Supplementary material for paper:
Optimal Transport for structured data and its application on graphs

1. Proofs

First we recall the notations from the paper :

Let two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ described respectively by their probability measure $\mu = \sum_{i=1}^{n} h_i \delta_{(x_i,a_i)}$ and $\nu = \sum_{i=1}^{m} g_j \delta_{(y_j,b_j)}$, where $h \in \Sigma_n$ and $g \in \Sigma_m$ are histograms with $\Sigma_n = \{h \in (\mathbb{R}_+^*)^n, \sum_{i=1}^n h_i = 1, \}$.

We introduce $\Pi(h,g)$ the set of all admissible couplings between h and g, *i.e.* the set

$$\Pi(h,g) = \{\pi \in \mathbb{R}_+^{n\times m} \; s.t. \; \sum_{i=1}^{n} \pi_{i,j} = h_j \; , \sum_{j=1}^{m} \pi_{i,j} = g_i\},$$

where $\pi_{i,j}$ represents the amount of mass shifted from the bin h_i to g_j for a coupling π.

Let (Ω_f, d) be a compact measurable space acting as the feature space. We denote the distance between the features as $M_{AB} = (d(a_i, b_j))_{i,j}$, a $n \times m$ matrix.

The structure matrices are denoted C_1 and C_2, and μ_X and μ_A (resp. ν_Y and ν_B) the marginals of μ (resp. ν) *w.r.t.* the structure and feature respectively. We also define the similarity between the structures by measuring the similarity between all pairwise distances within each graph thanks to the 4-dimensional tensor $L(C_1, C_2)$:

$$L_{i,j,k,l}(C_1, C_2) = |C_1(i,k) - C_2(j,l)|.$$

We also consider the following notations :

$$J_q(C_1, C_2, \pi) = \sum_{i,j,k,l} L_{i,j,k,l}(C_1, C_2)^q \pi_{i,j}\, \pi_{k,l} \quad (1)$$

$$H_q(M_{AB}, \pi) = \sum_{i,j} d(a_i, b_j)^q \pi_{i,j} \quad (2)$$

$$\begin{aligned} E_q(M_{AB}, C_1, C_2, \pi) &= \langle (1-\alpha) M_{AB}^q + \alpha L(C_1,C_2)^q \otimes \pi, \pi \rangle \\ &= \sum_{i,j,k,l} (1-\alpha) d(a_i,b_j)^q + \alpha L_{i,j,k,l}(C_1,C_2)^q \pi_{i,j}\pi_{k,l} \end{aligned} \quad (3)$$

Respectively J_q, H_q and E_q designate the Gromov-Wasserstein (GW) loss, the Wasserstein (W) loss and the FGW loss so that :

$$FGW_{q,\alpha}(\mu,\nu) = \min_{\pi \in \Pi(h,g)} E_q(M_{AB}, C_1, C_2, \pi) \quad (4)$$

$$W_q(\mu_A, \nu_B)^q = \min_{\pi \in \Pi(h,g)} H_q(M_{AB}, \pi) \quad (5)$$

$$GW_q(\mu_X, \nu_Y)^q = \min_{\pi \in \Pi(h,g)} J_q(C_1, C_2, \pi) \quad (6)$$

Please note that the minimum exists since we minimize a continuous function over a compact subset of $\mathbb{R}^{n\times m}$ and hence the FGW distance is well defined.

