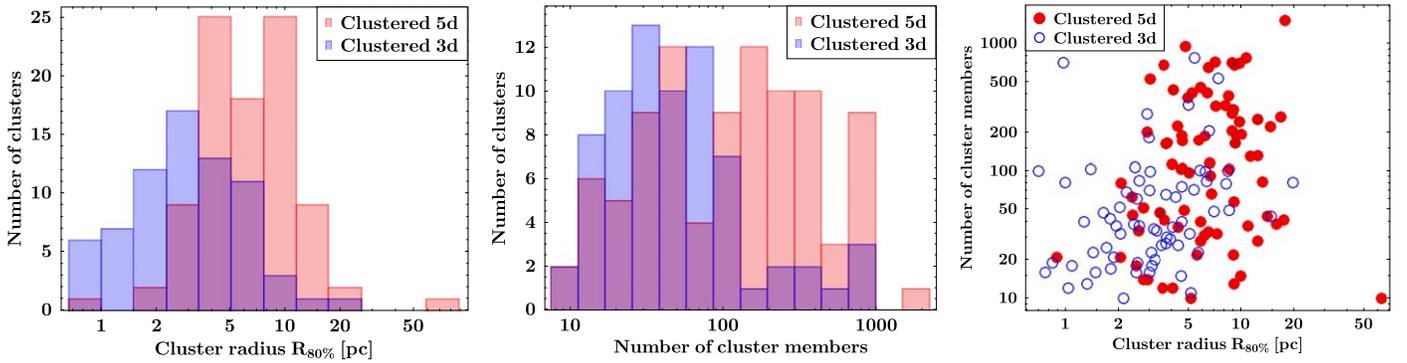
Figure 3. Left: histogram of cluster radii identified in 3D and 5D spaces. Cluster radius is defined as that containing 80% of the cluster's stars from the cluster geometrical center. Center: histogram of the number of sources in each cluster. Right: cluster radii vs. number of sources per cluster identified in 3D (blue) and 5D (red) spaces.

define 350–4000 pc as the region where our HDBSCAN methodology is most sensitive, and unless otherwise specified, we limit our results and analysis to it. There are both clustered and isolated stars throughout this distance range, containing 263 stars.

Of the 263 stars, 92 are clustered in 5D (35%), 71 are clustered in 3D (27%), 36 are clustered in both 3D and 5D (14%), and 127 are clustered in 3D or 5D (48%). A total of 136 sources are isolated (52%), not appearing clustered in any HDBSCAN configuration. The clustering nature, cluster properties, and HDBSCAN results for each intermediate- and high-mass YSO considered in this work between 350 and 4000 pc are summarized in Table 1.

3.3. Cluster Sizes and Number of Cluster Members

In this Section, we describe the properties of the identified clusters and associations. In Figure 3, the sizes of all identified clusters and the number of sources they contain are shown. To characterize the cluster size, we used a radius defined as the angular distance to the geometrical cluster center (in R.A. and decl.) that contains 80% of the cluster's stars. The angular radius was converted to physical radius ($R_{80\%}$) using the distance to the massive YSO contained in each cluster. A radius containing 80% of the clusters' stars was used instead of the more typical 50% value (Hunt & Reffert 2023) to better characterize the full extent of these clusters and associations (e.g., Meingast et al. 2021). We note that some clusters have been identified using both a "min_cluster_size" of 10 and 30 (Section 3.1). In those cases, we always used as reference the more populated cluster identified in each case (as it has more statistical meaning). We also note that in some cases HDBSCAN identifies clusters with tidal tails, or that are heavily nonspherical, so the $R_{80\%}$ radii should always be interpreted as the size of the clusters in their more elongated direction.

Figure 3 shows that the clusters identified in 5D space are larger on average than those identified in 3D space, with $R_{80\%}$ mean and standard deviation of 7.6 ± 6.8 and 3.7 ± 3.0 pc, respectively. These differences in radial sizes are caused by HDBSCAN often being able to identify larger clusters when proper-motion information is also provided. The large standard deviations show the large scatter in cluster sizes in both 3D and 5D. Overall, 90% of the identified clusters are smaller than 10 pc, and all but one are smaller than 20 pc. Figure 3 also shows that the clusters identified in 5D space are typically more

populated than those identified in 3D space, but again the spread is very large in either case. In 5D space, 45% of the clusters have less than 100 members, but this percentage is of 80% for the 3D case. Only 13% and 6% of clusters in 5D and 3D have more than 500 members, respectively.

From Figure 3, only two cases of cluster misidentification are suspected. The first one is the cluster with $R_{80\%} = 62$ pc and 10 members. The second one is the cluster with 706 members in less than a parsec. These two clusters might not be real associations, but we cannot rule out their existence. For consistency, we kept them in the analysis of Section 4, but in Table 1 they are marked as possible contaminants.

There are 24 YSOs that share a cluster with other YSOs of our list. None of these sources are likely to be binaries considering their Gaia astrometry, and they span the entire extent of stellar masses considered (Figure 1). We find that these clusters with more than one intermediate- and high-mass YSO are on the denser and more populated end of the cluster distribution (Figure 3).

The population of known open clusters (see Figure 12 of Hunt & Reffert 2023) have $R_{50\%}$ typical radii of 1.5–5 pc, and up to 15 pc. These sizes are very similar to the ones we find in our population of clusters and associations. The number of cluster members in the population of known open clusters also match well the results we retrieve for our sample (Hunt & Reffert 2023; see their Figure 2), even for the more populated clusters we obtain. Hence, we conclude that the clusters and associations identified in this work are similar in size and in number of members to the population of known open clusters (see Section 4.3 for further comparison).

4. Analysis

In this Section, we analyze the clustered and isolated intermediate- and high-mass YSOs identified in Section 3, and their relation to measured cluster properties.

From here onwards, we define as "clustered" those YSOs appearing clustered in 3D or in 5D (Section 3.2), unless stated otherwise. We note that this definition considers to be "clustered" whatever star that is not clearly isolated in Gaia space, when the literature usually defines clustered sources in the opposite way (as stars clearly belonging to a group; see Cantat-Gaudin et al. 2020; Hunt & Reffert 2023, and references therein). This approach was taken because Sections 3.2 and 3.3 have evidenced that, with our methodology, we are detecting different types of clusters in 3D and 5D. In Section 4.3 we
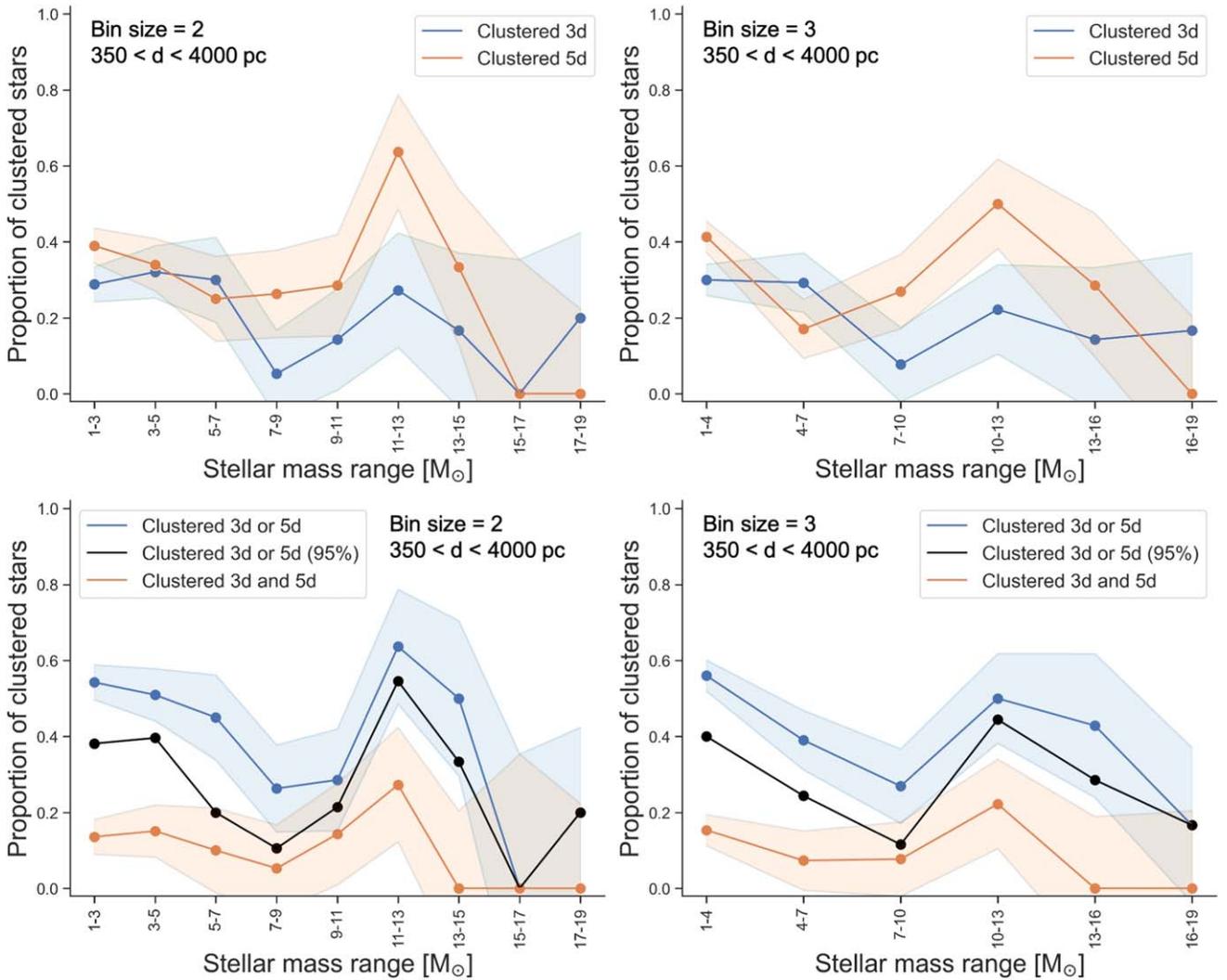
Figure 4. Proportion of clustered intermediate- and high-mass YSOs as a function of stellar mass. Top plots show the proportion of clustered stars in 3D and 5D HDBSCAN executions, left plot with 2 $M_{\odot}$ bins and right plot with 3 $M_{\odot}$ bins. Bottom plots show the proportion of clustered stars in either "3D or 5D" and both "3D and 5D." Shown in black are the "3D or 5D" results when only considering those sources for which HDBSCAN reports more than a 95% clustering probability. Unless otherwise specified, in this work we define as "clustered" the "3D or 5D" combination (Section 4). Complete error is $1/\sqrt{n}$, with n being the number of sources per bin.

compare the results of our approach with the population of known open clusters in the Galaxy.

4.1. Clustered Nature as a Function of YSO Mass

In Section 3.2, we concluded that 48% of the known Galactic population of intermediate- and high-mass YSOs appears clustered. In Figure 4 we show the fraction of clustered intermediate- and high-mass YSOs as a function of stellar mass. To minimize the impact of statistical flukes on the interpretation, different binning to the stellar mass were considered, as well as the four combinations of 3D, 5D, "3D and 5D," and "3D or 5D" HDBSCAN results (Section 3.2).

Regarding the "3D or 5D" case, Figure 4 shows that at the lower-mass end of the sample (1.5–4 $M_{\odot}$, mainly the Herbig Ae and IMTT regimes) ∼55% of the sources are clustered. This number nicely matches the proportion of clustered stars found in the T Tauri regime using similar methodologies (e.g., Winston et al. 2020; Kuhn et al. 2021b). However, as we go up in stellar mass (4–10 $M_{\odot}$), the proportion of clustered stars diminishes, down to ∼27% at the 7–10 $M_{\odot}$ range. The

proportion of clustered stars goes up again from 10 $M_{\odot}$ to a ∼65% clustering rate at 11–13 $M_{\odot}$, but it seems to decay abruptly shortly after that, and most stars above 15 $M_{\odot}$ appear isolated (albeit see below and Section 5.2). This trend of intermediate-mass YSOs in the range 4–10 $M_{\odot}$ being less clustered than lower-mass YSOs (<4 $M_{\odot}$) and the massive YSOs (10–13 $M_{\odot}$) is noticeable in all other HDBSCAN results (3D, 5D, 3D, and 5D; Figure 4) and mass binning. Furthermore, this trend is also apparent when only considering as clustered those sources for which HDBSCAN reports more than a 95% clustering probability (Section 3.2, black lines of Figure 4).

We perform a two-sample Kolmogorov–Smirnov (K-S) test to probe the significance of these variations in clustering rate as a function of stellar mass. The null hypothesis is that, in stellar mass, the groups of clustered and isolated YSOs are drawn from the same unknown probability distribution. When the K-S test is performed using the whole sample, we obtain a p-value of 0.025. Because above 10 $M_{\odot}$ we are affected by low-number statistics and some biases (see below and Section 5.2), we also

report a *p*-value of 0.041 when we limit the analysis to $10\,M_\odot$. We repeat this K-S analysis only considering as clustered those sources for which HDBSCAN reports more than a 95% clustering probability (Section 3.2). The *p*-values obtained in this latter case are 0.083 (whole sample) and 0.039 (limited to $10\,M_\odot$). We conclude that these K-S tests show that the clustering rate varies with stellar mass to within ~95% significance.

