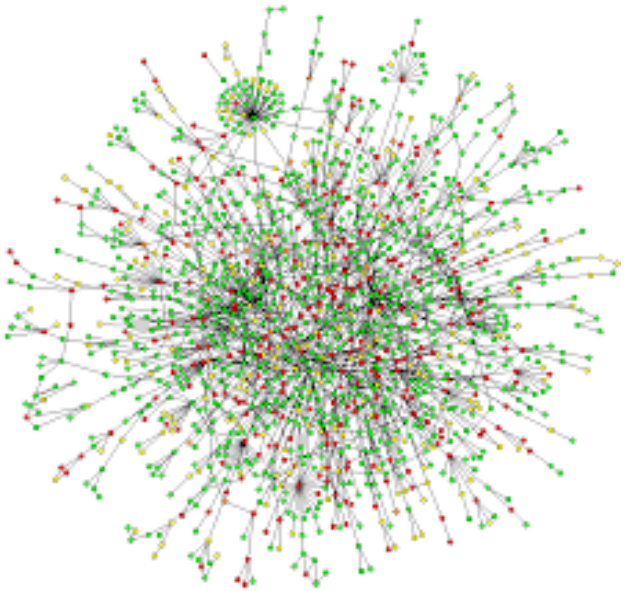
AligNet:Alignment of Protein-Protein Interaction Networks

A. Alcalá R. Alberich M. Llabrés F. Rosselló G. Valiente

University of the Balearic Islands

Bioinformatics Day @ DAIS
July 7, 2016

Protein-Protein Interaction Networks (PPI)



- Nodes are proteins
- Protein-Protein interactions are edges

Figure: Protein-Protein Interaction Network of *Saccharomyces cerevisiae*.

Xarxes d'interacció de proteïnes (PPI)

- PPI networks are huge: $\sim$ 6000 nodes, $\sim$ 20000 edges.

Network	Nodes	Edges
C. Elegans	2,805	4,495
D. Melanogaster	7,518	25,635
H. Sapiens	9,633	34,327
S. Cerevisiae	5,499	31,261

PPI networks from the IsoBase database.

- We need computer tools to analyze and compare PPI networks.

PPI networks comparison

Why do we need to compare PPI networks?

- Protein networks in disease, like for instance cancer, increase considerably their number of edges. That is, the amount of proteins that interact is much bigger.

  Ideker T, Sharan R *Protein networks in disease*. Genome Res. 18:644–652 (2008)

- To detect network similarities and differences is a key point to identify the causes of certain diseases.

AligNet: Aligner of PPI networks.

PPI networks aligners

Local alignments:

- Aimed at finding local regions with the same network structure: low coverage.
- Local alignments are mutually inconsistent.

Global alignment:

- Aimed at finding the best overall alignment between two protein interaction networks.
- Designed to either obtain a high number of conserved interactions or a functional consistence between aligned proteins.

PPI networks aligners

- The right balance between network topology and biological information, is one of the most difficult and key points in every protein interaction network alignment algorithm.

With this lack of well-balanced algorithms in mind, we designed **AligNet**.

AligNet

Let G and G' be two PPI networks, **AligNet** is divided in the following steps:

1. The computation of overlapping clusterings $C(G)$ and $C(G')$, respectively, of the input networks G, G'.
2. The computation of alignments between pairs of clusters in $C(G)$ and $C(G')$.
3. The computation of a matching between $C(G)$ and $C(G')$.
4. The computation of a local alignment of the input networks G, G'.
5. The extension of this local alignment to a global meaningful one.

1. Computation of overlapping clusterings

For every node $u \in V$, the *cluster* C_u in G *centered* at u is

$$C_u = \{v \in V \mid s(u, v) > \alpha\}.$$

where α be the third quartile of the distribution of the similarity score values $s(u, v)$ on $V \times V$ defined by

