

Caracterización de Redes Estelares

Stellar Networks Characterization

Presentación: 12/08/2024

Doctorando:

Martín Gustavo CASATTI

Facultad Regional Córdoba, Universidad Tecnológica Nacional - Argentina
mcasatti@frc.utn.edu.ar

Director:

Marcelo Martín MARCISZACK

Codirector:

Carlos FEINSTEIN

Resumen

El presente trabajo expone los resultados obtenidos a partir del pre procesamiento, estructuración y posterior análisis de datos astronómicos a los fines de generar una red de datos asociados a observaciones de estrellas en galaxias cercanas, más específicamente en las Nubes de Magallanes a fin de caracterizar dicha estructura y comprobar sus similitudes y diferencias con respecto a una estructura conocida y ampliamente analizada denominada "Redes de Mundo Pequeño", estructura que reviste una especial importancia a la hora de seleccionar algoritmos de detección de comunidades, los cuales funcionan de manera particularmente eficiente cuando el grafo subyacente presenta una estructura de "Mundo Pequeño". Se presentan todos los pasos previos de adecuación y estandarización de datos, los filtros aplicados para eliminar información redundante o innecesaria, la determinación de atributos y datos relacionables, la construcción del grafo subyacente y el posterior análisis del mismo. Por último, se exponen las conclusiones obtenidas y los próximos pasos a seguir en la investigación.

Palabras clave: Cluster, Patrón, Estructura de red, Mundo Pequeño, Grafo, Analítica.

Abstract

This paper presents the results obtained from the pre-processing, structuring and subsequent analysis of astronomical data in order to generate a data network associated with observations of stars in nearby galaxies, more specifically in the Magellanic Clouds, in order to characterize said structure and check its similarities and differences with respect to a known and widely analyzed structure called "Small World Networks", a structure that is of special importance when selecting community detection algorithms, which work particularly efficiently when the underlying graph has a "Small World" structure. All the previous steps of data adaptation and standardization are presented, the filters applied to eliminate redundant or unnecessary information, the determination of attributes and relatable data, the construction of the underlying graph and its subsequent analysis. Finally, the conclusions obtained and the next steps to follow in the research are presented.

Keywords: Cluster, Pattern, Network structure, Small world, Graph, Analytics

Introducción

Las agrupaciones estelares, también denominados cúmulos o clusters, han sido objetos reconocidos desde hace tiempo como laboratorios importantes para la investigación astrofísica, siendo muy útiles en varios aspectos, entre los que se pueden destacar los siguientes:

- Contienen muestras estadísticamente significativas de estrellas de aproximadamente la misma edad, con composiciones químicas similares, un amplio rango de masas estelares y localizadas en un volumen relativamente pequeño del espacio, haciéndolas un conjunto ideal para el análisis de características comunes y determinación de los patrones que rigen su surgimiento (Klessen, 2000).
- En relación con el proceso de formación estelar, los cúmulos jóvenes permiten esclarecer la forma y las escalas de tiempo en las que estos mecanismos están activos, así como también permiten analizar su dependencia de los distintos ambientes interestelares de la Vía Láctea o de otras galaxias (Fall, 2012).

Los trabajos mencionados se han focalizado en mejorar el conocimiento de nuestra propia Galaxia (y de las Nubes de Magallanes (Vazquez, 2008), pero actualmente hay varios factores que incrementan de forma importante tanto la cantidad de objetos a investigar como la metodología para hacerlo.

En la actualidad existe una gran cantidad de información de las galaxias cercanas (a varios Mpc¹ debido, en gran parte, a que el Telescopio Espacial Hubble (HST) ha permitido obtener datos con alta resolución espacial utilizando varias cámaras de campo amplio (WFPC2; ACS) (Dalcanton, 2009).

Se cuenta con una enorme cantidad de datos proveniente de las varias observaciones continuas que se están realizando y que se proyectan realizar en modo "survey"² (p.e. VVV³ o LSST⁴ que necesitan ser estudiados con métodos automáticos.

