

# Network Analysis with Stack Overflow Tags

Leija Estrada Shanelly Abisay  
González Paz Edgar

Datos Masivos II  
Ciencia de datos

# Stack Overflow

- Foro que recopila preguntas y respuestas de la comunidad acerca de temas de programación.



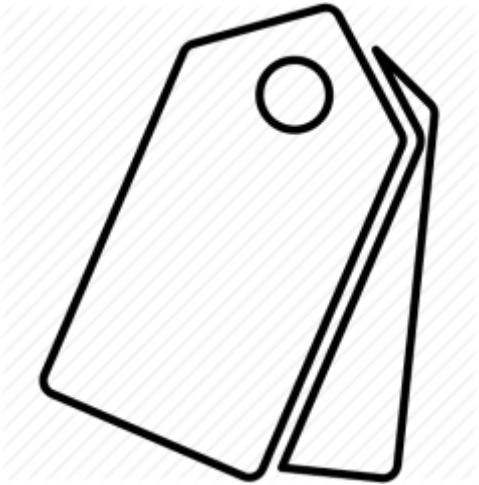
# Pregunta

- ¿Cuáles son los principales temas en los que la comunidad de Stack Overflow está interesada y cómo se relacionan entre sí?



# Metodología

- Análisis de las preguntas en el sitio, enfocándonos en las etiquetas/tags de las publicaciones



# Objetivo

- Ofrecerle al usuario patrones o áreas, según los temas o herramientas de su dominio



# Obtención de los datos

- Web Scrapping en el sitio: filtro sobre las preguntas más votadas de todos los tiempos
- Total de 20 páginas con 50 preguntas cada una



