

## Construcción de un pipeline para caracterización de redes estelares, utilizando Kedro

Martín Gustavo Casatti<sup>1</sup>, Marcelo Martín Marciszack<sup>1</sup>, and Carlos Feinstein<sup>2</sup>

<sup>1</sup> Universidad Tecnológica Nacional - Facultad Regional Córdoba

[mcasatti@frc.utn.edu.ar](mailto:mcasatti@frc.utn.edu.ar), [marciszack@gmail.com](mailto:marciszack@gmail.com)

<sup>2</sup> Universidad Nacional de La Plata, Argentina,

[cfeinstein@fcaglp.unlp.edu.ar](mailto:cfeinstein@fcaglp.unlp.edu.ar)

**Abstract.** El trabajo expone los pasos necesarios para implementar un pipeline de procesamiento de datos utilizando el framework Kedro, destinado al preprocesamiento de datos astronómicos, a fin de realizar una caracterización de grupos estelares, tendientes a determinar si el set de datos objetivo conforma una red de mundo pequeño. Se detallan las características de la herramienta utilizada, así como los criterios analizados para su adopción, a la vez que los pasos de consulta y descarga de datos, definición de etapas de preprocesamiento, implementación de las mismas, ejecución y resultados finales. Asimismo se analizará en las conclusiones, el valor que estas herramientas aportan a la repetibilidad y confiabilidad, tanto de las etapas de preprocesamiento de información, como de la analítica de datos que con posterioridad se realiza sobre dicha información, tendiente a obtener flujos de análisis de datos confiables, repetibles y de alta calidad analítica.

**Keywords:** ciencia de datos, cúmulos estelares, reconocimiento de patrones

## Building a pipeline for stellar network characterization using Kedro

**Abstract.** This paper presents the steps required to implement a data processing pipeline using the Kedro framework for preprocessing astronomical data to characterize stellar groups and determine whether the target dataset constitutes a small-world network. The tool's characteristics and the criteria analyzed for its adoption are detailed, along with the steps for querying and downloading data, defining preprocessing stages, their implementation, execution, and final results. The conclusions also analyze the value that these tools contribute to the repeatability and reliability of both the data preprocessing stages and the subsequent data analytics performed on said data, aimed at obtaining reliable, repeatable, and high-quality data analysis flows.

**Keywords:** data science, stellar clusters, pattern recognition

## 1 Introducción

### 1.1 El estudio de agrupaciones estelares

Las agrupaciones estelares, también denominadas cúmulos o clusters, han sido objetos reconocidos desde hace tiempo como laboratorios importantes para la investigación astrofísica, siendo muy útiles en varios aspectos, entre ellos:

- Contienen muestras estadísticamente significativas de estrellas de aproximadamente la misma edad, 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 & Burkert, 2000).
- 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 (Fall & Chandar, 2012).

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

En la actualidad existe una gran cantidad de información de las galaxias cercanas (a varios Mpc<sup>3</sup>) 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 (Dalcanton et al., 2009).

Se cuenta con una enorme cantidad de datos proveniente de las varias observaciones continuas que se están realizando en modo ‘survey’<sup>4</sup> (p.e. VVV<sup>5</sup> o LSST<sup>6</sup>) 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). Estos algoritmos se basan en analizar las posiciones espaciales para encontrar los clusters 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 (Rodriguez et al., 2019).

### 1.2 Comunidades y redes sociales

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

---

<sup>3</sup> Megaparsec, medida de distancia, aproximadamente 3.26 millones de años luz.

<sup>4</sup> Técnica que consiste en realizar un mapeo sistemático de una porción determinada de la esfera celeste sin concentrarse de manera puntual en ningún objeto.

<sup>5</sup> <https://vvvsurvey.org>

<sup>6</sup> <https://www.lsst.org>

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 et al., 2001). Diversos estudios, tanto de la topología de dichas redes (Barnes & Harary, 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 & Alsuhbi, 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 et al., 2015), “detección de anomalías” (Kaur & Singh, 2016), “determinación de subredes similares”, “clustering dinámico” (Boccaletti et al., 2007) y “predicción de enlaces más probables” (Kushwah & Manjvar, 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.

