A Streaming Algorithm for Graph Clustering
Supplementary Material

Alexandre Hollocou
INRIA
Paris, France
alexandre.hollocou@inria.fr

Julien Maudet
Ecole Polytechnique
Palaiseau, France
julien.maudet@polytechnique.edu

Thomas Bonald
Telecom-Paristech
Paris, France
thomas.bonald@telecom-paristech.fr

Marc Lelarge
INRIA-ENS
Paris, France
marc.lelarge@ens.fr

Abstract

We here provide the proofs of the main results of the paper.

1 Proof of Lemma 1

This lemma gives an expression of Q_{t+1} in function of Q_t when a new edge (i, j) arrives.

Lemma 1. *If $e_{t+1} = (i, j)$ and if $P_{t+1} = P_t$, Q_{t+1} can be expressed in function of Q_t as follows*

$$Q_{t+1} = Q_t + 2\left[\delta(i,j) - \frac{\mathrm{Vol}_t(C(i)) + \mathrm{Vol}_t(C(j)) + 1 + \delta(i,j)}{w}\right]$$

where $C(v)$ denotes the community of v in P_t, and $\delta(i, j) = 1$ if i and j belongs to the same community of P_t and 0 otherwise.

Proof. Given a new edge $e_{t+1} = (i, j)$, we have the following relation between quantities $\mathrm{Int}(C)$ and $\mathrm{Vol}(C)$ at times t and $t + 1$.

$$\mathrm{Int}_{t+1}(C) = \mathrm{Int}_t(C) + 1_{i\in C}1_{j\in C}$$

and

$$\mathrm{Vol}_{t+1}(C) = \mathrm{Vol}_t(C) + 1_{i\in C} + 1_{j\in C}.$$

This gives us the following equation for $(\mathrm{Vol}_{t+1}(C))^2$

$$(\mathrm{Vol}_{t+1}(C))^2 = (\mathrm{Vol}_t(C))^2 + \begin{cases} 0 & \text{if } C \neq C(i) \text{ and } C \neq C(j) \\ 2\mathrm{Vol}_t(C(i)) + 1 & \text{if } C = C(i) \\ 2\mathrm{Vol}_t(C(j)) + 1 & \text{if } C = C(j) \end{cases}$$

in the case $C(i) \neq C(j)$, and

$$(\mathrm{Vol}_{t+1}(C))^2 = (\mathrm{Vol}_t(C))^2 + \begin{cases} 0 & \text{if } C \neq C(i) \text{ and } C \neq C(j) \\ 4\mathrm{Vol}_t(C(i)) + 4 & \text{if } C = C(i) = C(j) \end{cases}$$

in the case $C(i) = C(j)$.

Finally, the definition of Q_{t+1}

$$Q_{t+1} = \sum_{C\in P_{t+1}} \left[2\mathrm{Int}_{t+1}(C) - \frac{(\mathrm{Vol}_{t+1}(C))^2}{w}\right]$$

gives us the wanted result. □

2 Proof of Lemma 2

Lemma 2 gives an expression for the variation of Q_t when node i joins community $C(j)$.

Lemma 2.

$$\Delta Q_t = Q_t^{(a)} - Q_t^{(c)} = 2\left[L_t(i, C(j)) - L_t(i, C(i)) - \frac{(w_t(i))^2}{w}\right]$$

where

$$L_t(i, C) = \sum_{(i',j') \in S_t} \left[1_{i' \in C}\left(1_{j'=i} - \frac{w_t(i)}{w}\right) + 1_{j' \in C}\left(1_{i'=i} - \frac{w_t(i)}{w}\right)\right].$$