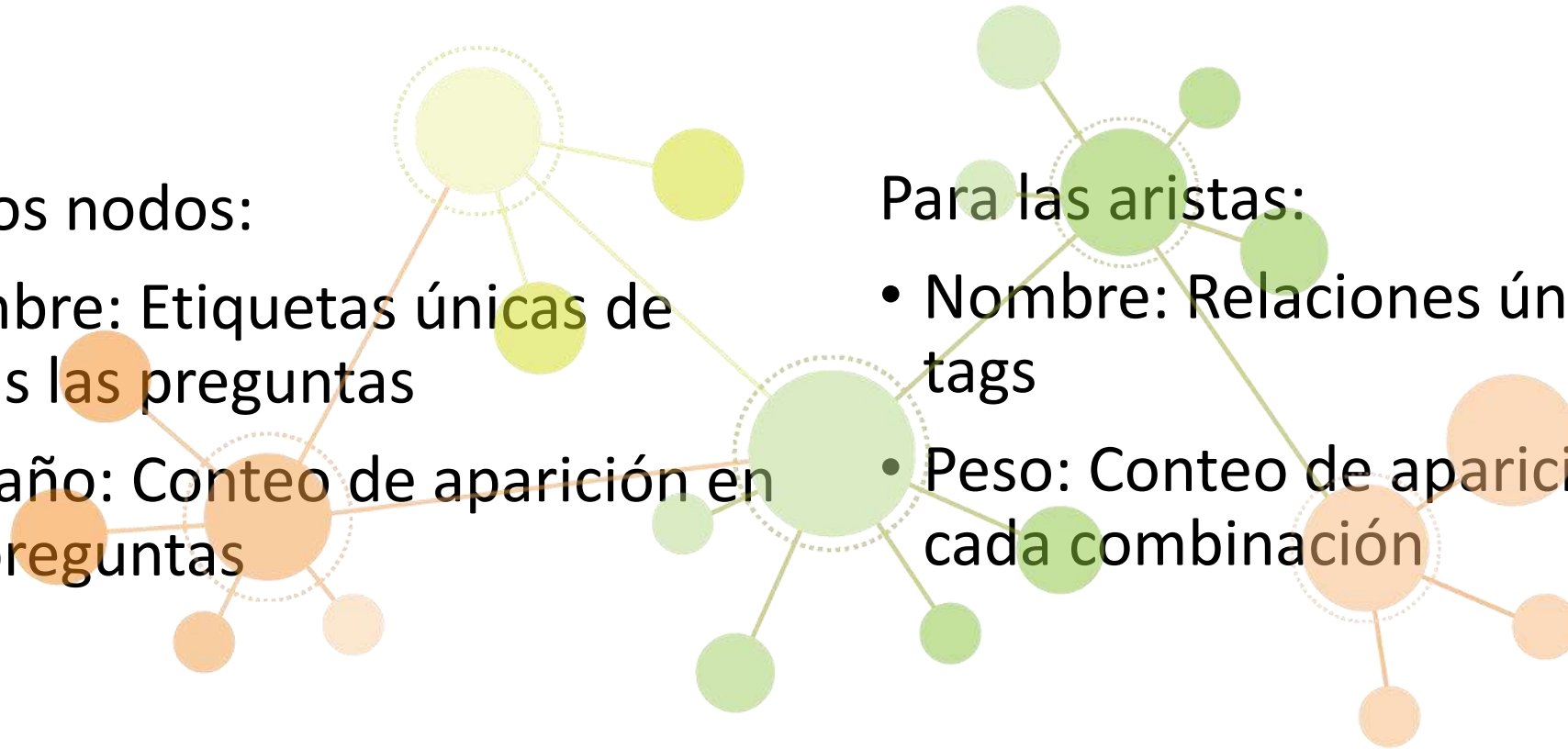
# Conjunto de datos

Para los nodos:

- Nombre: Etiquetas únicas de todas las preguntas
- Tamaño: Conteo de aparición en las preguntas

Para las aristas:

- Nombre: Relaciones únicas entre tags
- Peso: Conteo de aparición de cada combinación





# Modelo: Tag Network

Para asignar un grupo a la etiqueta:

