POKer: a Partial Order Kernel for Comparing Strings with Alternative Substrings

Maryam Abdollahyan

- String comparison
  - A fundamental operation in many areas, e.g. **bioinformatics**
  - Strings may contain alternative substrings that can be substituted for each other
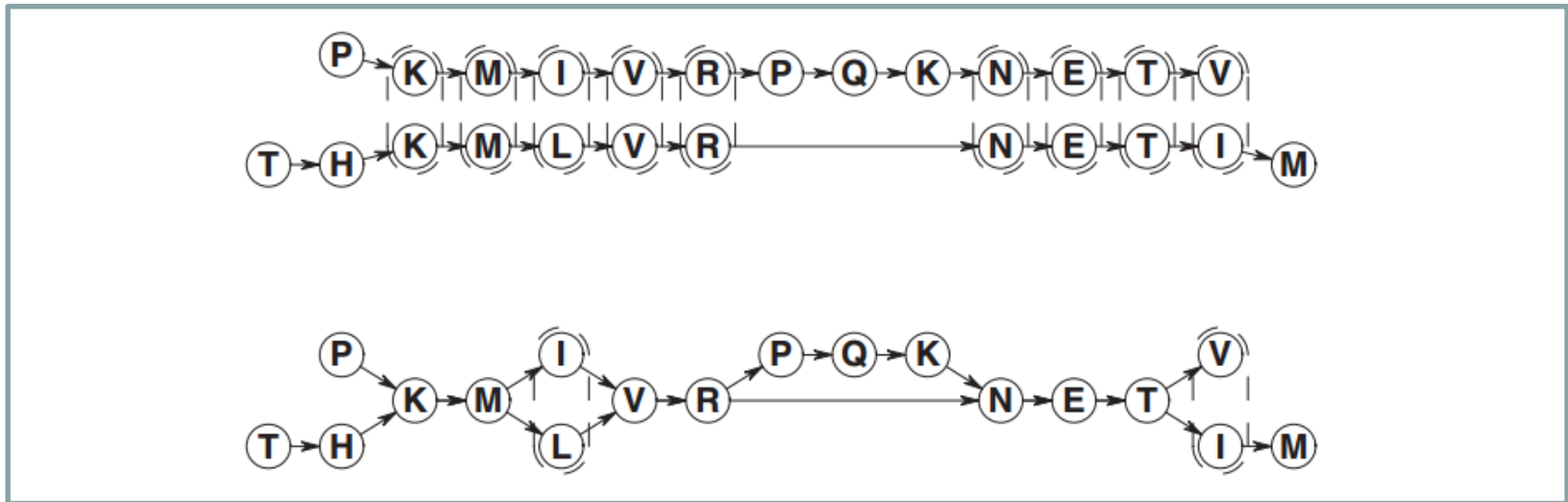 - Alternative substrings can be of *different* lengths

Figure from the "Multiple sequence alignment using partial order graphs" paper by C. Lee, C. Grasso and M.F. Sharlow

The Partial Order Kernel

- Represent each string as a directed acyclic **graph**
- Find the optimal **local alignment** between all possible choices of paths in the two graphs
- Sum the alignment scores
- Computed using **dynamic programming**
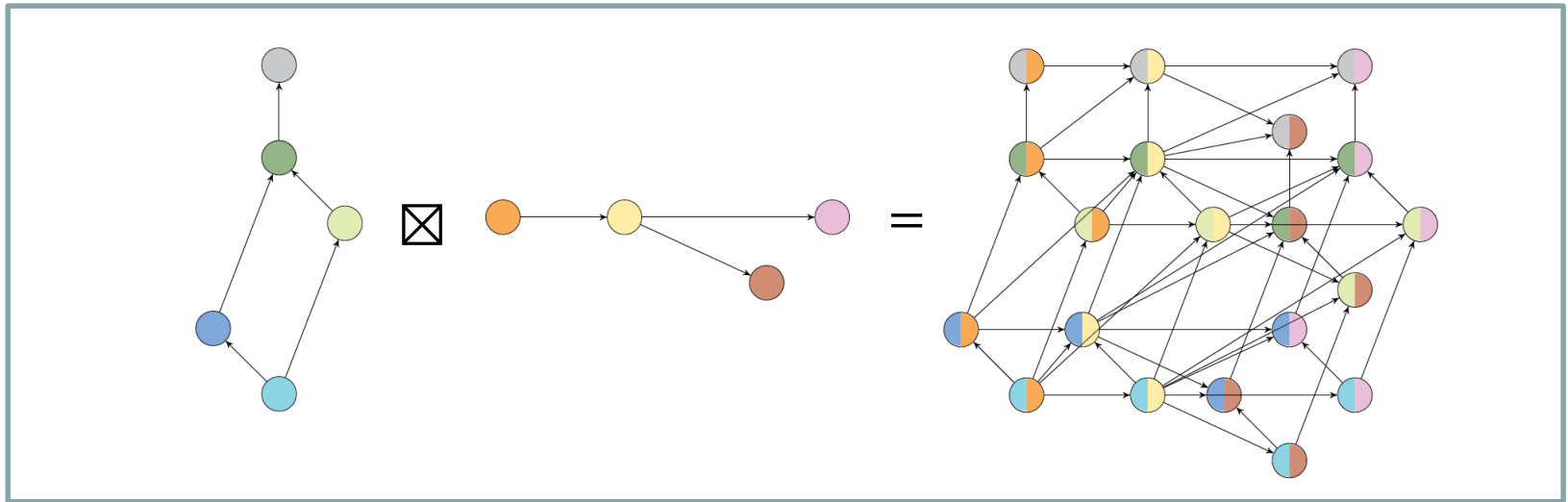
Results and conclusion

- Evaluated with SVMs in a multi-class classification scenario
- *Better* performance compared to a *k*-mer based kernel

Mean ROC Score

No. of classes	POKer	Spectrum kernel
5	**98.4%**	93.4%
10	**98.1%**	90.1%
20	**97.8%**	86.4%
40	**96.4%**	82%

- An **effective** algorithm for classification of strings, with **low computational complexity** and **rich mathematical properties** as a kernel

POKer: a Partial Order Kernel for Comparing Strings with Alternative Substrings
Maryam Abdollahyan and Fabrizio Smeraldi
European Symposium on Artificial Neural Networks (ESANN), 2017, Bruges – Belgium

CIS centre for intelligent sensing

Queen Mary
University of London