SGR: Self-Supervised Spectral Graph Representation Learning

Expressive graph comparison is hard

Four key properties for expressive graph similarity:
- **Permutation invariance**: reordering nodes does not change similarity
- **Scale-adaptivity**: structure is captured on both local and global scales
- **Size-invariance**: structure of the graph may not depend on its size
- **Scalability**: able to deal with both many and big graphs

Scalability is possible with a suitable representation to analyze large graph collections (e.g., classification, clustering).

Graph structure at different scales

In different applications, graphs are analyzed at different scales:
- **Local interactions**, e.g. molecular bonds in computational biology
- **Medium-scale structure**, e.g. core-periphery in economic networks
- **Global connectivity**, e.g. community structure in social networks

We argue that *scale* is a continuum, as in Kronecker graphs.

Learn more...

Enjoy SGR? Try NetLSD!

poster #39 oral tomorrow @ 4PM anton.tsitsulin@hpi.de

Rushing to dinner? Read this!

SGR is a learned graph descriptor that allows to compare graphs:
- **Fast**: in $O(1)$, with $O(n)$ precomputation;
- **Focused**: SGR learns which part of the spectrum to pay attention to

Heat kernel matrix involves pairs of nodes, so we use its *trace*:

$$h_t = \text{tr}(H_t) = \sum_j e^{-\lambda_j t}$$

We view $h_t$ as a family of low-pass filters on the spectrum. We can learn a filter with a neural network if we re-sample the spectrum $\lambda$ as

$$\sigma = \psi(W \lambda + b),$$

where $\psi$ is SelU non-linearity. However, NN parameters $\Theta = (W, b)$ still need to be trained.

We propose a self-supervised training that does not require any real-world data nor labels. We use a surrogate task of distinguishing between Erdős-Rényi and stochastic block model graphs.

Theory: computational geometry

Mémoli [1] suggests a spectral definition of the Gromov-Wasserstein distance between Riemannian manifolds. Matching a pair of points $(x, z')$ on manifold $\mathcal{M}$ to a pair of points $(y, y')$ on manifold $\mathcal{N}$ at scale $t$ costs

$$\Gamma(x, y, x', y', t) = |H_t^\mathcal{M}(x, x') - H_t^\mathcal{N}(y, y')|.$$ 

The distance $d(\mathcal{M}, \mathcal{N})$ is defined as the infimal measure coupling

$$d(\mathcal{M}, \mathcal{N}) = \inf_{\rho} \sup_{t>0} e^{-2(t+1)} \|\Gamma_t|_{L^2(\rho \times \rho)},$$

where the infimum is sought over all measures on $\mathcal{M} \times \mathcal{N}$ marginalizing to the standard measures on $\mathcal{M}$ and $\mathcal{N}$. This distance can be lower bounded by

$$d(\mathcal{M}, \mathcal{N}) \geq \sup_{t>0} e^{-2(t+1)} |h_t^\mathcal{M} - h_t^\mathcal{N}|.$$

We adopt this result mutatis mutandis to graphs, substituting the Laplace-Beltrami operator of the manifold with the normalized graph Laplacian.

Experiments

SGR is both versatile and expressive. Table 1 shows that SGR captures natural properties of real graphs. We do not use node and edge labels of graphs, and do not tune the classifiers.

<table>
<thead>
<tr>
<th>dataset</th>
<th>WL</th>
<th>NetSimile</th>
<th>FGSD</th>
<th>$\Lambda$</th>
<th>SGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>D&amp;D</td>
<td>68.27</td>
<td>70.02</td>
<td>64.88</td>
<td>64.54</td>
<td>76.12</td>
</tr>
<tr>
<td>ENZYMES</td>
<td>25.11</td>
<td>28.06</td>
<td>28.85</td>
<td>25.28</td>
<td>33.67</td>
</tr>
<tr>
<td>MUTAG</td>
<td>81.16</td>
<td>83.66</td>
<td>85.23</td>
<td>82.07</td>
<td>86.97</td>
</tr>
<tr>
<td>PROTEINS</td>
<td>72.33</td>
<td>70.59</td>
<td>63.27</td>
<td>71.32</td>
<td>73.83</td>
</tr>
<tr>
<td>COLLAB</td>
<td>78.52</td>
<td>74.26</td>
<td>70.66</td>
<td>66.15</td>
<td>71.98</td>
</tr>
<tr>
<td>IMDB-B</td>
<td>72.26</td>
<td>70.96</td>
<td>69.20</td>
<td>63.16</td>
<td>70.38</td>
</tr>
<tr>
<td>IMDB-M</td>
<td>50.75</td>
<td>46.80</td>
<td>48.88</td>
<td>41.14</td>
<td>47.97</td>
</tr>
<tr>
<td>REDDIT-B</td>
<td>71.97</td>
<td>86.84</td>
<td>87.12</td>
<td>76.25</td>
<td>87.45</td>
</tr>
<tr>
<td>REDDIT-M</td>
<td>48.57</td>
<td>44.96</td>
<td>48.51</td>
<td>48.02</td>
<td>53.22</td>
</tr>
</tbody>
</table>

Table 1: Accuracy of a logistic regression classifier.

References