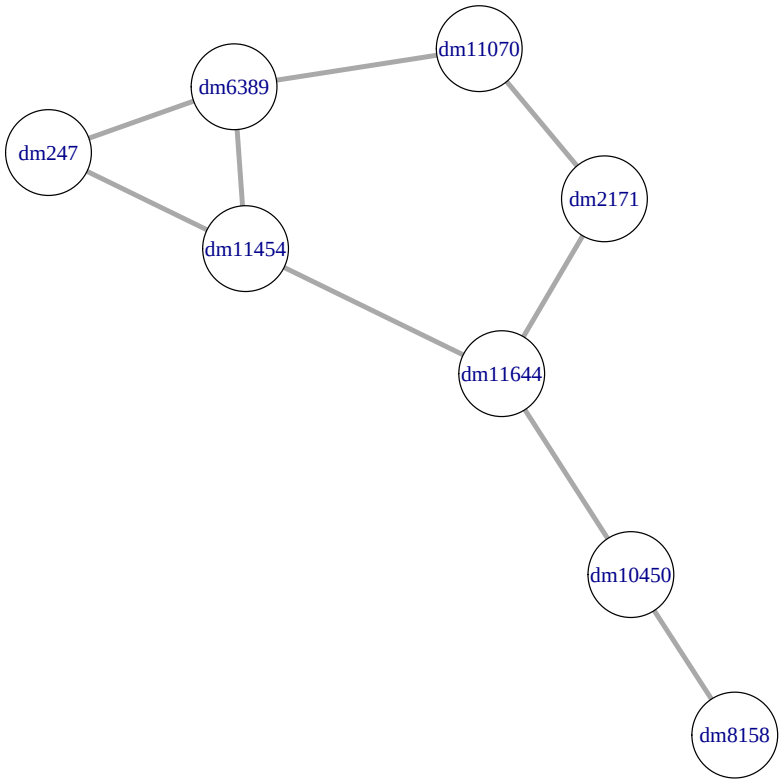
$$s(u, v) = \frac{B(u, v) + \frac{D(G)+1-d_G(u,v)}{D(G)+1}}{2}$$

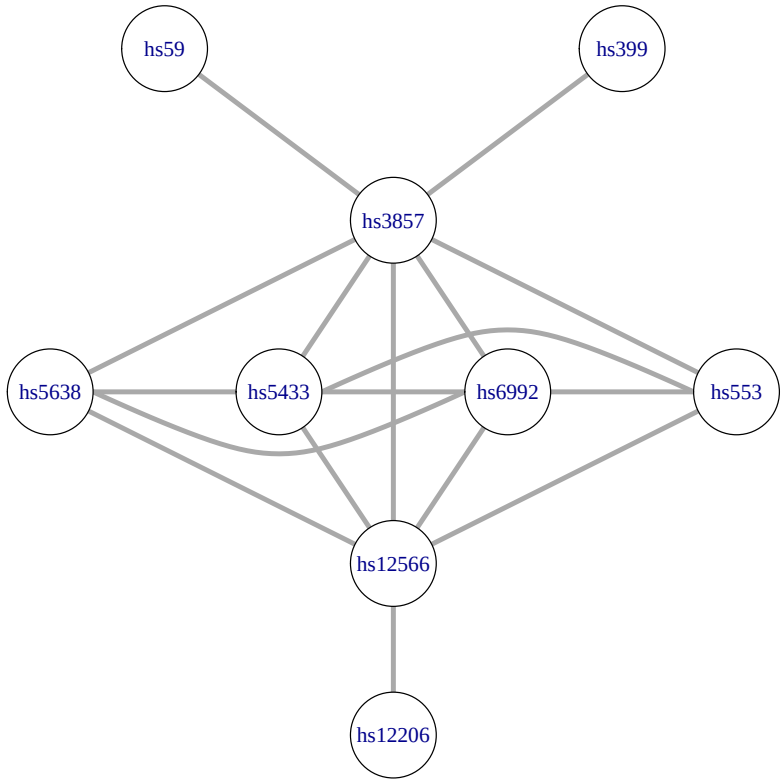
where:

- $D(G)$ is the diameter of the G and $d_G(u, v)$ is the distance between u and v.
- $B(u, v)$ is the *normalized bit score* of the proteins associated to the nodes u and v.