En este ámbito, los algoritmos de reconocimiento automático de patrones están teniendo una importante revisión y desarrollo tal como se puede apreciar en el análisis comparativo de Schmeja (Schmeja, 2011).

Tal como se desprende de esa publicación, estos algoritmos se basan en analizar sólo las posiciones espaciales para encontrar a los sistemas estelares por sobre-densidades contra el fondo estelar o por su equivalente relacionado con la distribución de distancias entre estrellas.

En otros ámbitos científicos se han aplicado con éxito diversos algoritmos de clustering, como por ejemplo "K-mean", "Birch", "Spectral Clustering", "DbSCAN", etc (Rodríguez, 2019).

Por otra parte, el auge que tiene desde hace algunos años el análisis de redes sociales nos ha brindado otro amplio campo de estudios en el que se pueden apreciar algunos de los atributos que son comunes al problema de la detección de cúmulos estelares, como, por ejemplo:

- En el ámbito de las redes sociales también se cuenta con una gran cantidad de datos.
- Existe un conjunto de relaciones no evidentes entre los mismos y
- Un nutrido grupo de atributos analizables a fin de guiar la detección de patrones.

La estructura inherente de dichas redes es la de un grafo, sobre el que se puede realizar multitud de análisis sustentados por la Teoría de Grafos (West, 2001).

Diversos estudios, tanto de la topología de dichas redes (Barnes, 1983) como de las características que presentan sus participantes, nos brindan un fértil campo para el estudio de algoritmos de detección de patrones estructurales, muchos de ellos asistidos por técnicas de Machine Learning (Alharbi, 2021).

En la actualidad el análisis de algoritmos y su aplicación para la determinación de las características de las redes sociales es un campo en permanente evolución.

Algoritmos como los de "detección de comunidades" (Wang, 2015), "detección de anomalías" (Kaur, 2016), "determinación de subredes similares", "clustering dinámico" (Boccaletti, 2007) y "predicción de enlaces más probables" (Kushwah, 2016), son un ámbito en donde las técnicas de aprendizaje supervisado están encontrando cada vez más aplicaciones.

Uno de los requisitos para la aplicación de varios de los algoritmos mencionados es que la red cumpla con los requisitos de ser una "red de mundo pequeño", característica que se analizará más adelante en el presente trabajo.

¹ Megaparsec, medida de distancia, aproximadamente 3.26 millones de años luz.

² Técnica que consiste en realizar un mapeo sistemático de una porción determinada de la esfera celeste sin concentrarse de manera puntual en ningún objeto.

³ <https://vvvsurvey.org/>

⁴ <https://www.lsst.org/>

Desarrollo

Las “redes de mundo pequeño” son un tipo especial de grafo, dirigido o no, cuyas características más importantes se pueden resumir en:

1. Tienen un coeficiente de agrupamiento alto (clustering coefficient)
2. Tienen una longitud de camino promedio corto (average path length)

Esto coincide con observaciones realizadas sobre diversos tipos de redes sociales, las cuales tienen comunidades muy conectadas separadas entre sí por enlaces débiles.

En las próximas secciones se analizarán los pasos para determinar si una red estelar puede considerarse una "red de mundo pequeño" con lo que implica para analizar patrones y comunidades.

Obtención y preprocesamiento de datos con TOPCAT

Para obtener los datos iniciales se consultó por nombre sobre el catálogo GAIA versión 3, con los datos asociados a la Pequeña Nube de Magallanes (SMC o NGC292), que estudia este trabajo.

Accediendo a las interfaces de consulta interactiva del catálogo se realizó una búsqueda por nombre la cual devolvió 3.024.418 registros, los cuales deberían ser preprocesados para reducir el set de datos y eliminar información redundante e innecesaria. Este preprocesamiento se realizó utilizando la herramienta TOPCAT de gestión de tablas astronómicas.

Como el análisis se realizaría sobre los datos del movimiento propio de las estrellas (proper motion o PM, en adelante) se debió diferenciar las estrellas más lejanas, con un movimiento propio imperceptible, de aquellas más cercanas, las que cuentan con un movimiento claramente detectable con los instrumentos apropiados.

