

¿Existe correlación entre importancia y centralidad? Evaluación de personajes con redes sociales en obras teatrales de la Edad de Plata

María Teresa Santa María

Universidad Internacional de La Rioja, Spain

José Calvo Tello

Göttingen State and University Library, Germany

Concepción María Jiménez

Universidad Internacional de La Rioja, Spain

Resumen

El objetivo de este estudio es responder a la pregunta: ¿Los nodos centrales de una red social de personajes corresponden con los personajes protagonistas de una obra de teatro? Para responder a esta cuestión, evaluamos diferentes medidas de centralidad y otros valores cuantitativos textuales en relación con metadatos anotados manualmente en un corpus de veinticinco obras del teatro español correspondientes al periodo de la Edad de Plata (1868-1936). Los resultados que hemos encontrado señalan que la centralidad está en correlación moderada con la importancia, siendo mayor la correlación con valores cuantitativos textuales.

Abstract

The objective of this study is to answer the question: Do the central nodes of a social network of characters correspond with the protagonists of a theater play? To answer this question we evaluate different measures of centrality along with other textual quantitative values in relation to other manually annotated metadata on a corpus of twenty five dramatic plays of the Spanish theatre of the Silver Age (1868-1936). The found results show that centrality correlates moderately with importance, but the correlation with the textual quantitative values is stronger.

Correspondence:

María Teresa Santa María, Av. de la Paz, 137 26006, Logroño (La Rioja) Spain.

E-mail:

teresa.santamaria@unir.net

1 Introducción

La utilización de grafos y redes sociales en el estudio de los personajes y sus relaciones aporta nuevas herramientas y perspectivas interesantes al estudio de las

obras literarias (Marcus, 1973; Moretti, 2011). La centralidad es uno de los valores más utilizados dentro de la teoría de grafos y de su aplicación para el estudio de redes sociales (Jannidis *et al.*, 2017) y constituye un término asumido por las Humanidades Digitales

(HD). Diferentes investigadores están aproximándose a una mejor comprensión sobre qué representan exactamente estos conceptos. Algee-Hewitt (2017) propone conclusiones sobre la historia de la literatura con base en la distribución de la centralidad *eigenvector* de los personajes. El grupo DLINA (Trilcke *et al.*, 2015 y 2016) está analizando en la actualidad las correlaciones entre los conceptos de centralidad y otras medidas textuales sobre cada personaje. Así mismo, RoCHAT (2014) expone el concepto de centralidad aplicada a redes de personajes textuales basados en la autobiografía del francés Rousseau, para lo que hace uso del índice de personas mencionadas en la obra y la página donde aparecen, lo que da lugar a la obtención de información relevante incluso para crear aristas.

Dentro de los estudios sobre Literatura Hispánica, se han utilizado enfoques cuantitativos para analizar la densidad versal en obras del Siglo de Oro (Hermenegildo, 1995 y Espejo, 2002), para estudiar tanto contenido simbólico y sociopolítico de los personajes de Galdós (Menéndez, 1983), así como para examinar el origen social y la caracterización de los personajes de Lope de Vega (Oleza, 2013). Ya empiezan a existir interesantes trabajos que analizan las relaciones entre los personajes a partir del estudio de las redes sociales que se visualizan con los grafos, como, por ejemplo, en las novelas de Galdós (Isasi, 2017); o bien, estudios que defienden que, a más relaciones directas entre los personajes, mayor centralidad (Rodríguez, 2016).

En este trabajo queremos evaluar si es correcta la interpretación intuitiva que cuanto más central sea un nodo más importante es el personaje en la obra de teatro. Además de las medidas de centralidad, evaluamos otros dos tipos de datos que pueden ser fácilmente rescatables de las obras de teatro. En primer lugar, valores textuales sobre cada personaje, como la cantidad de veces que habla. En segundo lugar, consideramos la posición del personaje en la lista (*dramatis personae*) que precede al texto teatral, ya que esperamos que los protagonistas o personajes más importantes aparezcan en los primeros puestos de esa lista. En concreto, nuestra investigación parte de cuatro preguntas capitales:

- (1) ¿Cuán estrecha o relevante resulta la correlación entre las medidas de centralidad y la importancia del personaje?