As described in Section 2, the considered sample of YSOs has two main components, a heterogeneous collection of historically known and studied intermediate- and high-mass YSOs (compiled from Vioque et al. 2018 and Guzmán-Díaz et al. 2021), and a set of intermediate- and high-mass YSOs homogeneously identified with all-sky surveys (Vioque et al. 2020, 2022). In Figure 5, we show the clustering ratio as a function of stellar mass of these two groups. It is noteworthy that, although the trend of clustering ratio going down with stellar mass until $10\,M_\odot$ holds for both samples, the two sets show opposed behaviors above $10\,M_\odot$. For masses $M > 10\,M_\odot$, the newly identified sources show a sharp increase of the clustering rate up to a 60%–100% clustering rate at $15\,M_\odot$, whereas the historically known sources show a smooth decline of their clustering rate with mass, with only 0%–40% of them above $10\,M_\odot$ appearing clustered. This trend is also apparent when only considering as clustered those sources for which HDBSCAN reports more than a 95% clustering probability (Section 3.2), or when considering only HDBSCAN results in 3D, 5D, and "3D and 5D." We note that the sample from Vioque et al. (2022) is limited to $15\,M_\odot$, whereas the historical sample extends to $20\,M_\odot$. We also note that the low number of sources at these stellar mass ranges do not allow for robust statistics, as is indicated by the large uncertainties of Figure 5 (see Section 5.2 for further discussion).

4.2. Cluster Properties as a Function of YSO Mass

In this Section, we study the properties of the identified clusters as a function of the properties of the intermediate- and high-mass YSOs they contain. Figure 6 shows the YSO stellar masses as a function of the number of stars per cluster (N_c), the cluster radii ($R_{80\%}$), and the cluster densities. The cluster density is defined as $0.8N_c/\pi R_{80\%}^2$, and thus can be understood as the average number of cluster stars per square parsec, assuming the clusters have a circular distribution in the sky (which is not always a good assumption; see Section 3.3). In the case of the clusters that were identified both in 3D and 5D, in Figure 6 we report the properties of the 5D clusters, which typically have more members and statistical significance (see Figure 3). Most clusters have 0.1–10 stars per squared parsec, with an average mean of $3.2\,\mathrm{stars\,pc^{-2}}$. As illustrated in Figure 6, we find no dependence of cluster properties with the stellar mass of the considered intermediate- or high-mass YSO.

We also study the location of the intermediate- and high-mass YSOs within the clusters, to trace possible mass segregation. In Figure 7, we show the angular distances from the intermediate- and high-mass YSOs to the geometric center of the clusters (in R.A. and decl.), converted to physical distances by using the distances to the YSOs. Again, we find no correlation between the stellar masses of the intermediate- and high-mass YSOs and their distances to the cluster centers. The only appreciable fact is that above $6\,M_\odot$, no YSO is beyond 10 pc of its cluster center, but there is no statistical significance in this given that most clusters are smaller than 10 pc
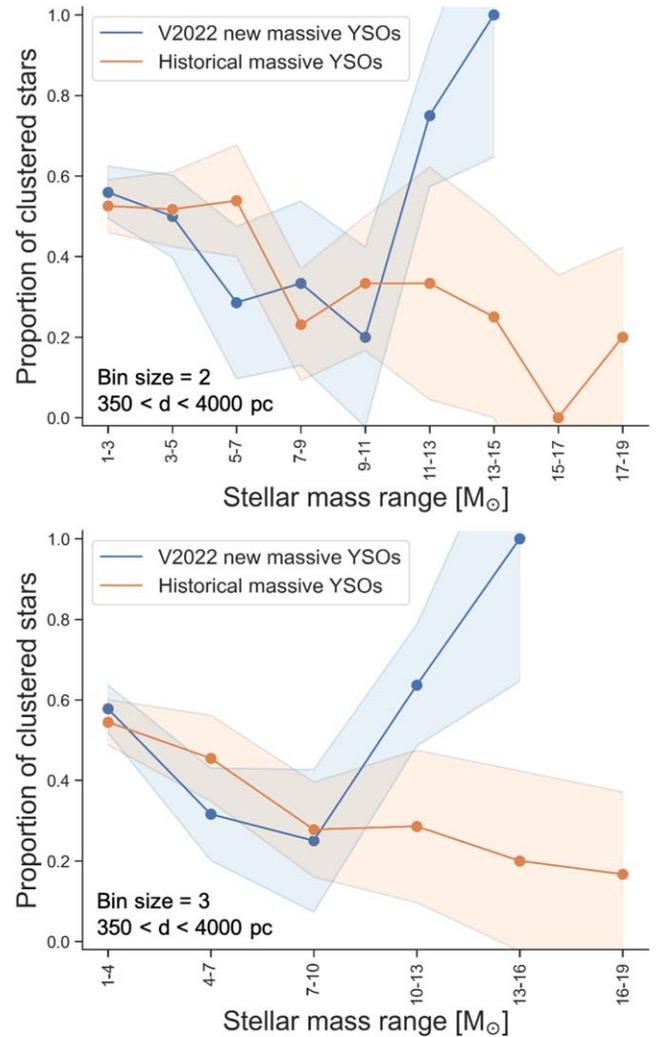

Figure 5. Proportion of clustered ("3D or 5D") intermediate- and high-mass YSOs as a function of stellar mass. Historically known YSOs have been separated from those identified homogeneously and in a position-unbiased way by Vioque et al. (2022). The top and bottom plots differ in the stellar mass bin used. Complete error is $1/\sqrt{n}$, with n being the number of sources per bin.

(Section 3.3, Figure 3). In 11% of the cases, the intermediate- or high-mass YSO is located beyond the $R_{80\%}$ radius. This percentage goes up to 46% and 86% for $R_{80\%}/2$ and $R_{80\%}/5$ radii, respectively (gray lines of Figure 7). Therefore, we estimate that only ~15% of the clustered intermediate- and high-mass YSOs are in the inner 20% region of the clusters, and that this percentage does not depend much on the stellar mass of the YSO (Figure 7). Again, only a weak trend can be reported, which is that the intermediate- and high-mass YSOs under $4\,M_\odot$ mostly appear as the only ones outside the $R_{80\%}$ radius.

4.3. Comparison with Literature Open Clusters and Associations

In this Section, we compare the clusters obtained with HDBSCAN in Section 3 with different compilations of open clusters and associations of the literature, most of them also based in Gaia DR3. We note that this comparison is limited to the 263 sources within 350–4000 pc. Unless otherwise specified, cross-matches with other catalogs were done with a 5″ aperture.
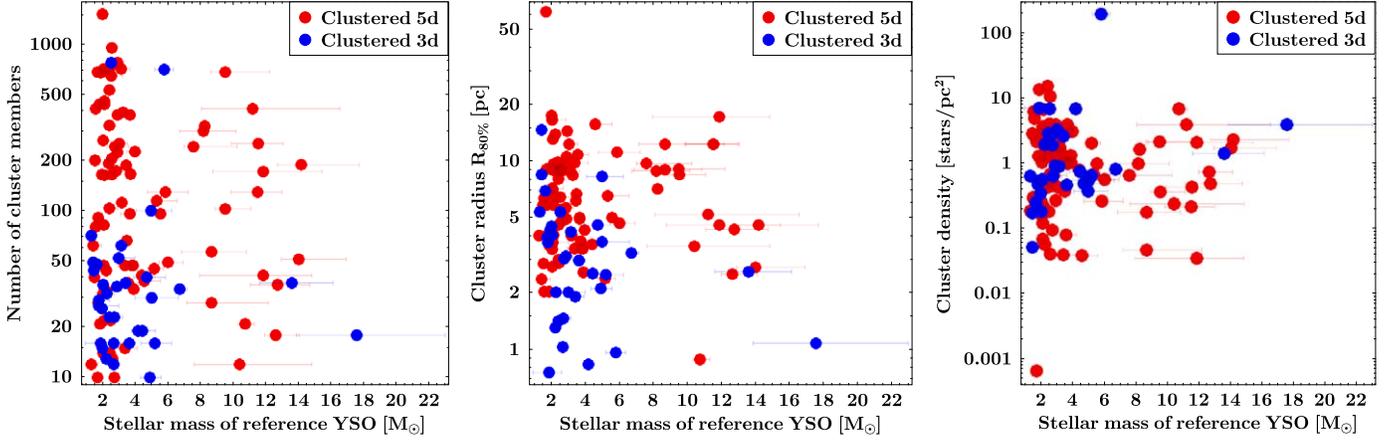
Figure 6. Cluster properties as a function of the stellar mass of the intermediate- or high-mass YSO contained in each cluster. Left: number of sources in each cluster (N_c). Center: cluster size ($R_{80\%}$). Right: cluster density ($0.8N_c/\pi R_{80\%}^2$).
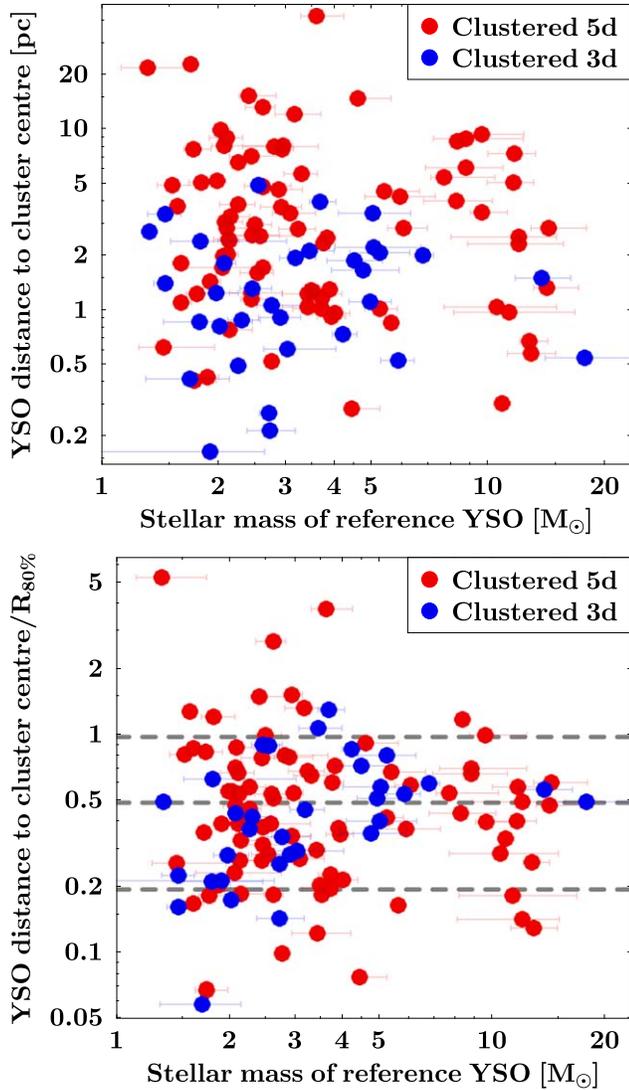

Figure 7. Top: distances of the intermediate- and high-mass YSOs to the geometrical center of their clusters as a function of their stellar mass. Bottom: distances of the intermediate- and high-mass YSOs to the geometrical center of their clusters divided by the cluster sizes ($R_{80\%}$) vs. the stellar mass of the intermediate- and high-mass YSOs.

1. Hunt & Reffert (2023) . We have 78 stars in common with this catalog of cluster members. Of these 78, we found 70 clustered and eight isolated.

2. Kerr et al. (2023). There are 41 stars in our sample which the SPYGLASS Catalog of Young Associations identified as clustered ($P_{\mathrm{mem}} > 0.95$). Of these 41, we identified 28 as clustered.

3. Castro-Ginard et al. (2022). We have only one star in common with this catalog of cluster members, and we also identify it in a cluster.

4. Prisinzano et al. (2022). This is an unbiased survey of YSO associations in the Milky Way. Only two stars of our YSO sample match their list of members. Of these two, one we detect as clustered and one as isolated.

5. Dias et al. (2021). Our sample of YSO sources has six stars matching this catalog of YSO clusters within $2'$. We also identify these six sources in a cluster.

6. Cantat-Gaudin et al. (2020). We have 37 stars in common with this catalog of cluster members. Of these 37, we have detected 30 as clustered.

7. Cantat-Gaudin & Anders (2020). We have 46 stars in common with this catalog of cluster members. Of these 46, we have detected 35 as clustered.

8. Kounkel & Covey (2019). We have 50 stars in common with this catalog of cluster members. Of these 50, we have detected 36 as clustered.

Regarding infrared-based catalogs of clusters and associations:

1. SPICY catalog of YSOs (Kuhn et al. 2021b). There are only 24 YSOs of our list in SPICY. Of these, SPICY reports 15 as clustered, of which we find 11. In contrast, we find 14 clustered stars and 10 isolated, and 11 and 4 of them, respectively, appear clustered in SPICY. We note that SPICY only considered clusters of suspected YSOs in its HDBSCAN implementation, and used different HDBSCAN hyperparameters to those of this work (see Section 3.1).