1.1. Bounds

We first introduce the following lemma:

Lemma 1.1. $FGW_{q,\alpha}(\mu,\nu)$ is lower-bounded by the straight-forward interpolation between $W_q(\mu_A,\nu_B)^q$ and $GW_q(\mu_X,\nu_Y)^q$:

$$FGW_{q,\alpha}(\mu,\nu) \geq (1-\alpha) W_q(\mu_A,\nu_B)^q + \alpha GW_q(\mu_X,\nu_Y)^q \quad (7)$$

Proof. Let π^α be the coupling that minimizes $E_q(M_{AB}, C_1, C_2, \cdot)$. Then we have:

$$\begin{aligned} FGW_{q,\alpha}(\mu,\nu) &= E_q(M_{AB}, C_1, C_2, \pi^\alpha) \\ &= (1-\alpha) H_q(M_{AB}, \pi^\alpha) + \alpha J_q(C_1, C_2, \pi^\alpha) \end{aligned}$$

But also:

$$\begin{aligned} W_q(\mu_A,\nu_B)^q &\leq H_q(M_{AB}, \pi^\alpha) \\ GW_q(\mu_X,\nu_Y)^q &\leq J_q(C_1, C_2, \pi^\alpha) \end{aligned}$$

The provided inequality is then derived. □

We also have two other straight-forward lower bounds for FGW:

$$FGW_{q,\alpha}(\mu,\nu) \geq (1-\alpha)\, W_q(\mu_A,\nu_B)^q \quad (8)$$

$$FGW_{q,\alpha}(\mu,\nu) \geq \alpha\, GW_q(\mu_X,\nu_Y)^q \tag{9}$$

1.2. Interpolation properties

We now claim the following theorem:

Theorem 1.2. *Interpolation properties.*

As α tends to zero, the FGW distance recovers $W_q(\mu_A,\nu_B)^q$ between the features, and as α tends to one, we recover $GW_q(\mu_X,\nu_Y)^q$ between the structures:

$$\lim_{\alpha\to 0} FGW_{q,\alpha}(\mu,\nu) = W_q(\mu_A,\nu_B)^q$$
$$\lim_{\alpha\to 1} FGW_{q,\alpha}(\mu,\nu) = GW_q(\mu_X,\nu_Y)^q$$

Proof. Let $\pi^W \in \Pi(h,g)$ be the optimal coupling for the Wasserstein distance $W_q(\mu_A,\nu_B)$ between μ_A and ν_B and let $\pi^\alpha \in \Pi(h,g)$ be the optimal coupling for the FGW distance $FGW_{q,\alpha}(\mu,\nu)$. We consider :

$$\begin{aligned}
&FGW_{q,\alpha}(\mu,\nu) - (1-\alpha)W_q(\mu_A,\nu_B)^q \\
&= E_q(M_{AB},C_1,C_2,\pi^\alpha) - (1-\alpha)H_q(M_{AB},\pi^W) \\
&\overset{*}{\leq} E_q(M_{AB},C_1,C_2,\pi^W) - (1-\alpha)H_q(M_{AB},\pi^W) \\
&= \sum_{i,j,k,l} \alpha |C_1(i,k) - C_2(j,l)|^q \pi^W_{i,j}\pi^W_{k,l} \\
&= \alpha J_q(C_1,C_2,\pi^W)
\end{aligned}$$

In (*) we used the suboptimality of the coupling π^W *w.r.t* the FGW distance. In this way we have proven :

$$FGW_{q,\alpha}(\mu,\nu) \leq (1-\alpha)W_q(\mu_A,\nu_B)^q + \alpha J_q(C_1,C_2,\pi^W) \tag{10}$$

Now let $\pi^{GW} \in \Pi(h,g)$ the optimal coupling for the Gromov-Wasserstein distance $GW_q(\mu_X,\nu_Y)$ between μ_X and ν_Y. Then :

$$\begin{aligned}
&FGW_{q,\alpha}(\mu,\nu) - \alpha GW_q(\mu_X,\nu_Y)^q \\
&= E_q(M_{AB},C_1,C_2,\pi^\alpha) - \alpha J_q(C_1,C_2,\pi^{GW}) \\
&\overset{*}{\leq} E_q(M_{AB},C_1,C_2,\pi^{GW}) - \alpha J_q(C_1,C_2,\pi^{GW}) \\
&= (1-\alpha)\sum_{i,j,k,l} (1-\alpha) d(a_i,b_j)^q \pi^{GW}_{i,j} \\
&= (1-\alpha) H_q(M_{AB},\pi^{GW})
\end{aligned}$$

where in (*) we used the suboptimality of the coupling π^{GW} *w.r.t* the FGW distance so that :

$$FGW_{q,\alpha}(\mu,\nu) \leq \alpha GW_q(\mu_X,\nu_Y)^q + (1-\alpha)H_q(M_{AB},\pi^{GW}) \tag{11}$$