### **1.3 Trabajos previos**

El trabajo “Bases de datos de grafos como soporte para la detección de estrellas jóvenes en cúmulos estelares cercanos” publicado en la 7º Conferencia de Big Data y Cloud Computing (Jornadas de Cloud Computing, 2019), sentó las bases iniciales para la línea de investigación del presente trabajo. El mismo utiliza algunos conceptos analizados sobre otro ámbito de aplicación, como son las bases de datos cienciométricas y las características que debe reunir un sistema de almacenamiento de información para permitir la detección de patrones de manera eficiente. Dicho enfoque se expone en los trabajos “Criterios para el diseño de una base de datos cienciométrica” (Muñoz et al., 2020) y “Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico” (Paz Menvielle et al., 2018)

### **1.4 Kedro como framework para pipelines de procesamiento de datos**

Por otra parte, Kedro (Alam et al., 2025) es un framework de desarrollo de código abierto para la creación de pipelines (flujos) de procesamiento de datos robusto y mantenible. Su estructura modular y basada en convenciones facilita la organización del flujo de trabajo, la gestión de dependencias y la reproducibilidad de los análisis.

El framework Kedro presenta varias características que lo hacen particularmente interesante para la creación de un pipeline de procesamiento de información astronómica

Característica	Descripción
Modularidad y Organización del Código	Fomenta una estructura de proyecto clara y modular, separando la lógica de procesamiento de datos en <b>nodos</b> y la gestión de datos en un <b>catálogo de datos</b> . Facilita la organización, comprensión del flujo de trabajo y colaboración.
Reproducibilidad	Asegura la <b>reproducibilidad</b> de experimentos al definir explícitamente dependencias entre nodos y gestionar entradas/salidas a través del catálogo de datos. Crucial para la verificación y replicación de resultados.
Gestión de Dependencias	Maneja <b>automáticamente las dependencias</b> entre los nodos del pipeline, ejecutándolos en el orden correcto y solo cuando sus dependencias están satisfechas, simplificando flujos de trabajo complejos.
Abstracción de la Capa de Datos	El <b>catálogo de datos abstrae</b> cómo se cargan y guardan los datos, permitiendo cambiar fácilmente las fuentes de datos sin modificar la lógica de procesamiento. Facilita el manejo de diferentes formatos y fuentes de datos.
Facilidad para la Experimentación y el Desarrollo Iterativo	La estructura modular y el enfoque en la configuración permiten realizar <b>experimentos y desarrollo iterativo</b> variando parámetros o sustituyendo nodos fácilmente.
Integración con Herramientas de Ciencia de Datos	Se integra bien con librerías estándar de Python como ‘numpy’, ‘pandas’, ‘scikit-learn’, ‘astropy’ y ‘networkx’.
Visualización del Pipeline	Ofrece herramientas para <b>visualizar el pipeline</b> como un grafo dirigido acíclico (DAG), proporcionando una representación clara del flujo de datos y dependencias para depuración y comprensión de la lógica interna.
Concurrencia y Escalabilidad (Implícita)	Su arquitectura permite la <b>ejecución paralela de nodos independientes</b> , beneficiosa para procesar conjuntos de datos más grandes en el futuro.

Tabla 1: Características destacadas del framework Kedro

y el análisis de redes complejas como el propuesto, se mencionan a continuación las más significativas, resumidas en la Tabla 1.

En resumen, Kedro proporciona un marco de trabajo robusto y organizado que simplifica la creación, gestión y reproducción de pipelines de procesamiento de datos complejos, haciéndolo una elección adecuada para el análisis de información astronómica y la construcción de grafos para el estudio de redes de mundo pequeño.