- (2) ¿Incluye siempre un dramaturgo a los personajes más importantes al comienzo del *dramatis personae*?
- (3) ¿Existe correlación entre la importancia de un personaje y los valores textuales (cantidades de unidades textuales de dicho personaje)?
- (4) ¿Qué valores podríamos utilizar para distinguir a los personajes protagonistas de los que no lo son?

2 Textos y Metadatos

A diferencia de otras lenguas europeas, el español no cuenta con un gran corpus teatral anotado en XML-TEI. El grupo de investigación sobre teatro y Humanidades Digitales (GHEDI) surgió en la Universidad Internacional de La Rioja en 2015 y continúa su labor dentro del proyecto HDATEATROUNIR cuyo objetivo es investigar la aplicación de las redes sociales con grafos en el estudio del teatro español del Siglo de Oro y de la Edad de Plata. Uno de sus principales proyectos es la *Biblioteca Electrónica Textual del Teatro en Español de la Edad de Plata* (1868-1936) (BETTE), donde ha publicado veinticinco obras en XML-TEI de Lorca, Valle, Galdós, Clarín o Muñoz Seca, como repositorio GitHub (Jiménez *et al.*, 2017). En la versión 2.0 cada personaje ha sido anotado con diferentes metadatos que hemos empleado para la evaluación posterior:

- Sexo
- Papel en la obra (protagonista, amante, antagonista u otro)
- Naturaleza (persona, animal, no humano para objetos animados. . .)
- Importancia (personaje primordial, secundario o terciario)
- Persona individual frente a grupo¹

Por otro lado, también se añadieron una serie de valores textuales cuantitativos de manera automática para cada personaje y que nos permiten obtener de manera rápida y sencilla información referida a:

- Posición en el *dramatis personae*
- Cantidad de texto que pronuncia
- Cantidad de intervenciones
- Cantidad de referencias a su nombre
- Cantidad de escenas en las que aparece

Se puede ver a continuación, en la [Figura 1](#), un ejemplo de toda esta información incluida en dicho corpus en lenguaje de marcado XML-TEI:

```
<listPerson cert="high">
  <person n="1" role="protagonist" sex="M" xml:id="max">
    <persName>Max Estrella</persName>
    <trait type="nature"><desc>person</desc></trait>
    <trait type="importance"><desc>primary</desc></trait>
    <listEvent>
      <event type="characters"><desc>15813</desc></event>
      <event type="sps"><desc>278</desc></event>
      <event type="rss"><desc>96</desc></event>
      <event type="scenes"><desc>11</desc></event>
    </listEvent>
  </person>
```

Fig. 1 Metadatos de personajes en XML-TEI en *Luces de bohemia* de Valle-Inclán, procedente del corpus de BETTE

El metadato que utilizamos para la correlación es el valor de importancia de un personaje. Los tres grados diferentes de importancia fueron asignados a partir de los resúmenes – incluidos también en el archivo TEI - que de cada pieza realizaron los especialistas en Filología Española que llevaron a cabo la edición en lenguaje de marcado de este corpus.² Se establecieron tres posibles valores ordinales según los siguientes criterios:

- Minor: si el personaje no aparece en el resumen
- Secondary: si el personaje aparece nombrado en el resumen
- Primary: si pertenece al grupo de entre dos y cuatro personajes esenciales, pues su nombre se repite o se resalta dentro del resumen en más de una ocasión

De esta manera tenemos de cada personaje (sobre un total de 516) cuatro informaciones que se complementan y pueden incidir en el mayor o menor grado de relevancia de dicho personaje:

- (1) Un valor de su importancia dentro de la obra (que puede ser utilizado como *ground truth* o verdad fundamental)
- (2) Diferentes valores cuantitativos textuales
- (3) Posición en *dramatis personae*
- (4) Diferentes valores según medidas de centralidad³

3 Metodología

La implementación para extraer, analizar, evaluar y visualizar los datos se realizó en Python mediante

bibliotecas como *lxml* y *networkx*. Por otro lado, para la creación de las redes sociales se definió la arista no direccional como la coaparición en escenas, mientras que los personajes constituyen los nodos en los grafos ([Fig. 2](#)):