Proof. Q_t is defined as a sum over all communities of partition P_t. Only terms depending on $C(i)$ and $C(j)$ are modified by action (a). Thus, we have:

$$\begin{aligned}\Delta Q_t = {} & 2\left[\mathrm{Int}_t(C(i) \setminus \{i\}) + \mathrm{Int}_t(C(j) \cup \{i\}) - \mathrm{Int}_t(C(i)) - \mathrm{Int}_t(C(i))\right] \\ & - \frac{(\mathrm{Vol}_t(C(i)) - w_t(i))^2 + (\mathrm{Vol}_t(C(j)) + w_t(i))^2 - (\mathrm{Vol}_t(C(i)))^2 - (\mathrm{Vol}_t(C(j)))^2}{w}.\end{aligned}$$

This leads to:

$$\begin{aligned}\Delta Q_t = {} & 2 \sum_{(i',j') \in S_t} \left[1_{j'=i}(1_{i' \in C(j)} - 1_{i' \in C(i)}) + 1_{i'=i}(1_{j' \in C(j)} - 1_{j' \in C(i)})\right] \\ & - 2\frac{w_t(i)\mathrm{Vol}_t(C(j)) - w_t(i)\mathrm{Vol}_t(C(i)) + (w_t(i))^2}{w}.\end{aligned}$$

Using the definition of Vol_t, we obtain the wanted expression for ΔQ_t. $\square$

3 Proof of Theorem 1

Theorem 1 gives a sufficient condition in order to have a positive variation ΔQ_{t+1} of the modularity when i joins $C(j)$.

Theorem 1. *If* $\mathrm{Vol}_t(C(i)) \leq \mathrm{Vol}_t(C(j))$*, then:*

$$\mathrm{Vol}_t(C(j)) \leq v_t(i, j) \quad \Longrightarrow \quad \Delta Q_{t+1} \geq 0$$

where

$$v_t(i, j) = \frac{1 - (w_t(i) + 1)^2/w}{l_t(i, C(i)) - l_t(i, C(j))}.$$

Proof. From Lemma 1, we obtain

$$\begin{aligned}\Delta Q_{t+1} = {} & Q_t^{(a)} + 2\left[1 - \frac{(\mathrm{Vol}_t(C(j)) + w_i(t)) + (\mathrm{Vol}_t(C(j)) + w_i(t)) + 2)}{w}\right] \\ & - Q_t^{(b)} - 2\left[0 - \frac{(\mathrm{Vol}_t(C(i)) - w_i(t)) + (\mathrm{Vol}_t(C(j)) + w_i(t)) + 1)}{w}\right],\end{aligned}$$

which gives us:

$$\Delta Q_{t+1} = \Delta Q_t + 2\left[1 - \frac{\mathrm{Vol}_t(C(j)) - \mathrm{Vol}_t(C(i)) + 2w_i(t) + 1}{w}\right] \tag{1}$$

Then, Equation (1) and Lemma 2 gives us the following expression for ΔQ_{t+1}

$$\begin{aligned}\Delta Q_{t+1} = 2\Bigg[& 1 + \left(l_t(i, C(j)) - \frac{1}{w}\right) Vol_t(C(j)) - \left(l_t(i, C(i)) - \frac{1}{w}\right) Vol_t(C(i)) \\ & - \frac{(w_t(i) + 1)^2}{w}\Bigg].\end{aligned}$$

Thus, $\Delta Q_{t+1} \geq 0$ is equivalent to

$$\left(l_t(i, C(i)) - \frac{1}{w}\right) Vol_t(C(i)) - \left(l_t(i, C(j)) - \frac{1}{w}\right) Vol_t(C(j)) \leq 1 - \frac{(w_t(i)+1)^2}{w}. \tag{2}$$

We use $u_t(i, j)$ to denote the left-hand side of this inequality. If $Vol_t(C(i)) \leq Vol_t(C(i))$, then we have

$$u_t(i, j) \leq [l_t(i, C(i)) - l_t(i, C(j))] Vol_t(C(j))$$

Thus, the following inequality

$$[l_t(i, C(i)) - l_t(i, C(j))] Vol_t(C(j)) \leq 1 - \frac{(w_t(i)+1)^2}{w}$$

implies inequality (2), which proves the theorem. $\square$