



Need for Speed: Proteinortho Nitro

Paul Klemm
University of Marburg

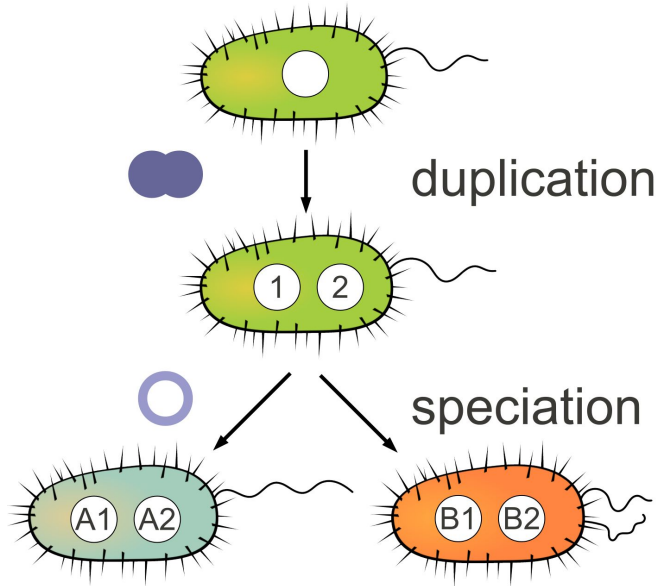
Introduction - Orthology Inference

The RBH and BLAST alternatives

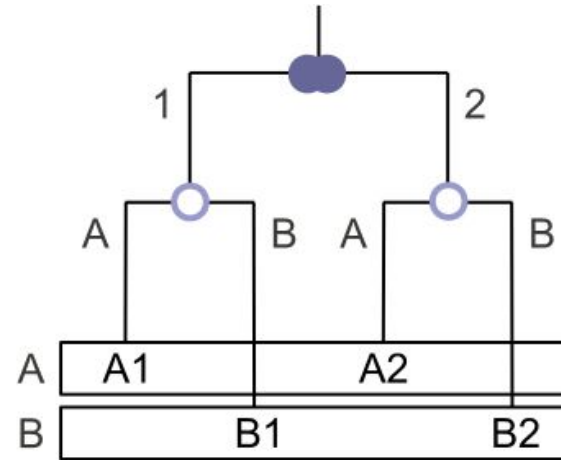
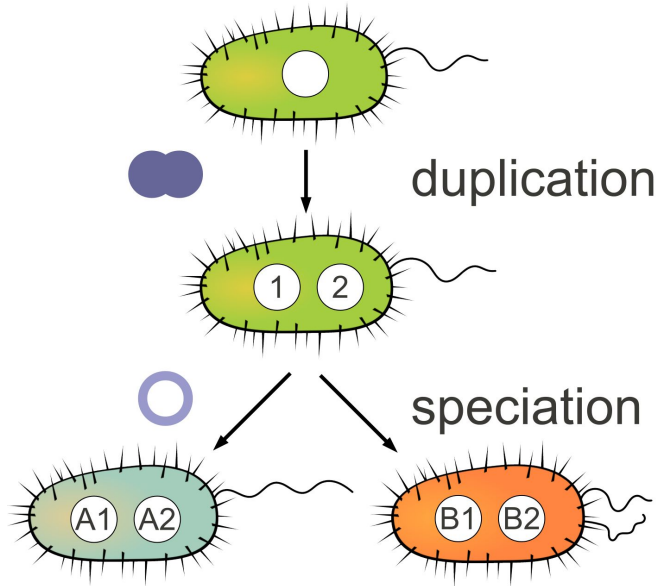
Spectral Graph Clustering



Introduction - Orthology Inference



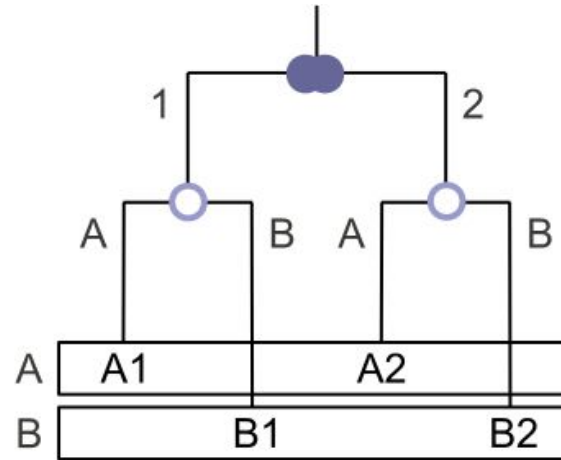
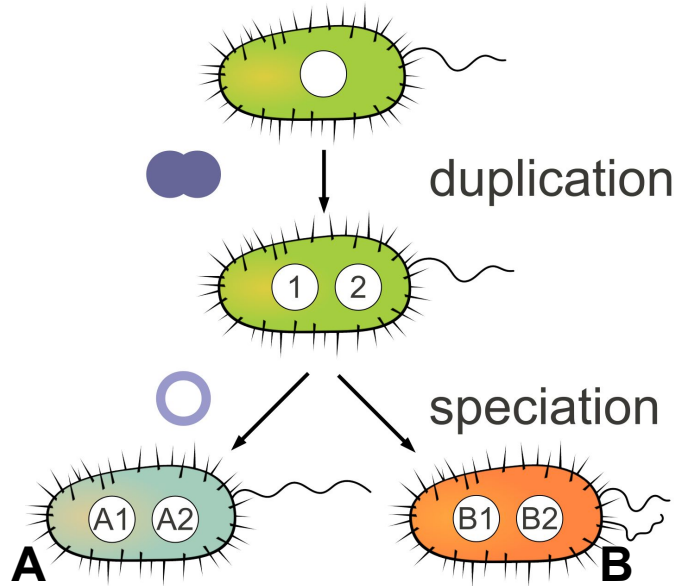
Introduction - Orthology Inference



● duplication
○ speciation
1/2: gene copy 1/2
A/B: species A/B



Introduction - Orthology Inference



● duplication
○ speciation
1/2: gene copy 1/2
A/B: species A/B

Orthologs : genes that originate in a speciation event

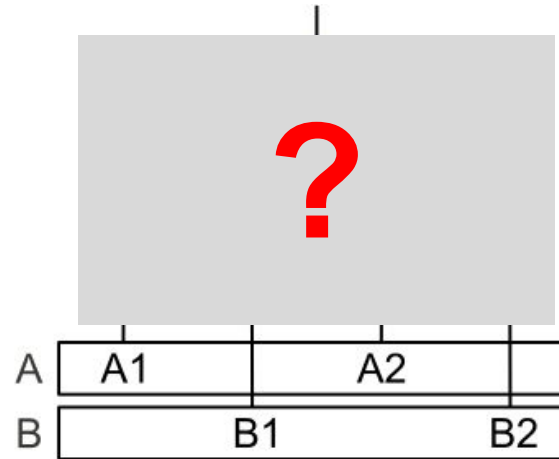
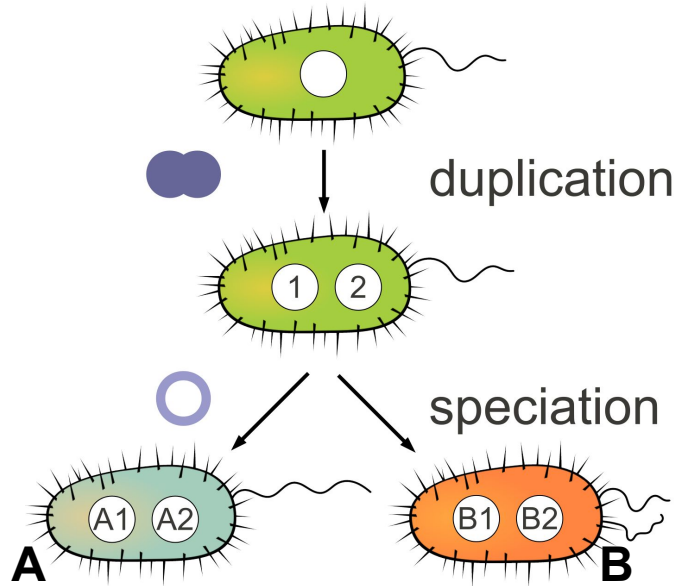
Paralogs : aka the rest



Walter M. Fitch



Introduction - Orthology Inference



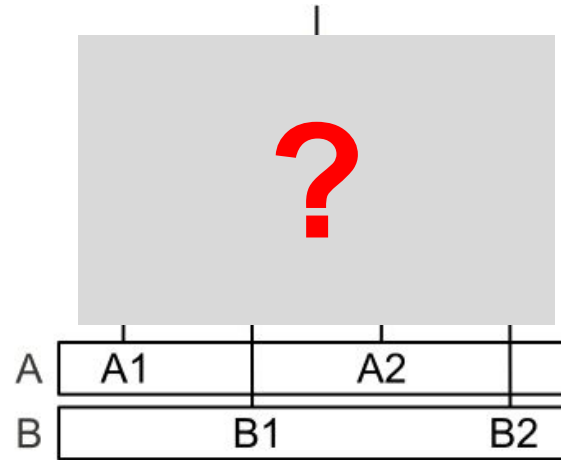
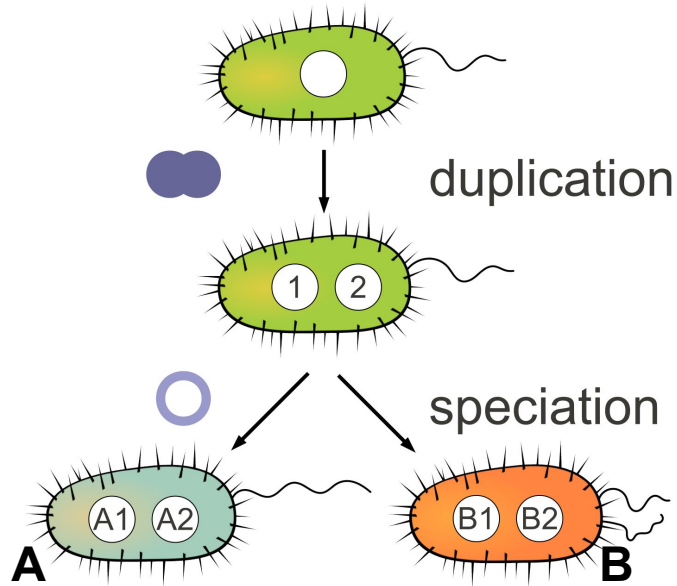
- duplication
- speciation
- 1/2: gene copy 1/2
- A/B: species A/B

Orthologs : genes that originate in a speciation event

Paralogs : aka the rest



Introduction - Orthology Inference



- duplication
- speciation
- 1/2: gene copy 1/2
- A/B: species A/B

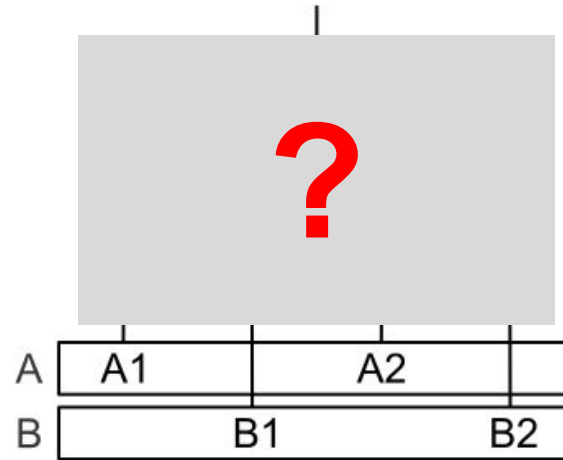
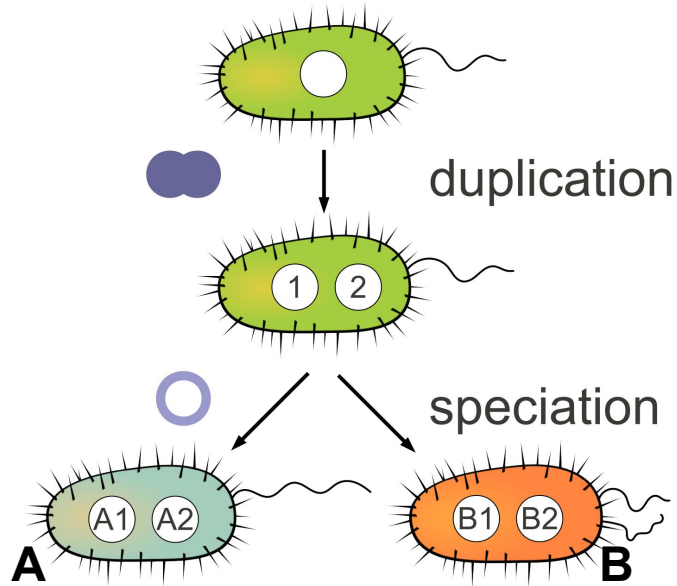
Orthologs : genes that originate in a speciation event