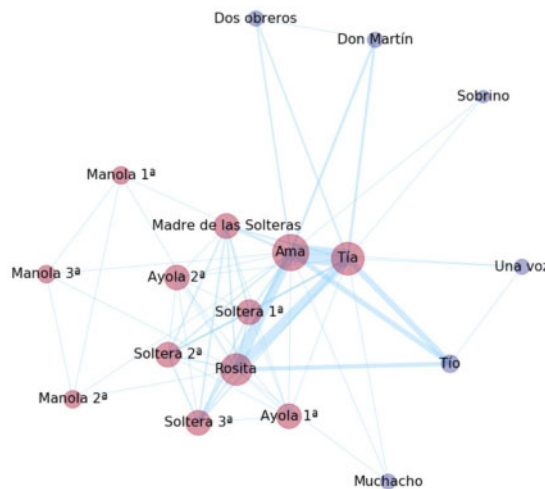


Fig. 2 Red social de personajes en *Doña Rosita la soltera* de Lorca

A partir de estas redes sociales calculamos diferentes medidas de centralidad e información sobre los nodos. En total, se establecen nueve medidas diferentes:⁴

- *Degree centrality* (centralidad de grado)
- *Betweenness centrality* (centralidad de intermediación)
- *Eccentricity*
- *Closeness centrality* (centralidad de cercanía)
- *Load centrality*
- *Current flow betweenness centrality*
- *Eigenvector centrality* (centralidad de vector propio)
- *Approximate current flow betweenness centrality*
- *Communicability centrality exp*

4 Resultados

En primer lugar, analizamos la dependencia entre la importancia asignada por especialistas y el resto de valores y calculamos su correlación (Spearman rho; [Evans, 1996, 152](#)):

Como se puede comprobar en la [figura 3](#), ninguna de las medidas de centralidad tiene una correlación fuerte (con valor absoluto mayor de 0.6 según [Evans, 1996](#)). El valor máximo (-0.5) corresponde a *current flow betweenness centrality*, también conocida como *information centrality* ([Brandes y Fleischer, 2005](#); [Stephenson y Zelen,](#)

1989), medida que no está entre el repertorio usual de las HD.

En cuanto a la hipótesis de que el dramaturgo sitúa a los personajes principales al principio del *dramatis personae*, la correlación entre la posición y la importancia es de 0.42, una correlación moderada similar a la de las medidas de centralidad. Conclusión

