

Tarea 02 - Tópicos Avanzados en Ciencia de Datos

Clustering sobre blogs usando métodos de análisis de redes sociales.

Objetivos.

El análisis de redes sociales con los métodos para describir y visualizar actores y enlaces mediante grafos ha permitido la descripción de un gran número de fenómenos de interés en la actualidad. Una de ellos fue el análisis de la estructura de los principales promotores de la opinión pública en el escenario electoral, la blogósfera política.

En el trabajo "La blogósfera política y los EE.UU. elecciones de 2004", se presentaron los datos de una red dirigida de hipervínculos entre los weblogs en la política de Estados Unidos, grabado en 2005 por L. A. Adamic and N. Glance. El objetivo de esta asignación es que, mediante técnicas de análisis de redes sociales, realice un estudio descriptivo sobre la estructura de las relaciones entre los blogs políticos de EEUU en el 2004 (categorizados como conservadores y liberales), según los enlaces a otros blogs contenidos en cada uno de ellos.

Este estudio debería darnos luces sobre la estructura de la "red de blogs políticos" así como la identificación de conglomerados y niveles de importancia de estos blogs a través de visualizaciones gráficas.

Archivo de datos.

Nombre del archivo	Formato	Uso
polblogs.gml	GLM	Datos del grafo de blogósfera política de EEUU en 2004
polblogs.paj	PAJ	Datos de vértices y enlaces (sin atributos adicionales)
polblogs.txt	TXT	Descripción de los datos

Actividades a realizar.

1. Cargue los datos entregados en el archivo **polblogs.gml** en una estructura de grafo. [Referencia](#)
2. Realice los cálculos de las métricas globales de grafo.
3. Calcule las métricas de centralidad para cada uno de los blogs, denomine al data frame resultante como "centralidad" y [almacénelo en un archivo de formato RData](#) llamado **Centralidad.rda**.
4. Describa la distribución de los grados de los blogs y grafique el histograma asociado.

5. Realice un clustering sobre la estructura de grafo, determine el número de cluster identificados y cuantos miembros tiene de cada grupo. Verifique si existe correspondencia entre los grupos identificados y el atributo de orientación política.
6. Identifique los blogs más importantes que se encuentran cerca del grupo opuesto. Para esto seleccione los blog que mantienen enlaces a blogs del grupo opuesto y ordene esta lista de forma descendiente según alguna medida de centralidad.
7. Realice una visualización gráfica de la “red de blogs políticos”, indicando los grupos a los que pertenecen los blogs mediante colores distintos para cada grupo y la medida de importancia de cada uno, por el tamaño de cada nodo. Identifique los enlaces hacia nodos de otro grupo con algún color distintivo. Almacene este gráfico en un archivo llamado **Blogosfera.png**

Condiciones.

- Las acciones propuestas deben implementarse en el lenguaje R.
- El funcionamiento correcto del algoritmo será probado con los datos de prueba entregados en esta tarea.
- Se puede realizar individualmente o en equipos de máximo 2 personas.
- Debe ser entregada hasta el lunes 10 de agosto de 2015.
- Elabore un breve informe con los resultados de las acciones propuestas, el código fuente y sus comentarios.