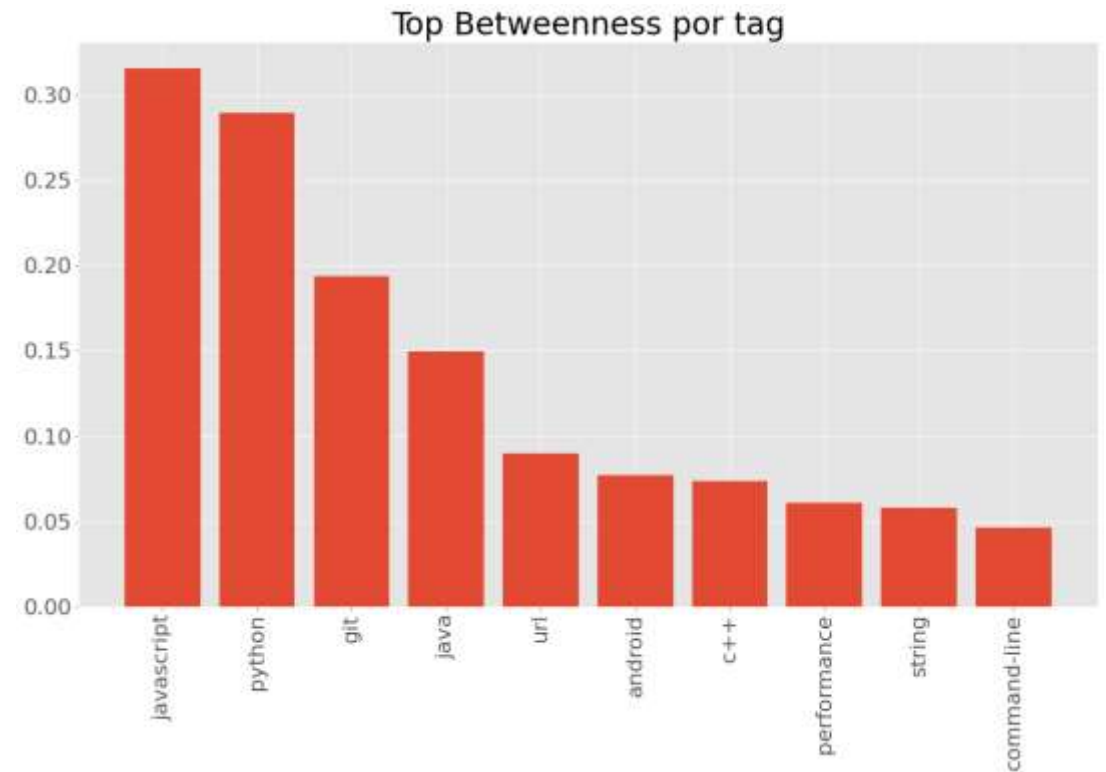
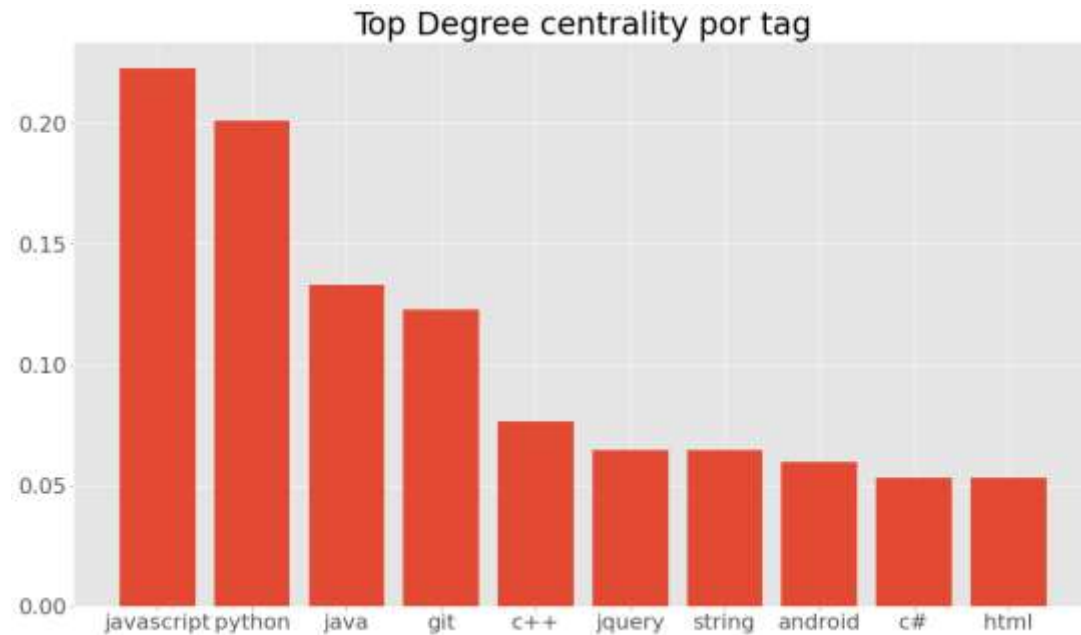
- Detección de comunidades:  
Algoritmo de Girvan-Newman