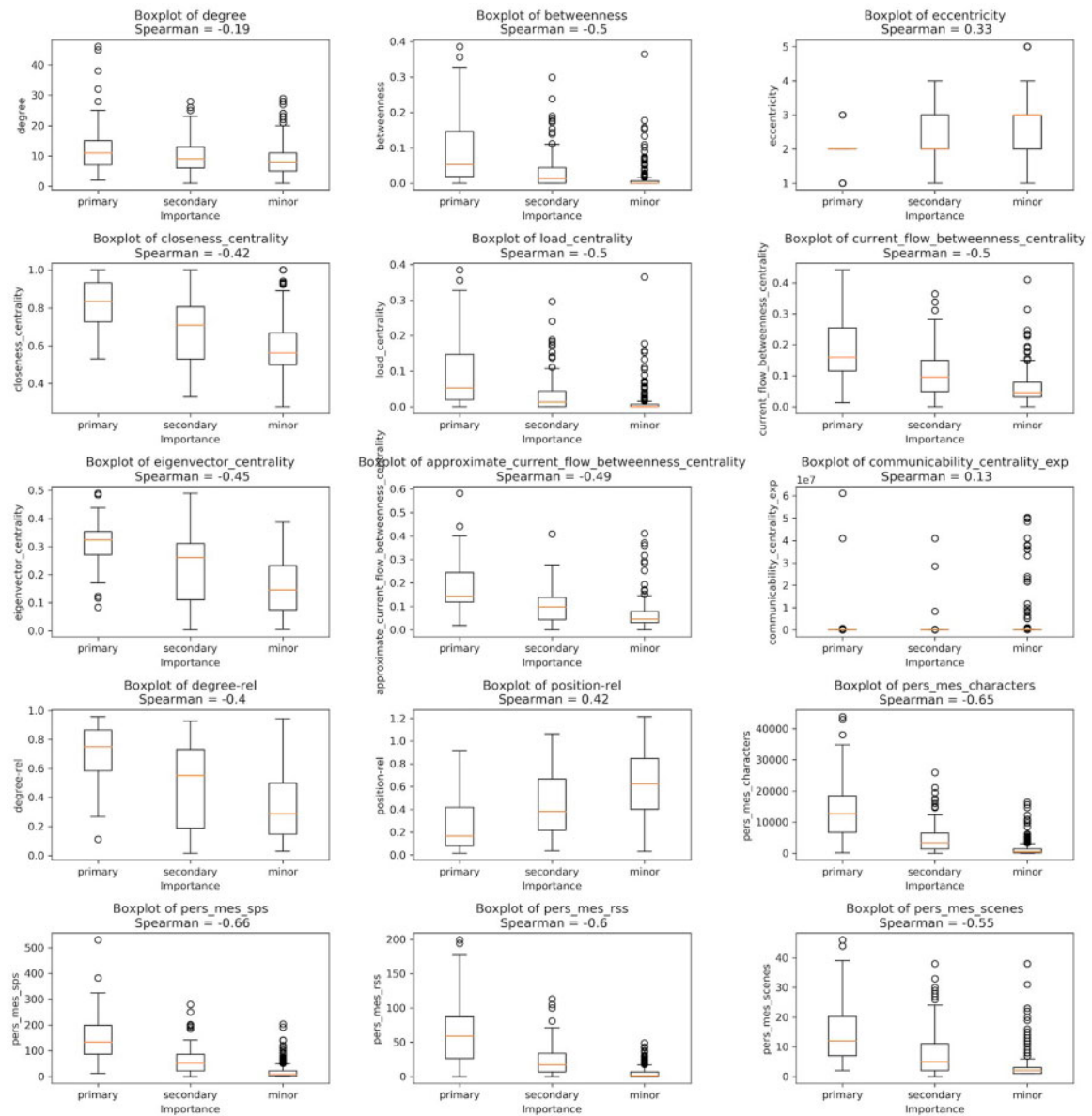


Fig. 3 Boxplots (diagramas de caja) y correlaciones con importancia de todas las obras de BETTE

corroborada porque en el *dramatis personae* a veces los escritores eligen criterios diferentes, como Muñoz Seca que los ordena según su sexo.

En relación a las medidas cuantitativas textuales, todas muestran correlaciones notablemente más altas, llegando hasta 0.66 en la cantidad de intervenciones, un resultado inesperado.

Ante estos datos nos hemos preguntado si las medidas cuantitativas textuales y las medidas de centralidad

están captando aspectos diferentes sobre la importancia de los personajes. Para investigar su posible diferenciación, calculamos la correlación entre el valor textual con valor más alto en el gráfico anterior (*pers_mes_sps*) con el resto. Si las correlaciones son similares a las del anterior gráfico, pensaremos que tanto los valores textuales como las medidas de centralidad captan matices similares de la importancia de los personajes, tal y como se puede observar en la [Figura 4](#):