Paralogs : aka the rest

should have similar function



Introduction - Orthology Inference



- duplication
- speciation
- 1/2: gene copy 1/2
- A/B: species A/B

Orthologs : genes that originate in a speciation event

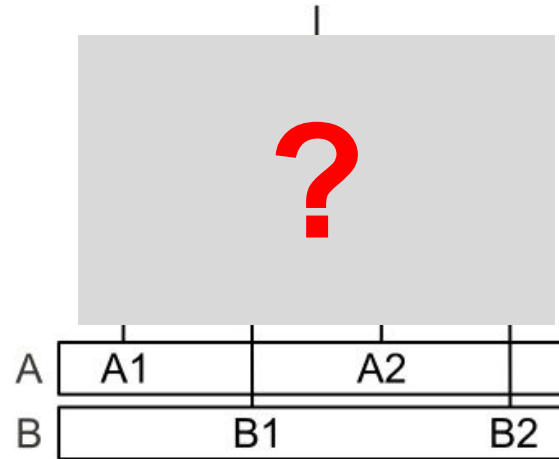
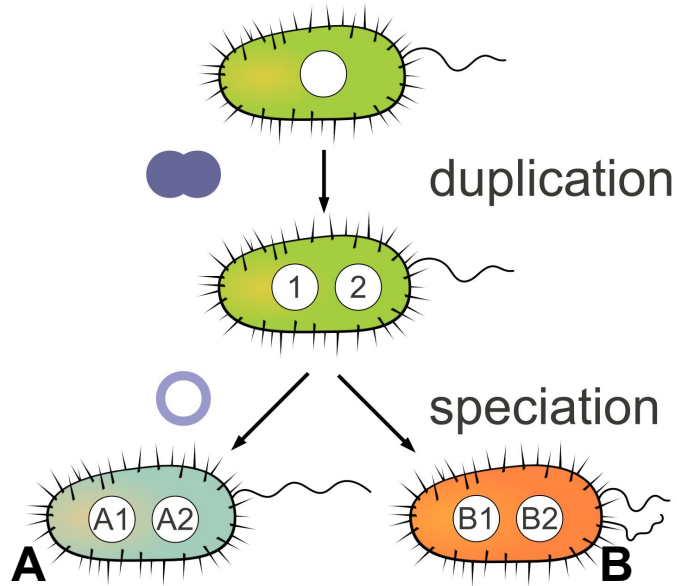
Paralogs : aka the rest

should have similar function

should have similar sequences



Introduction - Orthology Inference



- duplication
- speciation
- 1/2: gene copy 1/2
- A/B: species A/B

Orthologs : genes that originate in a speciation event

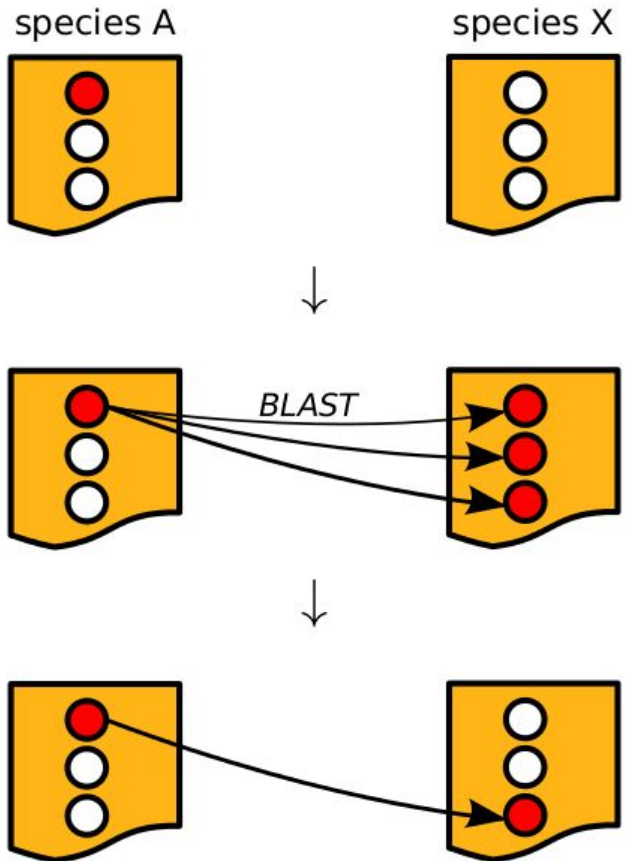
Paralogs : aka the rest

should have similar function

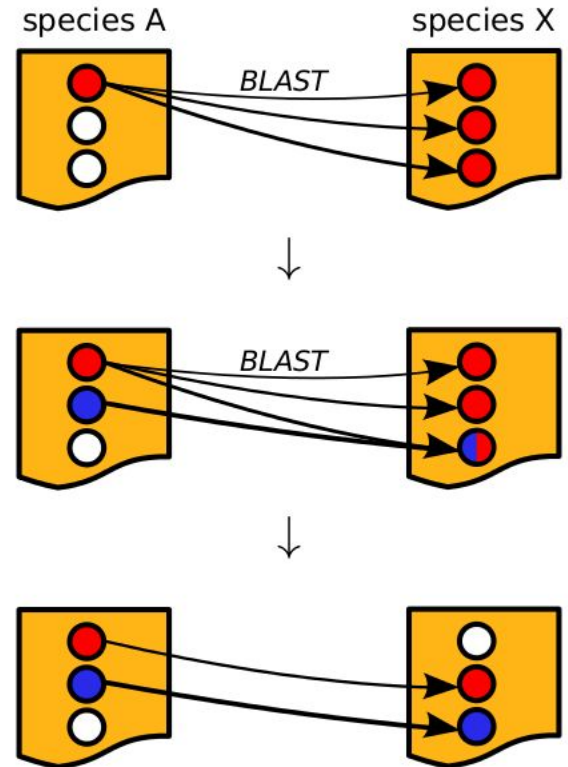
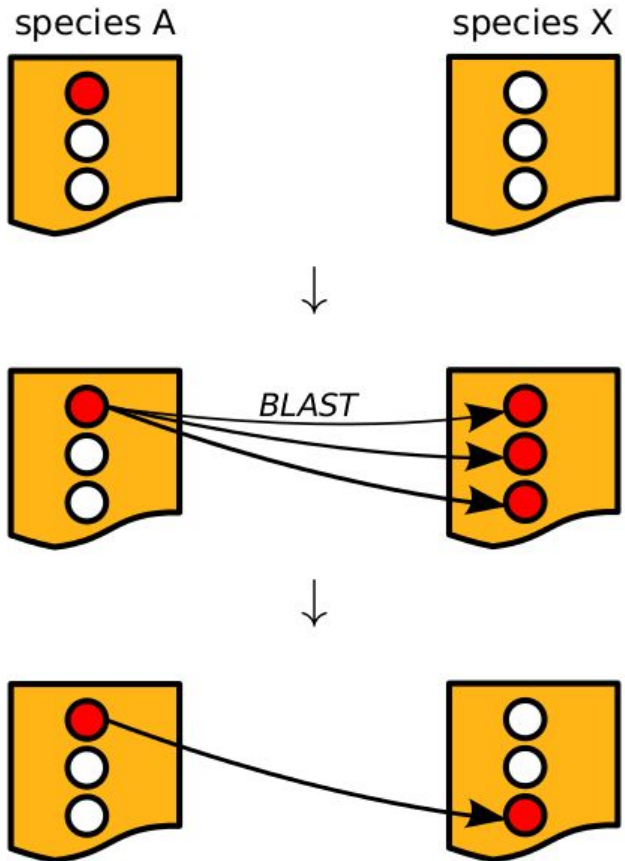
should have similar sequences



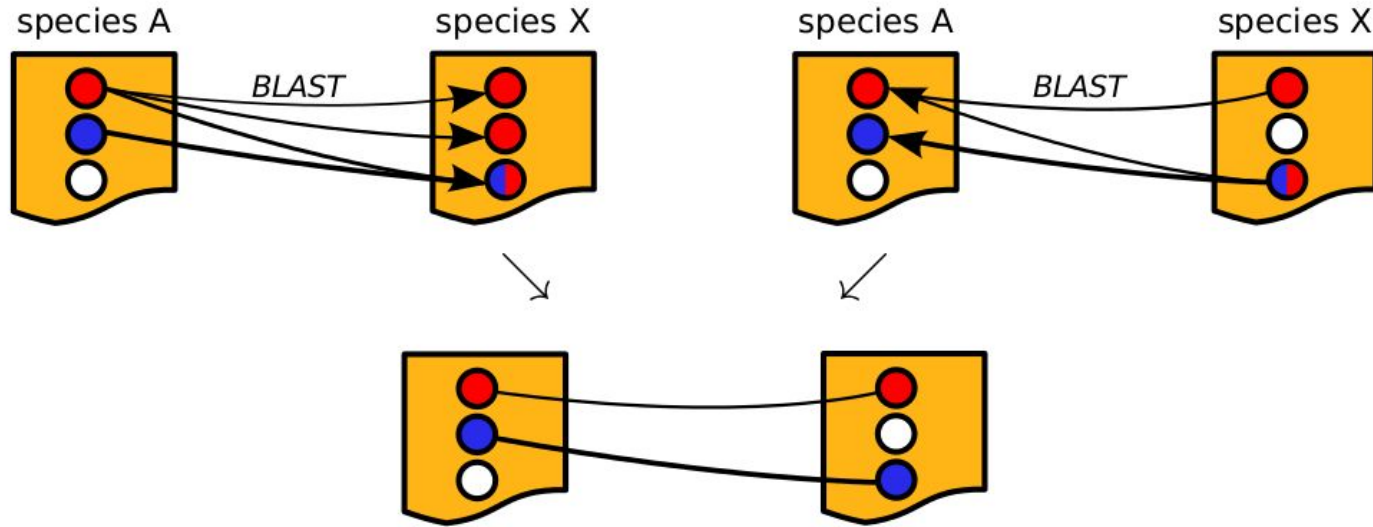
Introduction - Orthology Inference



Introduction - Orthology Inference



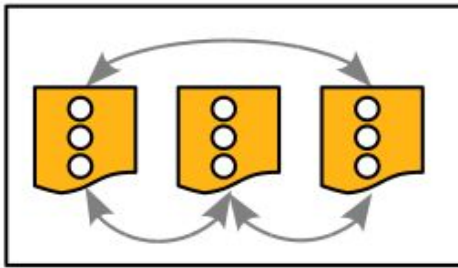
Introduction - Orthology Inference



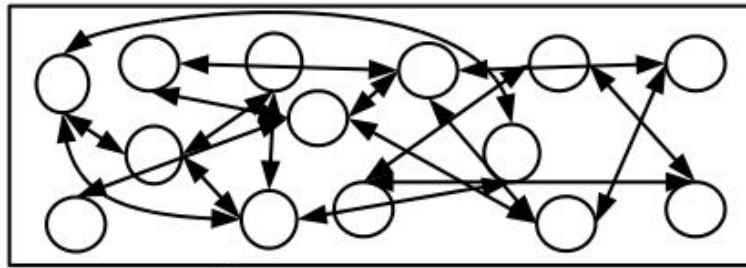
Introduction - Orthology Inference

Reciprocal Best (BLAST) Hit Heuristic

Step 1: all against all alignments



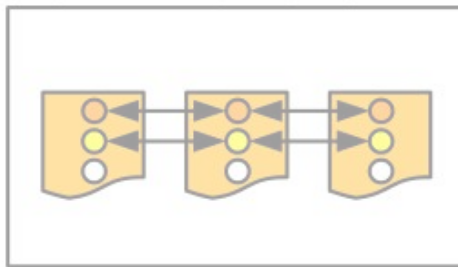
Step 2: filtering by similarity and graph conversion