As α goes to zero Eq. (10) and Eq. (8) give $\lim_{\alpha\to 0} FGW_{q,\alpha}(\mu,\nu) = W_q(\mu_A,\nu_B)^q$ and as α goes to one Eq. (11) and Eq. (9) give $\lim_{\alpha\to 1} FGW_{q,\alpha}(\mu,\nu) = GW_q(\mu_X,\nu_Y)^q$

□

1.3. FGW is a distance

For the following proofs we suppose that C_1 and C_2 are distance matrices, $n \geq m$ and $\alpha \in]0,...,1[$. We claim the following theorem :

Theorem 1.3. *FGW defines a metric for $q = 1$ and a semi-metric for $q > 1$.*

FGW defines a metric over the space of structured data quotiented by the measure preserving isometries that are also feature preserving. More precisely, FGW satisfies the triangle inequality and is nul iff $n = m$ and there exists a bijection $\sigma : \{1,..,n\} \to \{1,..,n\}$ such that :

$$\forall i \in \{1,..,n\},\ h_i = g_{\sigma(i)} \tag{12}$$

$$\forall i \in \{1,..,n\},\ a_i = b_{\sigma(i)} \tag{13}$$

$$\forall i,k \in \{1,..,n\}^2,\ C_1(i,k) = C_2(\sigma(i),\sigma(k)) \tag{14}$$

If $q > 1$, the triangle inequality is relaxed by a factor 2^{q-1} such that FGW defines a semi-metric

We first prove the equality relation for any $q \geq 1$ and we discuss the triangle inequality in the next section.

1.3.1. Equality relation

Theorem 1.4. *For all $q \geq 1$, $FGW_{q,\alpha}(\mu,\nu) = 0$ iff there exists an application $\sigma : \{1,..,n\} \to \{1,..,m\}$ which verifies (12), (13) and (14)*

Proof. First, let us suppose that $n = m$ and that such a bijection exists. Then if we consider the transport map π^* associated with $i \to i$ and $j \to \sigma(i)$ *i.e* the map $\pi^* = (I_d \times \sigma)$ with I_d the identity map.

By eq (12), $\pi^* \in \Pi(h,g)$ and clearly using (13) and (14):

$$\begin{aligned}
E_q(C_1,C_2,\pi^*) &= (1-\alpha)\sum_{i,k} d(a_i,b_{\sigma(i)})^q h_i g_{\sigma(i)} h_k g_{\sigma(k)} \\
&+ \alpha \sum_{i,k} |C_1(i,k) - C_2(\sigma(i),\sigma(k))|^q h_i g_{\sigma(i)} h_k g_{\sigma(k)} \\
&= 0
\end{aligned} \tag{15}$$

We can conclude that $FGW_{q,\alpha}(\mu,\nu) = 0$.

Conversely, suppose that $FGW_{q,\alpha}(\mu,\nu)=0$ and $q\geq 1$. We define :

$$\forall i,k \in \{1,...,n\}^2, \hat{C}_1(i,k) = \frac{1}{2}C_1(i,k) + \frac{1}{2}d(a_i,a_k) \tag{16}$$

$$\forall j,l \in \{1,...,m\}^2, \hat{C}_2(j,l) = \frac{1}{2}C_2(j,l) + \frac{1}{2}d(b_j,b_l) \tag{17}$$

To prove the existence of a bijection σ satisfying the theorem properties we will prove that the Gromov-Wasserstein distance $GW_q(\hat{C}_1,\hat{C}_2,\mu,\nu)$ vanishes.