Se calcularon las métricas estadísticas sobre las columnas de datos asociadas al movimiento propio en sus dos componentes, ascensión recta (rect ascension, RA) y declinación (declination, DEC).

Después se realizó un filtrado de datos descartando aquellas estrellas cuyo movimiento fuera inferior a la media de los movimientos del set de datos en general y aquellas cuyo error de lectura superase al error de lectura promedio del set de datos, en valor absoluto para evitar compensaciones por cambio de signos.

Este proceso redujo el set de datos a 180.900 datos, los cuales se cargaron en una notebook Jupyter para seguir su procesamiento por medio de Python y la librería AstroPy.

Procesamiento utilizando AstroPy

Una vez importados los datos a una Jupyter Notebook se volvieron a calcular los indicadores estadísticos descriptivos sobre la nueva muestra, entre ellos la media del movimiento propio, dato bajo estudio.

Una vez obtenido el valor promedio de movimiento de todas las estrellas del set de datos se tomó como referencia un rango de $\pm 20\%$ por encima y por debajo de dicho promedio y se individualizó las estrellas que cumplieran con este requisito, para su análisis detallado.

Una vez aplicados los criterios de filtro se determinó la siguiente cantidad de datos:

- Movimiento en RA dentro del rango: 21022 registros
- Movimiento en DEC dentro del rango: 20647 registros
- Movimiento SIMULTANEAMENTE dentro del rango (RA +

Siendo esta última cantidad sobre la que se realizarán los análisis.

Ya habiendo obtenido las estrellas con movimientos similares dentro del set de datos, se procedió a detectar cuáles de ellas estaban próximas entre sí. El procedimiento utilizado fue el siguiente:

1. Recorrer todas las estrellas del set de datos filtrado
2. Para cada una de ellas obtener un círculo, con centro en la estrella y radio equivalente a 5 minutos de arco y realizar el siguiente proceso:
 - a. Buscar las estrellas que se encuentran dentro del
 - b. Agregarlas a una lista de estrellas “vecinas”.
 - c. Asociar la lista de “vecinas” a la estrella sobre la cual se está realizando el análisis.
 - d. Proseguir con la siguiente estrella de la lista

Una vez completado el análisis se cuenta con una estructura de datos similar a la de la Tabla 1.

Identificador	Vecinas
6377284298571599744	[6379356298170174336]
6377285849055413888	[6379353613814721280, 6379353721189807488]

Tabla 1. Estructura de datos

Estos datos se utilizaron posteriormente para el armado del grafo a fin de realizar el análisis necesario sobre la estructura y atributos de la red.

Se determinó que aquellas estrellas que no cuentan con vecinas no aportan valor al análisis de la red por lo que los elementos asociados a estrellas aisladas también se eliminaron del set de datos.

Como paso final, se exportaron los datos de la notebook Jupyter en un formato adecuado para ser importado en el software de análisis de grafos, que, en nuestro caso, se llevó a cabo utilizando Gephi.

Importación y análisis en Gephi

Gephi es un software de análisis de redes, de código abierto y gratuito, ampliamente utilizado por su potencia y facilidad de uso. En el presente trabajo se utilizó la versión 0.10, actualizada a noviembre de 2023.

Los archivos exportados desde Jupyter tienen la estructura esperada por Gephi para su importación, la que se presenta en la Tabla 2.

source	target
6377284298571599744	6379356298170174336
6377285849055413888	6379353613814721280
6377285849055413888	6379353721189807488

Tabla 2. Estructura de importación de Gephi

Una vez generados los archivos desde Jupyter, los mismos se importan a Gephi, momento en el cual el software crea todos los nodos necesarios y establece las relaciones (aristas) entre los nodos “source” y “target”.

Para determinar las características propias del grafo estelar, se generó un grafo comparativo, utilizando información aleatoria y generando aproximadamente los mismos nodos y relaciones que el grafo analizado, para comparar sus estructuras y características.

