Optimal Transport for structured data with application on graphs

A novel distance between labeled graphs based on optimal transport

Titouan Vayer
Joint work with Laetitia Chapel, Remi Flamary, Romain Tavenard and Nicolas Courty
Contributions:

- Differentiable distance between labeled graphs. Jointly considers the features and the structures
Contributions:

• Differentiable distance between labeled graphs. Jointly considers the features and the structures

Optimal transport: soft assignment between the nodes

Distance = 1.41
Contributions:

- Differentiable distance between labeled graphs. Jointly considers the features and the structures

\[ \frac{1}{2} (Q + Q) = Q \]

Computing average of labeled graphs
Structured data as probability distribution
Structured data as probability distribution

Features \((a_i)_i\)
Structured data as probability distribution

Features \((a_i)_i\)  \(\alpha_i\)

nodes \((x_i)_i\) in the metric space of the graph
Structured data as probability distribution

Features \((a_i)_i\) in the metric space of the graph \(G\), nodes \((x_i)_i\) in the metric space of the graph, weighted by their masses \((h_i)_i\).
Optimal transport in a nutshell

Compare two probability distributions by transporting one onto another

Wasserstein distance

Gromov-Wasserstein distance
Optimal transport in a nutshell

Compare two probability distributions by transporting one onto another

Wasserstein distance

Gromov-Wasserstein distance
Optimal transport in a nutshell

Compare two probability distributions by transporting one onto another

Wasserstein distance

\[ \mu_A \quad \nu_B \quad d(a_i, b_j) \]

Gromov-Wasserstein distance

\[ |C_1(i, k) - C_2(j, l)| \]
Fused Gromov-Wasserstein distance

\[ FGW_{q,\alpha}(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \sum_{i,j,k,l} \left( (1 - \alpha)d(a_i, b_j)^q + \alpha |C_1(i, k) - C_2(j, l)|^q \right) \pi_{i,j} \pi_{k,l} \]

where \( \pi \) is the soft assignment matrix
\( \alpha \) is a trade-off features/structures
Fused Gromov-Wasserstein distance

Properties

• **Interpolate** between Wasserstein distance on features and Gromov-Wasserstein distance on the structures

• **Distance on labeled graph**: vanishes iff graphs have same labels and weights at the same place up to a permutation

Optimization problem

• Non convex Quadratic Program: hard!

• Conditional Gradient Descent (aka Frank Wolfe)

• Suitable for entropic regularization + Sinkhorn iteraterations
**Applications**

**Classification**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MUTAG</th>
<th>PTC</th>
<th>NCI1</th>
<th>IMDB-B</th>
<th>SYNTHETIC</th>
<th>PROTEIN</th>
<th>CUNEIFORM</th>
</tr>
</thead>
<tbody>
<tr>
<td>WL</td>
<td>86.21±8.15</td>
<td>62.17±7.80</td>
<td>85.13±1.61</td>
<td>Unapplicable(U)</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>GK</td>
<td>82.42±8.40</td>
<td>56.46±8.03</td>
<td>60.78±2.48</td>
<td>56.00±3.61</td>
<td>41.13±4.68</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>RW</td>
<td>79.47±8.17</td>
<td>55.09±7.34</td>
<td>58.63±2.44</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>SP</td>
<td>85.79±2.51</td>
<td>58.53±2.55</td>
<td>73.00±0.51</td>
<td>55.80±2.93</td>
<td>38.93±5.12</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>HOPPER</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>90.67±4.67</td>
<td>71.96±3.22</td>
<td>32.59±8.73</td>
</tr>
<tr>
<td>PROPA</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>64.67±6.70</td>
<td>61.34±4.38</td>
<td>12.59±6.67</td>
</tr>
<tr>
<td>PSCN k = 10</td>
<td>83.47±10.26</td>
<td>58.34±7.71</td>
<td>70.65±2.58</td>
<td>U</td>
<td>100.00±0.00</td>
<td>67.95±11.28</td>
<td>25.19±7.73</td>
</tr>
<tr>
<td>FGW</td>
<td>88.42±5.67</td>
<td>65.31±7.90</td>
<td>86.42±1.63</td>
<td>63.80±3.49</td>
<td>100.00±0.00</td>
<td>74.55±2.74</td>
<td>76.67±7.04</td>
</tr>
</tbody>
</table>

**Graph Barycenter + k-means clustering of graphs**
Check out our poster at Pacific Ballroom #133!!