

## Phase 1: Create a protein-similarity graph based on many-to-many alignment

>APWOR330  
MCCI VDSL  
>BOPPLA2..  
NAARLR..

Protein sequences

**Step 1**

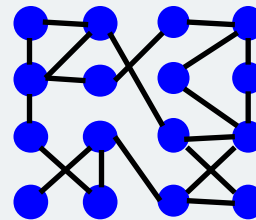
Many-to-many  
alignment

Seq 1	Seq 2	score
AP330	AP331	95
MCCIV	PPLAA	87
....	....	....

Alignment information

**Step 2**

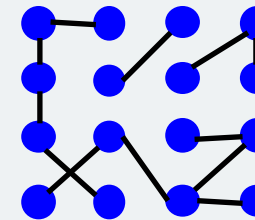
Generated graph  
by thresholding



Initial graph

**Step 3**

Sparsify  
graphs

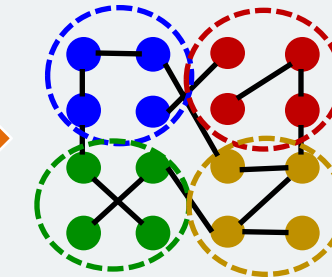


Sparse graph

## Phase 2: Classify proteins

**Step 4**

Learning or  
clustering  
algorithms



Protein families