

Twitter y una implementación del TT

Introducción

Twitter es una red social basada en la publicación mensajes cortos (*tweets*). Una de las características de esta red es la publicación de los temas más comentados del momento, llamados *trending topics* (TT). Los usuarios pueden contribuir a la popularidad de ciertos temas mencionándolos usando *hashtags* que se crean agregando un caracter numeral a una palabra, por ejemplo #Algoritmos2.

Objetivo:

- Modelar un sistema que calcule los TTs de una serie de tweets.

Counting Sketch Min

Nuestra simulacion de Twitter utiliza un estilo de TDA llamado Counting Sketch Min. Este almacena la cantidad de apariciones de cada hashtag (un contador) en un vector, al cual se accede por medio de una función de hashing. Es decir, nuestra estructura de TDA es una "matriz" que por cada vector que se genera, utiliza una función de hash diferente(almacenadas en otro vector). Por ende, podríamos decir que la cantidad de apariciones de una palabra es igual al mínimo de los contadores que le corresponden a cada una. Sin embargo, si se llegan a generar colisiones puede generar un problema. La solución a implementar es tener k contadores, y tomamos el mínimo entre ellos, entonces podemos decir que si bien puede haber aparecido menos veces, seguro no más. Sigue siendo una aproximación, pero es mejor.

Utilizando el TDA propuesto podemos estimar que el tiempo de procesamiento de cada tag es de $O(1)$ ya que, sus primitivas tardan eso. Por otro lado, la impresión de los k TT tarda $O(n \log k)$.