Los atributos de ambos grafos se resumen en la Tabla 3.

Grafo Aleatoria (referencia)		Grafo de Movimiento Propio (bajo análisis)	
Diámetro	16	Diámetro	9
Radio	1	Radio	1
Cantidad de nodos	2766	Cantidad de nodos	2936
Cantidad de aristas	4221	Cantidad de aristas	4076

Tabla 3. Características de los grafos bajo estudio

La visualización de ambos grafos se puede apreciar en las Figuras 1 y 2.

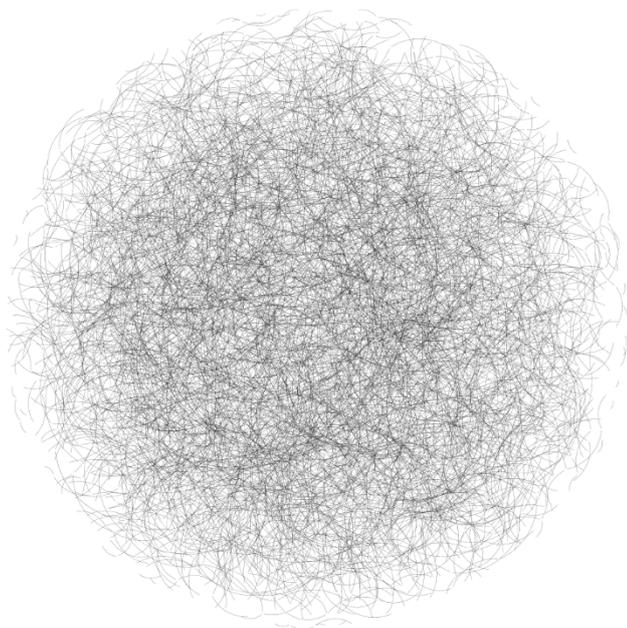


Figura 1. Grafo aleatorio de referencia

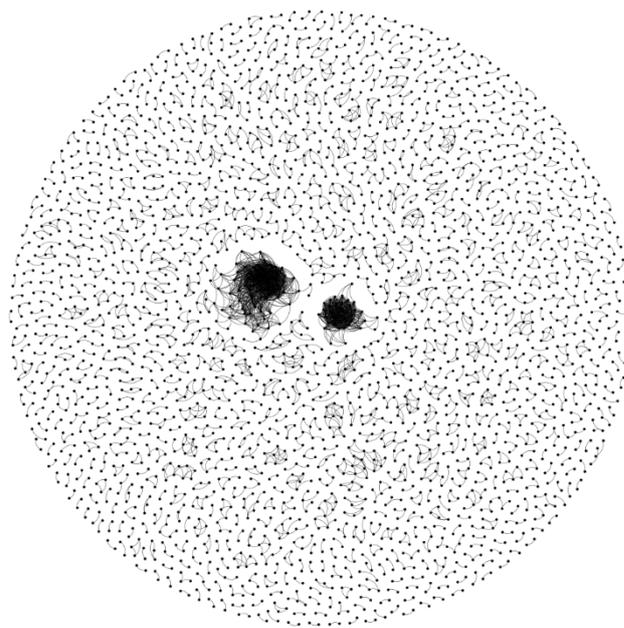


Figura 2. Grafo estelar bajo análisis

A simple vista se pueden apreciar algunos agrupamientos que serán analizado en mayor detalle en la sección de Resultados.

Resultados

El criterio más aceptado para determinar si una red dada puede considerarse una “red de mundo pequeño” radica en la comparación de dos valores típicos de las métricas de grafos (Newman, 2011).

La red tiene:

- Un coeficiente de agrupamiento superior a un grafo aleatorio de igual tamaño.
- Una longitud de camino promedio menor que en un grafo aleatorio de igual tamaño.

Dicho de otra manera, la red está más agrupada y presenta comunidades compactas, separadas unas de otras por caminos de longitud elevada.