### 1.5 Motivación y objetivos del trabajo

Este trabajo se motiva en la exploración de las posibles estructuras de mundo pequeño que puedan surgir a partir de las relaciones de similitud en atributos seleccionados de las estrellas. El objetivo principal es:

- Desarrollar un pipeline de procesamiento de datos astronómicos utilizando el framework Kedro para extraer información de atributos desde archivos FITS.
- Construir un grafo no dirigido donde los nodos representan estrellas y las aristas conectan estrellas con atributos similares según un criterio definido.
- Analizar las propiedades del grafo resultante, específicamente la distancia promedio de los caminos más cortos y el coeficiente de agrupamiento, para determinar si exhibe las características de una red de mundo pequeño.

## 2 Método

El diagrama general del pipeline se encuentra expresado gráficamente en la Figura 1, cuyos pasos son detallados a continuación.

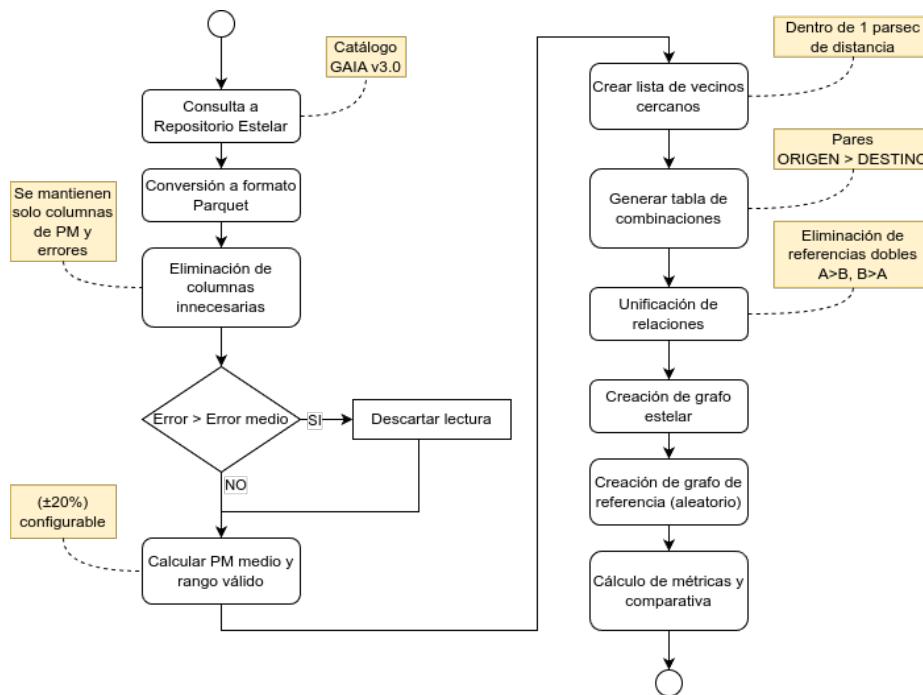


Figura 1: Diagrama general del pipeline Kedro

### 2.1 Obtención de datos

Para obtener los datos iniciales se consultó por nombre sobre el catálogo SIMBAD, el cual realiza el mapeo de objetos astronómicos conocidos, y obtiene sus correspondientes coordenadas. Una vez obtenidas dichas coordenadas se realizó una consulta con ADQL (Astronomical Data Query Language) al catálogo GAIA versión 3, para la Pequeña Nube de Magallanes (SMC o NGC292) (Westerlund, 1990), que estudia este trabajo. Accediendo a las interfaces de consulta del catálogo, se realizó una búsqueda que devolvió aproximadamente **3.000.000** de registros, los cuales deberían ser preprocesados para reducir el set de datos y eliminar información redundante e innecesaria, para lo cual se planteó el pipeline de proceso con la herramienta Kedro.

```

1 SELECT
2     solution_id, designation, source_id, ra, dec, parallax, pm,

```

```

3     pmra, pmra_error, pmdec, pmdec_error, pseudocolour,
4     pseudocolour_error, radial_velocity, radial_velocity_error
5 FROM gaiadr3.gaia_source AS g
6 WHERE
7     1 = CONTAINS(
8         POINT('ICRS', g.ra, g.dec),
9         CIRCLE('ICRS',
10            :RA, :DEC, :RADIUS
11        )
12    ) AND
13   g.phot_g_mean_mag < 20.5 AND
14   g.parallax IS NOT NULL