2. Bica et al. (2019). We have 12 stars in common with this catalog of cluster and association members. Of these 12, we have detected seven as clustered.

3. Kharchenko et al. ([2013](#)). Our sample of YSO sources has 18 stars matching this catalog of YSO clusters within 2′. We identify 13 of these sources in a cluster.
4. Gutermuth et al. ([2009](#)). We have nine stars in common with this catalog of cluster members. Of these nine, we have detected eight as clustered.

Hence, the general agreement with previous surveys of clusters and associations is of 70%–90%. This agreement holds for IR-based cluster surveys (e.g., 73% for SPICY and 89% for Gutermuth et al. [2009](#)). If we only consider the most recent and complete survey looking for clusters (Hunt & Reffert [2023](#)), the agreement is of 90%. When analyzing the YSOs that we report as isolated and other authors reported as clustered, we do not see any trend favoring more-massive objects. Hence, these comparisons show no bias in our methodology toward detecting less clusters at higher stellar masses.

The works of Hunt & Reffert ([2023](#)), Castro-Ginard et al. ([2022](#)), and Cantat-Gaudin et al. ([2020](#)) claim to be almost complete within 2 kpc, where ∼76% of our YSOs are (Figure [2](#)). It is hence noteworthy that 55 clustered stars in our analysis are not listed in these works. For example, 36 of the 94 clustered stars we have identified within 2000 pc do not appear clustered in Hunt & Reffert ([2023](#)). We believe this is due to the looser definition of "cluster" used in this work (Sections [3.1](#) and [4](#)). We retrieve most of the clusters reported in the literature, but also other less bound associations and stellar overdensities that were discarded by previous authors.

The main differences between our work and most other Gaia-based cluster works is that we use both a 3D and a 5D space to identify clusters, we do not attempt to fit the clusters photometrically in a color–magnitude diagram (because of circumstellar extinction and because, e.g., sequential star formation might have occurred; Nakajima et al. [2005](#)), and we do not perform a post-HDBSCAN case-by-case vetting (in order to keep the results homogeneous; see Section [4](#)). Indeed, an in-depth look shows that the clusters we report, which Hunt & Reffert ([2023](#)) did not identify, are typically the less dense ones, with larger radii, and fewer members. In particular, many of the clusters reported here and unidentified by Hunt & Reffert ([2023](#)) have less than 50 members. We report these 55 new clusters that we have found in Gaia space for future reference and analysis (Table [1](#)).

We emphasize that our approach was designed to trace exhaustively and homogeneously the population of clustered intermediate- and high-mass YSOs. The strength of our analysis is that our methodology has been applied homogeneously to all sources, without favoring any interpretation of the definition of "cluster," and that it includes both the 3D and 5D Gaia astrometric spaces for cluster selection. It is beyond the scope of this work to describe individual clusters, but we caution the reader that some of the clusters reported here can be loose or sparse associations, not fitting the canonical definition of open cluster. Indeed, many known YSO associations do not fit the definition of open cluster (e.g., the TW Hydrae sparse association; Luhman [2023a](#)). We refer the reader to the works of Hunt & Reffert ([2023](#)), Castro-Ginard et al. ([2022](#)), and Cantat-Gaudin et al. ([2020](#)) for a description, analysis, and discussion of open cluster identification in the Gaia era.

4.4. Comparison with Iglesias et al. (2022)

Iglesias et al. ([2023](#)) selected 220 intermediate-mass YSOs (1–7 $M_\odot$) within 400 pc as traced by their IR excess at 22 μm. It is interesting to compare Iglesias et al. ([2023](#)) sample with the one in this work because there are significant differences between them. Although both samples were selected by demanding a certain IR-excess level, our selection often resulted in the presence of emission lines (mostly tracing active accretion; e.g., Vioque et al. [2022](#); Brittain et al. [2023](#)) whereas the Iglesias et al. ([2023](#)) selection did not (only six of their 220 stars have detected emission lines). This indicates that the sample considered in this work, even though it is representative in mass distribution of the IMF (Section [2](#)), might be a subsample of the total Galactic population of intermediate- and high-mass YSOs; i.e., the subsample of sources with larger accretion rates. The results of Grant et al. ([2023](#)) seem to point in this direction. It is unclear whether the Iglesias et al. ([2023](#)) sample is more evolved than the one compiled for this work (Section [2](#)), or if we have been biased in the past toward mainly identifying YSOs undergoing accretion bursts (see Section [5](#) for further discussion).

In this Section, we apply the exact same HDBSCAN methodology of Section [3](#) to the sample of intermediate-mass YSOs of Iglesias et al. ([2023](#)). The Iglesias et al. ([2023](#)) sample is contained within 400 pc. As described in Section [3.2](#), this short distance is suboptimal for the HDBSCAN methodology of this work. Hence, to avoid any possible bias in the comparison with our sample, we also limited our sample to 400 pc in this Section. The clustering rate as a function of stellar mass for both samples is shown in Figure [8](#). For every mass bin, the clustering rates of the Iglesias et al. ([2023](#)) sample are lower. We report a ∼60% clustering rate for our sources, and a ∼25% clustering rate for Iglesias et al. ([2023](#)) sources. Both samples show roughly constant clustering rates as a function of stellar mass up to 5 $M_\odot$ (within error bars). Beyond 5 $M_\odot$, low-number statistics impede us to reach meaningful conclusions.

In Appendix [A.2](#), we describe the detected biases of applying our clustering methodology (Section [3](#)) to sources at short distances. We find a bias against identifying sources as isolated below 200 pc (Figure [12](#)). We note that 57% of Iglesias et al. ([2023](#)) sources are within 200 pc. In contrast, 38% of the sources considered in this Section from our sample are within 200 pc. Therefore, if correcting from this bias, we would expect to find an even larger difference in clustering rate between both samples.

4.5. Biases and Incompleteness

In Appendix [A](#) we study, in depth, the caveats, biases, and sources of incompleteness of the methodology and analyses described in Sections [3](#) and [4](#). This Section provides an informative summary of the conclusions of Appendix [A](#).

We have applied the same methodology homogeneously to most known and spectroscopically characterized intermediate- and high-mass YSOs. According to the IMF, the considered sample of YSOs is representative in stellar mass of the population of intermediate- and high-mass YSOs in the Galaxy (1.5–20 $M_\odot$; Figure [1](#)). This is further reinforced by the fact that the sample is well distributed in stellar mass across the Galaxy. Nevertheless, this sample has some selection biases. Mainly, we are limited to objects bright enough in the optical to
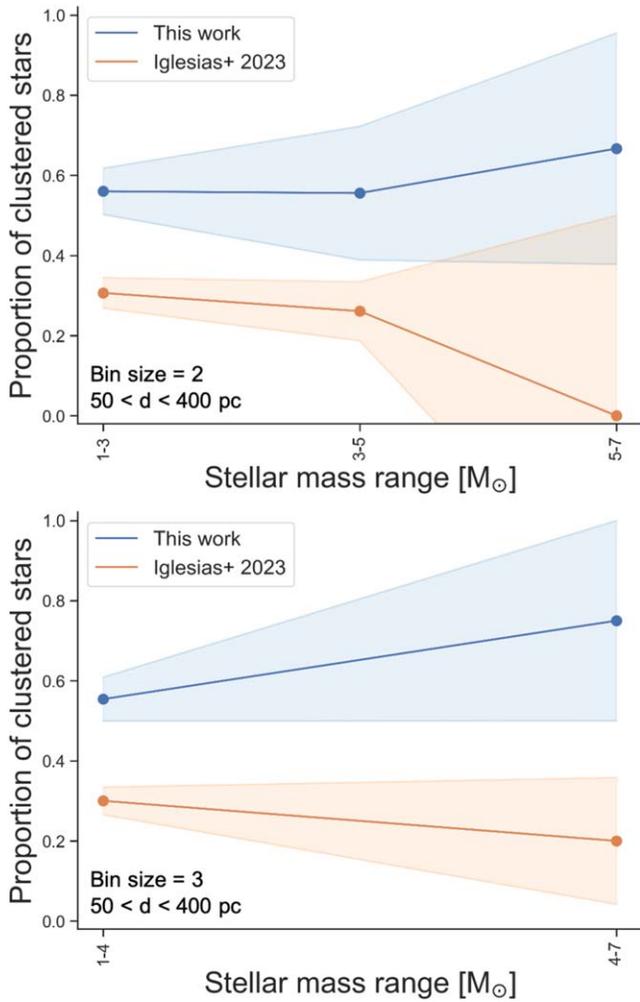
Figure 8. Comparison of the clustering rate between the YSO sample of this work and the YSO sample of Iglesias et al. (2023). Complete error is $1/\sqrt{n}$, with n being the number of sources per bin. For visualization purposes, the plots show a smaller range of stellar masses with respect to Figure 4, as the Iglesias et al. (2023) sample is limited to $7\,M_\odot$.

be detected by Gaia (faint limit $G \sim 21$ mag). Hence, we expect to be mostly tracing Class II and III objects, as the younger Class I sources are often too extincted for Gaia. This limitation also applies to the sources used to look for clusters with HDBSCAN, this resulting in incomplete Gaia five-parameter (5D) fields.

We have analyzed the Gaia 5D completeness by comparing with other optical and IR surveys and YSO catalogs. We conclude that more stars per mass bin would have appeared clustered if there were Gaia data for all embedded sources around each YSO. Hence, the results presented in Figures 4, 5, and 8 are likely lower limits. However, we do not see any dependence with stellar mass in this limitation caused by Gaia 5D incompleteness. Hence, the evolution of clustering rate as a function of stellar mass is robust (i.e., the shape of Figures 4, 5, and 8). We report no dependence of our results with the astrometric quality of the sources used for HDBSCAN.

We have also analyzed whether there are distance or sky location biases affecting our analyses and conclusions. We find that none of our results depend on biases of that sort. In particular, we observe the trend of clustering rate going down in the mass regime 4–10 $M_\odot$ at all distance ranges (Figure 13).

The reported cluster properties are also independent of the distance.

There is some ambiguity in the literature with respect to the definition of a "cluster." In this work, we claim that a YSO is in a "cluster" if HDBSCAN detects it as in a stellar overdensity in the Gaia 3D or 5D astrometric spaces, with respect to the surrounding field. In Section 3.3, we evidence that the clusters identified in this work have similar sizes and number of members to the population of open clusters identified with Gaia (see Hunt & Reffert 2023). In addition, we have identified most of the clusters reported in the literature containing the YSOs we have considered (Section 4.3). This evidences that our clustering detection methodology is robust and consistent with previous efforts looking for clusters and associations. In addition, we have identified 55 previously unreported clusters. These new clusters are in general of a lower stellar density, larger, and with fewer members than most of the known open clusters. We thus believe that our methodology is looser in the definition of cluster, and that these new clusters we have identified were likely discarded by previous authors' methodologies who aimed to achieve a high-purity in the canonical interpretation of open cluster (e.g., Cantat-Gaudin et al. 2020; Castro-Ginard et al. 2022; Hunt & Reffert 2023). We note that there are several known sparse associations of YSOs that also do not fit the standard definition of "open cluster," which are also typically larger and with fewer members (e.g., Luhman 2022, 2023a, 2023b). We thus provide a more exhaustive identification of the population of clustered or associated intermediate- and high-mass YSOs.

5. Discussion

In this work, we have used Gaia to study the clustering properties of the known and spectroscopically characterized population of intermediate- and high-mass YSOs (1.5–20 $M_\odot$; Section 2). We have limited our analysis to the 350–4000 pc range. In this Section, we discuss our results and the implications for star formation mechanisms at different stellar mass ranges.

5.1. Clustering Rate as a Function of Stellar Mass

Of the 263 stars considered in the analysis, 127 are clustered in 3D or 5D (48%) and 136 sources are isolated (52%), not appearing clustered in any HDBSCAN configuration. This fraction varies with stellar mass (Figure 4), and in Section 4.1 we prove this variation to be statistically significant. In particular, we find that the lower-mass YSOs (1.5–4 $M_\odot$) have clustering rates of ∼55%, a percentage similar to that found in the T Tauri regime (when a similar methodology and clustering tools are used). However, this clustering rate goes down with stellar mass in the 4–10 $M_\odot$ regime, down to 7–10 $M_\odot$ YSOs showing a ∼27% clustering rate. For sources in the 10–15 $M_\odot$ regime, the clustering rate goes up again to a value similar to or slightly higher than the one found in the lower-mass regime (see the discussion in Section 5.2). Although these percentages are likely lower limits due to optical extinction (Section 4.5), we note that this overall shape of the clustering rate as a function of stellar mass is independent of distance or Gaia completeness (Appendix A), and hence we suspect it is an intrinsic property of star formation. In addition, this result is independent of how we choose to select our clusters (i.e., (i) 3D, (ii) 5D, (iii) 3D and 5D, or (iv) 3D or 5D; Section 3 and

Figure 4). We hence conclude that intermediate-mass 4–10 $M_{\odot}$ YSOs are typically less clustered than low $M < 4\,M_{\odot}$ and high $M > 10$–15 $M_{\odot}$ mass YSOs (see Section 5.2 for some caveats on the high-mass regime).