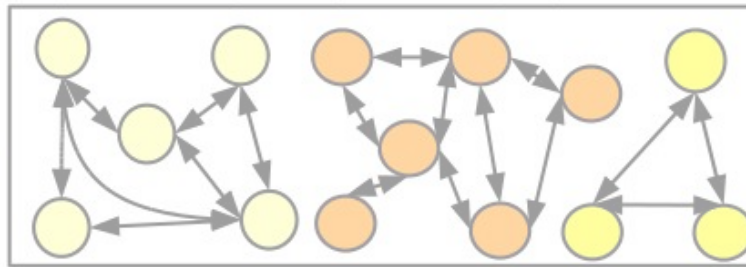
RBH



Step 4: orthologous groups



Step 3: clustering to acquire dense groups

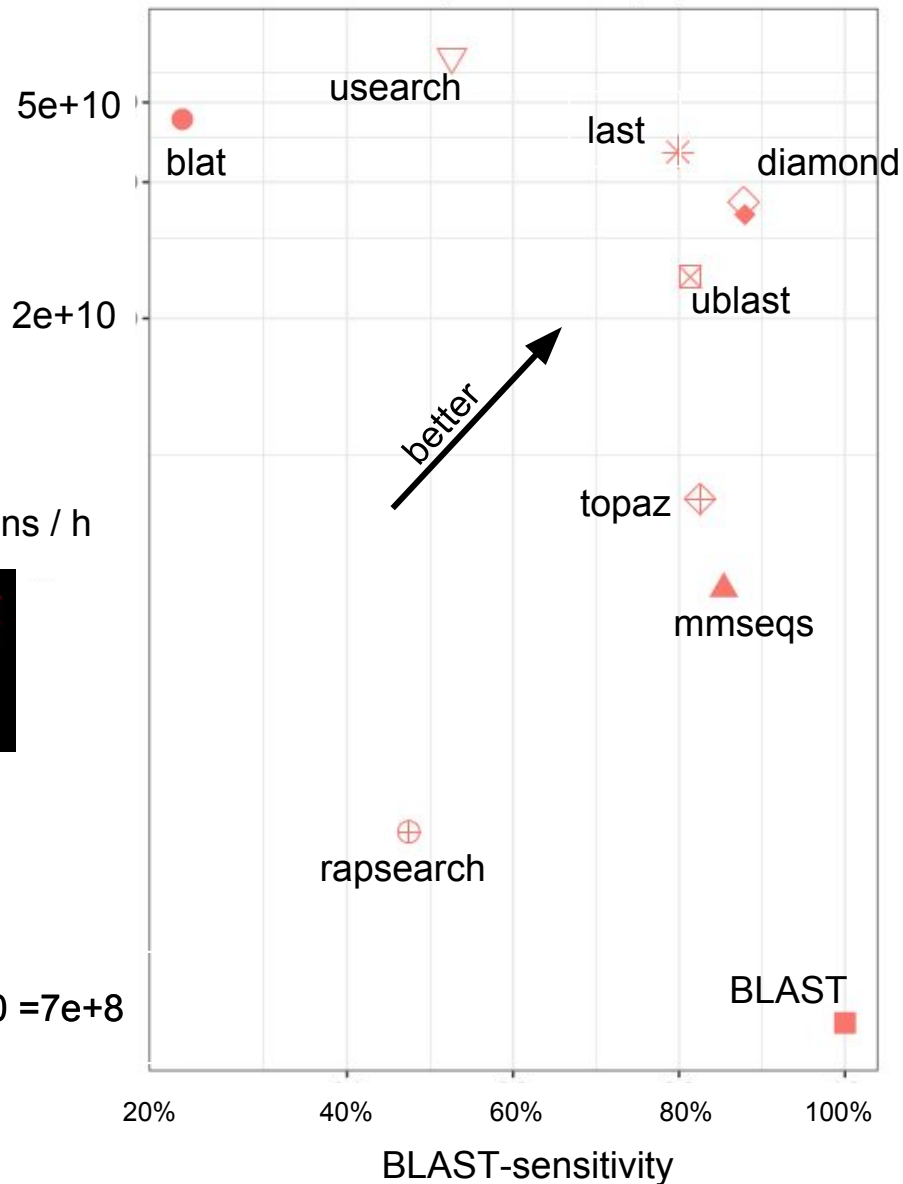


Legend:  species  gene  good alignment



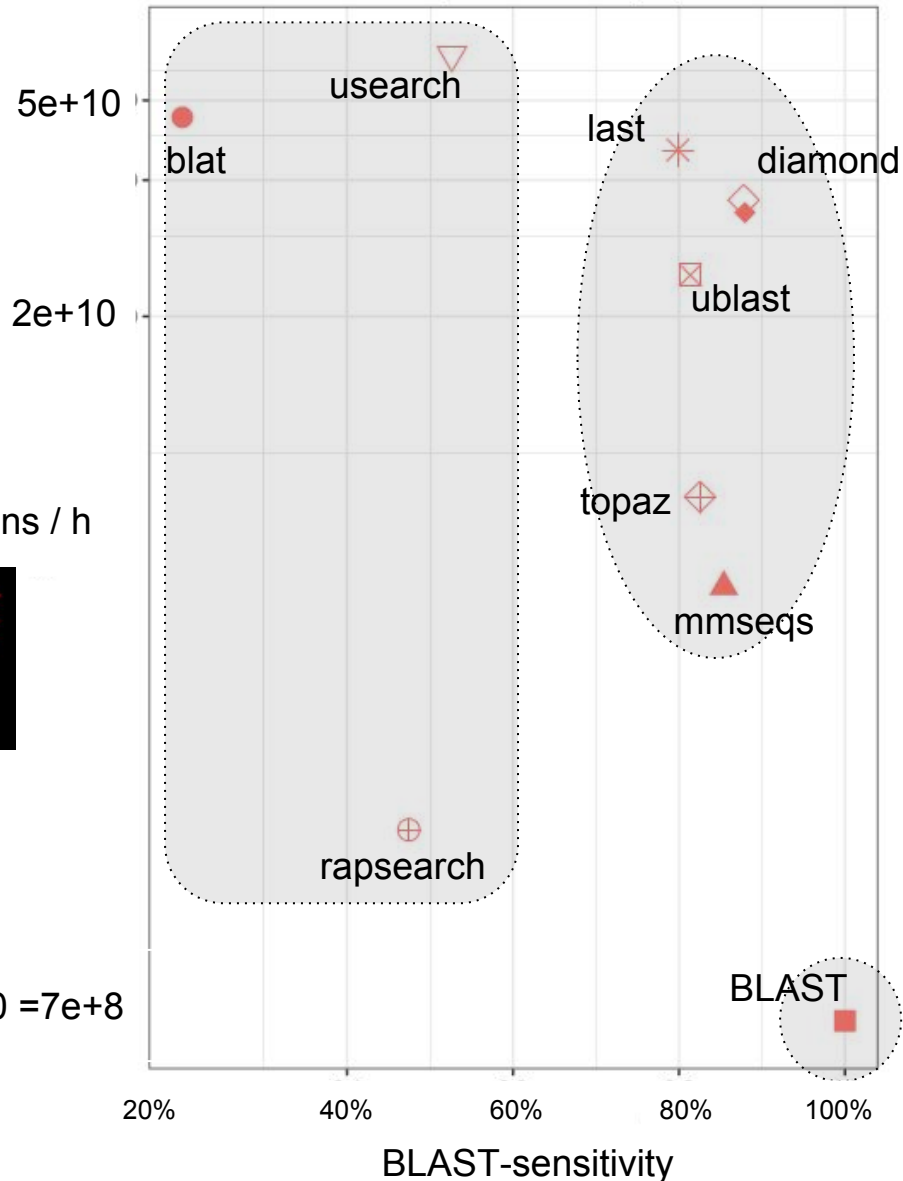
The RBH and BLAST alternatives

rank time and memory VS sensitivity , precision



The RBH and BLAST alternatives

rank time and memory VS sensitivity , precision



#comparisons / h

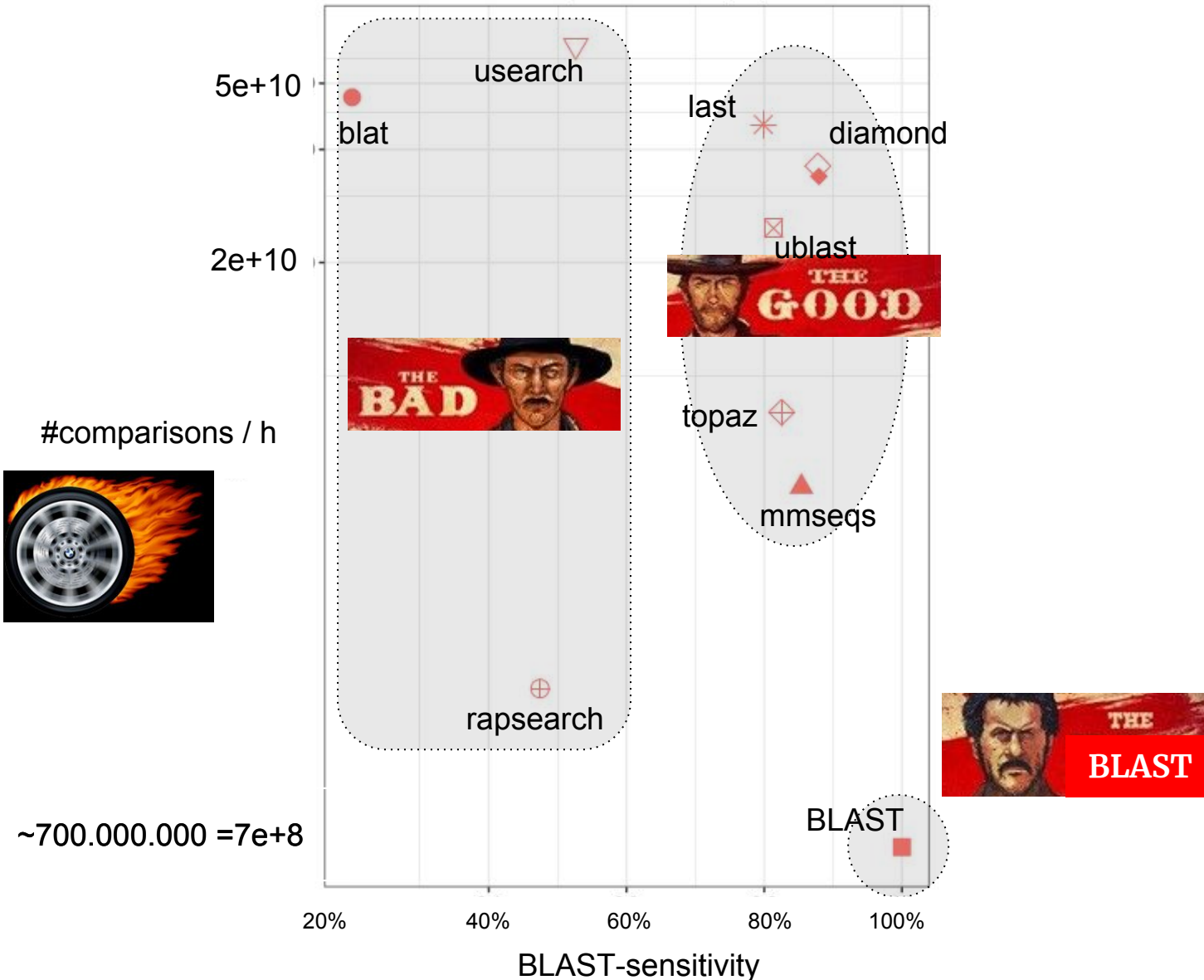


~700.000.000 = 7e+8



The RBH and BLAST alternatives

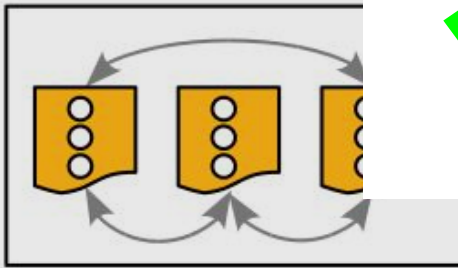
rank time and memory VS sensitivity , precision