Para asignar posición a los nodos en el grafo y visualizarlos

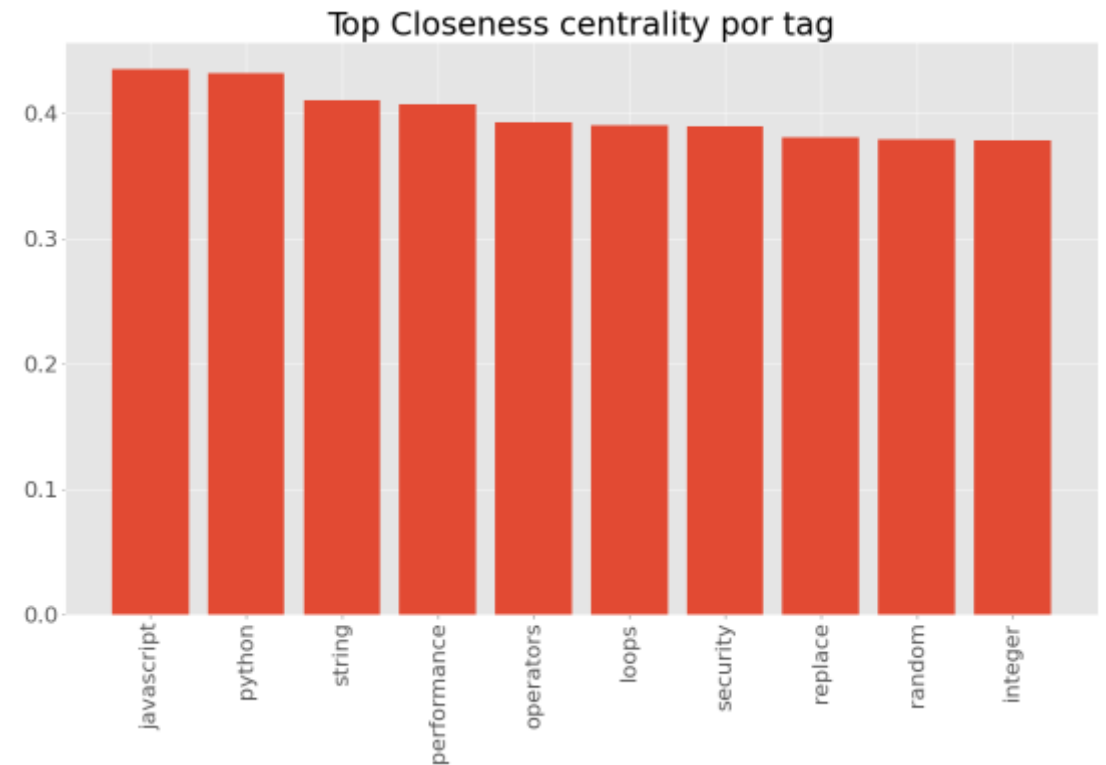
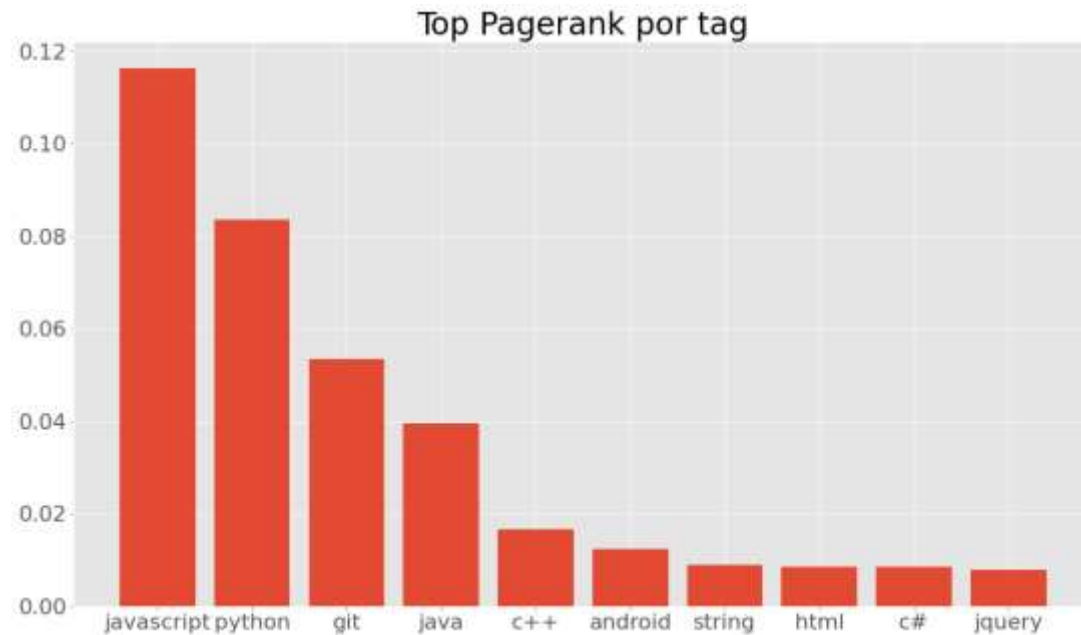
- Algoritmo de Fruchterman-Reingold



# Métricas: Degree centrality & Betweenness



# Métricas: Pagerank & Closeness centrality





# Tag network Stack Overflow