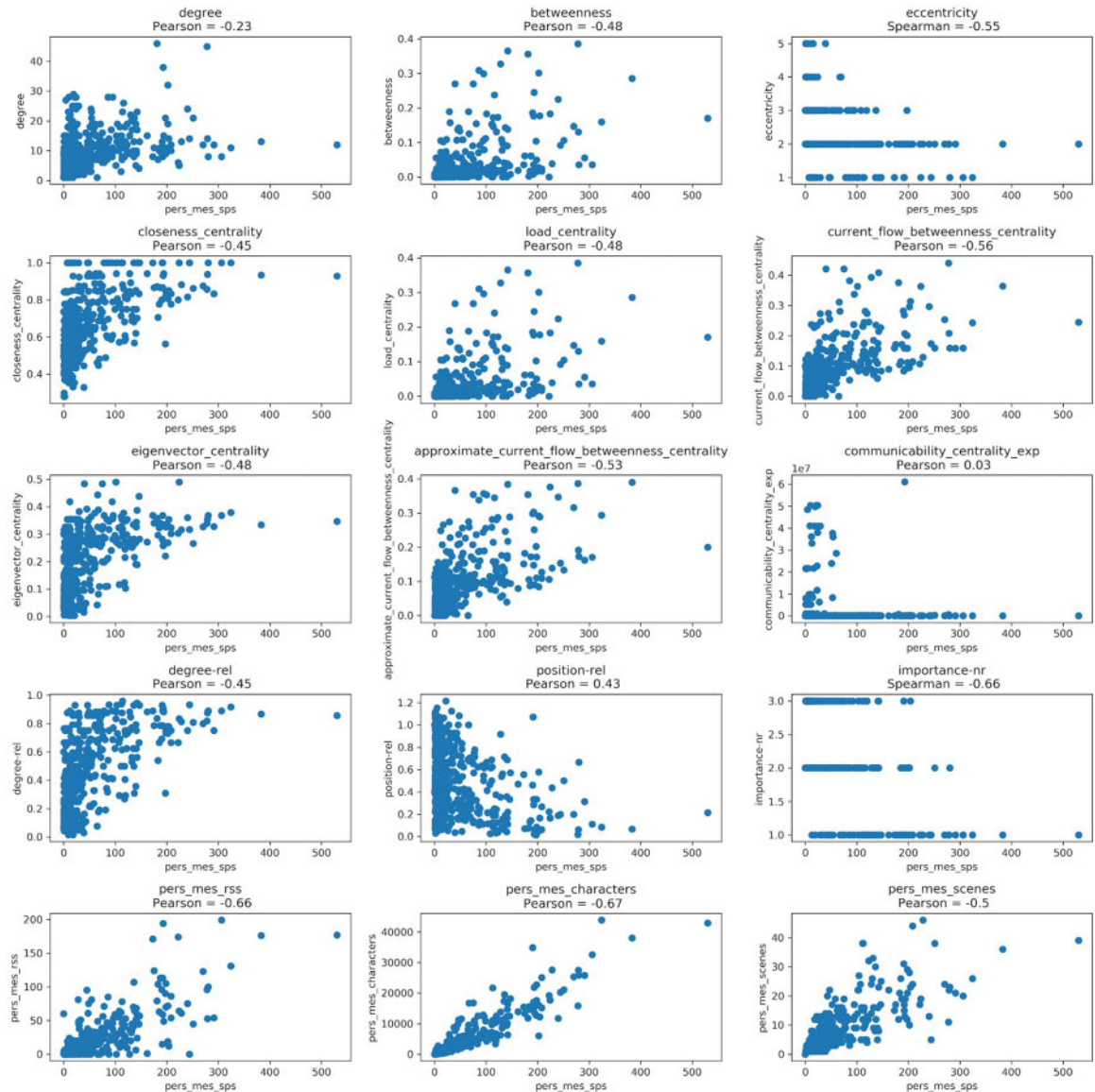


Fig. 4 Scatterplots (gráficos de dispersión) y correlación entre las veces que un personaje habla <sp> y otros valores

Se observa que la *information centrality*, de nuevo, es la medida de centralidad con la correlación más fuerte, junto con la cantidad de intervenciones, mientras que el resto de las medidas de centralidad tienen correlaciones similares a las del anterior gráfico. Es decir, no tenemos razones para pensar que los valores textuales y las medidas de centralidad estén captando

aspectos diferentes sobre la importancia de los personajes. Finalmente, nos hemos planteado si la distribución de centralidad o valores textuales son diferentes para los protagonistas de los del resto de personajes, que quedan reflejados en la [Figura 5](#):

La mayor diferenciación de ambos diagramas de caja entre las medidas de centralidad se consigue