It is not surprising that at the low-mass end ($<4\,M_{\odot}$) we see only 50%–60% of clustered stars. Chevance et al. (2023) showed that molecular clouds are very hierarchical and fractal systems, with short dynamical timescales. In addition, regions of filamentary collapse to hubs exist, but most stars do not form in hubs (only 10% do; Chevance et al. 2023). Hence, many low-mass stars might have formed relatively isolated. In contrast, Longmore et al. (2014) and Zucker et al. (2023) showed a high-degree of clustered low-mass YSOs, suggesting that those low-mass YSOs appearing isolated might be the result of fast cluster dispersion (e.g., Lada & Lada 2003; Hao et al. 2023) and ejection. In any case, many of the considered low-mass YSOs had time to be dispersed in the field.

However, it is noteworthy that for the intermediate-mass YSOs in the range 4–10 $M_{\odot}$, with lifetimes 1–2 orders of magnitude shorter, the clustering rate goes down with stellar mass. One explanation could be that more-massive YSOs form in more-massive clusters, and that these tend to disperse faster (e.g., Weidner et al. 2007; Oh & Kroupa 2016; Farias & Tan 2023; Hao et al. 2023). However, this would not explain why these intermediate-mass YSOs are often less clustered than the more-massive YSOs above 10 $M_{\odot}$. We suggest that intermediate-mass YSOs in the range 4–10 $M_{\odot}$ have typically a more isolated formation than lower-mass YSOs, and that this is more the case for the more-massive stars within this mass regime. To explain this, we put forward the idea that the most-massive YSOs ($>10\,M_{\odot}$) often form via a mechanism that demands that many stars are formed around them (like the competitive-accretion mechanism; Bonnell et al. 2001, 2004), producing their clustered nature. In contrast, intermediate-mass and low-mass YSOs typically form in an individual manner (e.g., via the monolithic collapse of an individual core; McKee & Tan 2003; Tan & McKee 2004). However, intermediate-mass YSOs in the 4–10 $M_{\odot}$ range do not appear often around very-massive YSOs. This could be due to the sharp decrease of the IMF, or due to some other effect. This would explain why low- and intermediate-mass YSOs have different clustering rates, and why the clustering rate decreases with stellar mass in the intermediate-mass regime.

5.2. Clustering Properties of the More-massive YSOs (M > 10 M⊙) and Evolutionary Trends

Figure 4 shows that the clustering rate goes up from 10 $M_{\odot}$ to 13 or even 15 $M_{\odot}$, but it is not trivial to describe the clustering rate of the more-massive YSOs ($M > 10\,M_{\odot}$). Only 31 stars of our sample within 350–4000 pc have masses above 10 $M_{\odot}$, and only 13 of these have masses above 13 $M_{\odot}$. Hence, we warn the reader that the clustering rates reported for the more-massive star formation regime ($>13\,M_{\odot}$) might suffer from low-number statistics and unknown population biases. We note that this does not affect the results of Section 5.1, as it is the decrease in clustering rate as a function of stellar mass in the intermediate-mass regime (4–10 $M_{\odot}$) that drives the conclusions of that Section.

In Section 4.1, we noted that there is a difference in clustering properties when we divide the sample in those intermediate- and high-mass YSOs identified homogeneously in Vioque et al. (2022) and those YSOs historically known and compiled from the literature. The difference only appears from 10 $M_{\odot}$ (Figure 5). For those sources from Vioque et al. (2022) there is a sharp increase of the clustering rate (up to 100% within error bars at 15 $M_{\odot}$, where the sample ends), whereas for the historically known sources, there is a monotonic decrease. We report no difference in the clusters identified for both sets, and there is no correlation between cluster properties and stellar mass in either sample (see Section 4.2). In addition to the way they were identified, the only difference between both sets of sources is that Vioque et al. (2022) YSOs are on average 2.7 mag fainter in Gaia G band for any given distance and stellar mass bin. This indicates that these YSOs are more embedded, which could be interpreted as a sign of youth. However, we note there are other explanations as to why the historically known massive YSOs we considered might be brighter (e.g., different evolution, envelope shredding, face-on disks, or pole-on sources exposing an outflow cavity). An argument supporting a younger nature of the Vioque et al. (2022) sample is that these sources typically have larger IR excess and Hα emission by construction (see the discussion of Vioque et al. 2020), which is classically understood as indicative of younger objects.

Hence, an interpretation for the different behavior of the two sets of massive YSOs ($>10\,M_{\odot}$) in Figure 5 is an evolutionary trend. The younger and more embedded massive YSOs are mostly clustered, whereas the older and optically brighter population is more isolated. However, we find it unlikely that in general massive YSOs would change their clustering properties so drastically in such a short timescale ($\lesssim 0.1$ Myr; Bressan et al. 2012). We believe some strong bias must also be at play. A possible scenario is that the historically known massive YSOs considered here, mostly first identified in old optical spectroscopic surveys (see Thé et al. 1994), are heavily biased toward bright, little obscured, and isolated objects. Thus, they might not be representative of the true population of massive YSOs. This idea is reinforced by the fact that most massive YSOs are optically faint, due to the extinction from their surrounding envelopes (Lumsden et al. 2013). In contrast, Vioque et al. (2022) sources were selected from Vioque et al. (2020) catalog, which increased the number of known intermediate- and high-mass YSOs with Gaia data by an order of magnitude, homogeneously identifying sources photometrically in an all-sky unbiased fashion with faint limit $G \sim 21$ mag. If we assume the Vioque et al. (2022) sources are more representative of the true population of massive YSOs, then the results of this work (Figure 5) indicate that the majority of the more-massive YSOs are clustered.

There is another evolutionary effect that deserves consideration. In Section 4.4 we compared the clustering rates of the sample of YSOs considered in this work with the clustering rates obtained via the same methodology for the sample of Iglesias et al. (2023, Figure 8). In this case, the comparison is limited to 7 $M_{\odot}$ and 400 pc. We see a clear difference in clustering rate between both samples. The typical values for our sample of YSOs are $\sim$60% and for Iglesias et al. (2023) are $\sim$25%. Iglesias et al. (2023) sources have, on average, lower IR excesses and emission line strengths than our sample, which implies smaller disk radii and masses and smaller accretion rates. Hence, we expect their sources to be typically older than those of this work. This would imply we are tracing a decrease of the YSO clustering rate with time. There are, however, alternative explanations. The disk radius in intermediate-mass

YSOs can be more dependent on the presence of giant planets than on time (see Pinilla et al. 2022; Rich et al. 2022; Stapper et al. 2022, 2023; Guzmán-Díaz et al. 2023), and although it is likely that the accretion rate decreases with time, episodic accretion seems to be the norm (Fischer et al. 2023; Grant et al. 2023).

In conclusion, our results suggest that a large fraction of the more-massive YSOs ($>10\,M_\odot$) are clustered. In addition, we have evidence to suspect that intermediate- and high-mass YSOs become less clustered with time, and—at least in the 1.5–$5\,M_\odot$ range—in a fashion that does not vary much with stellar mass. This evolution toward isolation is probably dominated by cluster expansion and dispersion (Gutermuth et al. 2009; Krumholz et al. 2019; Kuhn et al. 2019; Krause et al. 2020). Cluster dissipation and disruption is likely caused by Galactic evolution (Kruijssen et al. 2011; Messa et al. 2018). However, we note that star cluster survivability after gas expulsion is independent of the impact of the Galactic tidal field (Shukirgaliyev et al. 2019). Hence, we expect the increase of isolated intermediate- and high-mass YSOs with time to be independent of Galactic location (although see Ali et al. 2023).

5.3. Cluster Properties of YSOs of Different Stellar Masses

For those intermediate-0 and high-mass YSOs we detect clustered, we do not see any major correlation between cluster properties and the mass of the considered YSOs. There is also no correlation with the isochronal ages of the YSOs. The only result our analysis suggests is that YSOs above $4\,M_\odot$ are mostly in clusters larger than 2.5 pc, which are typically of low stellar density (<3 stars pc^{-2}; Figure 6), but there are exceptions.

In addition, Figure 7 shows that there is no major correlation between the mass of the YSOs and their projected distance to the geometrical center of the clusters (in R.A. and decl.). We can only note that above $6\,M_\odot$ none of the sources are at very large distances (>10 pc), and there are hints of a decreasing trend of distance to cluster center with YSO stellar mass. Above $12\,M_\odot$, all six clustered YSOs are within 3 pc of the cluster center. Comparing to the cluster radius, most intermediate- and high-mass YSOs are within $R_{80\%}$ (89%), and $\sim$15% of them are in the inner 20% region of the clusters. Again, no correlation with stellar mass is observed.

Following on from the conclusions of Section 5.1, we would expect in a competitive-accretion-like scenario the more-massive YSOs to appear in large clusters, and closer to the center of the clusters. Our results might indicate that the very-massive YSOs ($>10\,M_\odot$) are not in small clusters (see Bonnell & Clarke 1999), and that they are closer to the center of the clusters. However, more data points are needed to confirm this trend (see Section 5.2 for the caveats of the high-mass regime). Competitive accretion also predicts that the mass of the most-massive star in a cluster increases as the system grows in number of stars and in total mass (Bonnell et al. 2004). However, we see no correlation between the amount of stars per cluster and the mass of the considered YSO (Figure 6). In contrast, a monolithic-collapse-like scenario predicts strongly peaked density distributions. We can test this in 2D sky space by using the tidal radius (King 1962), or the radius at which the cluster has the best contrast to field stars. If a monolithic-collapse-like formation dominates at high masses, we would expect to see that low-mass YSOs are typically in clusters with larger tidal radii. However, we do not report any correlation between the tidal radii and the mass of the considered YSOs.

Therefore, the analysis of the properties of the clusters identified in this work is not conclusive of a competitive-accretion-like scenario acting in the most-massive YSOs ($>10\,M_\odot$). Likewise, we find no argument from this analysis supporting a more isolated, monolithic-type formation for the more-massive YSOs. An explanation for these results could be that mass segregation happens at later stages of cluster lifetimes (Della Croce et al. 2023; Farias & Tan 2023), and that cluster properties at early stages are determined by the initial cloud size and mass, and not the stellar content (Farias & Tan 2023; Morii et al. 2023). It is also an option that the YSOs considered in this work are not the most-massive YSOs in several of the detected clusters, which might contain heavily embedded massive YSOs that escaped the Gaia survey.

5.4. Comparison with Previous Results

Our results support the main conclusion of the seminal works of Hillenbrand (1995) and Testi et al. (1997, 1999). There is tentative evidence to suspect that YSOs above $10\,M_\odot$ appear more clustered than intermediate-mass YSOs. They based their results in a 2D near-IR K-band clustering analysis. This work was later extended in Habart et al. (2003), Qiu et al. (2008), and Faustini et al. (2009). We now analyze the clustering properties of intermediate- and high-mass YSOs in the Gaia era, providing an analysis in 3D and 5D. To our knowledge, we are the first to characterize the transition between the clustering properties of low-mass and intermediate-mass YSOs (the 1.5–$10\,M_\odot$ range). Our findings in the 4–$10\,M_\odot$ regime also support the idea that the formation of high-mass stars is influenced by dynamical interaction in a young cluster environment. However, we suggest that the transition in clustering properties at $\sim$10\,$M_\odot$ is not smooth, contrary to what was reported in Testi et al. (1999). In addition, we show that YSOs above $10\,M_\odot$ have similar or higher clustering rates than T Tauri stars, as the difference can be easily explained by the shorter timescale of massive YSO evolution. Contrary to Testi et al. (1999), we see no transition from low-density aggregates of T Tauri stars to dense clusters around massive stars.

Something to consider is the different scales analyzed by different studies. Hillenbrand (1995), Testi et al. (1999), Habart et al. (2003), Qiu et al. (2008), and Faustini et al. (2009) reported clusters with radii of 0.2–1 pc, whereas most of the clusters identified in this work have typical values of 2–15 pc (Figure 3). Another significant difference is the number of stars per cluster. We have identified clusters with a few tens to a few hundred members, whereas previous works mostly report clusters with an order-of-magnitude fewer members. We believe these differences are mainly caused by the difference in methodologies. Previous authors mostly sought clusters in the nearby environments of the intermediate- and high-mass YSOs. This has proven useful in many scientific cases (e.g., Saha et al. 2020; Arun et al. 2021), but this methodology favors peak-density results in the vicinity of the considered YSOs. In contrast, we have applied a position-independent methodology to the whole Gaia space, without favoring any specific mass range (Section 3). Maschberger & Clarke (2008) already alerted that previously reported clusters were too sparsely populated to match random drawing from the IMF. Our results fit the expected theoretical values of Maschberger & Clarke (2008),

and the clusters we have obtained have similar properties to the Galactic population of open clusters. One more significant difference with previous works is that our analysis is optically based, whereas previous works are mostly near-IR based. This does not mean a direct comparison, as in Gaia, we are missing sources fainter than $G \sim 21$ mag (see Appendix A). We conclude that in this work we have retrieved the complete scale of the clusters containing intermediate- and high-mass YSOs, whereas previous works have mostly provided a more complete (in number) yet limited (in volume) view of the vicinity of these YSOs. Both analyses are hence complementary and necessary to achieve high-purity studies like that of Kirk & Myers (2011) for low-mass stars in the vicinity of the Sun.

