

## Ampliació d'Algorismia: Approximation, Parameterization and Streaming

### QP 2023-2024: Sheet 3

29. Tenim un flux de dades que conté tots els nombres al conjunt  $\{1, \dots, n\}$  menys un. Dissenyeu un algorisme per trobar l'element que falta. El vostre algorisme pot llegir el flux només una vegada i fer servir  $O(\log n)$  espai a memòria. L'algorisme ha de proporcionar la resposta correcta després de processar tota l'entrada.
30. Considereu el següent algorisme que processa un graf  $G$  amb  $n$  nodes i donat per un flux d'arestes.
1. Selecciona una aresta  $(u, v)$  amb distribució uniforme sobre les arestes d' $E$ .
  2. Selecciona amb distribució uniforme un vèrtex de  $V \setminus \{u, v\}$
  3. Si  $(u, w)$  i  $v, w$  apareixen després de  $(u, v)$  en el flux, la sortida és  $m(n-2)$ , si no serà 0.
- (a) Proporcioneu una implementació com algorisme de streaming. El vostre algorisme només pot fer una pasada per el stream i utilitzar  $O(n \log n)$  memòria.
- (b) Demostreu que l'algorisme proporciona una estimació del nombre de triangles al graph  $G$ .
31. Hemos visto como usar reservoir sampling para muestreo de un stream del que no conocemos su longitud. Queremos extender este resultado a la estimación de valores de una función definida sobre el stream.
- Tenemos un stream  $s = a_1, a_2, \dots, a_m$  formado por valores enteros en  $[n] = \{0, \dots, n-1\}$ . Para  $i \in [n]$ ,  $f_i$  denota la frecuencia de  $i$  en  $s$ . Queremos estimar el valor

$$g(s) = \sum_{i \in [n]} g(f_i),$$

donde  $g$  es una función con valores reales, con  $g(0) = 0$ .

El siguiente algoritmo combina la obtención de una muestra con un conteo parcial:

- Obtener  $x$ , una muestra (con distribución uniforme) sobre  $[m]$
  - $r = |\{i \geq x \mid a_i = a_x\}|$
  - Return  $m(g(r) - g(r-1))$
- a) Demostrad que el algoritmo propuesto proporciona un estimador de la función  $g$ .
- b) Proporcionad una implementación del algoritmo propuesto para un stream de datos del que no conocéis su longitud  $m$ .