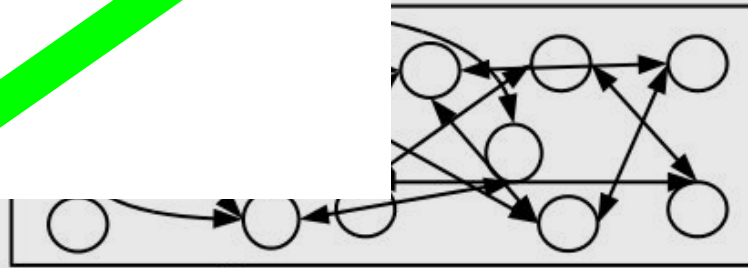
The RBH and BLAST alternatives

Reciprocal Best (BLAST)

Step 1: all against all align

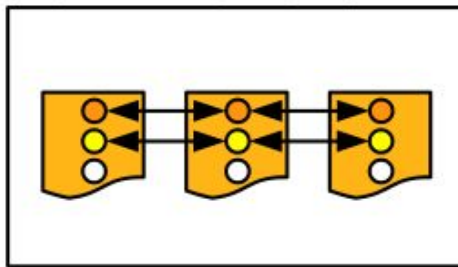


and graph conversion

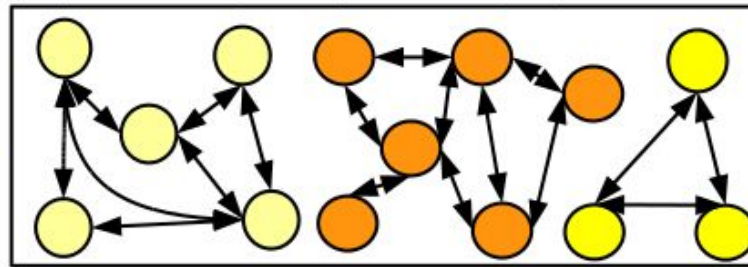


RBH

Step 4: orthologous groups



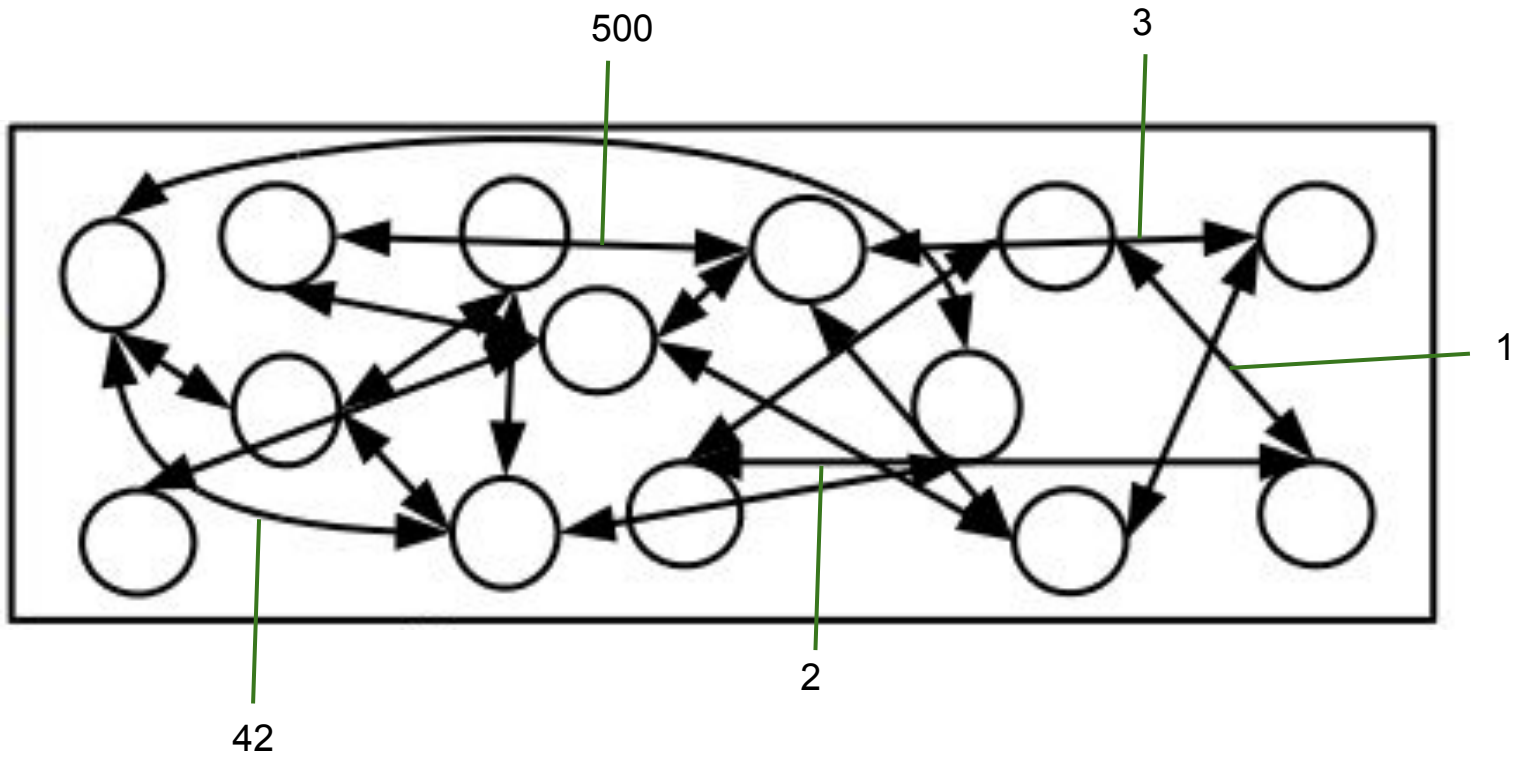
Step 3: clustering to acquire dense groups