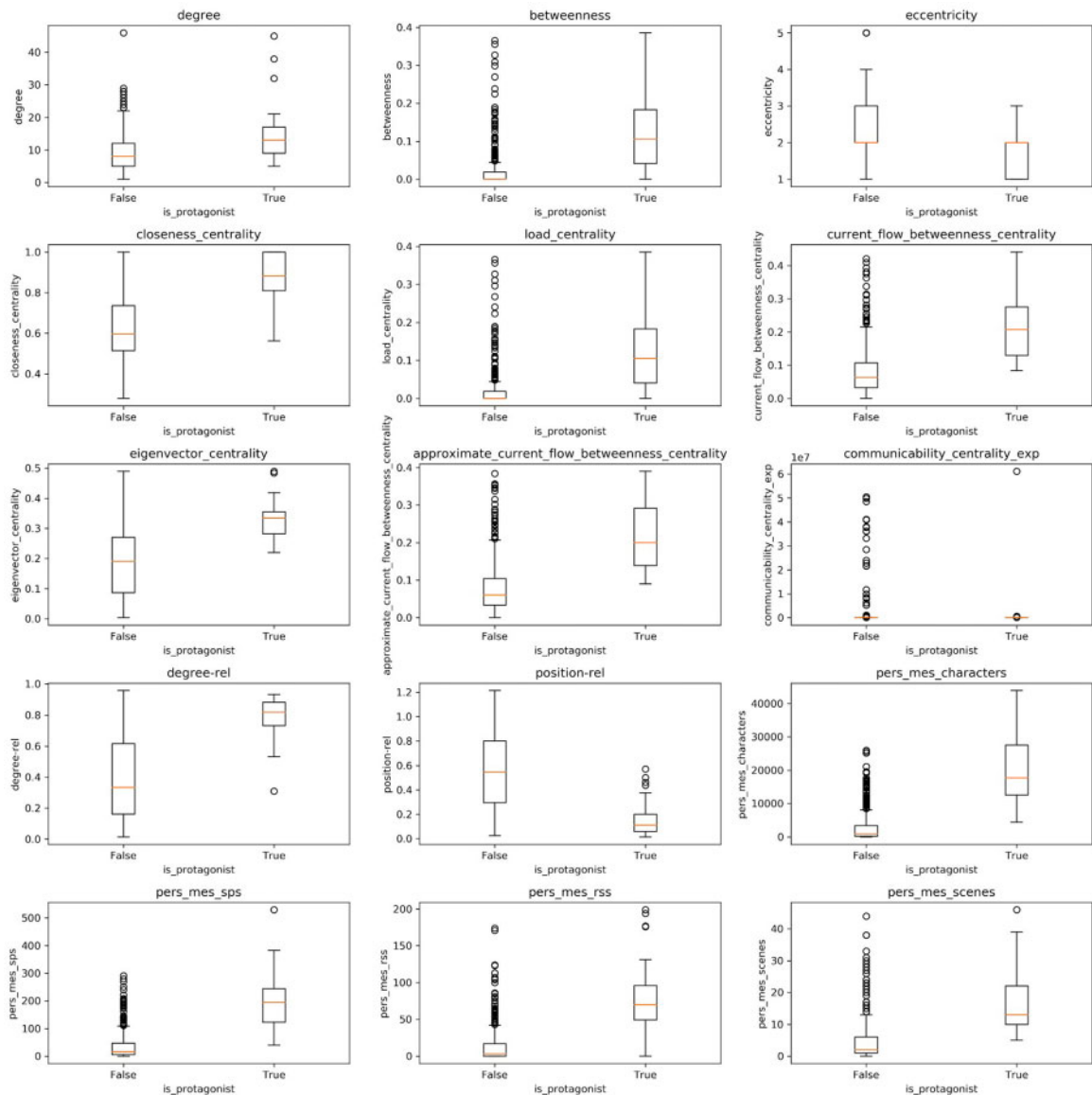


Fig. 5 Boxplots (diagramas de caja) de protagonistas frente al resto de personajes

mediante *approximate current flow betweenness centrality*, seguido de *current flow betweenness (o information) centrality*. El solapamiento menor se consigue mediante la cantidad de texto pronunciado (*pers_mes_characters*).

5 Conclusiones y futuros pasos

Coincidimos con la propuesta de Moretti (2013) respecto a la necesidad de distinguir tres tipos de personajes respecto a su relevancia. Ese tercer tipo de personaje - “*minor minor*”, para Moretti y “*minor*” o “terciario” dentro de nuestro análisis – implica una relevancia más baja que la del protagonista, pero también del personaje secundario o “*secondary*” y que coincide, en algunos casos, con los actores figurantes del cine o la televisión. Además, comparamos la idoneidad del género dramático para este análisis de redes sociales, pues el diálogo y el cómputo directo de las intervenciones de los personajes permite trabajar con los datos y metadatos de manera más fiable y sencilla que en otro tipo de obras, como puede ser la novela.

