

Análisis de redes y visualización con Gephi

Carlos B. Amat

INGENIO¹

Reseña

Cherven, K. (2013). *Network Graph Analysis and Visualization with Gephi*. Packt Publishing Ltd. 116 pág.

Gephi

El 31 de Julio de 2008, Mathieu Bastian lanzó por primera vez Gephi, que definía como 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 Sebastien Heymann y Mathieu Jacomy habían desarrollado Gephi para "importar, exportar, manipular, analizar, filtrar, representar, detectar comunidades y exportar grandes grafos y redes" (Bastian, Heymann, & Jacomy, 2009). Aquella primera versión (Gephi 0.6alpha1) alcanzó las 246 descargas; la última versión estable, Gephi.8.2beta, se ha descargado 311.742 veces desde su lanzamiento, en Diciembre de 2012, mientras el kit de desarrollo correspondiente ha alcanzado las 5019 descargas («timeline : Gephi», s. f.). Y siguen.

Gephi está construido sobre la plataforma NetBeans 7.0 y programado en Java y OpenGL. Se distribuye bajo una licencia dual GNU GPL v3 y Common Development and Distribution License (CDDL-1.0). Es, por tanto una aplicación de código abierto y multiplataforma (Bastian et al., 2009). Seguro que estas características son responsables de su éxito. Pero no de todo su éxito, hay más.

Gephi surgió en un panorama claramente dominado por aplicaciones mono-plataforma, desequilibradas y con una curva de aprendizaje realmente abrupta. Dos aplicaciones muy populares, Pajek² y Ucinet/Netdraw³, están limitadas al entorno

¹ Instituto de Gestión de la Innovación y el Conocimiento. CSIC y Universidad Politécnica de Valencia. Ciudad Politécnica de la Innovación, Campus de Vera, 46022 Valencia. cbenito@ingenio.upv.es

² <http://pajek.imfm.si/doku.php>

³ <https://sites.google.com/site/ucinetsoftware/home>



Windows y se inclinan hacia el análisis de grafos (Pajek) o hacia la visualización (Ucinet/Netdraw). NodeXL⁴ es la abreviatura de Network overview discovery exploration for Excel; NetworkX⁵ es una biblioteca de Python e iGraph⁶ un módulo de R, mientras que NetLogo⁷ es un entorno de modelado. No merece la pena mencionar otras aplicaciones, como EgoNet⁸, VOSviewer (Eck & Waltman, 2009) o sci2⁹ porque o son muy limitadas o excesivamente especializadas. Es posible que el gran competidor de Gephi sea Cytoscape¹⁰, en la medida en que se aparte de su origen como herramienta de análisis de procesos biológicos y se generalice. Una segunda característica que hace a Gephi destacado es que funciona en cualquier entorno y conecta con la mayoría de las aplicaciones mencionadas: en el campo del análisis y la visualización de redes, es una plataforma de propósito general. Sin embargo, comparte con los programas colaborativos, de código abierto, una debilidad que la acotación siguiente (Rees, 2010) ilustra:

“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 documentación y a su soporte. La comunidad de usuarios más activos, a pesar de su continuada producción de tutoriales, no basta 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.

Parece, entonces, que un texto que ofrezca una descripción de las características de Gephi y un relato de su funcionamiento ha de resultar oportuno y bienvenido. Y he aquí que Kenneth Cherven, un analista de mercados de General Motors, lo ha elaborado.