Let $\pi \in \Pi(h,g)$ be any admissible transportation plan. Then for $n\geq 1$, :

$$\begin{aligned} J_n(\hat{C}_1,\hat{C}_2,\pi) &= \sum_{i,j,k,l} L(\hat{C}_1(i,k),\hat{C}_2(j,l))^n \pi_{i,j}\pi_{k,l} \\ &= \sum_{i,j,k,l} \left|\frac{1}{2}(C_1(i,k)-C_2(j,l)) + \frac{1}{2}(d(a_i,a_k)-d(b_j,b_l))\right|^n \pi_{i,j}\pi_{k,l} \\ &\overset{*}{\leq} \sum_{i,j,k,l} \frac{1}{2}\left|C_1(i,k)-C_2(j,l)\right|^n \pi_{i,j}\pi_{k,l} \\ &+ \sum_{i,j,k,l} \frac{1}{2}\left|d(a_i,a_k)-d(b_j,b_l)\right|^n \pi_{i,j}\pi_{k,l} \end{aligned}$$

In (*) we used the convexity of $t\rightarrow t^n$ and Jensen inequality. We denote the first term (I) and (II) the second term. Combining triangle inequalities $d(a_i,a_k)\leq d(a_i,b_j)+d(b_j,a_k)$ and $d(b_j,a_k)\leq d(b_j,b_l)+d(b_l,a_k)$ we have :

$$d(a_i,a_k) \leq d(a_i,b_j) + d(a_k,b_l) + d(b_j,b_l) \tag{18}$$

We split (II) in two parts $S_1 = \{i,j,k,l\ ; d(a_i,a_k)-d(b_j,b_l)\geq 0\}$ and $S_2=\{i,j,k,l\ ; d(a_i,a_k)-d(b_j,b_l)\leq 0\}$ such that

$$\begin{aligned} (II) &= \sum_{i,j,k,l\in S_1} (d(a_i,a_k)-d(b_j,b_l))^n \pi_{i,j}\pi_{k,l} \\ &+ \sum_{i,j,k,l\in S_2} (d(b_j,b_l))-d(a_i,a_k))^n \pi_{i,j}\pi_{k,l} \end{aligned}$$

In the same way as Eq. (18) we have :

$$d(b_j,b_l) \leq d(a_i,a_k) + d(a_i,b_j) + d(a_k,b_l) \tag{19}$$

So Eq. (18) and (19) give :

$$\begin{aligned} (II) &\leq \sum_{i,j,k,l} \frac{1}{2}|d(a_i,b_j)+d(a_k,b_l)|^n \pi_{i,j},\pi_{k,l} \\ &\overset{def}{=} M_n(\pi) \end{aligned} \tag{20}$$

Finally we have shown that :

$$\forall \pi\in\Pi(h,g), \forall n\geq 1,\ J_n(\hat{C}_1,\hat{C}_2,\pi) \leq \frac{1}{2}J_n(C_1,C_2,\pi)+M_n(\pi) \tag{21}$$

Now let π^* be the optimal coupling for $FGW_{q,\alpha}(\mu,\nu)$. If $FGW_{q,\alpha}(\mu,\nu)=0$ then since $E_q(C_1,C_2,\pi^*)\geq \alpha J_q(C_1,C_2,\pi^*)$ and $E_q(C_1,C_2,\pi^*)\geq(1-\alpha)H_q(M_{AB},\pi^*)$, we have:

$$J_q(C_1,C_2,\pi^*) = 0 \tag{22}$$

and

$$H_q(M_{AB},\pi^*) = 0$$

Then $\sum_{i,j} d(a_i,b_j)^q \pi^*_{i,j}=0$. Since all terms are positive we can conclude that $\forall m\in\mathbb{N}^*, \sum_{i,j} d(a_i,b_j)^m\pi^*_{i,j}=0$.