La anotación en detalle de información sobre los protagonistas nos permite no solo visualizar y explorar los datos con medios digitales, sino también evaluar esos mismos métodos digitales. Las respuestas a las cuatro preguntas arriba planteadas, para el caso del corpus BETTE, quedan, por tanto, contestadas de esta forma:

- (1) La importancia tiene una correlación solamente entre débil y moderada con cualquier formalización de centralidad de redes sociales, teniendo la correlación más fuerte la *information centrality*, como hemos podido comprobar.
- (2) La posición en el *dramatis personae* tiene una correlación moderada con la importancia, similar a la que muestran las medidas de centralidad. Nos resulta sorprendente este dato pues, salvo en las obras de Muñoz Seca, donde los personajes femeninos preceden a los masculinos, en el resto de las estudiadas la lista de personajes sigue el criterio de relevancia dentro de la obra, para el autor, y no el orden alfabético o de aparición de los personajes en la pieza.
- (3) Los valores cuantitativos textuales tienen correlaciones más fuertes que el resto de

valores, algo sorprendente si se tiene en cuenta su sencillez a la hora de extraer dicha información.

- (4) Los valores cuantitativos textuales son también los que mejor distinguirían a los personajes entre protagonistas y no protagonistas (ya sean secundarios o terciarios)

El resultado más relevante de este trabajo estriba en que unidades textuales (tan sencillas como la cantidad de parlamentos de cada personaje) aportan más información sobre la importancia o protagonismo del personaje que las complejas medidas de la centralidad en redes, que resumen la interacción entre los personajes. Aun así, nos preguntamos si otros parámetros como la formalización de las aristas no podría arrojar resultados más beneficiosos para las medidas de centralidad. Por otro lado, el estudio no implica que los estudios con grafos puedan proporcionar otros datos interesantes para responder otras preguntas, como, por ejemplo, respecto a la estructura de la obra, un cambio en la tradicional visión de los papeles principales (cf. Martínez Carro, 2019) o el estudio por sexos de los personajes dentro de este corpus, que también están anotados.

Por último, como propuestas a futuro, querríamos indicar que, aunque el tamaño del corpus se asemeja a otros trabajos en redes sociales (cf. Moretti, 2011 y 2013; Rochat, 2014), nos gustaría ampliar el corpus y estamos ya trabajando en esa línea. Por otro lado, nos parece interesante comprobar si el comportamiento entre distintos géneros - dramas, tragedias y comedias - obedece a unas coordenadas comunes; o bien, matizar más la investigación por periodos (dramaturgia publicada en el siglo XIX, por un lado, y por el XX, en otra). Otra cuestión interesante puede ser analizar obras de un mismo autor, en la línea de los dos trabajos de Moretti antes comentados (2011 y 2013), donde el centro de interés se fija en menos textos, pero el análisis más pormenorizado permite visibilizar más y mejor la relación entre la trama y los personajes.

Financiación

This work was supported by the Universidad Internacional de La Rioja.

