



CRÍTICA DE LIBROS / BOOK REVIEWS

Gephi Cookbook

Devangana Khokhar

Birmingham, Packt publishing, 2015. 274 pags. ISBN 978-1-78398-740-5

Gephi y su contexto

El 31 de Julio de 2008, Mathieu Bastian lanzó por primera vez Gephi, una plataforma para la visualización interactiva y la exploración de todo tipo de redes, sistemas complejos y grafos dinámicos y jerárquicos. Junto a sus colaboradores Sébastien Heymann y Mathieu Jacomy la habían desarrollado para "importar, exportar, manipular, analizar, filtrar, representar, detectar comunidades y exportar grandes grafos y redes" (Bastian y otros, 2009). El propio Sébastien Heymann (2014) ofrece una descripción más actual y completa de la aplicación, cuya versión 0.9 se lanzó el pasado 21 de diciembre.

Gephi surgió en un panorama claramente dominado por aplicaciones mono-plataforma, parciales, desequilibradas y con una curva de aprendizaje realmente abrupta. Dos muy populares, Pajek y Ucinet/Netdraw, están limitadas al entorno Windows y se inclinan hacia el análisis de grafos (Pajek) o hacia la visualización (Ucinet/Netdraw, GUESS). NodeXL es la abreviatura de Network Overview Discovery Exploration for Excel; NetworkX es una biblioteca de Python e iGraph un módulo de R, mientras que NetLogo es un entorno de modelado. Entre las restantes aplicaciones destacan las destinadas a la exploración y visualización de redes de datos bibliométricos: HistCite, SciMAT (Cobo y otros, 2012) VOSviewer y CitNetExplorer (van Eck, 2014) CiteSpace (Chen, 2014) o Science of Science (SCI2) tool. Con la aparición de Bibliotools, achacar a Gephi la falta de una funcionalidad específica para el procesamiento de datos bibliométricos (van Eck 2014) ha dejado de tener sentido. Gephi es muy útil y cada vez más utilizado en el análisis y visualización de redes sociales y en bibliometría. De su popularidad dan cuenta algunos datos sencillos: de las 246 descargas de la primera versión

se ha pasado a un total de 1.272.293 descargas en septiembre de 2015. A pesar de que aún está en fase beta, ha recibido más de 630 citas en el WoS; además, 877 de los documentos de Scopus contienen el término "Gephi" en sus referencias bibliográficas. En palabras de Chaomei Chen (2014) "Gephi has become competitive even to the most mature and widely used software available from the earlier generations such as Pajek".

En España, tanto el grupo Scimago como el grupo SCI2S han producido interfaces gráficas propias para la visualización de datos bibliométricos, pero son excepciones en un panorama muy modesto. Casi 200 trabajos de investigación españoles incluidos en Scopus citan (han utilizado) Pajek; de ellos, 116 se han publicado desde 2011. La misma fuente ofrece 48 trabajos con contribución española publicados desde 2011 que citan Gephi. La rutina (García et al. 2015) y el recurso a aplicaciones muy especializadas para proyectos concretos (Nualart-Vilaplana y otros, 2014) pueden ser causas de ese desequilibrio pero, cuando se trata de un conjunto de aplicaciones, de una nueva plataforma en continuo desarrollo. Hay otra razón de peso: la falta de información y de formación.

La webografía y la bibliografía sobre Gephi

El problema con Gephi, el mismo que afecta a otras aplicaciones y plataformas colaborativas, se ha puesto de manifiesto de forma explícita (Rees, 2010): "The only thing keeping Gephi from being a professional grade tool is its bugginess and lack of documentation. If its development community can sustain it, Gephi should be able to improve and become a more robust software tool". En efecto, la naturaleza abierta puede resultar una desventaja para algunas aplicaciones en relación a su docu-

mentación y a su soporte. El grupo central de desarrolladores de Gephi y la comunidad de usuarios más activos, a pesar de su continuada producción de tutoriales, no bastan para abastecer a la muy heterogénea comunidad de usuarios y sus interrogantes, a pesar de que es precisamente la interacción entre unos y otros lo que favorece la continua producción de módulos y complementos, propios de una aplicación tan extensible.

Hasta hace poco, el conocimiento sobre el uso de Gephi dependía más de presentaciones y videos (Levallois 2014) entradas en blogs (Hirst 2010; Jacomy 2015) y artículos en wikis (<https://github.com/gephi/gephi/wiki>) que de uno o más textos "canónicos". Ante la falta de un manual oficial, se necesitaban obras que ofrecieran una descripción de las características de Gephi y un buen relato de su funcionamiento. Y esas obras llegaron.

