

Análisis de colaboración en comunidades por medio de analisis de redes sociales

Palabras clave: SNA, participación, comunidades de práctica

Introducción

Las comunidades de práctica son espacios donde las personas comparten conocimientos y contribuyen a objetivos individuales y de grupo. Conocer los tipos de miembros y formas de participación, junto con las colaboraciones que existen, y entre quiénes, es un insumo importante para entender la comunidad y tomar acciones para su mejora. Presentamos como implementar un análisis de redes sociales utilizando R en una comunidad de práctica.

Mapeando la comunidad como una red

La comunidad que analizamos es rOpenSci. Las formas de participación que se detallan en una guía. Las redes tienen como elementos nodos y vínculos, por lo que necesitamos datos que muestren algún tipo de conexión. Para iniciar el análisis seleccionamos de la guía las contribuciones que se pueden hacer colaborando y dejamos aquellas con información pública. De esta manera podemos mapear contribuciones en todos los proyectos de rOpenSci.

Obteniendo los datos

Daremos un ejemplo con la colaboración escribiendo blog posts. La web se construye con blogdown, cada artículo tiene un YAML con la información de co-autoría necesaria. La tabla 1 presenta el código necesario para recuperar toda la información del sitio web.

Tabla 1. Recuperando los datos de todos los artículos del sitio web de rOpenSci

<pre>1 file_list <- fs::dir_ls(path = "content/blog/", 2 recurse = TRUE, 3 type = "file", 4 glob = "*.md") 5 6 datos <- tibble(fecha = character(), 7 titulo = character(), 8 autor = character(), 9 year = character(), 10 contribution_type = character()) 11 12 for (documento in file_list){ 13 doc <- rmarkdown::yaml_front_matter(input = file.path(documento)) 14 datos <- tibble::add_row(datos, 15 fecha = doc\$date, 16 titulo = doc\$title, 17 autor = doc\$author, 18 year = as.character(year(date(doc\$date))), 19 contribution_type = 'blog post' 20) 21 } 22 23 write_csv(datos, "blog_post_authors_2023.csv") # ;-)</pre>	<div>1. Lee todos los archivos de la carpeta `content/blog/` con extensión `.md` 2. Crea un tibble con las variables a almacenar: `fecha`, `titulo`, `autor`, `año` y `tipo_contribución`. 3. Para cada documento markdown 4. Lee la cabecera YAML, extrae el valor de cada variable 5. y añadir una fila en el conjunto de datos con la información</div>
--	--

Generando la Red

La tabla 2 se muestra el código de R explicado para formatear los datos como una red y realizar un análisis. Usamos el paquete tidygraph, igraph y ggraph en esta etapa.

Tabla 2. Dando formato de red a los datos

<pre> 3 blog_post_net <- blog_post > 4 group_by(title, year) > 5 filter(n() > 1) > 6 summarise(as.data.frame(t(combn(author, 2)))) > 7 select(title, year, from=V1, to=V2) 8 9 10 the_nodes <- blog_post %>% 11 select(author) %>% 12 unique() 13 14 the_edges <- blog_post_net %>% 15 select(from, to, year) 16 17 bp_net <- tbl_graph(nodes = the_nodes, 18 edges = the_edges, 19 directed = FALSE) </pre>	<ol style="list-style-type: none"> 1. Toma la lista que creamos en el paso anterior. 2. Agrupar por título y año y conservar todas las entradas de blog con dos o más autores. 3. Para cada grupo, la función <code>combn</code> crea una matriz con dos filas y columnas que representan todas las combinaciones únicas de dos autores. Transponemos los datos para obtener 2 columnas que se convierten en <code>_from_</code> y <code>_to_</code>, que representan los nodos. 4. Generamos un dataset con los nodos y otros con los vínculos. 5. Creamos un objeto tidygraph.
---	---

Resultados

Visualizar la red y calcular una serie de metricas es lo que nos va a dar información valiosa. En la figura 1 vemos la evolución de la red del 2014 al 2022. También obtuvimos los miembros con mayor grado, centralidad y page rank, identificando personas claves de la comunidad, por ejemplo, Maëlle Salmon con la mayor cantidad de contribuciones y Yanina Bellini con la mayor cantidad en español. Al realizar calculos a nivel de red también podemos conocer mejor la y planificar acciones para su gestión. Por ejemplo sabemos que tenemos 224 miembros escribiendo para el blog con más de 900 vínculos.

Esto mismo se puede repetir para otro tipo de contribuciones como community calls, libros, charlas, revisiones de software por pares, desarrollo de paquetes, programa de campeones y mentoría.

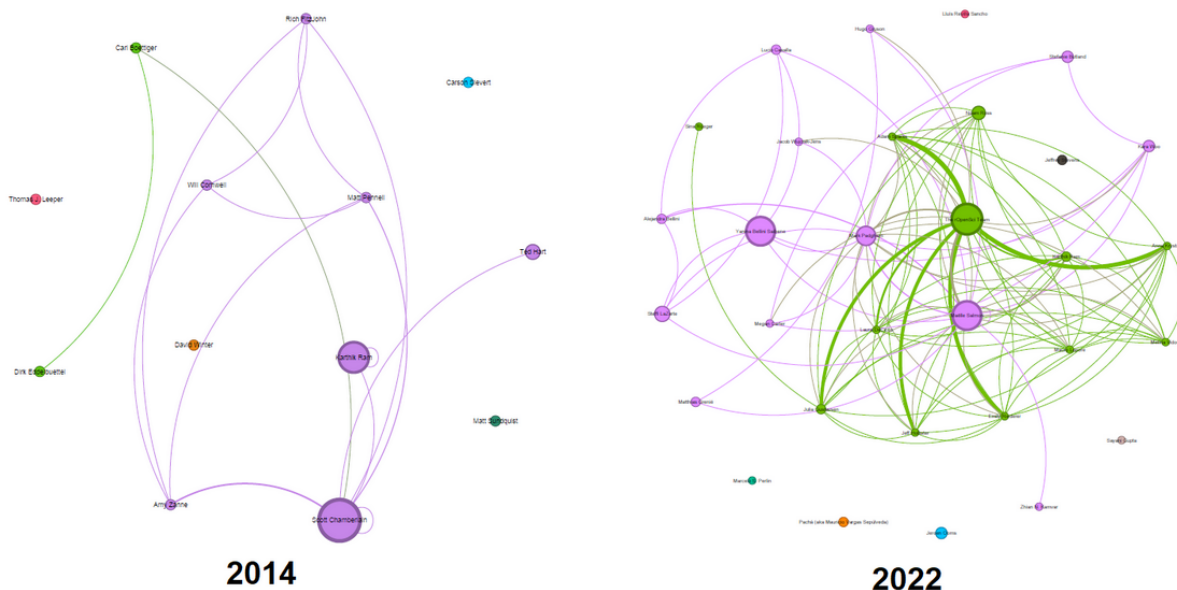


Figura 1. Ejemplo de visualización comparando dos años de autores de artículos de blog de rOpenSci