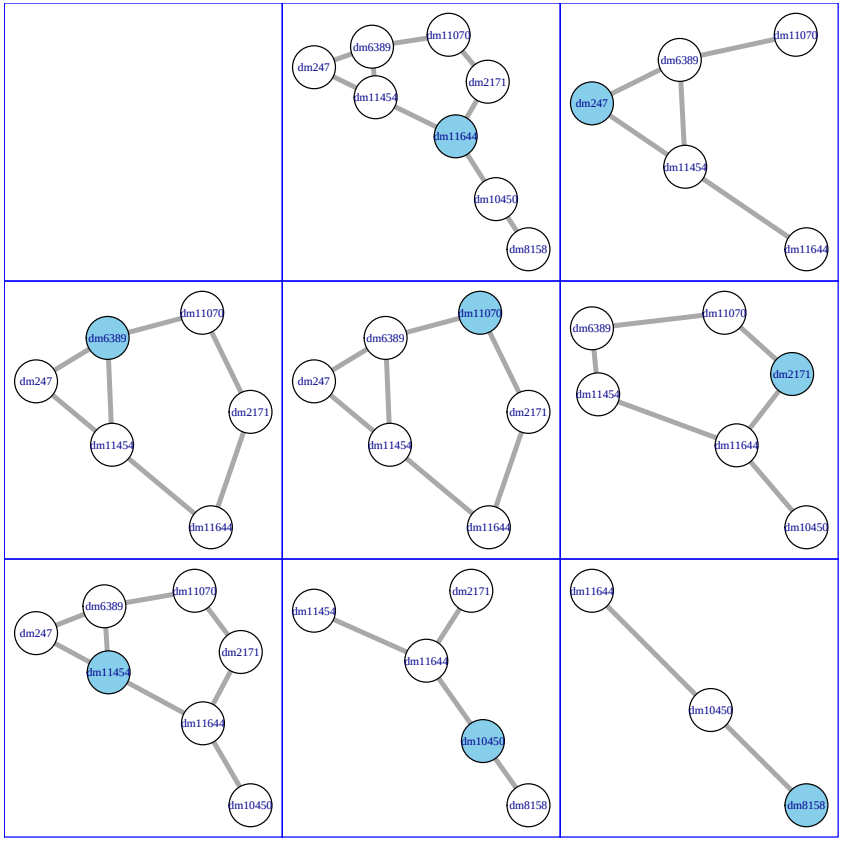
Running Example

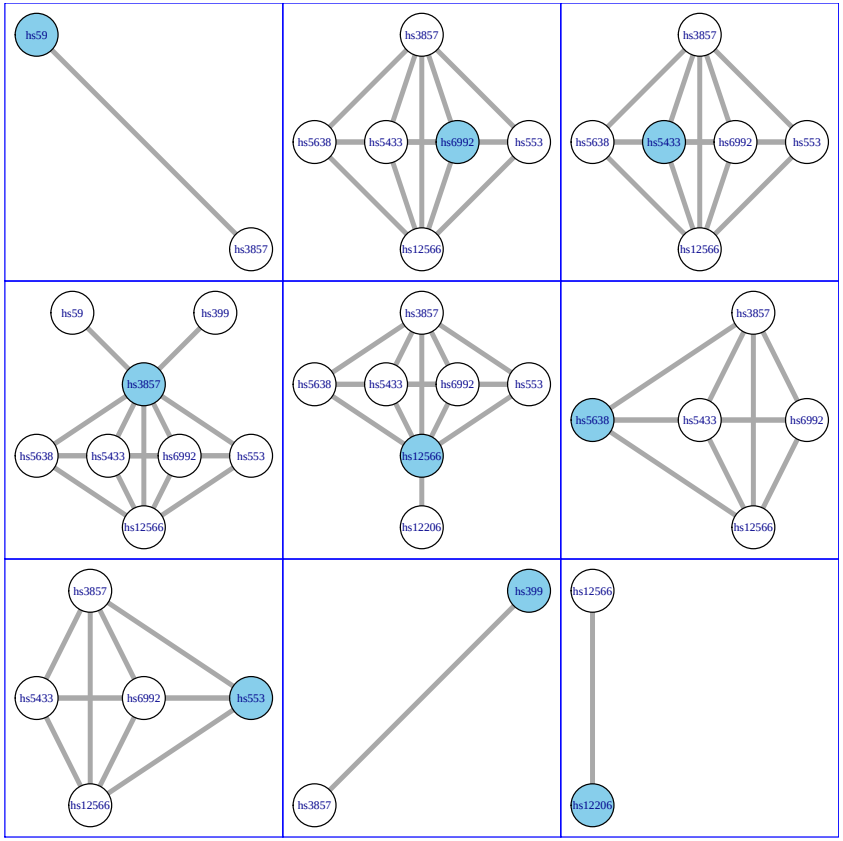
Subnetworks from to the *Drosophila melanogaster* (dme) and the *Homo sapiens* (hsa) PPI networks contained in the IsoBase database.