In this work, we report 18 isolated YSOs with $M > 10 \, M_\odot$ in Gaia astrometric space (four of them from Vioque et al. 2022). Regarding the possibility of truly isolated massive star formation (i.e., in the absence of a cluster environment, we note that $\sim 50\%$–70% of intermediate-mass YSOs are binaries; Baines et al. 2006; Thomas et al. 2023), several candidates have been put forward (e.g., Law et al. 2022; Fedriani et al. 2023), although in-depth analyses have found small clusters around most massive YSOs seemingly isolated (Stephens et al. 2017; Arun et al. 2021; Yan et al. 2023). Similarly, although isolated O stars have been reported (e.g., de Wit et al. 2005; Oey et al. 2013), this is disputed and it often seems to be a result of observational limitations (Parker & Goodwin 2007; Stephens et al. 2017). Theoretical analysis show that seemingly isolated O stars might be explained by low-mass clusters sampled randomly from a standard IMF (Parker & Goodwin 2007).

6. Conclusions

This study constitutes the first large-scale analysis of the clustering properties of intermediate- and high-mass YSOs (1.5–20 $M_\odot$), encompassing the classical groups of Herbig Ae/Be stars, MYSOs, and intermediate-mass T Tauris. We applied an HDBSCAN clustering methodology to 337 spectroscopically characterized intermediate- and high-mass YSOs with Gaia data. We analyzed the clustering properties of the 263 stars located in the distance range where our methodology works best (350–4000 pc). We present the resulting cluster parameters for these sources in Table 1. The main conclusions we report are as follows:

1. Intermediate-mass YSOs in the range 4–10 $M_\odot$ are more isolated than T Tauri stars and Herbig Ae/late-Be stars ($<4 \, M_\odot$) and MYSOs ($>10 \, M_\odot$). The clustering rate in this intermediate-mass regime decreases as a function of stellar mass. We propose that this is due to the fact that most YSOs above $10 \, M_\odot$ demand the presence of clusters of stars around them, which contain mostly low-mass stars, and hence both groups have high clustering rates. In contrast, YSOs in the 4–10 $M_\odot$ regime are too massive to appear often in these MYSO-induced clusters, but not massive enough to have their own clusters. We propose they tend to form similarly to T Tauri stars, via a monolithic-collapse-like scenario. Hence, intermediate-mass star formation in the range 4–10 $M_\odot$ is more isolated than star formation at other mass regimes.

2. Of the 263 stars, 71 are clustered in 3D (27%, position and parallax), 92 are clustered in 5D (35%, position,

parallax, and proper motion), 36 are clustered in both 3D and 5D (14%), and 127 are clustered in 3D or 5D (48%). A total of 136 intermediate- and high-mass YSOs are isolated (52%), not appearing clustered in any Gaia astrometric space. These clustering rates vary with stellar mass. In particular, we report a $\sim 55\%$–60% clustering rate for the lower-mass YSOs (1.5–4 $M_\odot$), and a clustering rate as low as $\sim 27\%$ for intermediate-mass YSOs in the range 4–10 $M_\odot$. These percentages are likely lower limits of the true clustering rates due to the optical limitation of the Gaia mission. However, we show that the conclusions of this work are not affected by this limitation.

3. We observe that for YSOs $>10 \, M_\odot$, there is a significant difference in clustering properties between the YSOs homogeneously identified in Vioque et al. (2022) and those historically considered. Vioque et al. (2022) MYSOs have a high clustering rate, whereas the historically known sources are mostly isolated. Vioque et al. (2022) sources show signs of a younger population. Hence, we believe this difference is due to a mixture of an evolutionary effect and a historical bias. If we assume that the Vioque et al. (2022) sample is more representative of the population of massive YSOs in the Galaxy, then we can conclude that most-massive YSOs appear clustered, and that this clustered nature starts at around $\sim 10 \, M_\odot$. We report 18 isolated YSOs with $M > 10 \, M_\odot$ in Gaia astrometric space (four of them from Vioque et al. 2022). We warn the reader that above $13 \, M_\odot$, low-number statistics and unknown population biases might be affecting the reported clustering rates.

4. We find that intermediate- and high-mass YSOs become less clustered with decreasing disk emission and accretion rate. This points toward an evolution with time. In the 1.5–5 $M_\odot$ range, this evolution with time toward isolation does not vary much with stellar mass.

5. We find no major correlation between the stellar properties of the YSOs considered and the properties of their clusters. We find no conclusive evidence from the cluster properties that can support either competitive-accretion or monolithic-collapse theories at high stellar masses. We can only report a weak trend of more-massive stars being closer to cluster centers above $6 \, M_\odot$, and that YSOs above $4 \, M_\odot$ are often in clusters larger than 2.5 pc in radius. These clusters tend to have medium to low stellar densities (<3 stars pc^{-2}, in Gaia space).

6. We present 55 new clusters not previously reported in the literature. We note that most of these new clusters are larger and less populated than what is normal in open clusters. Hence, they often do not fit the classical definition of "open cluster," as they are more similar to YSO sparse associations. The other clusters detected in this work have been previously identified with similar sizes and number of members. Eleven clusters contain more than one intermediate- or high-mass YSO.

The results of this work are based on almost the complete population of intermediate- and high-mass YSOs, which currently have spectroscopic characterization (337 sources). As is discussed in the text, this sample might be affected by some selection biases, and it is limited in the high-mass range ($>13 \, M_\odot$). Many of the intermediate- and high-mass YSO candidates of Vioque et al. (2020) catalog will be observed by

the WEAVE spectrograph (Jin et al. 2023). This will provide the community with an order-of-magnitude more sources to conduct studies like the one presented in this work (note that Vioque et al. 2022 estimated a catalog accuracy of >90%). This will provide key statistical insights to some of the discussions presented here.

Appendix A
Biases and Incompleteness

In this Appendix, we discuss in detail the possible sources of biases and incompleteness present in this work.

A.1. Assessment of Gaia 5D Space Incompleteness

We have based our analysis on the sources with a five-parameter astrometric solution in Gaia DR3. Here, we evaluate how many sources we are missing by not selecting all of the sources present in Gaia (i.e., with 2D solutions, R.A. and decl.). We also consider the possible impact of the Gaia astrometric uncertainties.

To have a meaningful estimation of the sources missed around every intermediate- and high-mass YSO, we limit the radius of comparison to $3'$. This angular radius roughly corresponds to half $R_{80\%}$ radius for most clusters considering typical distances of 1000–2000 pc (Figure 2). The results of this study are shown in the left panel of Figure 9. On average, we have missed 12% of the Gaia sources by demanding a five-parameter astrometric solution. However, there is no dependence between YSO clustered nature or stellar mass and the amount of Gaia sources missed by demanding a five-parameter astrometric solution. We conclude that demanding a five-parameter solution does not affect our classification of clustered or isolated YSO nature, and neither does it affect the trends with stellar mass detected in Section 4.1.

The Gaia astrometry of the considered sample is of high quality, i.e., there are no intermediate- or high-mass YSOs in our sample with large astrometric uncertainties (Section 2). This is true for relative errors ($>3\sigma$), and for absolute uncertainty values (e.g., proper-motion errors are under 0.5 mas yr^{-1} in all cases). In addition, we report no correlation in our sample of YSOs between stellar mass and astrometric uncertainties. In the right panel of Figure 9, we evaluate how many neighboring sources to each intermediate- and high-mass YSO have astrometric values larger than three times their uncertainty (in R.A., decl., parallax, proper-motion R.A., and proper-motion decl.). While we find that, in general, only $\sim$30% of the sources used to find clusters with HDSBCAN have more than 3σ astrometric quality, and again we report no significant correlation between the astrometric quality of the sources used for HDBSCAN and the intermediate- or high-mass YSOs stellar masses or derived clustered nature. Hence, we claim that if astrometric uncertainties are blurring structures together, this effect is very small in our derivations, and in any case affecting equally all of the stellar mass ranges considered.

Gaia has another important limitation for clustering studies. This is that it is limited to optically bright sources ($G < 21$ mag), and it does not contain the more embedded IR-bright YSOs. Hence, some of the seemingly isolated intermediate- and high-mass YSOs of our sample might be surrounded by many low-mass stars that Gaia cannot detect. Indeed, cluster membership analysis with the near-IR VVV survey (VISTA Variables in the Via Lactea; Minniti et al. 2010; Saito et al. 2012) have found $\sim$45% more member candidates in open clusters than those found only using Gaia (Peña Ramírez et al. 2021, 2022).

In order to assess the impact of this limitation in our analysis, we use the VVV Data release 2. There are 29 sources of our sample in the footprint of VVV, roughly $\sim$11% of our sample. As before, we made a $3'$ search around each of our YSOs, and studied the fraction of sources detected by VVV, which also have a Gaia five-parameter solution. The result is shown in the left panel of Figure 10. In Figure 10, we present an equivalent analysis for the UKIDSS-DR6 Galactic Plane Survey (Lucas et al. 2008; 58 sources match with our sample) and the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006; all of the sources of our sample are in 2MASS). These three surveys span from the J band (1.2 μm), to the K_s band (2.2 μm), with different optical depths. In the 2MASS comparison, we obtain that, on average, 95% of the sources around our YSOs are also in Gaia five-parameter space. However, in the comparison with VVV and UKIDSS, we see that, on average, only 20%–30% of the sources around each YSO have a Gaia 5D counterpart. In any case, we see in Figure 10 that for the VVV, UKIDSS, and 2MASS comparisons there is no significant correlation between Gaia 5D incompleteness and the clustered and isolated populations, and neither is there a correlation with the stellar mass of the intermediate- or high-mass YSOs. In addition to this, the depth and wavelength coverage of these surveys imply that most of the sources Gaia is not detecting are likely unrelated to the intermediate- or high-mass YSOs of interest. It is hard to assess what fraction of those sources that we are missing are real companions of the considered YSOs, as these IR surveys go very deep (e.g., VVV faint limit is $K_s = 17$ mag), and we lack parallax information for their sources. We note that these IR surveys are limited to the Galactic center, where the extinction is largest, and hence they suppose an upper limit to the Gaia incompleteness.