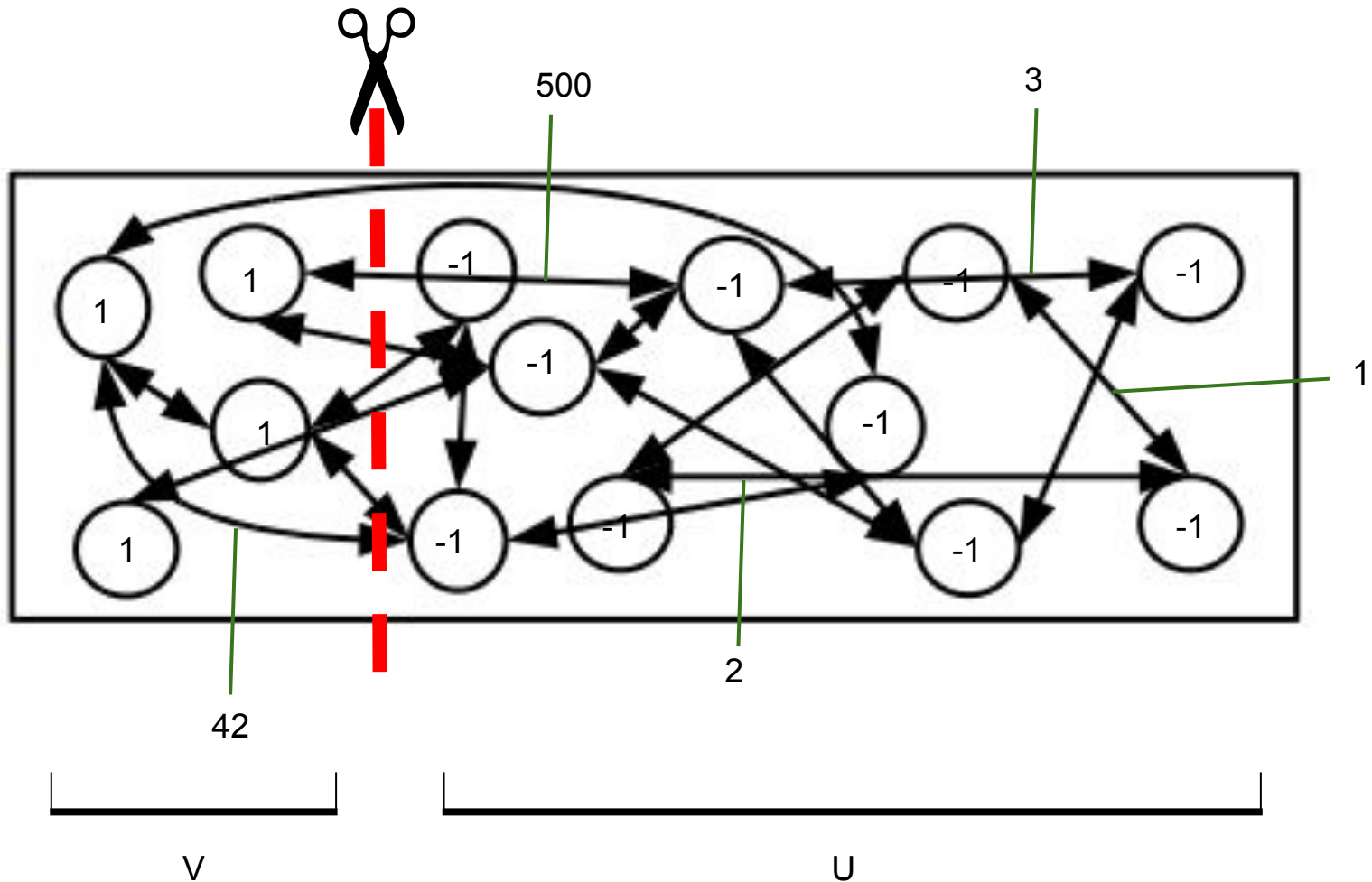
Legend:  species  gene  good alignment



Spectral Graph Clustering



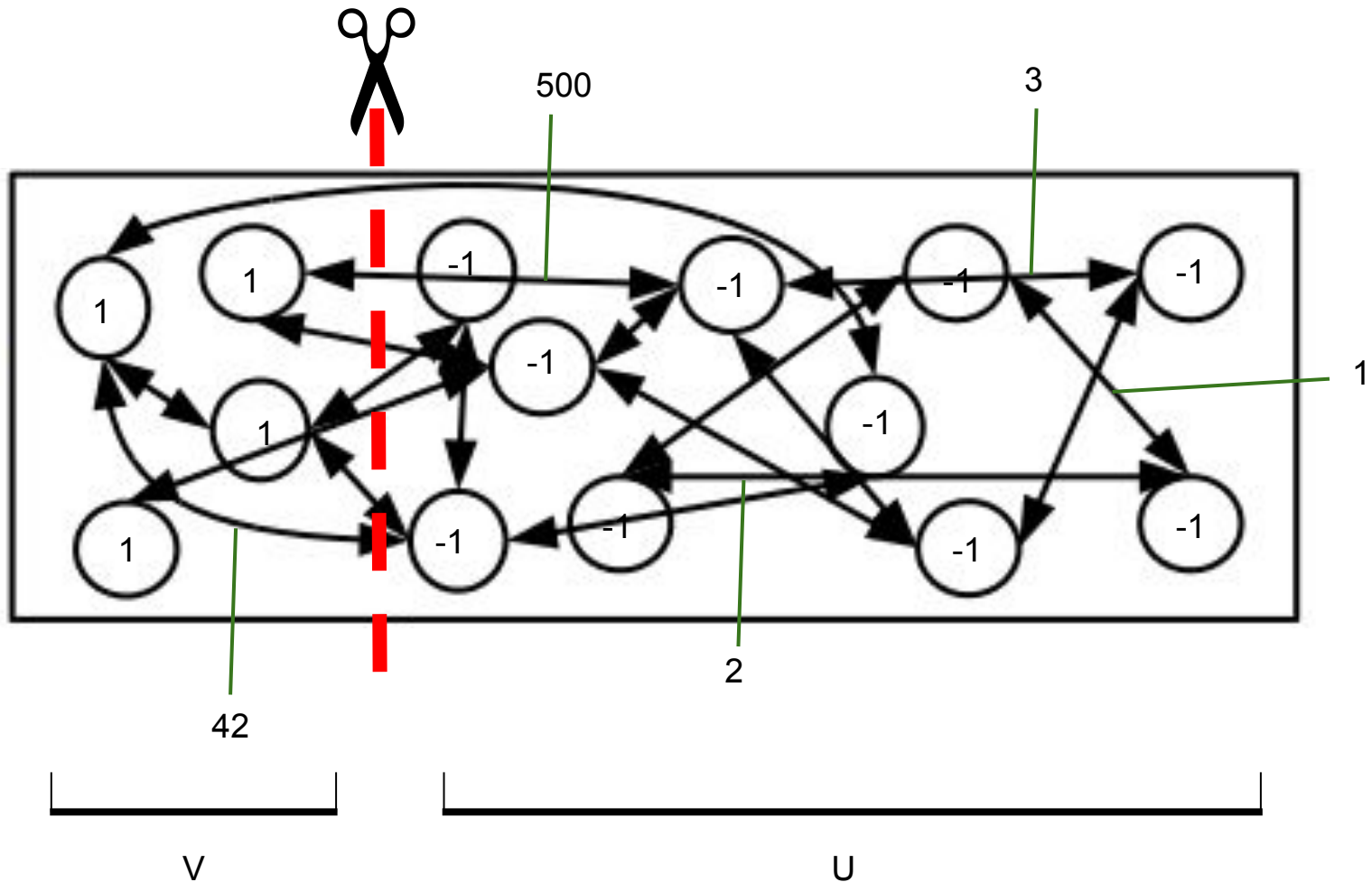
Spectral Graph Clustering



→ Min-Cut Problem



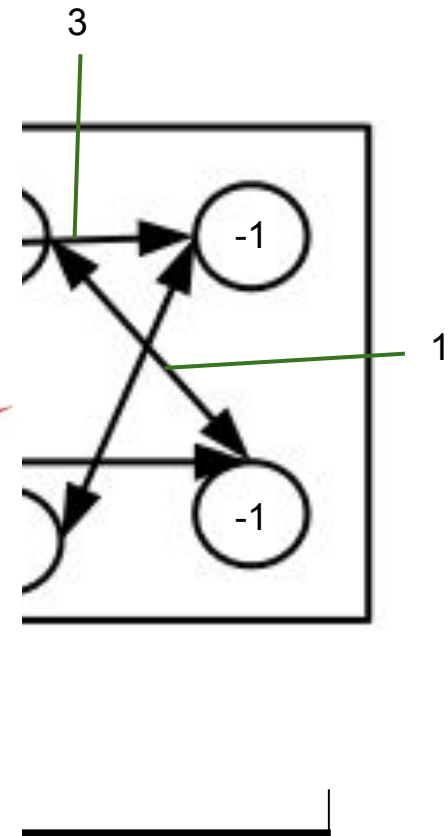
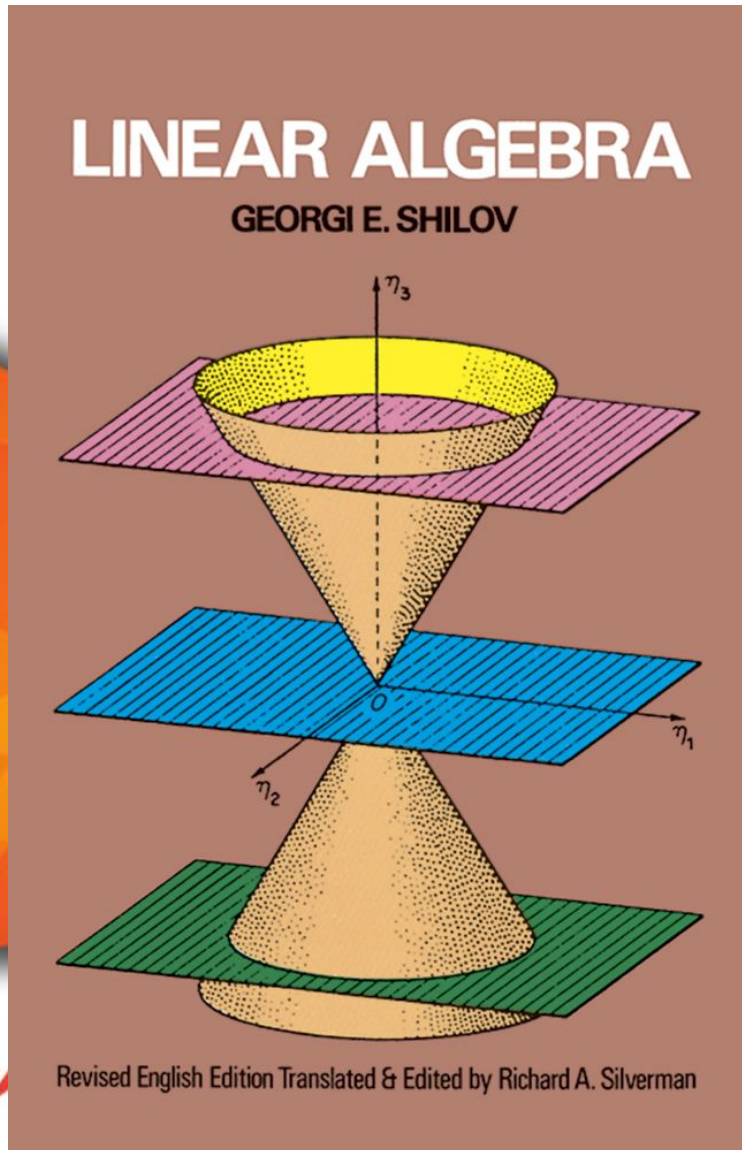
Spectral Graph Clustering



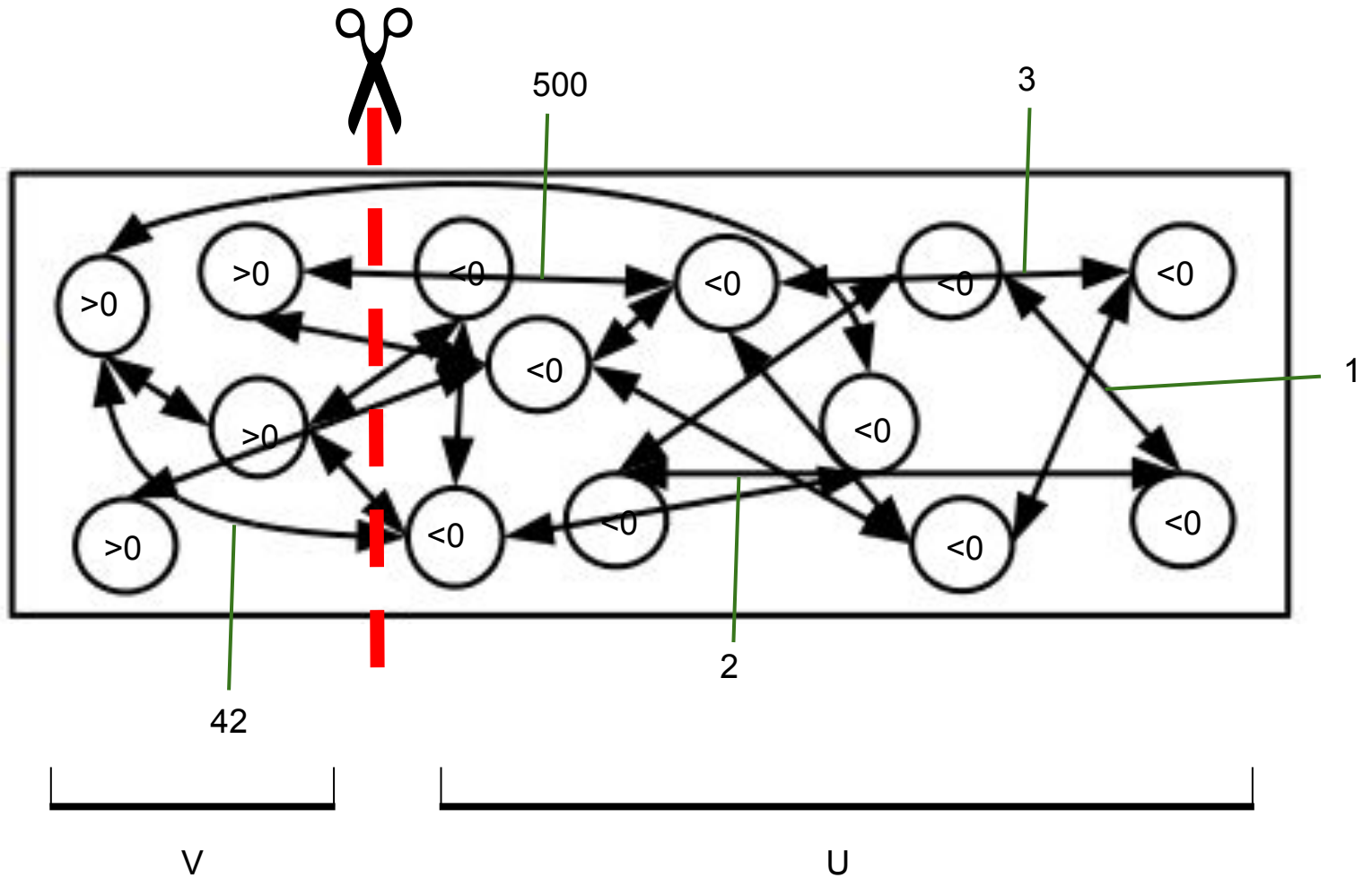
→ Min-Cut Problem (**NP-Complete**)