Running Example: Overlapping clusterings





2. Alignments between pairs of clusters

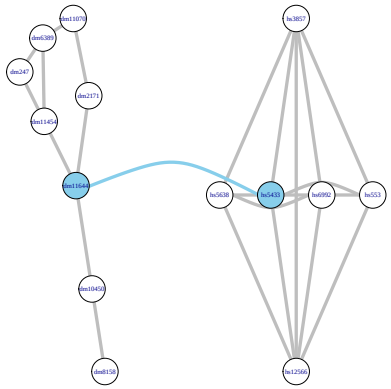
For every pair of clusters $C_u \in C(G)$ and $C_{u'} \in C(G')$ (with $B(u, u') > 0$)

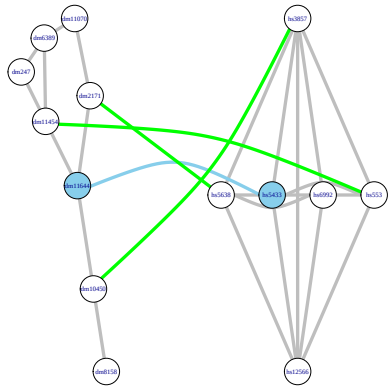
(i) Match u with u'. .

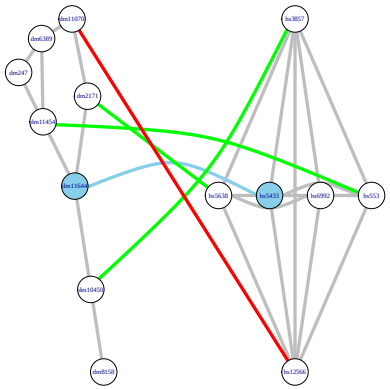
(ii) Iteratively match the neighbors of u to the neighbors of u' taking into account their degree and sequence similarity.

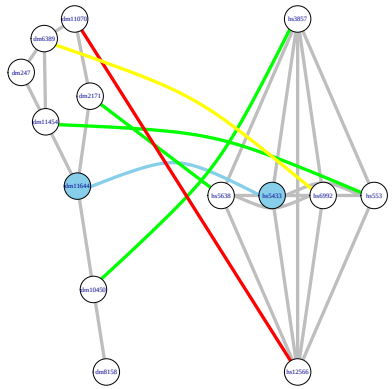
Idea: a node v in C_u connected by a path to u should be matched to a node v in $C_{u'}$ connected by a path to u' when they have a similar topological role in the cluster and similar sequences.

Running Example: Alignments between pairs of clusters









3. Matching between families of clusters

Let

$$\mathcal{A} = \{\eta_{u,u'} \mid u \in V,\ u' \in V', B(u,u') > 0\}$$

be the set of alignments obtained in step 2.

The score of every alignment $\eta_{u,u'} \in \mathcal{A}$ is defined as

$$Score(\eta_{u,u'}) = \frac{\sum_{v \in Dom\,\eta_{u,u'}} B(v, \eta_{u,u'}(v))}{|Dom\,\eta_{u,u'}|} + \frac{|Dom\,\eta_{u,u'}|}{max_{\eta_{w,w'} \in \mathcal{A}} |Dom\,\eta_{w,w'}|}.$$