```

Listing 1: Consulta SQL a la base de datos Gaia DR3

## 2.2 Pipeline de procesamiento con Kedro

Como el análisis se tuvo que realizar sobre los datos del movimiento propio de las estrellas (proper motion o PM, en adelante) (Piatek et al., 2008) 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, para lo cual se calcularon las métricas estadísticas sobre las columnas de datos asociadas al movimiento propio en sus dos componentes, ascensión recta (Rect Ascention, RA) y declinación (Declination, DEC).

Todo el procesamiento de los nodos se realizó utilizando las librerías Pandas<sup>7</sup> y Astropy<sup>8</sup>, mientras que el procesamiento de los grafos se realizó utilizando NetworkX<sup>9</sup>.

El pipeline que se definió en Kedro estuvo compuesto por las siguientes tareas (con sus nodos asociados):

**Conformar tabla de estrellas** Se eliminaron las columnas innecesarias del conjunto de datos original para minimizar el uso de memoria y optimizar los cálculos posteriores.

**Filtrar por error de PM (Proper Motion)** Se calculó el error medio del movimiento propio (PM) en las direcciones de ascensión recta (RA) y declinación (DEC). Posteriormente, se eliminaron todas las lecturas de estrellas cuyos errores en RA o DEC superaron sus respectivos errores medios, conservando así las mediciones más precisas disponibles.

**Filtrar por PM (Proper Motion)** Se determinó el valor medio del movimiento propio. Se estableció un rango de similitud, configurable por el investigador, alrededor de este valor medio. Dos estrellas se consideraron con movimientos propios similares si sus valores de movimiento propio se encontraban dentro de este rango. En este estudio particular, el rango se fijó en  $\pm 20\%$  del valor medio calculado.

---

<sup>7</sup> <https://pandas.pydata.org/>

<sup>8</sup> <https://www.astropy.org/>

<sup>9</sup> <https://networkx.org/>

**Encontrar vecinos cercanos** Para cada estrella en el conjunto de datos filtrado, se identificaron sus vecinos cercanos dentro de un radio espacial específico. En este caso, se utilizó un radio de 1 segundo de arco, que corresponde aproximadamente a 3.26 años luz (1 parsec), una unidad estándar en la medición de distancias astronómicas.

Una vez completado el análisis se cuenta con una estructura de datos similar a la de la Tabla 2.

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

Tabla 2: Tabla de vecinos cercanos

Estos datos se utilizaron posteriormente para la construcción 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.

**Generar tabla de combinaciones** Se creó una tabla que contenía todas las posibles combinaciones de pares de estrellas donde una estrella era considerada vecina cercana de la otra según el criterio espacial definido en el nodo anterior. Cada fila de esta tabla representaba un posible enlace en el grafo, con una estrella como origen y la otra como destino, tal como se aprecia en la tabla 3.

source	target
6377284298571599744	6379356298170174336
6377285849055413888	6379353613814721280
6377285849055413888	6379353721189807488

Tabla 3: Tabla de combinaciones

**Remoción de referencias mutuas** Dado que el objetivo era construir un grafo no dirigido, se eliminaron las referencias mutuas redundantes de la tabla de combinaciones. Si existía una entrada para la relación  $A \rightarrow B$  y otra para  $B \rightarrow A$ , se conservó solo una de ellas para evitar la duplicidad de aristas en el grafo final.

**Crear grafo estelar** Utilizando la tabla de combinaciones procesada, se construyó el grafo no dirigido final. Cada estrella se representó como un nodo en el grafo, y cada entrada en la tabla de combinaciones (después de la remoción de referencias mutuas) se convirtió en una arista que conectaba los dos nodos correspondientes.