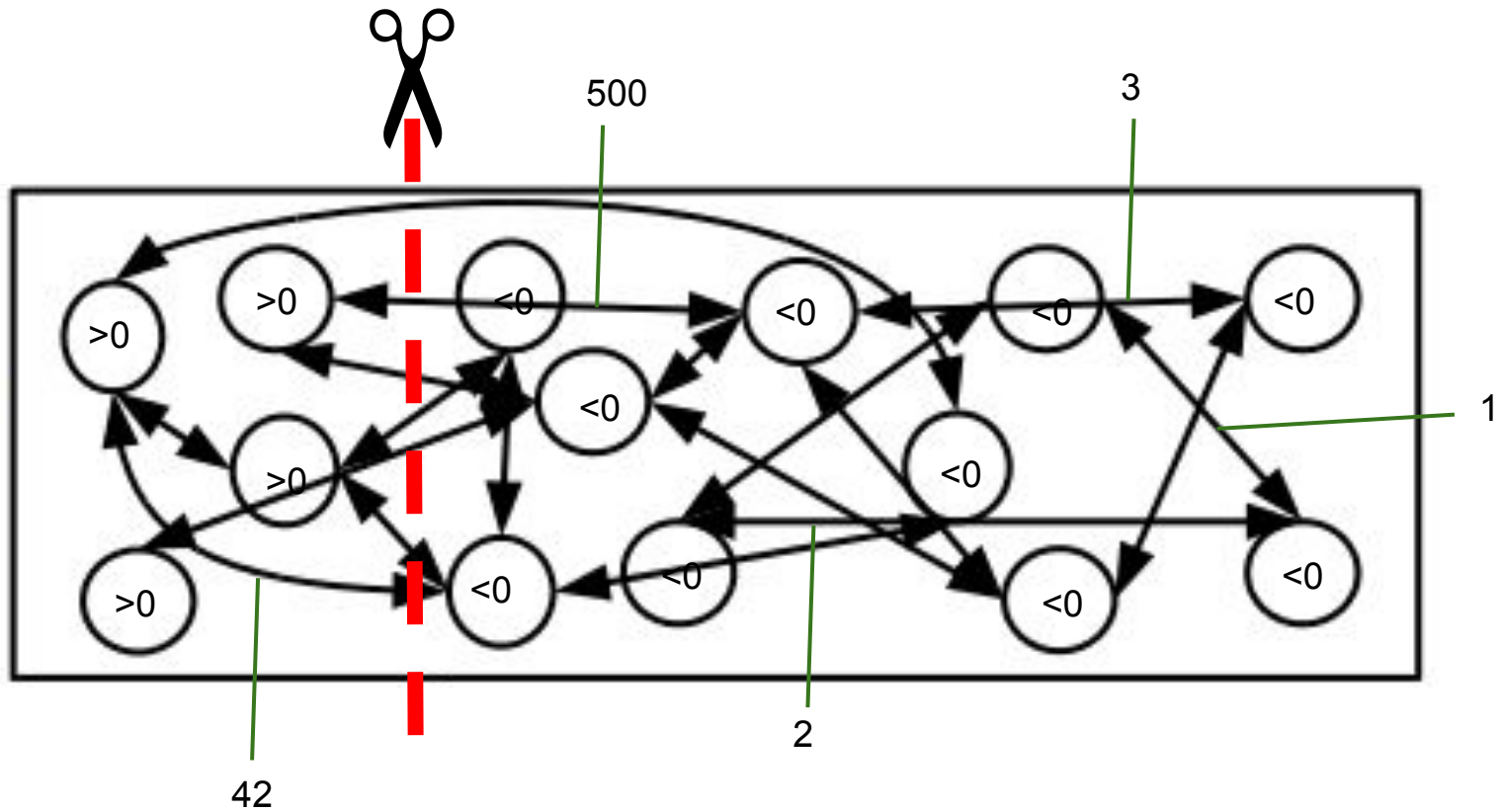
Spectral Graph Clustering



Spectral Graph Clustering



Spectral Graph Clustering



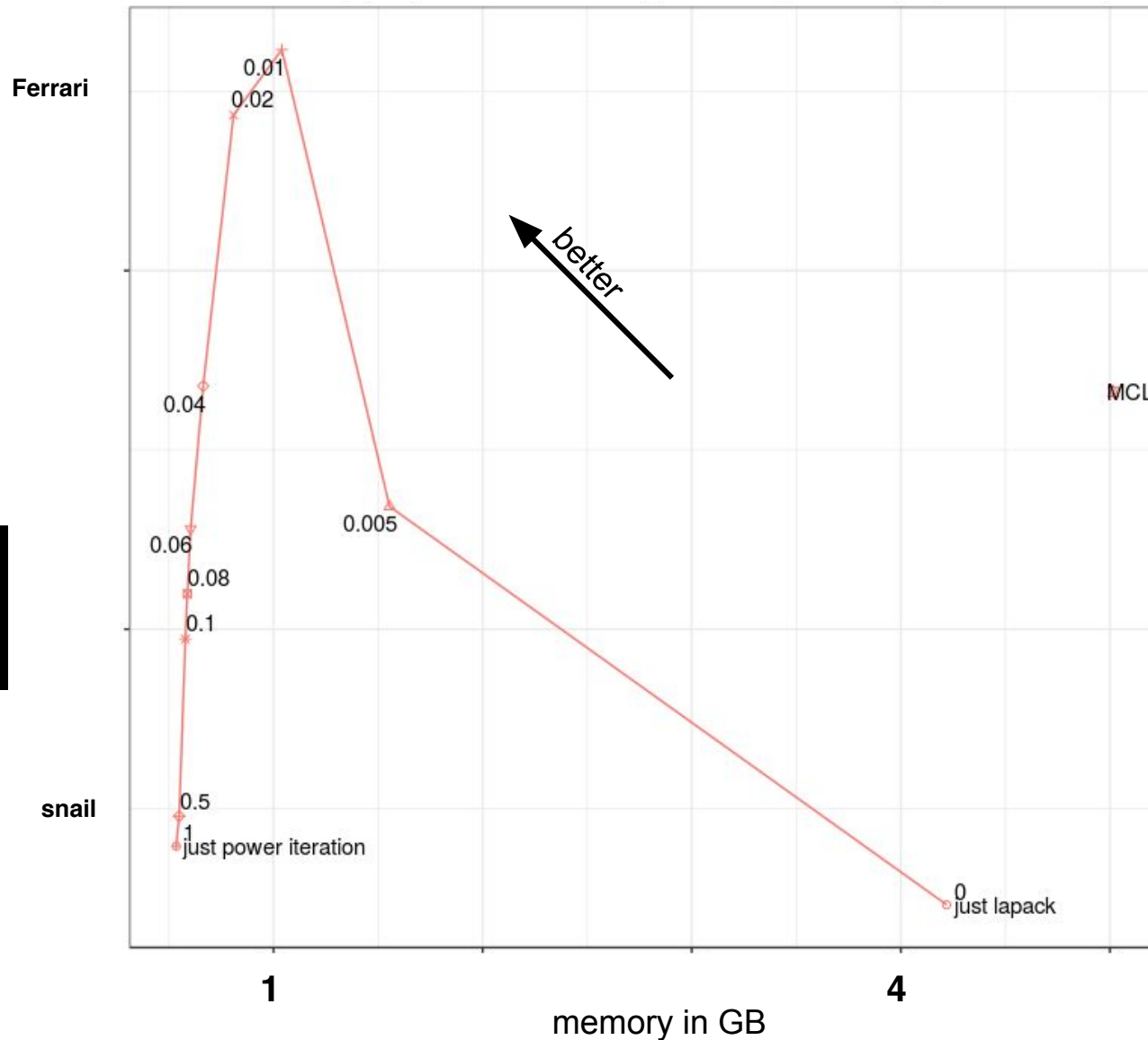
→ Eigenvalue Problem

- ❖ Power Iteration
- ❖ dsyevr (LAPACK Library)