In this way :

$$\begin{aligned} M_q(\pi^*) &= \frac{1}{2}\sum_h \binom{q}{p} \big(\sum_{i,j} d(a_i,b_j)^p \pi^*_{i,j}\big)\big(\sum_{k,l} d(a_k,b_l)^{q-p}\pi^*_{k,l}\big) \\ &= 0 \end{aligned} \tag{23}$$

Using equations (21) and (22) we have shown :

$$J_q(\hat{C}_1,\hat{C}_2,\pi^*) = 0$$

So π^* is the optimal coupling for $GW_q(\hat{C}_1,\hat{C}_2,\mu,\nu)$ and $GW_q(\hat{C}_1,\hat{C}_2,\mu,\nu)=0$. By virtue to Gromov-Wasserstein properties (see (Memoli, 2011)), there exists an isomorphism between the metric spaces associated with μ and ν. In the discrete case this results in the existence of a function $\sigma:\{1,..,m\}\rightarrow\{1,..,n\}$ which is a weight preserving isometry and thus bijective. In this way, we have $m=n$ and σ verifiying Eq (12). The isometry property leads also to :

$$\forall i,k\in\{1,..,n\}^2,\ \hat{C}_1(i,k)=\hat{C}_2(\sigma(i),\sigma(k)) \tag{24}$$

Moreover, since π^* is the optimal coupling for $GW_q(\hat{C}_1, \hat{C}_2, \mu, \nu)$ leading to a zero cost, then π^* is supported by σ, in particular $\pi^* = (I_d \times \sigma)$

So $H_q(M_{AB}, \pi^*) = \sum_i d(a_i, b_{\sigma(i)})^q h_i g_{\sigma(i)}$. Since $H_q(M_{AB}, \pi^*) = 0$ and all the weights are strictly positive we can conclude that $a_i = b_{\sigma(i)}$.

In this way, $d(a_i, a_k) = d(b_{\sigma(i)}, b_{\sigma(k)})$, so using the equality (24) and the definition of $\hat{C}_1$ and $\hat{C}_2$ in (16) and (17) we can conclude that :

$$\forall i, k \in \{1, .., n\} \times \{1, .., n\},\ C_1(i, k) = C_2(\sigma(i), \sigma(k))$$

which concludes the proof.

□

1.3.2. Triangle Inequality

Theorem 1.5. *For all $q = 1$, FGW verifies the triangle inequality.*

Proof. To prove the triangle inequality of FGW distance for arbitrary measures we will use the gluing lemma (see (Villani, 2008)) which stresses the existence of couplings with a prescribed structure. Let $h,\ g,\ f \in \Sigma_n \times \Sigma_m \times \Sigma_k$. Let also $\mu = \sum_{i=1}^n h_i \delta_{a_i, x_i}$, $\nu = \sum_{j=1}^m g_j \delta_{b_j, y_j}$ and $\gamma = \sum_{p=1}^k f_p \delta_{c_p, z_p}$ be three structured data as described in the paper. We note $C_1(i, k)$ the distance between vertices x_i and x_k, $C_2(i, k)$ the distance between vertices y_i and y_k and $C_3(i, k)$ the distance between vertices z_i and z_k.

Let P and Q be two optimal solutions of the FGW transportation problem between μ and ν and ν and γ respectively.

We define :

$$S = P\text{diag}(\frac{1}{g})Q$$

(note that S is well defined since $g_j \neq 0$ for all j). Then by definition $S \in \Pi(h, f)$ because :

$S1_m = P\text{diag}(\frac{1}{g})Q1_m = P\text{diag}(\frac{g}{g}) = P1_m = h$ (same reasoning for f).

We first prove the triangle inequality for the case $q = 1$.