**Crear grafo aleatorio** Para comparar las propiedades del grafo estelar construido, se generó un grafo aleatorio, siguiendo el modelo Erdős-Rényi, con el mismo número de nodos que el grafo estelar. La creación de este grafo aleatorio se basó en una probabilidad fija de que dos nodos cualesquiera estuvieran conectados por una arista. En este estudio, la probabilidad de conexión se estableció en un 1 %.

**Cálculo de métricas y comparación** Finalmente, se calcularon diversas métricas de ambos grafos (el grafo estelar y el grafo aleatorio) para analizar sus propiedades estructurales. Estas métricas se utilizaron para determinar si el grafo de movimiento propio estelar exhibía las características típicas de una red de mundo pequeño, comparando sus valores con los obtenidos para el grafo aleatorio de referencia.

Se calcularon las siguientes métricas de los grafos resultantes utilizando networkx:

- **Distancia promedio de los caminos más cortos ( $L$ ):** La media de las distancias de los caminos más cortos entre todos los pares de nodos en el grafo.
- **Coeficiente de agrupamiento ( $C$ ):** Una medida de la tendencia de los nodos a formar grupos o cliques, calculado como el coeficiente de agrupamiento promedio del grafo.

### 2.3 Análisis de “mundo pequeño”

Para determinar si el grafo construido presentaba características de “mundo pequeño”, se compararon sus métricas ( $L$  y  $C$ ) con las de un grafo aleatorio con el mismo número de nodos y el mismo grado promedio. Un grafo de mundo pequeño típicamente cumple las siguientes condiciones:

- Su distancia promedio de los caminos más cortos ( $L$ ) es similar a la de un grafo aleatorio ( $L \approx L_{rand}$ ).
- Su coeficiente de agrupamiento ( $C$ ) es significativamente mayor que el de un grafo aleatorio ( $C \gg C_{rand}$ ).

Grafo Aleatorio (referencia)		Grafo de Movimiento Propio (bajo análisis)	
Diámetro	16	Diámetro	9
Radio	1	Radio	1
Cantidad de nodos	2664	Cantidad de nodos	2936
Cantidad de aristas	4233	Cantidad de aristas	4076
Longitud de camino promedio ( $L_{rand}$ )	7.13	Longitud de camino promedio ( $L$ )	2.084
Coeficiente de agrupamiento ( $C_{rand}$ )	0.004	Coeficiente de agrupamiento ( $C$ )	0.606

Tabla 4: Características del grafo bajo estudio

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 siguientes secciones se analizarán los pasos para determinar si una red estelar puede considerarse una “de mundo pequeño” con lo que esto implica para el análisis de patrones y comunidades.

Los atributos de ambos grafos se resumen en la Tabla 4.

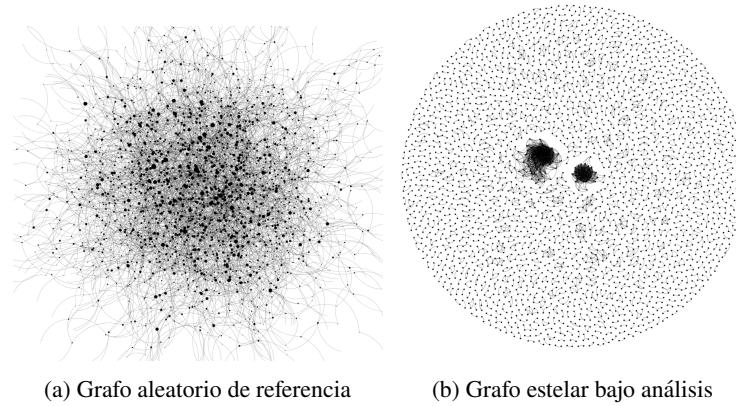


Figura 2: Comparativa de grafos para análisis

Un análisis visual de los dos grafos nos permite observar algunas características interesantes.

Por una parte el grafo bajo análisis, constituido a partir de los datos obtenidos por medio del pipeline Kedro, muestra claramente dos grandes componentes conexos, con gran cantidad de nodos y aristas, y un conjunto de agrupamientos de menor tamaño, bastante consistentes en cuanto a la cantidad de nodos (que varía en líneas generales entre 4 y 8 nodos).