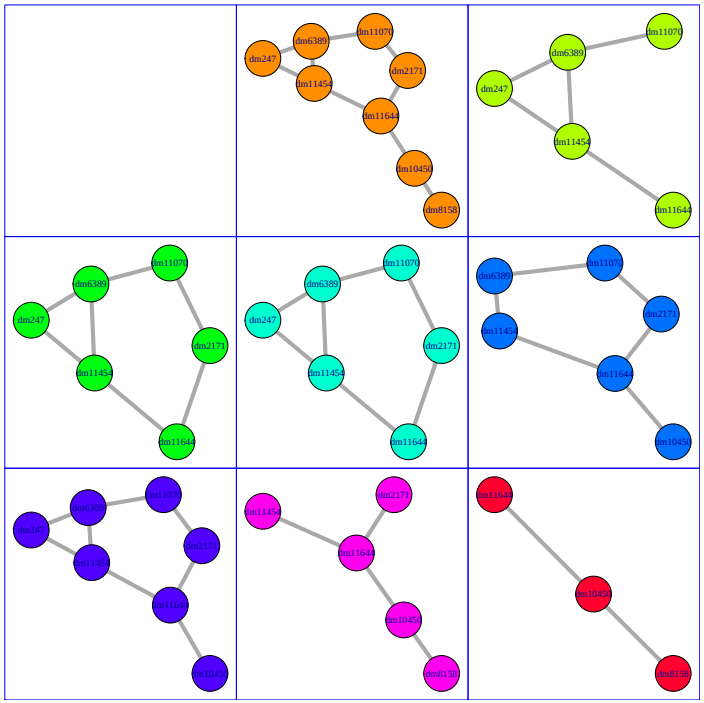
3. Matching between families of clusters

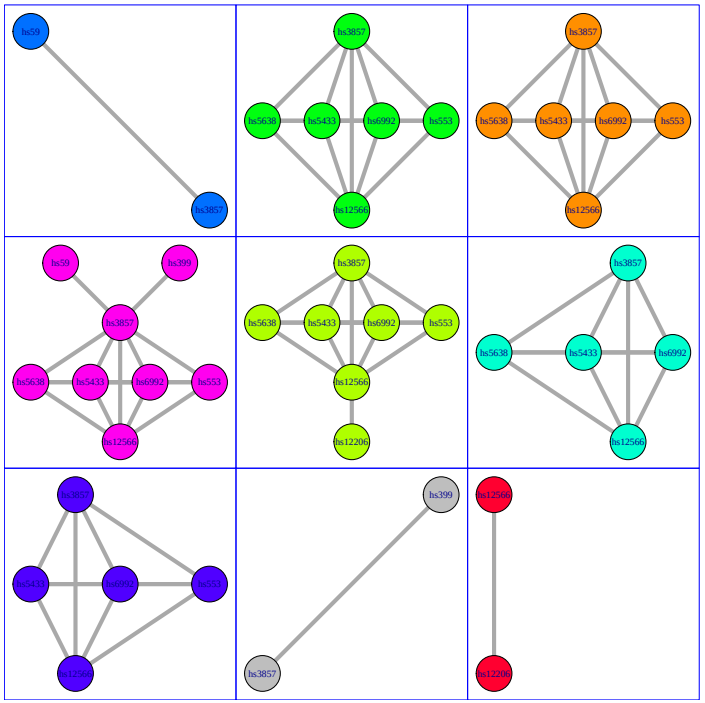
AligNet obtains a matching between $C(G)$ and $C(G')$ by considering a bipartite graph such that

- the nodes are the clusters $C(G)$ and $C(G')$
- the edges correspond to alignments $\eta_{u,u'} \in \mathcal{A}$
- the weight of an edge connecting C_u with $C_{u'}$ is the corresponding $Score(\eta_{u,u'})$

The matching between $C(G)$ and $C(G')$ is then obtained by applying the maximum weighted bipartite matching algorithm to this bipartite graph.

Running Example: Matching between families of clusters





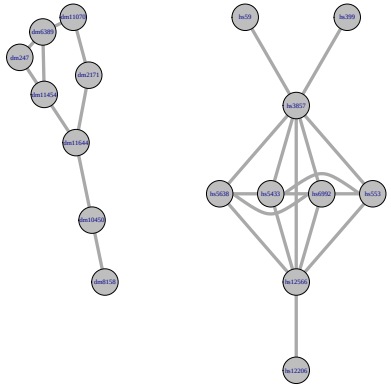
4. Local alignment of PPI networks

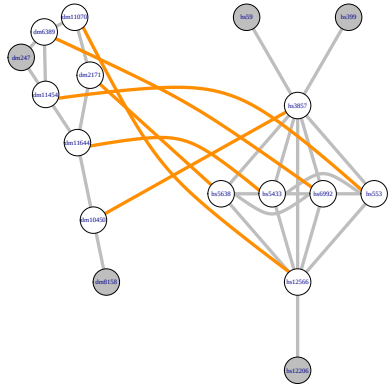
AligNet builds a set of *appropriate* alignments $\mathcal{R} \subseteq \mathcal{C}$ recursively such that $\eta_{w_0,w_0'} \in \mathcal{R}$ if