⁴ <https://nodexl.codeplex.com/Wikipage?ProjectName=nodexl>

⁵ <http://networkx.github.io/>

⁶ <http://cran.r-project.org/web/packages/igraph/index.html>

⁷ <http://ccl.northwestern.edu/netlogo/>

⁸ <http://sourceforge.net/projects/egonet/>

⁹ <https://sci2.cns.iu.edu/user/index.php>

¹⁰ <http://www.cytoscape.org/>

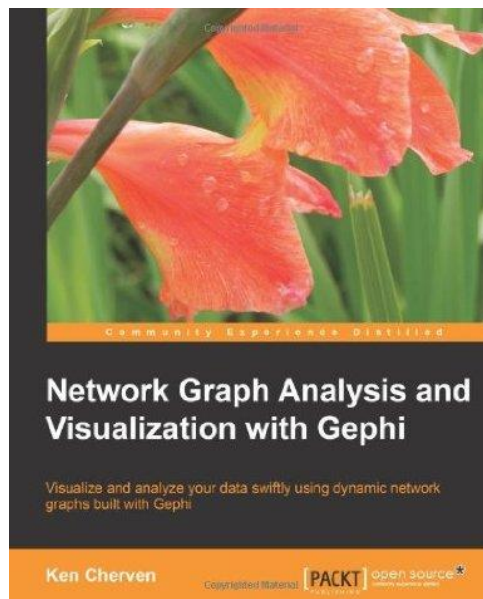


Gráfico 1. Cubierta del libro de Kenneth Cherven

El contexto del libro de Cherven

Packt es una editorial inglesa especializada en la publicación de libros sobre tecnologías de la información, tutoriales en video y otros materiales para programadores, administradores y usuarios de aplicaciones informáticas. Su actividad editorial pasa por la impresión bajo demanda de ejemplares de obras normalmente breves. De hecho, mientras el libro de Cherven ocupa 99 páginas numeradas, el título de la misma editorial dedicado a Cytoscape se queda en 76.

El libro se define como una guía práctica que dota al lector de todas las herramientas necesarias para visualizar y analizar sus datos usando redes generadas con Gephi y dice estar dirigido a cualquiera que pretenda explorar interacciones entre conjuntos de datos interrelacionados.

Volvamos por un momento a referirnos a la creciente comunidad de usuarios de Gephi. Sus foros, wikis y grupos revelan una maraña de estudiantes, investigadores en ciencias físicas, biológicas y sociales y una informe cantidad de simples curiosos entremezclados con profesionales con habilidades matemáticas. Hasta el momento todos han contado con muy poco apoyo para sacar partido de la potencia de Gephi. Un rápido inventario de recursos accesibles públicamente podría incluir algunos tutoriales “oficiales”¹¹ junto con otras valiosas presentaciones del Atelier Iceberg

¹¹ <https://gephi.org/users/>

distribuidos en SlideShare. Además, tanto en YouTube como en Vimeo se distribuyen videos (2570 y 80 respectivamente en Febrero de 2014) de origen académico o elaborados por usuarios avanzados de Gephi. La transcripción del seminario que Martin Grandjean (Grandjean, 2013) impartió el pasado año en la Digital Humanities Summer School en Berna, el curso de Nader Afzalan¹² en la plataforma Planetizen de acceso por suscripción y los cursos, también de pago, organizados por el mencionado Atelier Iceberg¹³ parecen ser los únicos relatos completos sobre Gephi hasta el momento, demasiado breves y limitados como para remediar la escasez de recursos de aprendizaje comprensivos sobre Gephi.

El contenido del libro

La forma más sencilla de introducir a los recién llegados al uso de una aplicación es organizar una guía paso-a-paso. ¿Es el caso del libro de Cherven? La obra se estructura en siete capítulos, un apéndice y un índice temático.

El primer capítulo es una típica introducción que describe el proceso de descarga, instalación y eventual descarte de la aplicación y, así, ocupa varias páginas reproduciendo en gran formato los sucesivos pantallazos del proceso. A pesar del carácter multiplataforma de Gephi, la descripción se limita al entorno Windows. Por otra parte, este capítulo inicial ofrece una descripción de la interfaz de usuario de Gephi, algo que merecería un capítulo completo en sí mismo y que hace que el título de esa primera parte "Installing Gephi" resulte incompleto. Además, la descripción de la interfaz dista mucho de ser completa o alcanzar un grado suficiente de detalle: de los tres componentes- vista general, laboratorio de datos y previsualización- sólo se realiza un recorrido por sus controles que se aplican al diseño del grafo. El recorrido es tan rápido que se limita a transcribir los textos explicativos que la aplicación despliega al deslizar el puntero sobre los controles. Además, no los presenta todos. Este capítulo inicial también olvida el resto de los elementos. Ni el bloque de estadísticas y filtrados; ni el bloque de particiones y ordenaciones, ni el de representación. Por extraño que parezca, esta descripción de la interfaz tampoco incluye las entradas del menú principal del programa. Pero es que tampoco se mencionan en el resto del libro!

¹² <http://courses.planetizen.com/course/gephi-network-analysis>

¹³ <http://ateliericeberg.fr/index.php?p=formations2>

A estas carencias cabe añadir otro problema de enfoque. La forma más sencilla de presentar una interfaz es ofrecer el desarrollo del análisis de un caso simple. Las presentaciones y videos recurren a veces al más querido por los creadores de Gephi: el de los personajes de Los Miserables y su aparición conjunta en escenas del folletón de Victor Hugo. Esta red, como la de las alianzas entre familias florentinas de Pajek, es simple, pero lo bastante completo como para ofrecer una rápida visión de las funcionalidades de ese componente de la interfaz.

Si el lector a quien se dirige este primer capítulo es tan novato como para tener que recibir una explicación de cómo se instala una aplicación, también debería de contar con información básica sobre las redes y su análisis.