El primer libro comercial sobre Gephi apareció en Septiembre de 2013 (Cherven, 2013). A pesar de escribir con fluidez y cierta espontaneidad, su autor, un analista de mercados de General Motors, abusaba de la llamada a la experimentación, al "prueba tu mismo". Este recurso ocultaba la falta de sistematización, de un plan para el desarrollo de la obra que, así, contenía muchas más omisiones que aciertos y parecía picotear nerviosamente aquí y allá, a lo largo de la interfaz de la aplicación (Amat, 2014). Quizá se trataba de una primera toma de contacto puesto que el mismo autor anunció, y finalmente ha publicado en enero de este mismo año, una nueva entrega en apariencia más elaborada. Casi coincidiendo con el segundo libro de Cherven, ha aparecido el "recetario" sobre Gephi. Su autora, Devangana Khokhar, es una ingeniera informática tan amiga de los grandes números (big data) como de las grandes distancias: es consultora de ThoughtWorks en Illinois y miembro del capítulo en Bangalore de Women Who Code. Su experiencia es palpable a lo largo de todo el libro.

Las recetas con Gephi

Los recetarios sobre aplicaciones informáticas se han popularizado. Su organización es casi invariable: una serie de unidades de estructura homogénea que representan operaciones y se agrupan por su finalidad en secciones o capítulos. En el caso de este recetario con Gephi, comprenden:

- Los preparativos (getting ready, los ingredientes)
- Las operaciones (how to do it..., la receta propiamente dicha)
- Los detalles (how it works, la descripción de los ingredientes)

- Para saber más (there's more..., variaciones y operaciones adicionales)
- Ver también (see also, reenvíos a otros capítulos, webliografía y referencias a otros materiales).

Las unidades se organizan después en capítulos que definen operaciones, desde el inicial sobre la instalación y la descripción de la interfaz básica del programa hasta el octavo, dedicado a un repaso de las redes dinámicas y las redes jerárquicas, con el añadido de dos capítulos finales dedicados a los conjuntos de datos ya elaborados y las extensiones de la aplicación. Los repaso a continuación.

Lo que en el libro de Cherven era un capítulo inicial incompleto, con olvido de elementos centrales en la interfaz (Amat, 2014) aquí se resuelve de manera muy ordenada y lógica, con la presentación de la red que va a acaparar la exploración de casi todas las operaciones que el libro presenta, la popular red que conecta los personajes de "Los Miserables". A partir de su carga, se describen someramente el laboratorio de datos, la pantalla de previsualización y la interfaz de trabajo (llamada vista general en la versión española) con sus módulos asociados.

El segundo capítulo, que contiene 14 recetas, se inicia con una sobre la generación de grafos aleatorios que parece estar fuera de lugar, y cuya utilidad se despacha con referencias al modelo original de Erdős y Rényi y a su revisión por Newman (2002). Salvando esta primera unidad, el capítulo ofrece un repaso muy completo de las posibilidades que Gephi ofrece para la exploración visual de una red, sus vértices (nodos) y sus aristas (o arcos, conexiones). Las recetas aparecen organizadas y en un orden lógico que responde al título de esta primera serie de operaciones, "Basic Graph Manipulations".

La representación gráfica y el análisis estadístico

Los capítulos del tres al cinco presentan las demás operaciones que Gephi puede realizar sobre las redes que explora. La presentación de los algoritmos para la representación de las redes, en el capítulo tres, no se limita a relacionar los principales. Además, la autora ofrece la suficiente información adicional en cada apartado "there's more" como para entender el mecanismo de representación. Hay un exceso de detalle en algunos casos; por ejemplo, cuando se dedican recetas diferentes al algoritmo de rotación en el sentido de las agujas del reloj y al correspondiente a la rotación en sentido inverso, al igual que también se separan los algoritmos de contracción y expansión de las gráficas. Y al contrario, no se menciona OpenOrd (Martin y otros, 2011) uno de los algoritmos más rápidos desarrollados e incorporados a Gephi re-

cientemente. Destaca, eso sí, en este capítulo las referencias externas a las descripciones completas de los algoritmos de representación.