Referencias

- Algee-Hewitt, M. A.** (2017). *Distributed Character: Quantitative Models of the English Stage, 1500-1920*. Montréal: McGill University & Université de Montréal, pp. 119–21.
- Brandes, U. and Fleischer, D.** (2005). *Centrality Measures Based on Current Flow. Theoretical Aspects of Computer Science (STACS '05)*. Springer-Verlag, pp. 533–44. <http://www.inf.uni-konstanz.de/algo/publications/bf-cmbcf-05.pdf>.
- Espejo, J.** (2002). Algunos aspectos sobre la construcción del personaje en el teatro conservado de Hernán López de Yanguas (1487-¿?). *Scriptura*, 17: 113–32.
- Evans, J. D.** (1996). *Straightforward Statistics for the Behavioral Sciences*. Pacific Grove: Brooks/Cole Pub. Co.
- Gómez, S., Calvo Tello, J., González, J. M. and Vilches, R.** (2015). Hacia una biblioteca electrónica textual del teatro en español de 1868-1936 (BETTE). *Texto Digital*, 11(2): 171–84.
- Hermenegildo, A.** (1995). Personaje y teatralidad: la experiencia de Juan del Encina en la *Égloga de Cristino y Feba*. In Pedraza F.B., y González R. (eds), *Los albores de teatro español: actas de las XVII Jornadas de teatro clásico Almagro, julio de 1994*. Almagro: Universidad de Castilla-La Mancha, pp. 90–113.
- Herrero, R.** (2000). La terminología del análisis de redes. Problemas de definición y de traducción. *Política y sociedad*, 33: 199–206.
- Isasi, J.** (2017). Acercamiento al análisis del sistema de los personajes en la narrativa escrita en español: el caso de *Zumalacárregui y Mendizábal* de Pérez Galdós. *Caracteres* 6 (2): 107–37.
- Jannidis, F., Kohle, H. and Rehbein, M.** (eds). (2017). *Digital Humanities: eine Einführung*. Stuttgart: J.B. Metzler Verlag.
- Jiménez, C., Martínez Carro, E., Santa María, M. T., Calvo Tello, J., Simón Parra, M., Martínez Nieto, R. B. and García Sánchez, M.** (2017). BETTE: Biblioteca Electrónica Textual del Teatro en Español de la Edad de Plata. *Sociedad, Políticas, Saberes*. Málaga: HDH, pp. 88–91. <http://hdh2017.es/wp-content/uploads/2017/10/Actas-HDH2017.pdf>.
- Marcus, S.** (1973). *Mathematische Poetik*. (Trans.) Mándroiú, E. București; Frankfurt/Main: Editura Academiei; Athenäum Verlag.
- Martínez Carro, E.** (2019). Aproximación al teatro lorquiano desde la teoría de las redes sociales: *La casa de Bernarda Alba*. *Artnodes*, 24: 134–41.
- Menéndez, C.** (1983). *Introducción al teatro de Benito Pérez Galdós*. Madrid: CSIC.
- Moretti, F.** (2011). Network Theory, Plot Analysis. *The New Left Review*, (68): 80–102.
- Moretti, F.** (2013). “Operationalizing”: or, the function of measurement in modern literary theory. *The New Left Review* (84): 103–19.
- Oleza, J.** (2013). *Biblioteca Digital Arte Lope*. Valencia: Universitat de València. artelope.uv.es/biblioteca.
- Rochat, Y.** (2014). *Character Networks and Centrality*. N.p. Web.
- Rodríguez, D.I.** (2016) *Análisis de grafos en paralelo mediante Graphx*. Trabajo de titulación. Universidad Católica de Loja. Ecuador.
- Stephenson, K. and Zelen, M.** (1989). Rethinking centrality: Methods and examples. *Social Networks*, 11(1): 1–37. doi:10.1016/0378-8733(89)90016-6.
- Trilcke, P., Fischer, F., Göbel, M. and Kampkaspar, D.** (2016). Dramen als small worlds? Netzwerkdaten zur Geschichte und Typologie deutschsprachiger Dramen 1730-1930. In Burr, E. (ed.), *DHD 2016 Modellierung, Vernetzung, Visualisierung*. Leipzig: DHd/nisaba, pp. 254–7.
- Trilcke, P., Fischer, F. and Kampkaspar, D.** (2015). Digitale Netzwerkanalyse dramatischer Texte. *DHD-Tagung*. Graz. <http://gams.uni-graz.at/o:dhd2015.v.040>.

Notes

- 1 En algunas obras, algunos personajes hablan y aparecen en escena exclusivamente de manera conjunta, funcionando como un personaje coral. Para este tipo de casos, utilizamos el elemento personGrp. Como ejemplo podemos mencionar el grupo de modernistas de la obra *Luces de bohemia*, de Valle-Inclán.
- 2 El listado de especialistas puede consultarse junto a las obras en el repositorio. <https://github.com/GHEDI/BETTE>.
- 3 Utilizamos la terminología en inglés mencionada en la documentación de la biblioteca referenciada, pero aportamos la traducción ya asentada de algunos términos al español (Martínez Carro 2019 y Herrero, 2000).
- 4 Los datos, el código (como Jupyter Notebook) y todas las visualizaciones pueden consultarse en el repositorio GitHub específico para este artículo, a través del siguiente enlace: https://github.com/GHEDI/BETTE/tree/master/publicacions/2018_DSH/ipybn_checkpoints.