Hay que esperar al segundo capítulo, "Creating Simple Network Graphs" para generar un primer gráfico partiendo de la utilización de un conjunto de datos y para obtener una descripción más detallada de la interfaz. La elección de los datos no ha podido resultar más desafortunada porque se sugiere un archivo en formato Pajek (.net) que ni siquiera cuenta con los nodos etiquetados, de modo que el "jugueteo" con los algoritmos de representación no surte mayor efecto que rellenar las páginas del capítulo con inexpresivas representaciones de las relaciones entre supuestos intérpretes de jazz. En cuanto al avance en la descripción de la interfaz, el autor no hace un solo intento de definir particiones o de ordenar los nodos o las relaciones y su exposición del laboratorio de datos se limita a enumerar las columnas de atributos, naturalmente idénticas, del archivo .net mencionado.

El hecho de no contar con un ejemplo provoca que la descripción del bloque de partición y ordenación ("particionamiento" y "clasificación" en la traducción al español de la interfaz) sea excesivamente parca, carezca de representación gráfica en el libro y escape también al mecanismo acción->efecto.

Otra vez se echan en falta las explicaciones básicas: ¿Cuál es la estructura de los archivos .net? ¿Qué otros tipos de archivos gráficos se pueden importar a Gephi? ¿Qué sentido tienen los controles que rodean el área del gráfico en la vista general?

Estas cuestiones tampoco se contestan en el siguiente capítulo, "Exploring Additional Layout Options". En él continua el tiovivo de figuras sin numerar y el repaso rápido de diferentes algoritmos de representación a la voz de "feel free to experiment" y se avanza de forma apresurada hasta dos extremos. El primero es el empleo de una red de afiliación como ejemplo; el segundo es un repentino llamamiento a la instalación de extensiones que no están contenidas en la instalación inicial y ofrecen más posibilidades de representación gráfica. Si Cherven hubiera explorado el menú principal de Gephi hubiera descubierto que la entrada

Tools>Plugins es mucho más directa y fácil para el usuario que el proceso en ocho pasos! que el texto propone para la instalación y actualización de complementos. El desfile desde el algoritmo de Fruchterman-Reingold hasta el OpenOrd no tiene, en este punto, mayor interés para el lector.

El cuarto capítulo, "Creating a Gephi Dataset" es el más importante del libro y el que debería situarse en cabeza, sólo precedido por la descripción de la interfaz. Por fin, bajo el apartado "Basic data requirements" el lector halla una sucinta descripción sobre nodos, aristas y sus propiedades y, en la sección siguiente, una explicación sobre el tamaño de los nodos y las aristas y una descripción del proceso para añadir manualmente nodos y aristas a través del laboratorio de datos. Ni en este capítulo ni en el resto de la obra se presenta una descripción o una mera noticia sobre el formato de intercambio habitual entre aplicaciones de análisis de redes, los archivos .net. En lugar de eso siguen dos secciones que describen la importación de datos procedentes de hojas de cálculo y de tablas MySQL.

El laboratorio de datos es un componente esencial de la interfaz y el manejo de Gephi. Sin embargo, ésta es la única ocasión en que se menciona en el libro y, así, la manipulación de columnas, su combinación y transformación, la edición de los valores de atributos, la agrupación de nodos, la exportación de tablas y otras tantas operaciones centrales en la exploración de los datos quedan en el más absoluto olvido. Resulta difícil averiguar por qué Cherven ni siquiera ha transcrito algunas de las páginas que el propio Wiki de Gephi contiene sobre este fundamental componente. En particular, la omisión del menú contextual de edición de las propiedades y relaciones de los nodos es grave, sobre todo porque determinadas operaciones (la agrupación, por ejemplo, combinada con la exportación a nuevos espacios de trabajo) son básicas en los análisis medianamente ambiciosos.

Del usuario lego del primer capítulo se ha pasado al usuario experto en el procesamiento de datos que, además, es capaz de detectar en el capítulo cinco los módulos necesarios para hacer más llevadero su análisis. Y ello en apenas cuatro capítulos.

A pesar de su título, "Advanced Features" el capítulo 6 presenta elementos básicos de exploración y manejo de las redes. Lo hace sin respetar el orden natural, ofreciendo primero un recuento muy incompleto de las operaciones de filtrado de datos según sus propiedades, después refiriéndose a la obtención de esas mismas propiedades mediante el panel de estadísticas y, finalmente, indicando el efecto sobre la representación gráfica del empleo de los valores obtenidos. En este capítulo sí se ofrece el efecto de las operaciones sobre el aspecto de la

representación gráfica de la red, aunque el ejemplo elegido (la red de afiliación de alumnos a sus clases otra vez) no es bueno y se utiliza la gradación de colores (o grises en el ejemplar impreso del libro) para mostrar una ordenación que hubiera resultado mucho más clara si se hubiera expresado por diferencias de tamaño de los nodos.