By suboptimality of S :

$$\begin{aligned}
&FGW_{1,\alpha}(C_1, C_3, \mu, \gamma) \\
&\leq \sum_{i,j,k,l} (1-\alpha)d(a_i, c_j) + \alpha L(C_1(i,k), C_3(j,l)) S_{i,j} S_{k,l} \\
&= \sum_{i,j,k,l} ((1-\alpha)d(a_i, c_j) + \alpha L(C_1(i,k), C_3(j,l)) \\
&\times (\sum_e \frac{P_{i,e}Q_{e,j}}{g_e})(\sum_o \frac{P_{k,o}Q_{o,l}}{g_o}) \\
&= \sum_{i,j,k,l} \big((1-\alpha)d(a_i, c_j) + \alpha |C_1(i,k) - C_3(j,l)|\big) \\
&\times (\sum_e \frac{P_{i,e}Q_{e,j}}{g_e})(\sum_o \frac{P_{k,o}Q_{o,l}}{g_o}) \\
&\overset{*}{\leq} \sum_{i,j,k,l,e,o} \big((1-\alpha)(d(a_i, b_e) + d(b_e, c_j)) \\
&+ \alpha |C_1(i,k) - C_2(e,o) + C_2(e,o) - C_3(j,l)|\big) \\
&\times (\frac{P_{i,e}Q_{e,j}}{g_e})(\frac{P_{k,o}Q_{o,l}}{g_o}) \\
&\overset{**}{\leq} \sum_{i,j,k,l,e,o} \big((1-\alpha)d(a_i, b_e) + \alpha L(C_1(i,k), C_2(e,o))\big) \\
&\times \frac{P_{i,e}Q_{e,j}}{g_e} \frac{P_{k,o}Q_{o,l}}{g_o} \\
&+ \sum_{i,j,k,l,e,o} \big((1-\alpha)d(b_e, c_j) + \alpha L(C_2(e,o), C_3(j,l))\big) \\
&\times \frac{P_{i,e}Q_{e,j}}{g_e} \frac{P_{k,o}Q_{o,l}}{g_o}
\end{aligned}$$

where in (*) we use the triangle inequality of d and in (**) the triangle inequality of $|.|$

Moreover we have :

$$\sum_j \frac{Q_{e,j}}{g_e} = 1,\ \sum_l \frac{Q_{o,l}}{g_o} = 1,\ \sum_i \frac{P_{i,e}}{g_e} = 1,\ \sum_k \frac{P_{k,o}}{g_o} = 1$$

So,

$$\begin{aligned}
&FGW_{1,\alpha}(C_1, C_3, \mu, \gamma) \\
&\leq \sum_{i,k,e,o} \big((1-\alpha)d(a_i, b_e) + \alpha L(C_1(i,k), C_2(e,o))\big) P_{i,e} P_{k,o} \\
&+ \sum_{l,j,e,o} \big((1-\alpha)d(b_e, c_j) + \alpha L(C_2(e,o), C_3(j,l))\big) Q_{e,j} Q_{o,l}
\end{aligned}$$

Since P and Q are the optimal plans we have :

$$\begin{aligned}
FGW_{1,\alpha}(C_1, C_3, \mu, \gamma) \leq\ & FGW_{1,\alpha}(C_1, C_2, \mu, \nu) \\
&+ FGW_{1,\alpha}(C_2, C_3, \nu, \gamma)
\end{aligned}$$

which prove the triangle inequality for $q = 1$.

□

Theorem 1.6. *For all $q > 1$, FGW verifies the relaxed triangle inequality :*

$$FGW_{q,\alpha}(C_1, C_3, \mu, \gamma) \leq 2^{q-1}\big(FGW_{q,\alpha}(C_1, C_2, \mu, \nu) + FGW_{q,\alpha}(C_2, C_3, \nu, \gamma)\big)$$

Proof. Let $q > 1$, We have :

$$\forall x, y \in \mathbb{R}_+,\ (x+y)^q \leq 2^{q-1}\big(x^q + y^q\big) \tag{25}$$