El cuarto capítulo quizá se debería haber aplazado hasta después de presentar las opciones de análisis cuantitativo de las redes. Como está dedicado a la exploración estructural y ésta se basa en la representación de los atributos de nodos y aristas, además de su agrupamiento, se recurre a atributos cuantitativos (grado, modularidad) que se anticipan al capítulo siguiente. En todo caso, la red de protagonistas de *Los Miserables*, que tiene formato gexf (lo comento más adelante) ya incorpora como atributo la clase de cada nodo y algunas de sus propiedades de visualización como el tamaño, el color y la posición. También el valor de cada enlace. El capítulo presenta las posibilidades de manipulación de los colores y tamaños (o grosores) de nodos y aristas, así como el etiquetado de unos y otras en función de sus atributos.

La métrica de las redes se trata a lo largo del capítulo cinco. Las propiedades estadísticas de los nodos, las aristas y la red en su conjunto se presentan en 14 recetas que, si bien no superan el nivel descriptivo de las anteriores, ofrecen con sencillez las definiciones de algunos de los parámetros que se presentan. El capítulo contiene algunas imprecisiones sobre el PageRank, algunos descuidos sobre el cálculo de la modularidad o del valor de HITS y concede poca relevancia al coeficiente de agrupamiento (clustering coefficient, transitivity) pero es el más instructivo de toda la obra. Y ello a pesar de que no explora el hecho de que cualquier cálculo sobre nodos y aristas se traduce automáticamente en la aparición de nuevas columnas en el laboratorio (tabla) de datos. Y tampoco se refiere a que cualquier filtro de la red también se refleja allí.

Estos efectos se podrían haber comentado en el sexto capítulo, "Working in the Data Laboratory Mode" pero la autora lo estructura con una receta introductoria sobre la importación de datos de red en formato tabular seguida de otras 10 que repasan las operaciones de manipulación de nodos y relaciones y sus atributos. Apenas se comenta las repercusiones que estas operaciones tienen en la topología de la red, vista a través de la interfaz de trabajo. De este modo, se pierde la posibilidad de presentar las opciones de importación de archivos de red, que son muy flexibles en Gephi y justifican su interoperabilidad con otras aplicaciones.

El capítulo dedicado a la visualización previa de las redes trazadas es un mero repaso a los diseños preinstalados en la aplicación y ni siquiera contiene una descripción de los controles para la preparación de la imagen.

Las redes dinámicas y el formato gexf

En el octavo capítulo, se tratan las redes dinámicas en Gephi, pero con demasiada brevedad y con excesiva simpleza. Utilizando el generador de redes dinámicas de Gephi, se exploran su visualización animada y la representación gráfica de algunos valores y atributos en las celdas correspondientes del laboratorio de datos. Pero no se explora el proceso para generar e importar (o exportar) redes dinámicas ni tampoco la estructura del formato que permite generarlas. Una de las mayores fortalezas de Gephi es el Graph Exchange XML Format (gexf) que sus mismos creadores desarrollaron a partir de 2007. La plasticidad de este formato es extraordinaria, soporta tanto las redes estáticas como las dinámicas y sirve de soporte al intercambio con otros programas (no con Pajek) al tiempo que permite el diseño de una red con el mismo procesador de textos con el que se crea cualquier archivo XML. Habida cuenta del auge que experimenta el análisis de redes dinámicas, tanto las redes como el formato que las describe hubieran merecido una presentación más completa, como la que les dedica en su segunda obra Cherven (2015).

Notas finales

Este Gephi Cookbook se define como "una guía para el aprendizaje de la exploración y visualización interactivas de redes, acompañada de los conceptos de teoría de grafos que las soportan". Y se dirige a "aquellos que, sin experiencia previa, desean aprender tanto esas operaciones como los conceptos subyacentes".

La obra cumple con creces esos objetivos y es recomendable como un manual de introducción al uso de Gephi. Su lectura, combinada con la de una obra de divulgación como la breve introducción a las redes de Guido Caldarelli y Michele Catanzaro (2012) puede favorecer una toma de contacto realmente útil con el dominio del análisis de redes, sea cual sea su campo de aplicación.

Cualquier análisis de redes comienza con datos y esos datos se encuentran en soportes con diferentes formatos, sean tablas, bases de datos o meros ficheros planos. Incorporar esos datos a Gephi es la tarea inicial en cualquier proyecto que recurra a esa magnífica aplicación. Por esta razón se hace necesario completar la lectura de las recetas con el seguimiento de otros materiales, por ejemplo la presentación "Gephi focus on data import" (Levallois, 2014). Del mismo modo, para obtener una idea sobre el uso y análisis de tablas dinámicas es conveniente recurrir al reciente texto de Cherven (2015) o al tutorial de Eduardo Ramos (el desarrollador zaragozano de Gephi) sobre importación de datos dinámicos (Ramos, 2015).