Cabría esperar de un capítulo así titulado la presentación de elementos y operaciones realmente avanzados, pero no. Las estadísticas que Gephi ofrece son muy completas (McSweeney, 2009) pero aquí se limitan a la presentación de algunos cálculos sobre la red general y no se menciona siquiera la posibilidad de exportar los informes gráficos generados (o la distribución completa si se recurre del laboratorio de datos) Tampoco se presenta la agregación o el colapso de nodos basándose en sus atributos y sus posibles particiones. El uso de diversos espacios de trabajo, ya mencionado, y el filtro de unión tampoco se tratan en este capítulo.

El último capítulo, "Deploying Gephi Visualizations" está centrado en el tercer componente de la interfaz de Gephi, la visualización previa, y describe la visualización y exportación de las imágenes obtenidas como un simple archivo gráfico o en formato HTML. Aunque sin presentarlo en figura, relata el manejo del bloque de configuración de la imagen y baraja algunos criterios para mejorar la visualización de las redes generadas.

Conclusiones

La obra de Cherven, escrita con fluidez y cierta espontaneidad, abusa sin embargo de la llamada a la experimentación, al "prueba tu mismo". Este recurso oculta la falta de sistematización, de un plan para el desarrollo de la obra que, así, contiene muchas más omisiones que aciertos y parece picotear aquí y allá a lo largo de la interfaz de la aplicación.

Ken Cherven no ha tenido en cuenta el exquisito equilibrio de Gephi y ha inclinado demasiado la balanza hacia la representación gráfica de las redes, en contra de lo que el propio título prometía: análisis y visualización. La insistencia en los algoritmos de representación a partir del final del segundo capítulo y durante toda la extensión del tercero y el quinto y el contenido del séptimo capítulo dan fe de ese sesgo, que va en detrimento del empleo de Gephi en la exploración cuantitativa de grafos.

El autor también se muestra errático en su consideración de los lectores. Un libro que declara estar dirigido a "personas que pretenden saber más de visualización de

redes en general y de Gephi en particular” (pag. 2) trata a sus lectores como analfabetos informáticos en el primer capítulo, les reenvía a las páginas de Yifan Hu (que en el ínterin ha cambiado de los laboratorios de AT&T a Yahoo) en el segundo y les ofrece la posibilidad de procesar datos procedentes de servidores de base de datos en el cuarto. El lector de una obra técnica, aunque sea tan breve como ésta, sigue una evolución página a página y Cherven no ha trazado una línea clara desde conceptos básicos a procedimientos avanzados.

La imagen de precipitación, casi de improvisación, que ofrecen los contenidos del libro se refuerza si se tienen en cuenta los descuidos formales. ¿Cómo es posible que ni una sola de las figuras esté numerada? ¿Por qué se ofrecen referencias de algunos algoritmos de visualización y no de otros?

Los usuarios de Gephi esperábamos un libro, pero no éste. Su precipitación, su descuido y su desorden lo rebajan desde manual introductorio no ya a guía paso a paso sino a meros apuntes. Kenneth Cherven ya ha anunciado la firma de un nuevo contrato con Packt Publishing para elaborar otra obra: esta vez sobre el uso avanzado de Gephi. Como buen estratega comercial, ha empezado por consultar públicamente sobre su contenido. Sería de desear que se tomase tiempo para elaborarla, al menos el tiempo necesario para dar paso a la versión 0.9 de Gephi, que parece próxima. Gephi merece algo mucho mejor que el libro que aquí se ha criticado.

Bibliografía

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi : An Open Source Software for Exploring and Manipulating Networks. Presentado en Third International ICWSM Conference, San Jose, California: Association for the Advancement of Artificial Intelligence. Recuperado a partir de gephi.org

Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. doi:10.1007/s11192-009-0146-3

Grandjean, M. (2013). Introduction to Network Visualization with GEPHI. *Martin Grandjean*. Recuperado a partir de <http://www.martingrandjean.ch/introduction-to-network-visualization-gephi/>

McSweeney, P. J. (2009). Gephi Network Statistics. Presentado en Google Summer of Code. Recuperado a partir de <http://gephi.org/google-soc/gephi-netalgo.pdf>

Rees, K. (2010). Data Visualization Review: Gephi, Free Graph Exploration Software. *Information Aesthetics*. Recuperado 1 de abril de 2014, a partir de http://infosthetics.com/archives/2010/07/review_gephi_graph_exploration_software.html

Timeline :Gephi. Recuperado 31 de marzo de 2014, a partir de <https://launchpad.net/gephi/+series>