En los dos grafos presentados anteriormente se puede apreciar este fenómeno, corroborado por las métricas que se muestran en la Tabla 4.

	Grafo aleatorio	Red estelar bajo estudio
Coefficiente de agrupamiento	0.001	0.606 (>)
Longitud media de camino	7.4032	2.084433 (<)

Tabla 4. Métricas comparativas del grafo aleatorio vs. Red estelar

Conclusiones

En vista a los resultados aquí expuestos, podemos afirmar que la red estelar, conformada por estrellas con similar movimiento propio, dentro del conjunto de datos de la Nube Pequeña de Magallanes, tiene una estructura que se condice con los criterios de una “red de mundo pequeño”, y por consiguiente se pueden utilizar sobre la misma algoritmos de detección de comunidades y otras analíticas asociadas al estudio de redes sociales, por tratarse estas últimas también de “redes de mundo pequeño”.

Estos resultados amplían en gran medida los métodos a utilizar para la detección de agrupaciones galácticas, las cuales hasta el momento se han analizado preponderantemente por medio de métodos numéricos tradicionales.

Como parte de los trabajos futuros a desarrollarse en el marco de la tesis de doctorado, se ampliará el análisis a otros atributos, además del movimiento propio, tales como la emisión de frecuencias de los cuerpos,

ya sea en espectro visible o infrarrojo, la masa, el brillo y otras características a fin de determinar si los resultados obtenidos son extrapolables a otros atributos o son particulares del movimiento propio.

También se prevé analizar diversos algoritmos de detección de comunidades, sobre redes sociales, pero ajustando los mismos para utilizar información astronómica asociada al dominio bajo estudio.

Referencias

- Alharbi, A., & Alsubhi, K. (2021). Botnet detection approach using graph-based machine learning. *Ieee Access*, 9, 99166-99180.
- Barnes, J. A., & Harary, F. (1983). Graph theory in network analysis. *Social networks*, 5(2), 235-244.
- Boccaletti, S., Ivanchenko, M., Latora, V., Pluchino, A., & Rapisarda, A. (2007). Detecting complex network modularity by dynamical clustering. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 75(4), 045102.
- Dalcanton, J. J., Williams, B. F., Seth, A. C., Dolphin, A., Holtzman, J., Rosema, K., ... & Quinn, T. (2009). The ACS nearby galaxy survey treasury. *The Astrophysical Journal Supplement Series*, 183(1), 67.
- Fall, S. M., & Chandar, R. (2012). Similarities in populations of star clusters. *The Astrophysical Journal*, 752(2), 96.
- Kaur, R., & Singh, S. (2016). A survey of data mining and social network analysis based anomaly detection techniques. *Egyptian informatics journal*, 17(2), 199-216.
- Klessen, R. S., & Burkert, A. (2000). The Formation of Stellar Clusters: Gaussian Cloud Conditions. I. *The Astrophysical Journal Supplement Series*, 128(1), 287.
- Kushwah, A. K. S., & Manjhvar, A. K. (2016). A review on link prediction in social network. *International Journal of Grid and Distributed Computing*, 9(2), 43-50.
- Newman, M., Barabási, A. L., & Watts, D. J. (2011). *The structure and dynamics of networks*. Princeton university press.
- Vázquez, R. A., May, J., Carraro, G., Bronfman, L., Moitinho, A., & Baume, G. (2008). Spiral structure in the outer galactic disk. I. The third galactic quadrant. *The Astrophysical Journal*, 672(2), 930.
- Rodríguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PloS one*, 14(1), e0210236.
- Schmeja, S. (2011). Identifying star clusters in a field: A comparison of different algorithms. *Astronomische Nachrichten*, 332(2), 172-184.
- Wang, C., Tang, W., Sun, B., Fang, J., & Wang, Y. (2015, December). Review on community detection algorithms in social networks. In *2015 IEEE international conference on progress in informatics and computing (PIC)* (pp. 551-555). IEEE.
- West, D. B. (2001). *Introduction to graph theory (Vol. 2)*. Upper Saddle River: Prentice hall.