Todos los recetarios, hasta los excelentes ejemplares de las colecciones de O'Reilly, se ven consuetudinados por sus propios esquemas y nunca resultan totalmente exhaustivos. Los olvidos de Devangana Khokhar son comprensibles. Su libro es recomendable y Gephi merece su lectura.

Carlos Benito Amat

Ingenio (CSIC-UPV)

REFERENCIAS

- Amat, C. B. (2014). Reseña de Network Graph Analysis and Visualization with Gephi by Ken Cherven. *Redes: revista hispana para el análisis de redes sociales*, 25(1), 201-209. <http://dx.doi.org/10.5565/rev/redes.499>
- Bastian, M.; Heymann, S.; Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. San Jose, Third International AAAI Conference on Weblogs and Social Media, 8, 361-362. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewFile/154/1009>
- Caldarelli, G.; Catanzaro, M. (2012). *Networks: a very short introduction*. Oxford University Press. La versión española (*Redes: Una breve introducción*. Madrid, Alianza) apareció en 2014. <http://dx.doi.org/10.1093/actrade/9780199588077.001.0001>
- Chen, C. (2014). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*, 2nd ed. New York: Springer. <http://dx.doi.org/10.1007/978-1-4471-5128-9>
- Cherven, K. (2013). *Network graph analysis and visualization with Gephi*. Birmingham, Packt Publishing Ltd.
- Cherven, K. (2015). *Mastering Gephi Network Visualization*. Birmingham, Packt Publishing Ltd.
- Cobo, M. J.; López-Herrera, A. G.; Herrera-Viedma, E.; Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, 63(8), 1609-1630. <http://dx.doi.org/10.1002/asi.22688>
- García, A. G.; Ibáñez, A. P.; Sapena, A. F.; Mancebo, M. F. P.; Moreno, L. M. G. (2015). Herramientas de análisis de datos bibliográficos y construcción de mapas de conocimiento: Bibexcel y Pajek. *BiD: Textos universitaris de biblioteconomia i documentació*, (34), 11. <http://dx.doi.org/10.1344/BiD2015.34.22>
- Levallois, C. (2014): Gephi Tutorials and Plugins. <http://www.clementlevallois.net/gephi.html> [consulta 15-9-2015].
- Heymann, S. (2014): Gephi. En Alhajj, R.; Rokne, J. (Eds.) *Encyclopedia of social network analysis and mining*. Springer. pp. 612-625. http://dx.doi.org/10.1007/978-1-4614-6170-8_299
- Hirst, T. (2010). Getting Started With The Gephi Network Visualisation App – My Facebook Network, Part I. Recuperado a partir de <http://blog.ouseful.info/2010/04/16/getting-started-with-gephi-network-visualisation-app-my-facebook-network-part-i/>
- Jacomy, M. (2015). Gephi blog. <https://gephi.wordpress.com/> [consulta 10-9-2015]
- Martin, S; Brown, W. M.; Klavans, R.; Boyack, K. W. (2011). OpenOrd: An Open-Source Toolbox for Large Graph Layout. En P. C. Wong y otros. (Eds.), *Proc. SPIE 7868: Visualization and Data Analysis*, p. 786806). San Francisco. <http://doi.org/10.1117/12.871402>
- Newman, M. E. (2002). Random graphs as models of networks. arXiv preprint <http://arxiv.org/abs/cond-mat/0202208v1> / <http://dx.doi.org/10.1002/3527602755.ch2>
- Nualart-Vilaplana, J.; Pérez-Montoro, M.; Whitelaw, M. (2014). Cómo dibujamos textos: Revisión de propuestas de visualización y exploración textual. *El profesional de la información*, 23(3), 221-235. http://www.elprofesionaldelainformacion.com/contenidos/2014/mayo/02_esp.pdf / <http://dx.doi.org/10.3145/epi.2014.may.02>
- Ramos, E. Import Dynamic Data. <https://github.com/gephi/gephi/wiki/Import-Dynamic-Data> [consulta 16-9-2015].
- Rees, K. (2010). Data Visualization Review: Gephi, Free Graph Exploration Software. Information Aesthetics. http://infosthetics.com/archives/2010/07/review_gephi_graph_exploration_software.html. [Consulta 10-9-2015].
- van Eck, N. J.; Waltman, L. (2014). Visualizing bibliometric networks. En Ding, Y.; Rousseau, R.; Wolfram, D. (Eds.) *Measuring Scholarly Impact: Methods and Practice*. New York: Springer. pp 285-320. http://dx.doi.org/10.1007/978-3-319-10377-8_13