Indeed,

$$(x+y)^q = \Big(\big(\tfrac{1}{2^{q-1}}\big)^{\frac{1}{q}} \frac{x}{\big(\frac{1}{2^{q-1}}\big)^{\frac{1}{q}}} + \big(\tfrac{1}{2^{q-1}}\big)^{\frac{1}{q}} \frac{y}{\big(\frac{1}{2^{q-1}}\big)^{\frac{1}{q}}}\Big)^q \leq \Big[\big(\tfrac{1}{2^{q-1}}\big)^{\frac{1}{q-1}} + \big(\tfrac{1}{2^{q-1}}\big)^{\frac{1}{q-1}}\Big]^{q-1}\Big(\frac{x^q}{\frac{1}{2^{q-1}}} + \frac{y^q}{\frac{1}{2^{q-1}}}\Big) = \frac{x^q}{\frac{1}{2^{q-1}}} + \frac{y^q}{\frac{1}{2^{q-1}}}$$

Last inequality is a consequence of Hölder inequality. Then using same notations :

$$\begin{aligned}
&FGW_{q,\alpha}(C_1, C_3, \mu, \gamma) \\
&\leq \sum_{i,j}\sum_{k,l}(1-\alpha)d(a_i, c_j)^q + \alpha L(C_1(i,k), C_3(j,l))^q S_{i,j}S_{k,l} \\
&= \sum_{i,j}\sum_{k,l}((1-\alpha)d(a_i, c_j)^q + \alpha L(C_1(i,k), C_3(j,l))^q \\
&\times \Big(\sum_e \frac{P_{i,e}Q_{e,j}}{g_e}\Big)\Big(\sum_o \frac{P_{k,o}Q_{o,l}}{g_o}\Big) \\
&= \sum_{i,j,k,l} \big((1-\alpha)d(a_i, c_j)^q + \alpha|C_1(i,k) - C_3(j,l)|^q\big) \\
&\times \Big(\sum_e \frac{P_{i,e}Q_{e,j}}{g_e}\Big)\Big(\sum_o \frac{P_{k,o}Q_{o,l}}{g_o}\Big) \\
&\overset{*}{\leq} \sum_{i,j,k,l,e,o} \big((1-\alpha)(d(a_i, b_e) + d(b_e, c_j))^q \\
&+ \alpha|C_1(i,k) - C_2(e,o) + C_2(e,o) - C_3(j,l)|^q\big) \\
&\times \Big(\frac{P_{i,e}Q_{e,j}}{g_e}\Big)\Big(\frac{P_{k,o}Q_{o,l}}{g_o}\Big) \\
&\overset{**}{\leq} 2^{q-1} \sum_{i,j,k,l,e,o} \big((1-\alpha)d(a_i, b_e)^q + \alpha L(C_1(i,k), C_2(e,o))^q\big) \\
&\times \frac{P_{i,e}Q_{e,j}}{g_e}\frac{P_{k,o}Q_{o,l}}{g_o} \\
&+ 2^{q-1} \sum_{i,j,k,l,e,o} \big((1-\alpha)d(b_e, c_j)^q + \alpha L(C_2(e,o), C_3(j,l))^q\big) \\
&\times \frac{P_{i,e}Q_{e,j}}{g_e}\frac{P_{k,o}Q_{o,l}}{g_o}
\end{aligned}$$

where in (*) we use the triangle inequality of d and in (**) the triangle inequality of $|.|$ and (25).

Since P and Q are the optimal plans we have :

$$FGW_{q,\alpha}(C_1, C_3, \mu, \gamma) \leq 2^{q-1}\big(FGW_{q,\alpha}(C_1, C_2, \mu, \nu) + FGW_{q,\alpha}(C_2, C_3, \nu, \gamma)\big)$$

Which prove that $FGW_{q,\alpha}$ defines a semi metric for $q > 1$ with coefficient 2^{q-1} for the triangle inequality relaxation.