Another way to assess the limitations of the Gaia 5D space is to compare it with infrared-based YSO catalogs. For this, we use the SPICY (Kuhn et al. 2021b) and Winston et al. (2020)
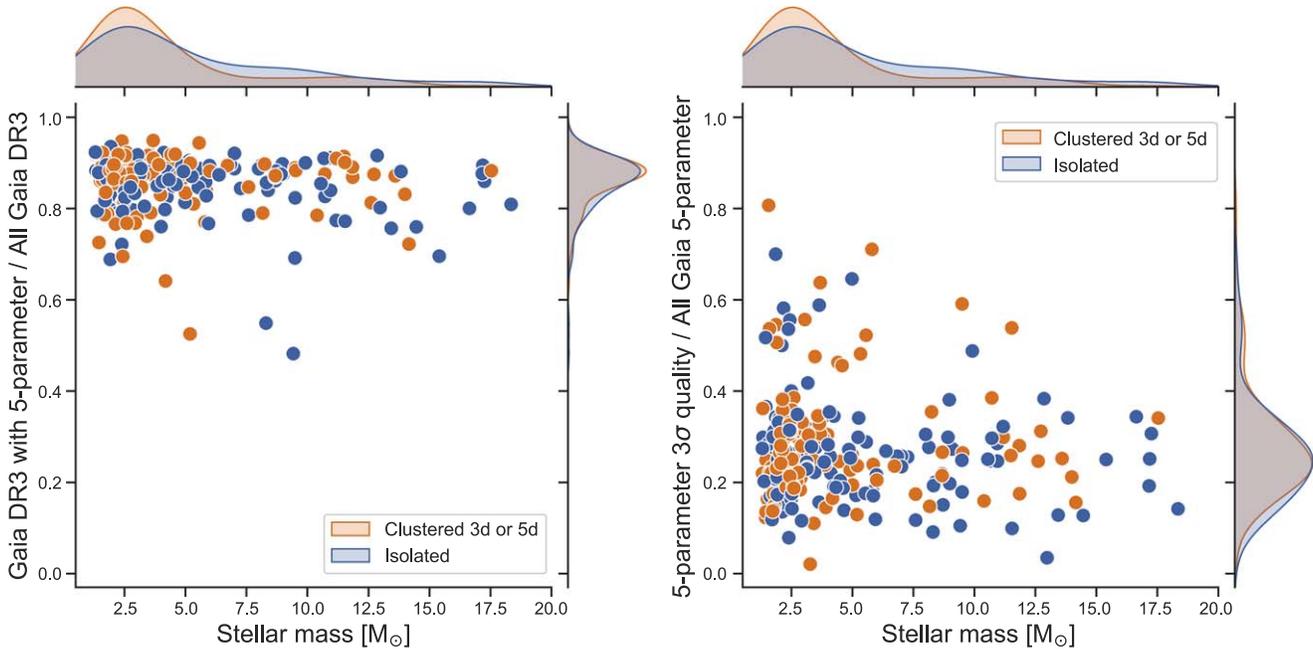
Figure 9. Within 3′ of each intermediate- or high-mass YSO. Left: fraction of sources considered in our analysis (with Gaia DR3 five-parameter solution) with respect to all Gaia DR3 sources. Right: fraction of Gaia DR3 sources with all five astrometric quantities having values over three times their uncertainty with respect to the sources considered in our analysis. The stellar mass of the intermediate- or high-mass YSOs is shown in the horizontal axes. Density curves of clustered and isolated sources are shown at the top and right.
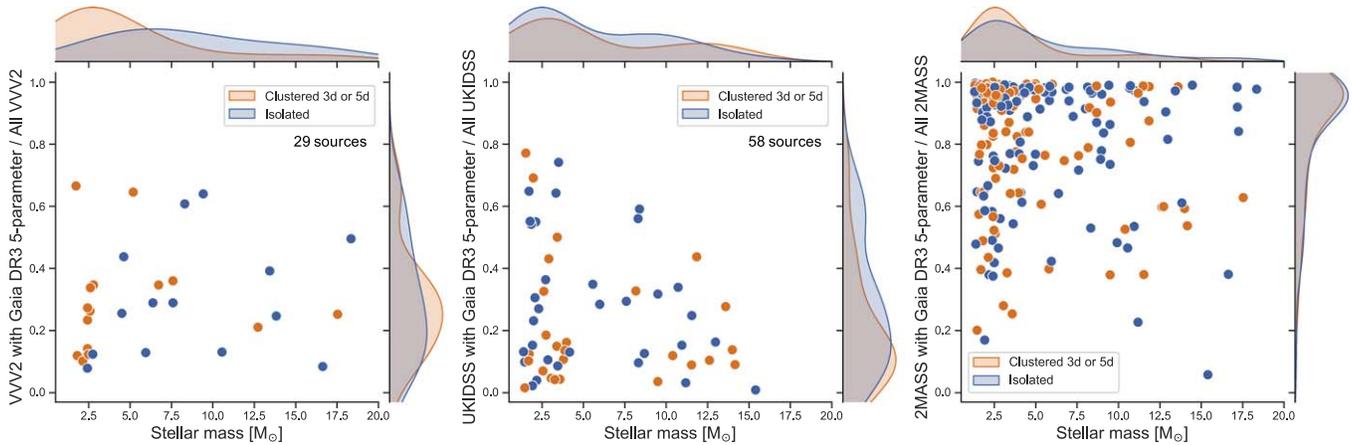
Figure 10. Within 3′ of each YSO, this plot shows the fraction of sources detected by IR surveys that were considered in our analysis (those with Gaia DR3 five-parameter solution). From left to right, VVV, UKIDSS, and 2MASS are shown. The horizontal axes show the stellar mass of each intermediate- or high-mass YSO. Density curves of clustered and isolated sources are shown at the top and right.

catalogs. The SPICY catalog contains 117,446 YSOs in the inner Galactic midplane. This catalog was produced using mainly Spitzer data (but also 2MASS, VVV, and UKIDSS data), as it is the most comprehensive catalog of YSOs from mid-infrared observations. A similar catalog was produced by Winston et al. (2020; with 43,094 sources), in this case targeting the outer Galaxy, making these two catalogs complementary. The accuracy of these catalogs has been evidenced by spectroscopic surveys (e.g., Kuhn et al. 2023). Of our sample of 263 sources, 51 sources overlap with SPICY and 29 with the Winston et al. (2020) catalog. We study the fraction of sources appearing in these IR-based catalogs, which were considered in our analysis. Given that these catalogs have a limited footprint, and for consistency with the previous analyses, we also consider a 3′ aperture to look for missing

sources in the Gaia five-parameter space. The results of this study are shown in Figure 11. In this comparison, we also get clustered stars at high and low levels of Gaia 5D completeness, and the means of the distributions of clustered and isolated stars are within 1σ of each other. In addition, no bias with respect to the stellar mass of the intermediate- or high-mass YSOs can be appreciated. We note that these IR surveys might be showing unrelated background or foreground YSOs. Indeed, only 67% of the SPICY sources with Gaia parallaxes are on average within 1000 pc of the corresponding intermediate- or high-mass YSOs.

Our results in this Appendix give us no reason to believe we would have different incompleteness levels around sources of different masses. We conclude that the evolution of the clustering ratio as a function of stellar mass presented in
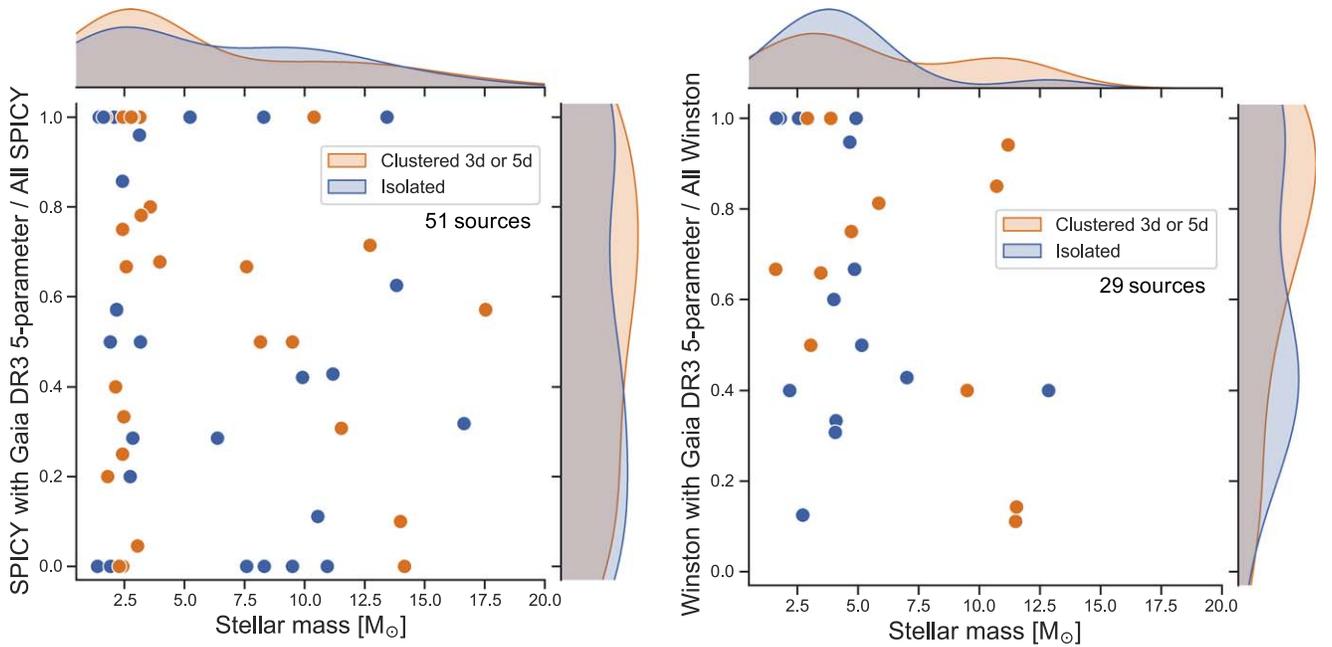
Figure 11. Within 3′ of each YSO, this plot shows the fraction of sources appearing in the IR-based catalogs of Kuhn et al. (2021a; SPICY) and Winston et al. (2020), which were considered in our analysis (with Gaia DR3 five-parameter solution). The horizontal axes show the stellar mass of each intermediate- or high-mass YSO. Density curves of clustered and isolated sources are shown at the top and right.

Section 4.1 is robust (Figures 4 and 5), and it is not affected by Gaia 5D incompleteness or astrometric uncertainty. We, however, cannot claim the absolute clustering rate values presented in Sections 3 and 4, and Figures 4, 5, and 8 to be accurate, as they are lower limits, given that we are missing an unknown fraction of sources around each YSO because they are not present in Gaia five-parameter astrometric space.

A.2. Assessment of Possible Distance and Sky Position Biases

According to Hunt & Reffert (2021) and Section 3.2, our HDBSCAN methodology should be efficient and consistent with the distance range considered in this work (350–4000 pc), and it should be independent of sky location. In this Appendix, we evaluate possible distance and sky location effects in our clustering derivations.

We start by noting that our YSO sample is well distributed in stellar mass across the whole distance range considered (Figure 12). In the distance range mainly analyzed in this work (350–4000 pc), we have both clustered and isolated stars across all distances. This is also true for stars that appear clustered in both 3D and 5D. In addition, there is no significant difference between clustered and isolated stars as a function of distance (Figure 12). We report no distance effect on the measured cluster properties: number of stars per cluster (N_c), cluster radii ($R_{80\%}$), and cluster densities. In addition, we detect clustered stars at all proper-motion ranges. Hence, the methodology is not biased against high proper-motion stars.

In the distance range mainly analyzed in this work (350–4000 pc), there are two noticeable effects. First, we lack massive YSOs (above $8\,M_\odot$) below 500 pc. This is a real, purely IMF effect; massive stars form fast and are rare, so there are very few of them nearby with Gaia data. The other effect is artificial; we lack low-mass YSOs (1.5–3 $M_\odot$) above 2500 pc. This is due to the fact that these sources are too faint to have fallen in the spectroscopic surveys used to construct the

sample. In order to assess how much the two previous distance effects influence our results, we repeat the analysis of Section 4.1 for different distance ranges (350–1000, 350–2500, 1000–2000, 2000–3000, and 3000–4000 pc; Figure 13). The general shape of the evolution of clustering rate with stellar mass described in Section 4.1 is held independently of the distance range considered. It goes down from lower-mass sources to the intermediate-mass regime, and then goes up again for the more-massive YSOs. The mass bin of minimum clustering rate varies, but it is always contained within the 7–10 $M_\odot$ range. At the more-massive end ($M > 13$–$15\,M_\odot$), low-number statistics and population biases affect the results, and the clustering rates presented are dominated by very few sources (see Section 5.2 for further discussion). Outside the distance range mainly analyzed in this work (350–4000 pc), Figure 12 shows that there is a strong bias favoring cluster detection for distances under 200 pc.