Por otro lado, el grafo aleatorio presenta un tamaño de agrupamiento menor, más cantidad de grupos de menor tamaño, y una conectividad más distribuida.

La visualización de ambos grafos se puede apreciar en las Figuras 2a y 2b, en donde se pueden distinguir algunos agrupamientos que serán analizados en mayor detalle en la sección 3.

### 3 Resultados

#### 3.1 Kedro como herramienta de analítica de datos

Kedro demostró ser una herramienta eficiente para la definición de flujos de análisis de datos, permitiendo descomponer un flujo continuo en unidades autocontenidoas (denominadas “nodos”), con entradas y salidas definidas y un proceso atómico que transforma dichas entradas en salidas.

La interconexión de estos bloques de análisis en un pipeline permite modelar un proceso analítico con poco esfuerzo y una simplicidad y economía de conceptos importantes.

Asimismo, la posibilidad de visualizar el flujo de manera gráfica (tal como se aprecia en la Figura 3) brinda una visión clara y eficaz tanto de los datos como de los procesos y parámetros involucrados, permitiendo detectar rápidamente inconsistencias y posibles puntos de fallo.

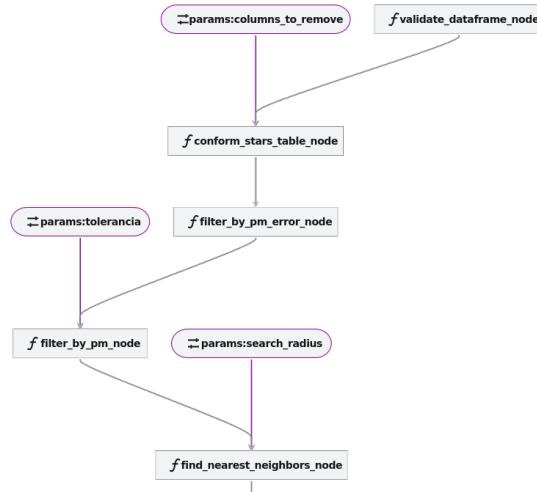


Figura 3: Visualización de un pipeline de Kedro

Otro punto a destacar en el uso de Kedro es la implementación del concepto abstracto Dataset (o fuente de datos). Este concepto permite especificar diferentes tipos de archivos y almacenamientos como origen de los datos, que son tratados de manera uniforme por la herramienta. De esta manera, cambiar una fuente de datos de un archivo local en formato Parquet a una fuente remota de datos que descarga información en formato JSON es tan simple como cambiar dos parámetros en un archivo de configuración.

Esto acelera en gran medida el análisis sobre diferentes sets de datos, sin tener que preocuparse de los pormenores de procesar cada formato de archivo por separado o de adecuar las rutinas de carga y lectura.

En líneas generales el uso de Kedro como herramienta de analítica de datos redujo el tiempo total de definición del flujo de análisis en un 30 % y el tiempo de ejecución y análisis de resultados, de los distintos escenarios, en más de un 70 %, en promedio.

### 3.2 Los grafos estelares como “redes de mundo pequeño”

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 et al., 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 5.

	Grafo aleatorio	Red estelar bajo estudio
Coeficiente de agrupamiento	0.004	0.606 (>)
Longitud media de camino	7.13	2.084433 (<)

Tabla 5: Métricas comparativas del grafo aleatorio vs. Red estelar

## 4 Conclusiones

El uso de herramientas de automatización de analítica de datos es una tendencia creciente en el mundo de Big Data. La cantidad y complejidad de los análisis necesarios hacen que sea imprescindible el uso de ayudas para el planteo de procesos analíticos repetibles, parametrizables y escalables.

Asimismo, muchas de estas herramientas permiten, además, experimentación por medio del planteo de escenarios paramétricos, los cuales pueden ser ejecutados y comparados en sus resultados.