□

Table 1. Percentage of α chosen in $]0, ..., 1[$ compared to $\{0, 1\}$ for discrete labeled graphs

DISCRETE ATTR.	MUTAG	NCI1	PTC
FGW RAW SP	100%	100%	98%
FGW WL H=2 SP	100%	100%	88%
FGW WL H=4 SP	100%	100%	88%

Table 2. Percentage of α chosen in $]0, ..., 1[$ compared to $\{0, 1\}$ for vector attributed graphs

VECTOR ATTRIBUTES	BZR	COX2	CUNEIFORM	ENZYMES	PROTEIN	SYNTHETIC
FGW SP	100 %	90%	100%	100%	100%	100%

2. Comparaison with W and GW

Cross validation results During the nested cross validation, we divided the dataset into 10 and use 9 folds for training, where α is chosen within $[0, 1]$ *via* a 10-CV cross-validation, 1 fold for testing, with the best value of α (with the best average accuracy on the 10-CV) previously selected. The experiment is repeated 10 times for each dataset except for MUTAG and PTC where it is repeated 50 times. Table 2 and 2 report the average number of time α was chose within $]0, ...1[$ without 0 and 1 corresponding to the Wasserstein and Gromov-Wasserstein distances respectively. Results suggests that both structure and feature pieces of information are necessary as α is consistently selected inside $]0, ...1[$ except for PTC and COX2.

Nested CV results We report in tables 3 and 4 the average classification accuracies of the nested classification procedure by taking W and GW instead of FGW (*i.e* by taking $\alpha = 0, 1$). Best result for each dataset is in bold. A (*) is added when best score does not yield to a significative improvement compared to the second best score. The significance is based on a Wilcoxon signed rank test between the best method and the second one.

Results illustrates that FGW encompasses the two cases of W and GW, as scores of FGW are usually greater or equal on every dataset than scores of both W and GW and when it is not the case the difference is not statistically significant.

Table 3. Average classification accuracy on the graph datasets with vector attributes.

VECTOR ATTRIBUTES	BZR	COX2	CUNEIFORM	ENZYMES	PROTEIN	SYNTHETIC
FGW SP	85.12±4.15	**77.23±4.86***	**76.67±7.04**	71.00±6.76	74.55±2.74	**100.00±0.00**
W	**85.36±4.87***	77.23±3.16	61.48±10.23	**71.16±6.32***	**75.98± 1.97***	34.07±11.33
GW SP	82.92±6.72	77.65±5.88	50.66±8.91	23.66±3.63	71.96± 2.40	41.66±4.28

Table 4. Average classification accuracy on the graph datasets with discrete attributes.

DISCRETE ATTR.	MUTAG	NCI1	PTC-MR
FGW RAW SP	83.26±10.30	72.82±1.46	55.71±6.74
FGW WL H=2 SP	86.42±7.81	85.82±1.16	63.20±7.68
FGW WL H=4 SP	**88.42±5.67**	**86.42±1.63***	65.31±7.90
W RAW SP	79.36±3.49	70.5±4.63	54.79±5.76
W WL H=2 SP	87.78±8.64	85.83±1.75	63.90±7.66
W WL H=4 SP	87.15±8.23	86.42±1.64	**66.28±6.95***
GW SP	82.73±9.59	73.40±2.80	54.45± 6.89

Timings In this paragraph we provide some timings for the discrete attributed datasets. Table 5 displays the average timing for computing FGW between two pair of graphs.

Table 5. Average timings for the computation of FGW between two pairs of graph

DISCRETE ATTR.	MUTAG	NCI1	PTC-MR
FGW	2.5 MS	7.3 MS	3.7 MS

References

Memoli, F. Gromov wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, pp. 1–71, 2011. ISSN 1615-3375. URL http://dx.doi.org/10.1007/s10208-011-9093-5. 10.1007/s10208-011-9093-5.

Villani, C. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer, 2009 edition, September 2008. ISBN 3540710493.