There is one more distance effect we have detected for the lower-mass YSOs. Beyond 1500 pc, most low-mass stars belong to the Vioque et al. (2022) sample, which, by construction, has a Gaia G brightness faint limit of ∼14.5 mag. Therefore, at larger distances, the lower-mass objects we are including in the analysis are forced to be optically more luminous than those at short distances. This implies that they are on average less extincted. Our methodology finds it easier to detect clusters around YSOs in low extinction environments, as more sources will appear in Gaia 5D space (Appendix A.1). We hence expect to detect a higher fraction of clustered low-mass YSOs at distances above 1500 pc. This is indeed what we see in Figure 13. Lower-mass YSOs have higher clustering rates at distances of 2000–4000 pc (of ∼80%) than at lower distances. We note that we report a clustering rate of ∼55%–60% for the lower-mass stars until 2000 pc, which matches that reported for the complete sample. Hence, this bias affects a small fraction of the lower-mass sources and does not have an impact on our conclusions.
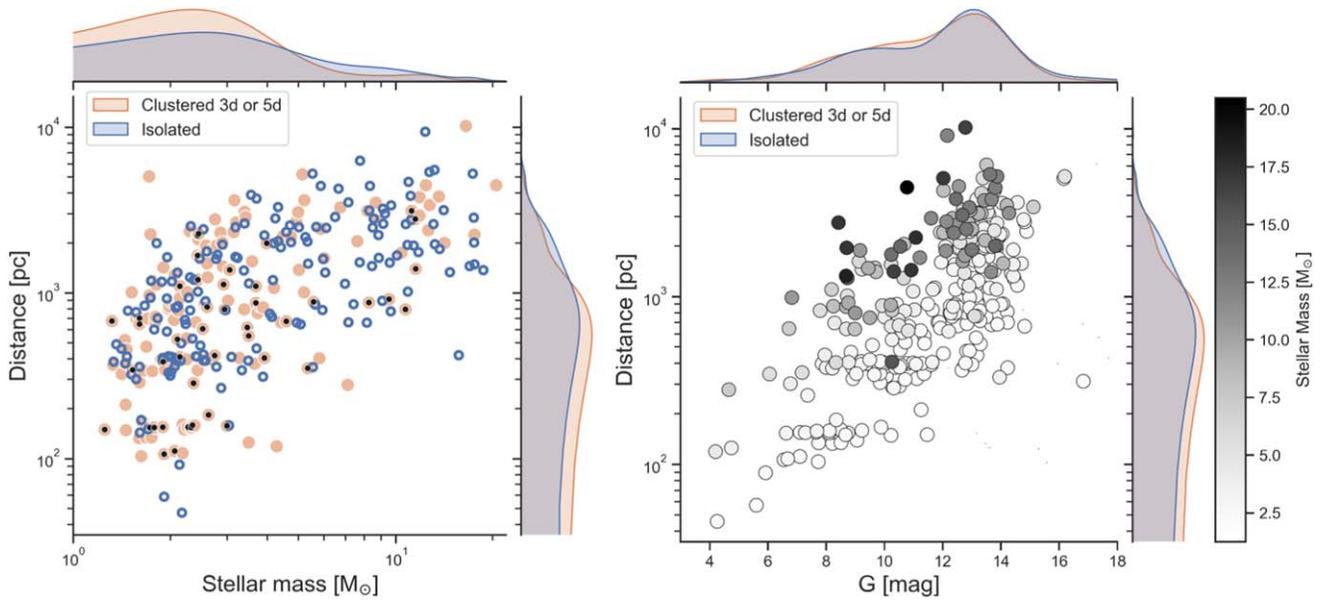
Figure 12. Left: distance vs. stellar mass for all YSOs compiled in this work (337 sources). Density curves of clustered and isolated sources are shown at the top and right. Black dots indicate those sources that appear clustered both in 3D and 5D. Right: distance vs. Gaia G-band magnitude for all YSOs compiled in this work. Color scheme indicates stellar mass, with darker sources being more massive (linear scale). For the analysis, we have only considered sources between 350 and 4000 pc.
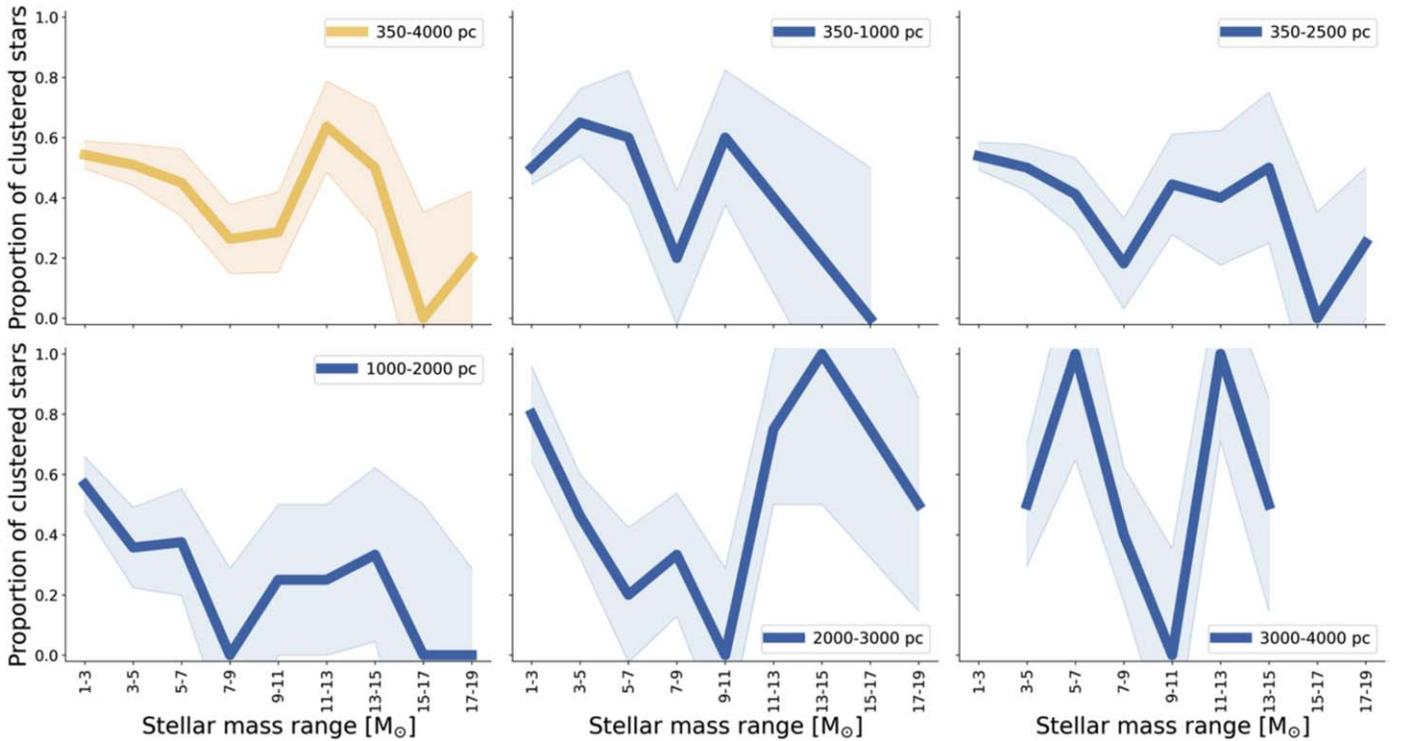

Figure 13. Same as Figure 4 "3D or 5D" with 2 $M_\odot$ bins but considering only certain distance ranges. The top-left panel is identical to the Figure 4 bottom-left "3D or 5D" panel. We show it here for comparison purposes. Complete error is $1/\sqrt{n}$, with n being the number of sources per bin.

No correlation between clustering properties or YSO stellar mass and sky coordinates or Galactic latitude or longitude was found. Hence, we do not expect the Gaia selection function to affect our results (Cantat-Gaudin et al. 2023). The only sky-related difference in the methodology appeared in Section 3.1, where we decided to use HEALPix 6 instead of HEALPix 5 for sources that were in very crowded fields, to retrieve similarly sized clusters as for the rest of sources. Of the sample of 263 sources considered in Section 3.2, we have 191 stars with

HEALPix 5 and 72 with HEALPix 6. Eighty-five percent of these 72 HEALPix 6 stars are beyond 1 kpc and hence, in general, we can safely assume the HEALPix choice does not make a difference as to whether a given source was identified as part of a cluster or not. In any case, we have made a comparison between both HEALPix subsets. These sets have 49% (HEALPix 5) and 47% (HEALPix 6) of clustered stars. Hence, HDBSCAN is equally efficient in both cases. As the more-massive sources are typically farther away and hence

more likely situated against the crowded Galactic plane, we have also evaluated whether this efficiency varies as a function of stellar mass. We observe that this is not the case. The cluster properties obtained with HEALPix 5 and HEALPix 6 are identical.

We thus conclude that the distance and sky distribution of the considered sample of intermediate- and high-mass YSOs do not affect the analysis and conclusions of this work.

Appendix B
Assessment of Other HDBSCAN Parameters

For a subsample of 235 sources, we have applied HDBSCAN to the 5D astrometric space using a wider grid of hyperparameters ("min_cluster_size" and "min_samples"). We used a grid defined as [5, 5], [5, 10], [10, 5], [10, 10] (or [5–10, 5–10] step 5), and [60–100, 60–100] (step 5). We ran HDBSCAN at each field following the grid order, stopping if the cluster probability for the massive YSO was different to zero. This allowed us to assess the limitations of Section 3.1 methodology, considering smaller and bigger cluster sizes as well as variations of "min_samples."

For 197 of the 235 sources (84%), there is agreement between the grids of Section 3.1 and that of this Appendix, whether the stars are clustered or unclustered. There are only three stars (1%) that were identified with the [30, 30] setting that were not identified by this Appendix's grid. However, there are 35 stars (15%) that were identified as part of a cluster with this Appendix's grid that were not detected in Section 3.1. Of those 35, 16 appear clustered with a minimum cluster size of

5, and 10 required of a minimum cluster size of 100 to be detected as part of a cluster. These two extremes are likely not indicative of physical clusters. A minimum cluster size of 5 as traced by HDBSCAN cannot possibly be statistically robust (the minimum cluster size is often considered to be 10; Krumholz et al. 2019; Krause et al. 2020). Similarly, if the clusters detected with a minimum cluster size of 100 were real, they should have been detected at smaller cluster sizes (see Hunt & Reffert 2021, which proves "min_cluster_size" from 10–80). Therefore, there are only nine sources that can be truly considered as missed clusters in Section 3.1 (seven with [10, 5] and two with [60, 60]).

Hence, the choice of hyperparameters in Section 3.1 is not affecting the determination of whether a star is clustered or not to within 95% confidence. Therefore, we conclude that the hyperparameters chosen in this work are appropriate for identifying clusters and associations with HDBSCAN around intermediate- and high-mass YSOs.

Appendix C
Table of Clustered and Isolated Stars

In Table 1, we detail the clustering properties of the 263 intermediate- and high-mass YSOs analyzed between 350 and 4000 pc. When it corresponds, the cluster properties are also tabulated. Here we present a portion of the table for guidance regarding its form and content. The full table will be made available within the CDS in the VizieR archive service and published with this article.

Table 1
Clustering Properties of the 263 Intermediate- and High-mass YSOs Analyzed between 350 and 4000 pc

SIMBAD Name	Gaia DR3 Source ID	R.A. hh:mm:ss	Decl. dd:mm:ss.s	Stellar Mass ($M_\odot$)	Distance (pc)	Gaia G (mag)	HDBSCAN Probability	HDBSCAN Configuration	$R_{80\%}$ (pc)	Members	Density (stars pc^{-2})	Distance to Center (pc)	Notes
HD 36917	3017264007761349504	05:34:47	−05:34:14.6	$3.64^{0.58}_{0.29}$	$447.95^{11.32}_{10.21}$	7.96	0	Isolated	...	...	...	...	...
HD 245185	3338005774514542720	05:35:10	+10:01:51.4	$2.14^{0.26}_{0.05}$	$410.26^{4.94}_{5.71}$	9.86	0.75, 0.84	3D, 5D	9.66	702	1.92	3.26	...
HD 36982	3017360348171372672	05:35:10	−05:27:53.2	$5.80^{0.58}_{0.58}$	$404.00^{3.58}_{4.06}$	8.34	0.68	3D	0.97	706	191.69	0.53	cont
NV Ori	3017346952181227776	05:35:31	−05:33:08.9	$1.90^{0.12}_{0.08}$	$383.61^{2.87}_{3.06}$	9.7	0.49, 0.84	3D, 5D	3.58	675	13.38	1.44	...
T Ori	3017347673735214592	05:35:50	−05:28:34.9	$2.38^{0.16}_{0.10}$	$398.95^{5.34}_{4.46}$	11.29	0	Isolated	...	...	...	...	...
BD-06 1253	3016923017420354688	05:36:25	−06:42:57.7	$2.50^{0.31}_{0.11}$	$379.64^{14.12}_{16.00}$	10.53	0	Isolated	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...

Note. When it corresponds, the cluster properties are also tabulated.

Coordinates and G magnitudes were obtained from Gaia DR3. Distances to sources are the geometric distances of Bailer-Jones et al. (2021). The rest of the parameters were derived in this work. In the *Notes* column, "cont" means possible cluster misidentification, and "new" means a new cluster identification previously unreported in the literature (see Hunt & Reffert 2023). Here we present a portion of the table for guidance regarding its form and content. The full table will be made available within the CDS in the VizieR archive service and published with this article. The four HDBSCAN resulting cluster probabilities from Section 3.1 appear tabulated in different columns, and Gaia DR3 coordinates in decimal degrees are also provided.

(This table is available in its entirety in machine-readable form.)

