
GENERACIÓN DISTRIBUIDA DE GRAFOS CON HADOOP – MAPREDUCE

**FERNANDA SOLEDAD LÓPEZ GALLEGOS
INGENIERO CIVIL EN COMPUTACIÓN**

RESUMEN

Los grafos son una herramienta muy utilizada debido a su gran capacidad de modelar redes complejas. Desafortunadamente, encontrar conjuntos de datos representativos de la realidad, es un problema para muchos profesionales de diversas áreas. Esto, ha motivado la construcción de herramientas que permitan generar grafos de manera artificial. En ese sentido, R3MAT es un método diseñado para producir grafos sintéticos cuyas características se asemejan a las que ocurren en el mundo real (ej. distribución de ley de potencia). A pesar de su buen desempeño en comparación con otros generadores, R3MAT tiene problemas para generar grafos de muy gran tamaño. Además, el grafo más grande que se puede generar, se encuentra limitado por la cantidad de datos que se puedan gestionar en la memoria principal de un computador. En este trabajo se estudian variantes distribuidas basadas en R3MAT, con el objetivo de disminuir el tiempo de ejecución y al mismo tiempo soportar la generación de grafos más grandes que R3MAT. Para ello, se diseñan e implementan métodos usando Hadoop - MapReduce, los cuales posteriormente son evaluados en términos de eficiencia (tiempo de ejecución), escalabilidad (tamaño del grafo) y realismo (ley de potencia). Los resultados obtenidos muestran que: (i) para grafos con más de diez millones de nodos, los nuevos métodos (distribuidos) son más rápidos que R3MAT (secuencial), (ii) los nuevos métodos soportan la generación de grafos más grandes que el método secuencial, (iii) los grafos producidos con los nuevos métodos presentan la propiedad de distribución de ley de potencia, y (iv) los nuevos métodos son mejores que los métodos distribuidos que se encuentran en el estado del arte, en el sentido que: presentan un mejor ajuste a una distribución de ley de potencia, permiten diferenciar la generación de un grafo dirigido de uno no dirigido, y aseguran la producción de una cantidad determinada de aristas, sin generar aristas repetidas.