Los resultados expuestos en el presente trabajo demuestran que el uso de herramientas analíticas de definición y ejecución de pipelines mejora en gran medida los tiempos de desarrollo y simplifica notablemente los proyectos de analítica de datos.

De esta manera el analista puede definir procesos más compactos y predecibles por medio de la programación de partes reutilizables (los “nodos”), las que pueden ser probadas de manera individual e incluso intercambiadas por otras, que utilicen diferentes técnicas o enfoques, sin tener que replantear el proceso en general.

Por otra parte, 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”.

La posibilidad de utilizar algoritmos de fines específicos, como los de detección de comunidades, o redes de influencia, de amplia aplicación sobre redes sociales, en un ámbito por el momento no explorado, es una buena noticia tanto para astrónomos como para científicos de datos y otros investigadores, ya que amplía de manera importante la cantidad de herramientas disponibles.

#### 4.1 Trabajos futuros

En vista de los presentes resultados se propone avanzar en la elaboración de una técnica estandarizada de análisis de características estelares, con el soporte de herramientas como Kedro, la cual debería incluir tanto la determinación de atributos, el preprocesamiento de los mismos y la creación del grafo estelar asociado, como la caracterización final de dicho grafo como etapa previa a la selección de algoritmos de detección de comunidades (clusters), tales como los mencionados en (Kumar et al., 2018) y (Wang et al., 2015), entre otros.

Como parte de los trabajos futuros a desarrollarse, 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.

Asimismo se efectuará el presente análisis pero sobre distintas secciones del cielo para determinar si las características detectadas son generalizables o son particulares de la zona seleccionada para este estudio.

### Referencias

- Alam, S., Chan, N. L., Couto, L., Dada, Y., Danov, I., Datta, D., DeBold, T., Gundaniya, J., Honoré-Rougé, Y., Kaiser, S., Kanchwala, R., Katiyar, A., Pilla, R. K., Nguyen, H., Cano Rodríguez, J. L., Schwarzmann, J., Sorokin, D., Theisen, M., Zablocki, M., & Brugman, S. (2025, marzo). *Kedro* (Ver. 0.19.12). <https://github.com/kedro-org/kedro>
- 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*, 75(4), 045102.
- Dalcanton, J. J., Williams, B. F., Seth, A. C., Dolphin, A., Holtzman, J., Rosema, K., Skillman, E. D., Cole, A., Girardi, L., Gogarten, S. M., et al. (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.
- Jornadas de Cloud Computing, B. D. E. T. (-B.). (2019). *Cloud Computing and Big Data*. Springer International Publishing. <https://link.springer.com/book/10.1007/978-3-030-27713-0>
- 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.
- Kumar, P., Chawla, P., & Rana, A. (2018). A review on community detection algorithms in social networks. *2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 304-309.
- 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.
- Muñoz, R. M., Casatti, M. G., Guzmán, A., Damiano, L. E., & Cuevas, J. C. (2020). Criterios para el diseño de una base de datos cienciométrica. *CoNaIISI 2020*.
- Newman, M., Barabási, A.-L., & Watts, D. J. (2011). *The structure and dynamics of networks*. Princeton university press.
- Paz Menivelle, M. A., Corso, C. L., Ligorria, K., Guzmán, A., & Casatti, M. (2018). Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico. *XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste)*.
- Piatek, S., Pryor, C., & Olszewski, E. W. (2008). Proper motions of the Large Magellanic Cloud and Small Magellanic Cloud: re-analysis of Hubble space telescope data. *The Astronomical Journal*, 135(3), 1024.
- Rodriguez, 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.
- 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.
- Wang, C., Tang, W., Sun, B., Fang, J., & Wang, Y. (2015). Review on community detection algorithms in social networks. *2015 IEEE international conference on progress in informatics and computing (PIC)*, 551-555.
- West, D. B., et al. (2001). *Introduction to graph theory* (Vol. 2). Prentice hall Upper Saddle River.
- Westerlund, B. (1990). The Magellanic Clouds: their evolution, structure and composition. *The Astronomy and Astrophysics Review*, 2, 29-78.