ORCID iDs

Miguel Vioque ⓘ https://orcid.org/0000-0002-4147-3846
Manuel Cavieres ⓘ https://orcid.org/0009-0002-2978-8383
Michelangelo Pantaleoni González ⓘ https://orcid.org/0000-0001-9933-1229
Álvaro Ribas ⓘ https://orcid.org/0000-0003-3133-3580
René D. Oudmaijer ⓘ https://orcid.org/0000-0001-7703-3992
Ignacio Mendigutía ⓘ https://orcid.org/0000-0002-0233-5328
Lena Kilian ⓘ https://orcid.org/0000-0002-0885-1198
Héctor Cánovas ⓘ https://orcid.org/0000-0001-7668-8022
Michael A. Kuhn ⓘ https://orcid.org/0000-0002-0631-7514

References

Ali, A. A., Dobbs, C. L., Bending, T. J. R., Buckner, A. S. M., & Pettitt, A. R. 2023, MNRAS, 524, 555
Arun, R., Mathew, B., Maheswar, G., et al. 2021, MNRAS, 507, 267
Arun, R., Mathew, B., Manoj, P., et al. 2019, AJ, 157, 159
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
Baines, D., Oudmaijer, R. D., Porter, J. M., & Pozzo, M. 2006, MNRAS, 367, 737
Barnes, A. T., Longmore, S. N., Avison, A., et al. 2019, MNRAS, 486, 283
Bica, E., Pavani, D. B., Bonatto, C. J., & Lima, E. F. 2019, AJ, 157, 12
Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 2001, MNRAS, 323, 785
Bonnell, I. A., & Clarke, C. J. 1999, MNRAS, 309, 461
Bonnell, I. A., Vine, S. G., & Bate, M. R. 2004, MNRAS, 349, 735
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Bressert, E., Bastian, N., Evans, C. J., et al. 2012, A&A, 542, A49
Brittain, S. D., Kamp, I., Meeus, G., Oudmaijer, R. D., & Waters, L. B. F. M. 2023, SSRv, 219, 7
Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in Advances in Knowledge Discovery and Data Mining, ed. J. Pei et al. (Berlin: Springer), 160
Cánovas, H., Cantero, C., Cieza, L., et al. 2019, A&A, 626, A80
Cantat-Gaudin, T., & Anders, F. 2020, A&A, 633, A99
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1
Cantat-Gaudin, T., Fouesneau, M., Rix, H.-W., et al. 2023, A&A, 669, A55
Carpenter, J. M. 2000, AJ, 120, 3139
Castro-Ginard, A., Jordi, C., Luri, X., et al. 2022, A&A, 661, A118
Chevance, M., Kruijssen, J. M. D., & Longmore, S. N. 2021, ApJL, 910, L19
Chevance, M., Krumholz, M. R., McLeod, A. F., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S. Inutsuka et al. (San Francisco, CA: ASP), 1
Concha-Ramírez, F., Wilhelm, M. J. C., & Zwart, S. P. 2022, MNRAS, 520, 6159
Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, A&A, 674, A26
Cuello, N., Ménard, F., & Price, D. J. 2023, EPJP, 138, 11
de Wit, W. J., Testi, L., Palla, F., & Zinnecker, H. 2005, A&A, 437, 247
Della Croce, A., Dalessandro, E., Livernois, A., et al. 2023, A&A, 674, A93
Dias, W. S., Monteiro, H., Moitinho, A., et al. 2021, MNRAS, 504, 356
Fairlamb, J. R., Oudmaijer, R. D., Mendigutía, I., Ilee, J. D., & van den Ancker, M. E. 2015, MNRAS, 453, 976
Farias, J. P., & Tan, J. C. 2023, MNRAS, 523, 2083
Faustini, F., Molinari, S., Testi, L., & Brand, J. 2009, A&A, 503, 801
Fedriani, R., Tan, J. C., Telkamp, Z., et al. 2023, ApJ, 942, 7
Fischer, W. J., Hillenbrand, L. A., Herczeg, G. J., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII (San Francisco, CA: ASP), 355
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28
Fujii, M. S., & Portegies Zwart, S. 2016, ApJ, 817, 4
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Gouliermis, D. A., Schmeja, S., Dolphin, A. E., et al. 2012, ApJ, 748, 64
Grant, S. L., Espaillat, C. C., Brittain, S., Scott-Joseph, C., & Calvet, N. 2022, ApJ, 926, 229
Grant, S. L., Stapper, L. M., Hogerheijde, M. R., et al. 2023, AJ, 166, 147
Grasha, K., Calzetti, D., Adamo, A., et al. 2019, MNRAS, 483, 4707
Grudić, M. Y., Diederik Kruijssen, J. M., Faucher-Giguère, C.-A., et al. 2021, MNRAS, 506, 3239
Grudić, M. Y., Guszejnov, D., Offner, S. S. R., et al. 2022, MNRAS, 512, 216
Guszejnov, D., Markey, C., Offner, S. S. R., et al. 2022, MNRAS, 515, 167
Gutermuth, R. A., Megeath, S. T., Myers, P. C., et al. 2009, ApJS, 184, 18

Guzmán-Díaz, J., Mendigutía, I., Montesinos, B., et al. 2021, A&A, 650, A182
Guzmán-Díaz, J., Montesinos, B., Mendigutía, I., et al. 2023, A&A, 671, A140
Habart, E., Testi, L., Natta, A., & Vanzi, L. 2003, A&A, 400, 575
Hao, C. J., Xu, Y., Hou, L. G., Lin, Z. H., & Li, Y. J. 2023, RAA, 23, 075023
Hillenbrand, L. A. 1995, PhD thesis, University of California System; University of Texas, Austin, Department of Astronomy
Hunt, E. L., & Reffert, S. 2021, A&A, 646, A104
Hunt, E. L., & Reffert, S. 2023, A&A, 673, A114
Iglesias, D. P., Panić, O., van den Ancker, M., et al. 2023, MNRAS, 519, 3958
Jin, S., Trager, S. C., Dalton, G. B., et al. 2023, MNRAS, in press
Kerr, R., Kraus, A., & Rizzuto, A. 2023, ApJ, 954, 134
Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R. D. 2013, A&A, 558, A53
King, I. 1962, AJ, 67, 471
Kirk, H., & Myers, P. C. 2011, ApJ, 727, 64
Kirk, H., & Myers, P. C. 2012, ApJ, 745, 131
Koumpia, E., de Wit, W. J., Oudmaijer, R. D., et al. 2021, A&A, 654, A109
Kounkel, M., & Covey, K. 2019, AJ, 158, 122
Krause, M. G. H., Offner, S. S. R., Charbonnel, C., et al. 2020, SSRv, 216, 64
Kruijssen, J. M. D. 2012, MNRAS, 426, 3008
Kruijssen, J. M. D., Longmore, S. N., & Chevance, M. 2020, ApJL, 905, L18
Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., Portegies Zwart, S. F., & Icke, V. 2011, MNRAS, 414, 1339
Krumholz, M. R., & McKee, C. F. 2020, MNRAS, 494, 624
Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, ARA&A, 57, 227
Ksoll, V. F., Gouliermis, D., Sabbi, E., et al. 2021, AJ, 161, 257
Kuhn, M. A., Benjamin, R. A., Zucker, C., et al. 2021a, A&A, 651, L10
Kuhn, M. A., de Souza, R. S., Krone-Martins, A., et al. 2021b, ApJS, 254, 33
Kuhn, M. A., Feigelson, E. D., Getman, K. V., et al. 2014, ApJ, 787, 107
Kuhn, M. A., Getman, K. V., & Feigelson, E. D. 2015, ApJ, 802, 60
Kuhn, M. A., Hillenbrand, L. A., Sills, A., Feigelson, E. D., & Getman, K. V. 2019, ApJ, 870, 32
Kuhn, M. A., Saber, R., Povich, M. S., et al. 2023, AJ, 165, 3
Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57
Law, C.-Y., Tan, J. C., Gorai, P., et al. 2022, ApJ, 939, 120
Lindegren, L., Bastian, U., Biermann, M., et al. 2021a, A&A, 649, A4
Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021b, A&A, 649, A2
Lodato, G., & Manara, C. F. 2023, EPJP, 138, 675
Longmore, S. N., Chevance, M., & Kruijssen, J. M. D. 2021, ApJL, 911, L16
Longmore, S. N., Kruijssen, J. M. D., Bastian, N., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 291
Lucas, P. W., Hoare, M. G., Longmore, A., et al. 2008, MNRAS, 391, 136
Luhman, K. L. 2022, AJ, 164, 151
Luhman, K. L. 2023a, AJ, 165, 269
Luhman, K. L. 2023b, AJ, 165, 37
Lumsden, S. L., Hoare, M. G., Urquhart, J. S., et al. 2013, ApJS, 208, 11
Marcos-Arenal, P., Mendigutía, I., Koumpia, E., et al. 2021, A&A, 652, A68
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
Maschberger, T., & Clarke, C. J. 2008, MNRAS, 391, 711
McInnes, L., Healy, J., & Astels, S. 2017, JOSS, 2, 205
McKee, C. F., & Tan, J. C. 2003, ApJ, 585, 850
Meingast, S., Alves, J., & Rottensteiner, A. 2021, A&A, 645, A84
Mendigutía, I., Solano, E., Vioque, M., et al. 2022, A&A, 664, A66
Messa, M., Adamo, A., Calzetti, D., et al. 2018, MNRAS, 477, 1683
Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, NewA, 15, 433
Morii, K., Sanhueza, P., Nakamura, F., et al. 2023, ApJ, 950, 148
Nakajima, Y., Kato, D., Nagata, T., et al. 2005, AJ, 129, 776
Nuñez, E. H., Povich, M. S., Binder, B. A., Townsley, L. K., & Broos, P. S. 2021, AJ, 162, 153
Oey, M. S., Lamb, J. B., Kushner, C. T., Pellegrini, E. W., & Graus, A. S. 2013, ApJ, 768, 66
Oh, S., & Kroupa, P. 2016, A&A, 590, A107
Panić, O., Haworth, T. J., Petr-Gotzens, M. G., et al. 2021, MNRAS, 501, 4317
Pantaleoni González, M., Maíz Apellániz, J., Barbá, R. H., & Reed, B. C. 2021, MNRAS, 504, 2968
Parker, R. J., & Goodwin, S. P. 2007, MNRAS, 380, 1271
Peña Ramírez, K., González-Fernández, C., Chené, A. N., & Ramírez Alegría, S. 2021, MNRAS, 503, 1864
Peña Ramírez, K., Smith, L. C., Ramírez Alegría, S., et al. 2022, MNRAS, 513, 5799
Pinilla, P., Garufi, A., & Gárate, M. 2022, A&A, 662, L8
Prisinzano, L., Damiani, F., Sciortino, S., et al. 2022, A&A, 664, A175
Qiu, K., Zhang, Q., Megeath, S. T., et al. 2008, ApJ, 685, 1005
Rich, E. A., Monnier, J. D., Aarnio, A., et al. 2022, AJ, 164, 109
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3

Robitaille, T. P., Meade, M. R., Babler, B. L., et al. 2008, AJ, 136, 2413
Rodríguez, M. J., Baume, G., & Feinstein, C. 2019, A&A, 626, A35
Saha, P., Gopinathan, M., Kamath, U., et al. 2020, MNRAS, 494, 5851
Saito, R. K., Hempel, M., Minniti, D., et al. 2012, A&A, 537, A107
Salpeter, E. E. 1955, ApJ, 121, 161
Schmeja, S., Gouliermis, D. A., & Klessen, R. S. 2009, ApJ, 694, 367
Shridharan, B., Mathew, B., Nidhi, S., et al. 2021, RAA, 21, 288
Shukirgaliyev, B., Parmentier, G., Berczik, P., & Just, A. 2019, MNRAS, 486, 1045
Sitko, M. L., Carpenter, W. J., Kimes, R. L., et al. 2008, ApJ, 678, 1070
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Stapper, L. M., Hogerheijde, M. R., van Dishoeck, E. F., & Mentel, R. 2022, A&A, 658, A112
Stapper, L. M., Hogerheijde, M. R., van Dishoeck, E. F., & Paneque-Carreño, T. 2023, A&A, 669, A158
Stephens, I. W., Gouliermis, D., Looney, L. W., et al. 2017, ApJ, 834, 94
Sun, J., Gutermuth, R. A., Wang, H., Zhang, S., & Long, M. 2022, MNRAS, 516, 5258
Tan, J. C., & McKee, C. F. 2004, ApJ, 603, 383
Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
Testi, L., Palla, F., & Natta, A. 1999, A&A, 342, 515
Testi, L., Palla, F., Prusti, T., Natta, A., & Maltagliati, S. 1997, A&A, 320, 159
Thé, P. S., de Winter, D., & Perez, M. R. 1994, A&AS, 104, 315
Thomas, S. J., Rodgers, B., van der Bliek, N. S., et al. 2023, AJ, 165, 135

Valegård, P. G., Waters, L. B. F. M., & Dominik, C. 2021, A&A, 652, A133
Vioque, M., Oudmaijer, R. D., Baines, D., Mendigutía, I., & Pérez-Martínez, R. 2018, A&A, 620, A128
Vioque, M., Oudmaijer, R. D., Schreiner, M., et al. 2020, A&A, 638, A21
Vioque, M., Oudmaijer, R. D., Wichittanakom, C., et al. 2022, ApJ, 930, 39
Wang, P., Li, Z.-Y., Abel, T., & Nakamura, F. 2010, ApJ, 709, 27
Ward, J. L., Kruijssen, J. M. D., & Rix, H.-W. 2020, MNRAS, 495, 663
Weidner, C., Kroupa, P., Nürnberger, D. E. A., & Sterzik, M. F. 2007, MNRAS, 376, 1879
Wichittanakom, C., Oudmaijer, R. D., Fairlamb, J. R., et al. 2020, MNRAS, 493, 234
Winston, E., Hora, J. L., & Tolls, V. 2020, AJ, 160, 68
Winter, A. J., Ansdell, M., Haworth, T. J., & Kruijssen, J. M. D. 2020a, MNRAS, 497, L40
Winter, A. J., & Haworth, T. J. 2022, EPJP, 137, 1132
Winter, A. J., Kruijssen, J. M. D., Longmore, S. N., & Chevance, M. 2020b, Natur, 586, 528
Wright, N. J., Goodwin, S., Jeffries, R. D., Kounkel, M., & Zari, E. 2022, arXiv:2203.10007
Yan, Z., Jerabkova, T., & Kroupa, P. 2023, A&A, 670, A151
Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, A&A, 620, A172
Zhang, Y.-J., Hou, W., Luo, A. L., et al. 2022, ApJS, 259, 38
Zucker, C., Alves, J., Goodman, A., Meingast, S., & Galli, P. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S. Inutsuka et al. (San Francisco, CA: ASP), 43