

OPTIMALITY OF SPECTRAL SPARSIFICATION IN DISTRIBUTED SYSTEMS WITH APPLICATION TO DATA CLUSTERING

Autor: FABRICIO AUGUSTO MENDOZA GRANADA

Tutor: Prof. Dr. MARCOS DANIEL VILLAGRA

RESUMEN

La técnica de Dispersión Espectral es utilizada en una variedad de aplicaciones tales como resolución de Sistemas Laplacianos y búsqueda de multicortes en un grafo mediante sus propiedades espectrales. Esencialmente, aproxima el espectro de un grafo por un factor constante reponderando algunas de sus aristas y eliminando otras. En la literatura existen algoritmos determinísticos así como probabilísticos para encontrar dispersores espectrales de grafos. Dado las múltiples aplicaciones en sistemas distribuidos, es de gran interés encontrar dispersores espectrales en tales sistemas. En este trabajo, demostramos que la dispersión espectral funciona bajo la suposición de que los datos de entrada están distribuidos en diferentes sitios. Para lograr esto, introducimos una herramienta matemática que captura la noción de solapamiento de datos entre diferentes sitios en un sistema distribuido. Además, describimos un protocolo que computa un dispersor espectral de un grafo $G = (V, \cup_{i=1}^s E_i)$ cuyas aristas, representadas por $\{E_i\}_{i \leq s}$, están alojadas en s diferentes sitios. El resultado del protocolo es un grafo $H = (V, \cup_{i=1}^s \hat{E}_i)$ donde H es un dispersor espectral de G y \hat{E}_i es el conjunto de aristas reponderadas en el sitio i .

La idea de datos que se solapan entre diferentes sitios nos ha inspirado a estudiar modelos de comunicación, los cuales funcionan como herramientas teóricas para estudiar algoritmos que trabajan con datos distribuidos. En particular, nos hemos enfocado en el modelo *Number-On-Forehead*, el cual es una poderosa herramienta con aplicaciones teóricas en complejidad de circuitos. Además, desarrollamos un protocolo que detecta si una familia dada de conjuntos de puntos constituye una estructura combinatorial particular llamada *Sistema- Δ* . Un *Sistema- Δ* es una familia de conjuntos con intersección única. Nuestro protocolo requiere a lo más 3 bits de comunicación.

Finalmente, desarrollamos además un protocolo que aproxima el multicorte de un grafo dado en el modelo *Number-On-Forehead* bajo la suposición de que la familia de subconjuntos de aristas constituye un *Sistema- Δ* . Nuestro protocolo tiene un costo de comunicación de $O(\log(\frac{n}{\epsilon^2} \sqrt{1 - \delta}))$ donde ϵ es el factor de aproximación del dispersor espectral y δ es un coeficiente que captura la mayor cantidad de datos solapados entre los sitios.

Palabras Clave: 1. dispersión espectral. 2. grafos densos. 3. algoritmos distribuidos. 4.