Spectral Graph Clustering

LAPACK vs Power Iteration

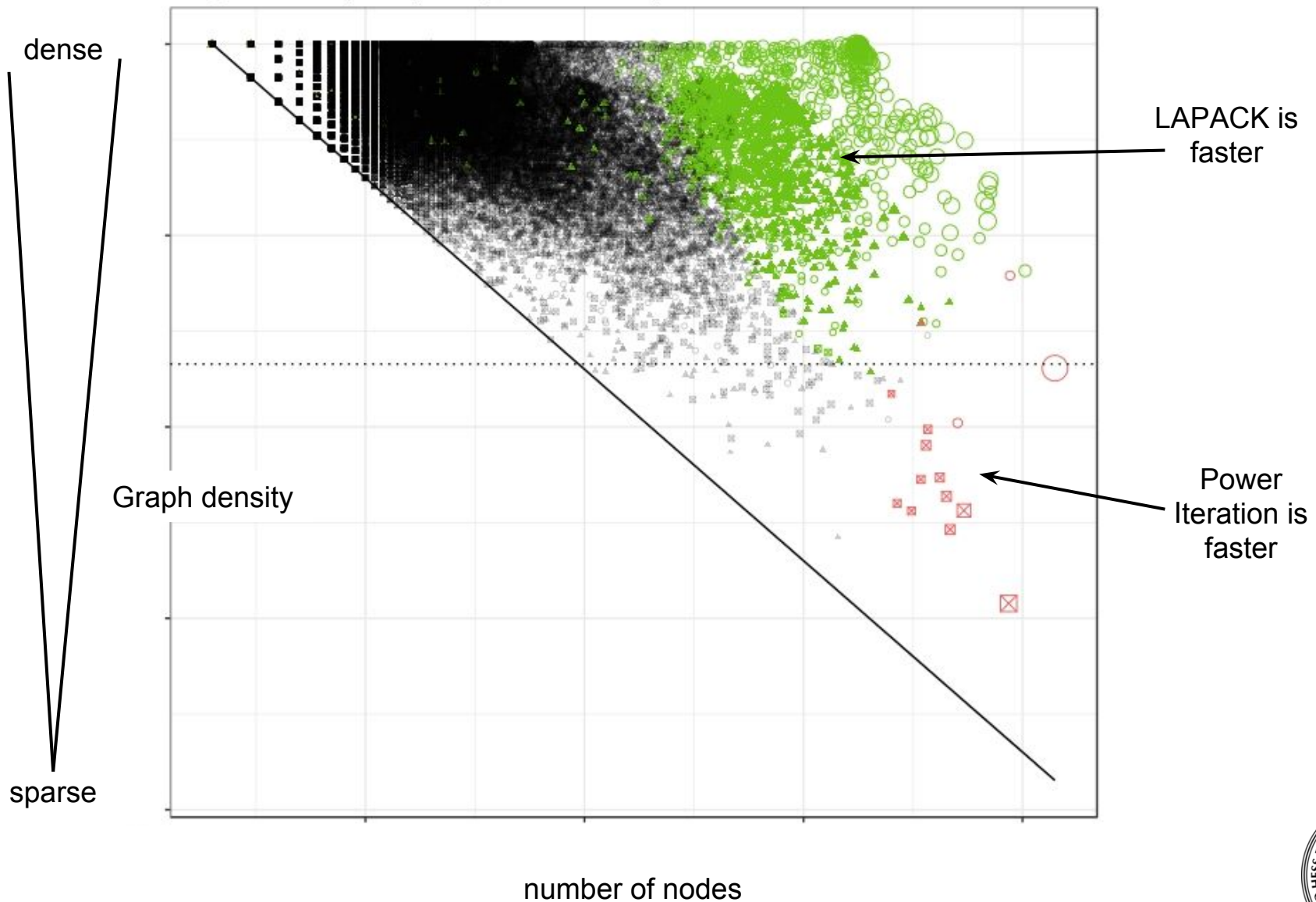


#edges / min

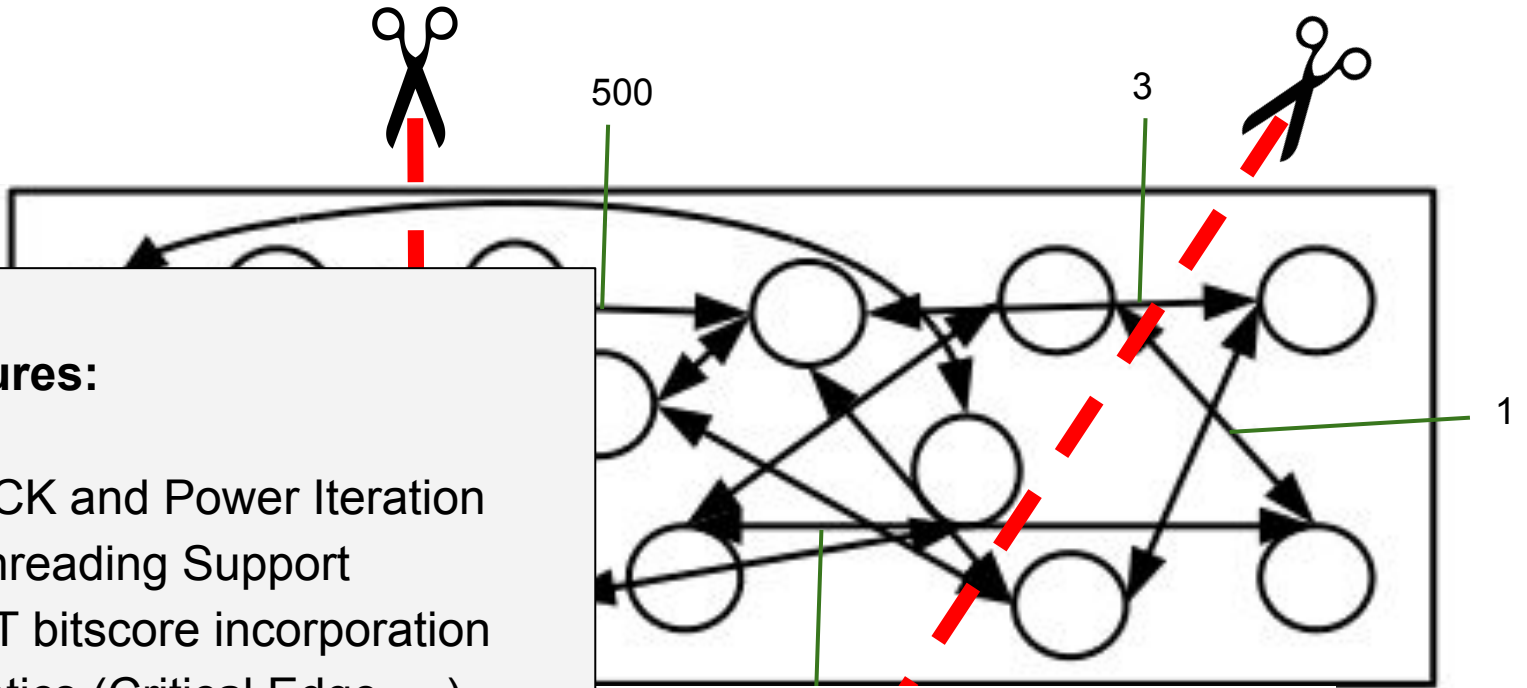


Spectral Graph Clustering

LAPACK vs Power Iteration

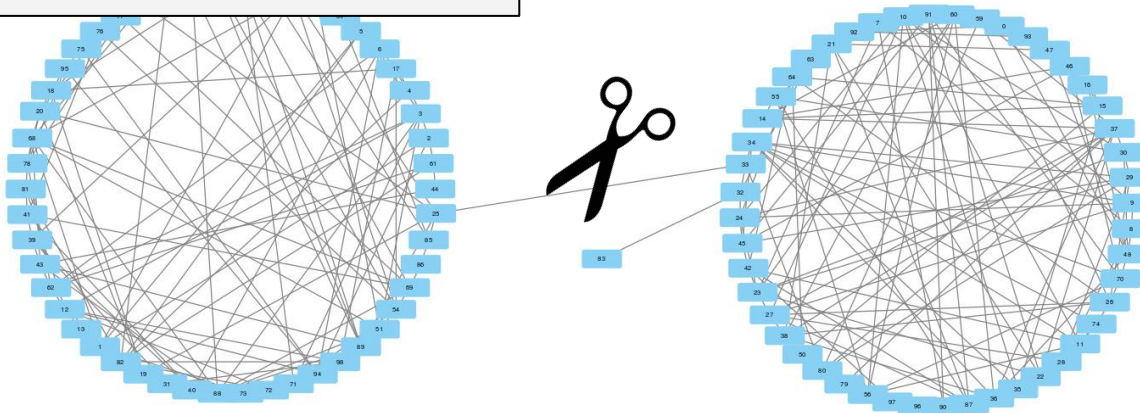


Spectral Graph Clustering



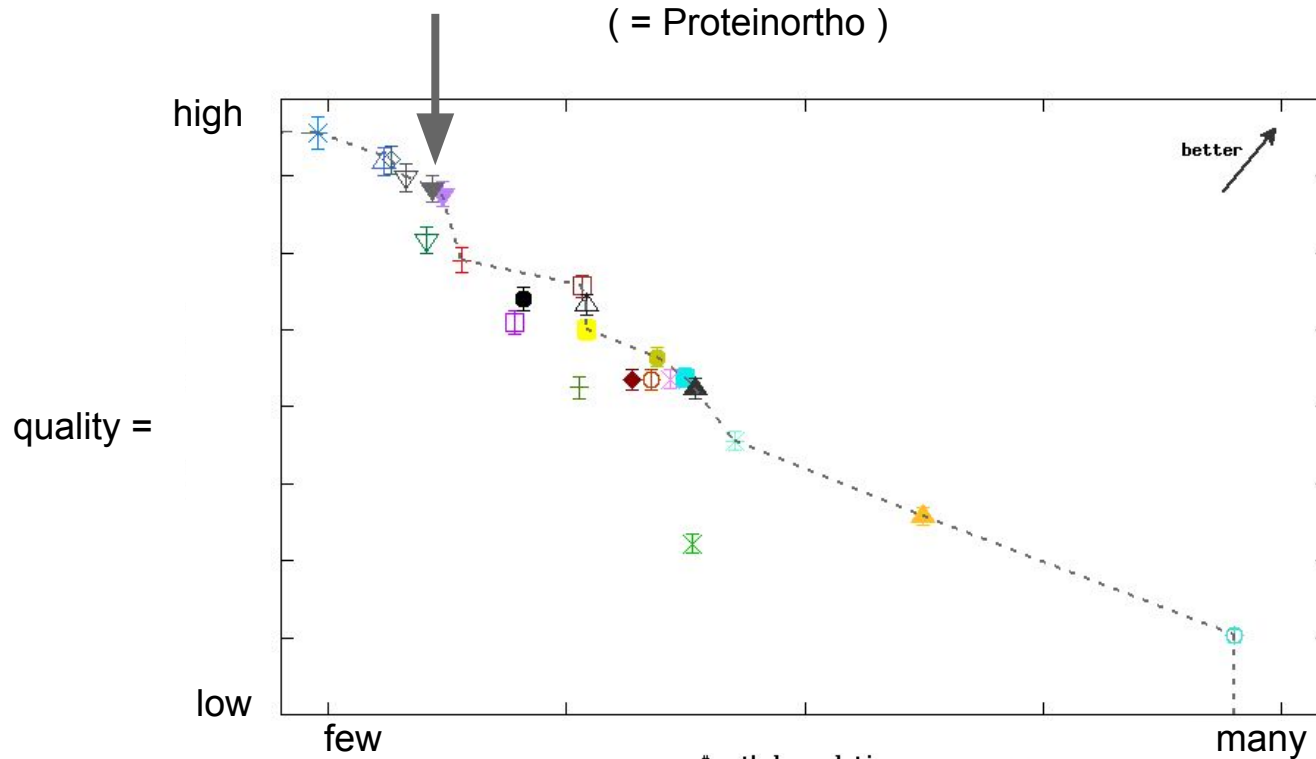
New Features:

- + LAPACK and Power Iteration
- + Multithreading Support
- + BLAST bitscore incorporation
- + Heuristics (Critical Edge, ...)



(= Proteinortho)



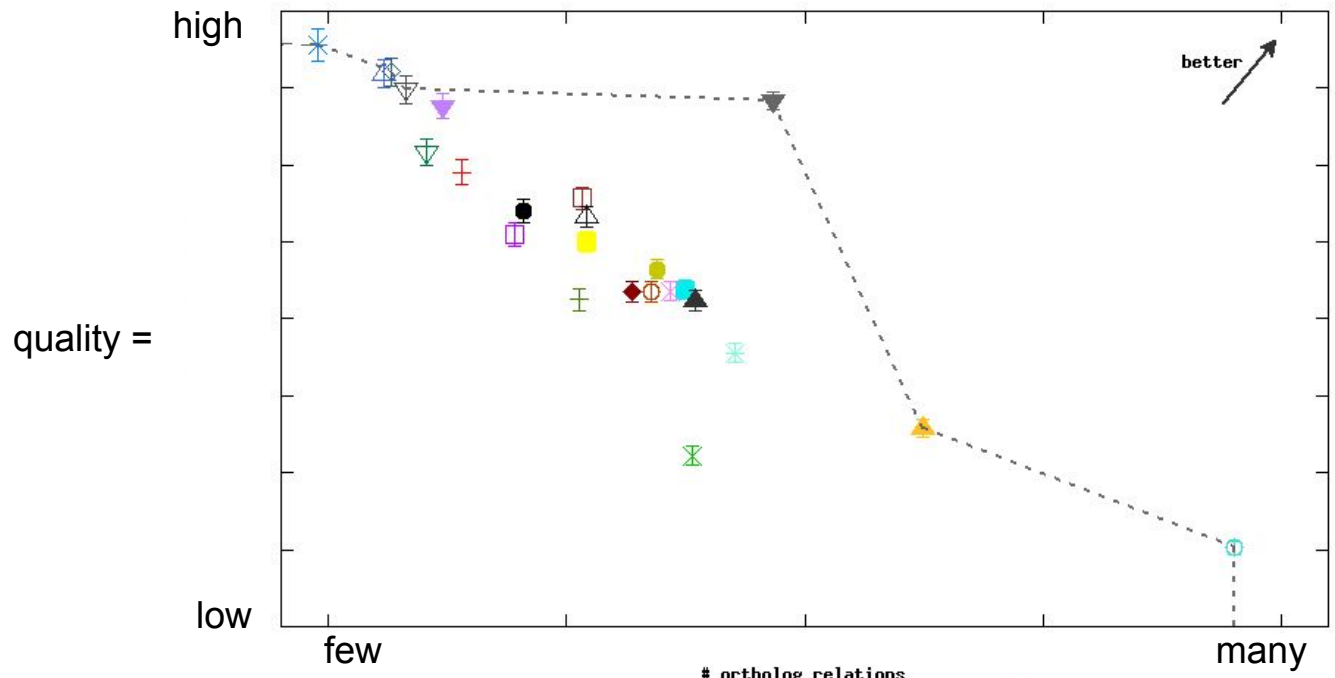


sensitivity =

- EggNOG
- GETHOgs 2.0
- OMA Groups 2.0
- OMA Pairs 2.0
- OrthoFinder 2.0 (Defaults, with BLAST)
- OrthoFinder 2.0 (Defaults, with DIAMOND)
- OrthoFinder 2.0 (MSA, with BLAST)
- PANTHER 8.0 (LDO only)
- PANTHER 8.0 (all)
- EggNOG 5.0 (Fine-grained)
- RBH / BBH
- RSD 0.8 1e-5 Deluca
- SonicParanoid (default)
- SonicParanoid (fast)
- SonicParanoid (sensitive)
- metaPhOrs
- orthoinspector 1.30 (blast threshold 10-9)
- phyloneDB
- EggNOG 5.0 (Groups)
- Ensembl Compara (e81)
- Hieranoid 2
- InParanoid
- InParanoidCore
- 2011 blastp blast-graph



(= Proteinortho)

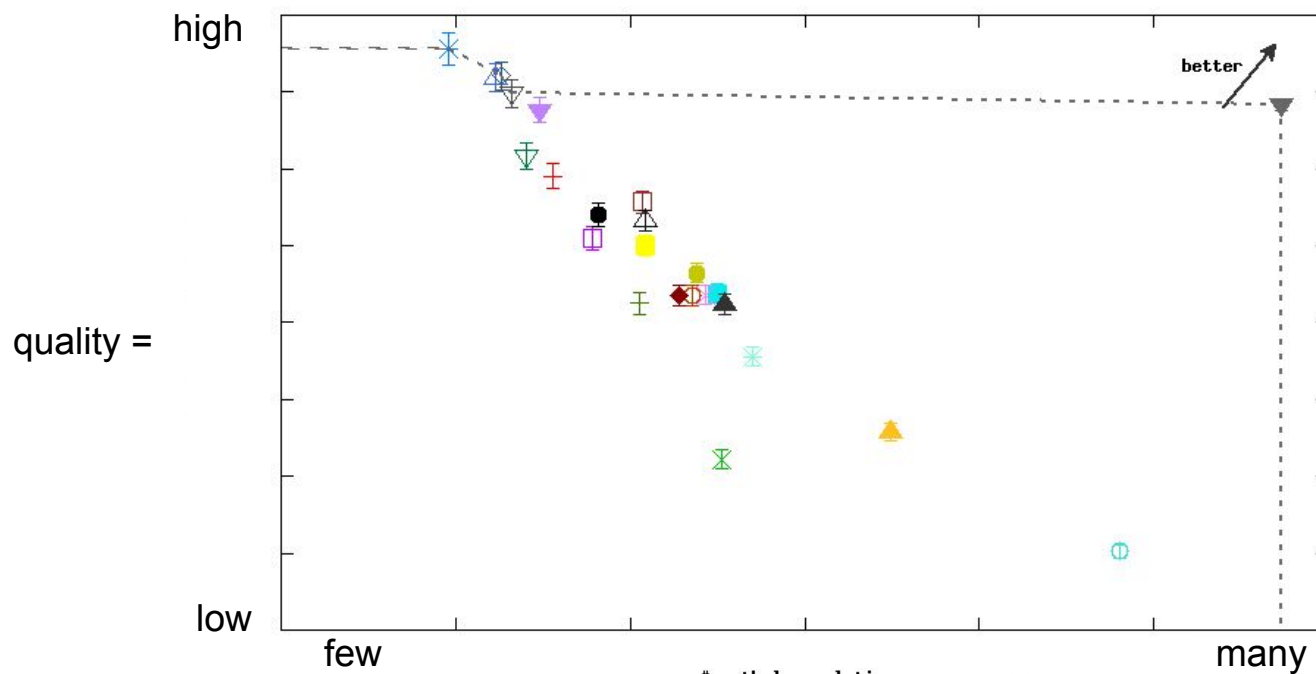


sensitivity =

- EggNOG
- GETHOgs 2.0
- OMA Groups 2.0
- OMA Pairs 2.0
- OrthoFinder 2.0 (Defaults, with BLAST)
- OrthoFinder 2.0 (Defaults, with DIAMOND)
- OrthoFinder 2.0 (MSA, with BLAST)
- PANTHER 8.0 (LDO only)
- PANTHER 8.0 (all)
- EggNOG 5.0 (Fine-grained)
- RBH / BBH
- RSD 0.8 1e-5 Deluca
- SonicParanoid (default)
- SonicParanoid (fast)
- SonicParanoid (sensitive)
- metaPhOrs
- orthoinspector 1.30 (blast threshold 10-9)
- phyloneDB
- EggNOG 5.0 (Groups)
- Ensembl Compara (e81)
- Hieranoid 2
- InParanoid
- InParanoidCore
- 2011 blastp bothdirections edge



(= Proteinortho)



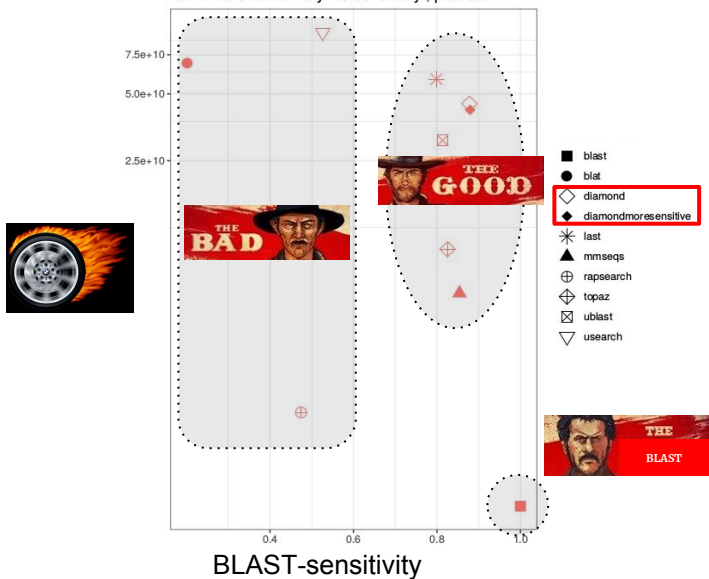
sensitivity =

- EggNOG
- GETHOgs 2.0
- OMA Groups 2.0
- OMA Pairs 2.0
- OrthoFinder 2.0 (Defaults, with BLAST)
- OrthoFinder 2.0 (Defaults, with DIAMOND)
- OrthoFinder 2.0 (MSA, with BLAST)
- PANTHER 8.0 (LDO only)
- PANTHER 8.0 (all)
- EggNOG 5.0 (Fine-grained)
- RBH / BBH
- RSD 0.8 1e-5 Deluca
- SonicParanoid (default)
- SonicParanoid (fast)
- SonicParanoid (sensitive)
- metaPhOrs
- orthoinspector 1.30 (blast threshold 10-9)
- phyloneDB
- EggNOG 5.0 (Groups)
- Ensembl Compara (e81)
- Hieranoid 2
- InParanoid
- InParanoidCore
- 2011 blastp bothdirections edge 2times



RBH All-Versus-All BLAST

rank time and memory VS sensitivity , precision

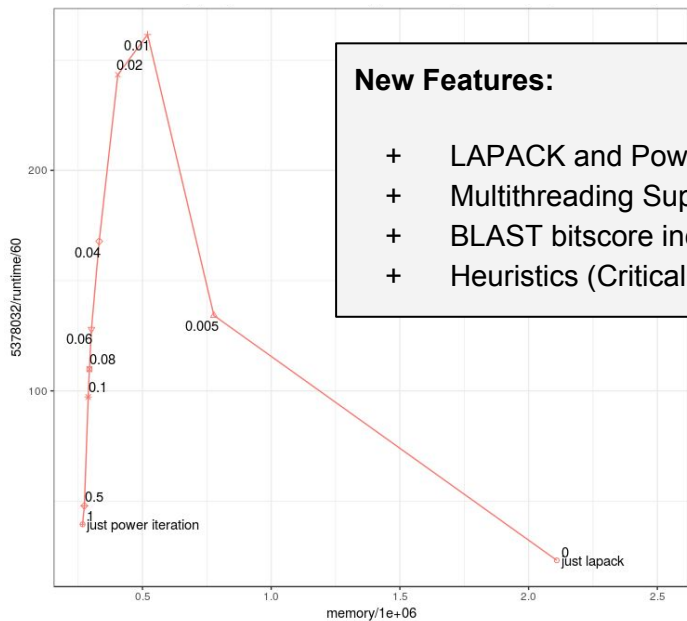


Thank you

and

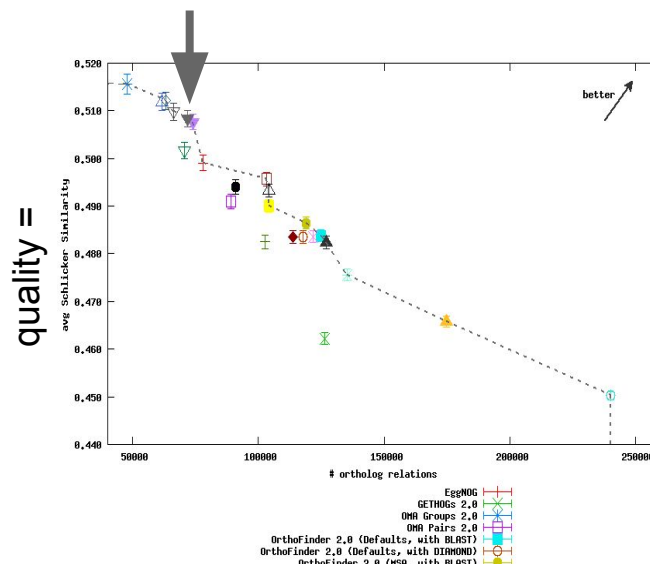


Spectral Graph Clustering

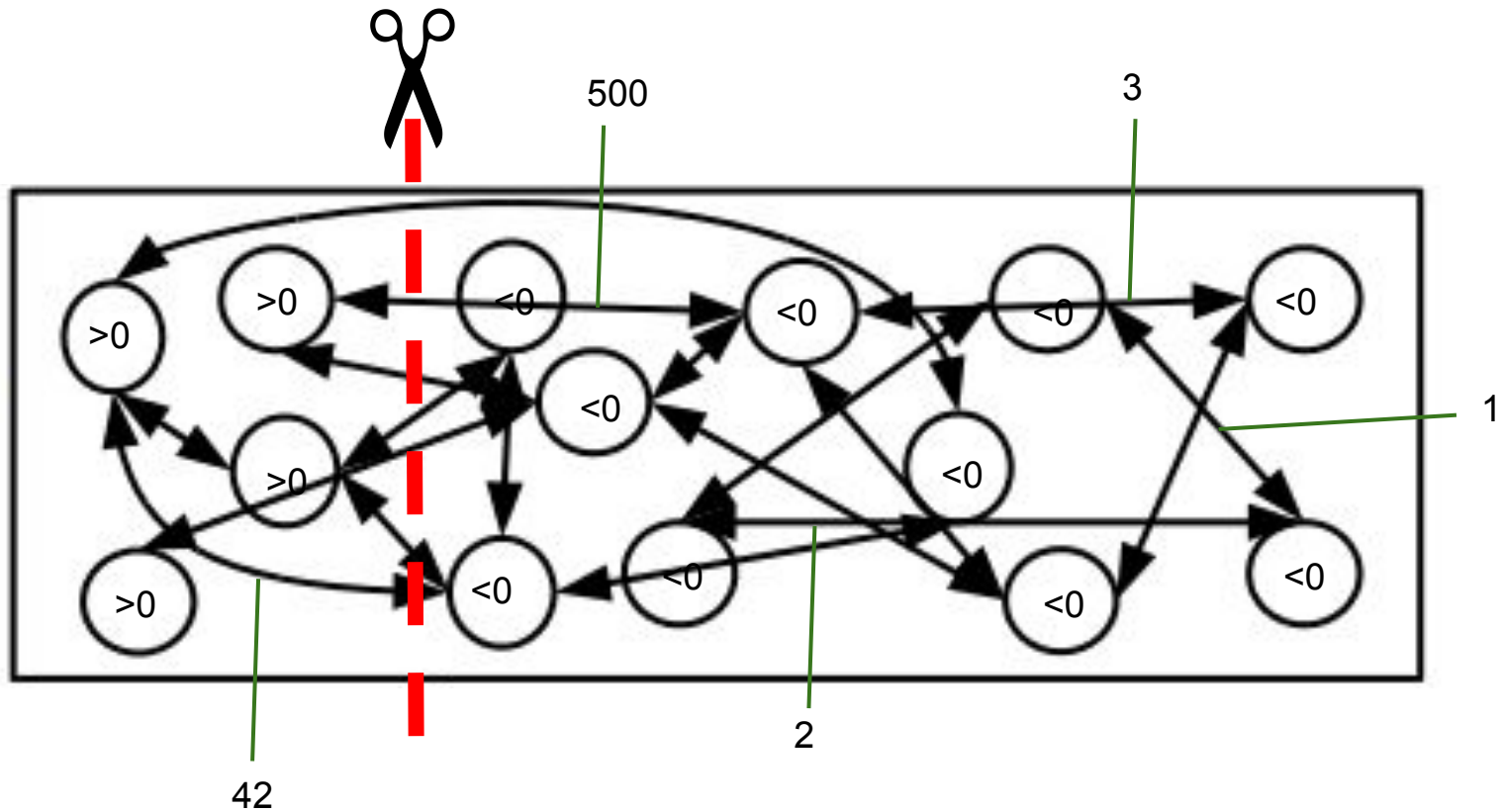


- New Features:**
- + LAPACK and Power Iteration
 - + Multithreading Support
 - + BLAST bitscore incorporation
 - + Heuristics (Critical Edge, ...)

RBH All-Versus-All BLAST AND Spectral Graph Clustering



Spectral Graph Clustering



→ Eigenvalue Problem

- ❖ Power Iteration $O(\text{number of nodes} * \text{number of edges})$
- ❖ dsyevr (LAPACK Library) $O(\text{number of nodes})$





(Need for Speed) Proteinortho Nitro

Paul Klemm
University of Marburg

Introduction - Orthology Inference

The RBH and BLAST alternatives

Spectral Graph Clustering



(Need for Speed)
Proteinortho Nitro