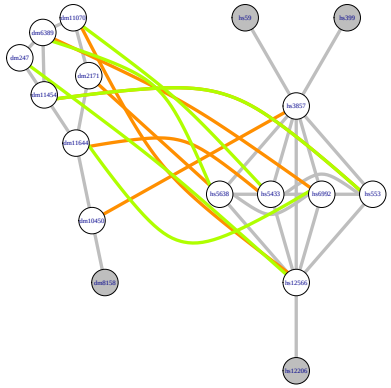
- w_0 not belonging to the union of the domains of the mappings already in $\mathcal{R}$.
- $Score(\eta_{w_0,w_0'})$ maximum among all such mappings.

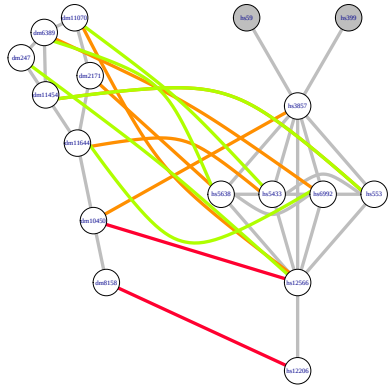
This procedure is iterated until every node in $\bigcup_{\eta_{u,u'} \in \mathcal{C}} Dom\, \eta_{u,u'}$ belongs to the domain of some mapping in $\mathcal{R}$.

Running Example: Set of *appropriate* alignments









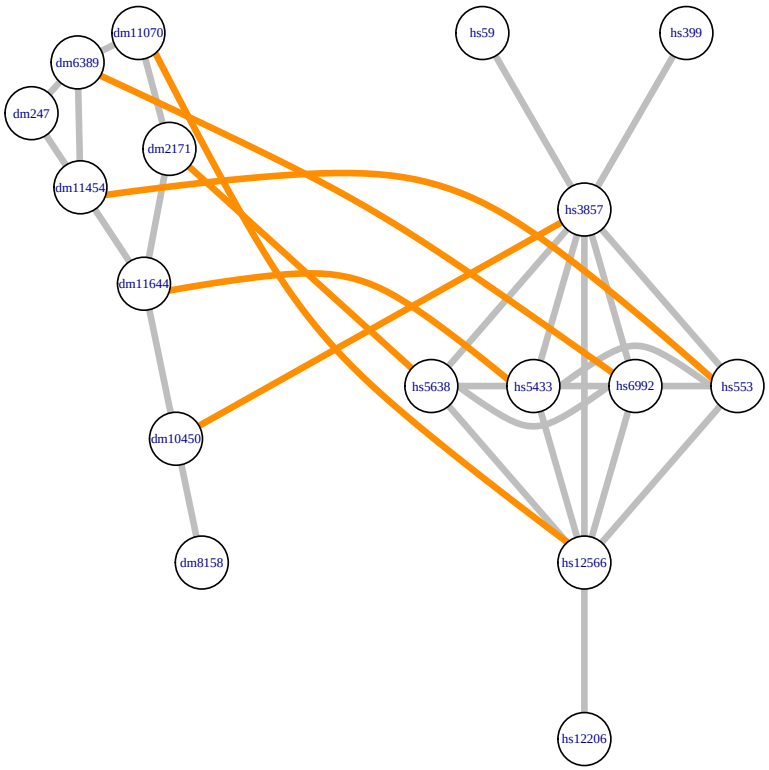
4. Local alignment of PPI networks

H a directed hypergraph where

- nodes are $V \cup V'$
- hyperarcs are the mappings $\eta_{w,w'} \in \mathcal{R}$: each $\eta_{w,w'}$ is understood as a hyperarc with source its domain and target its image.

Then, **AligNet** obtains from this hypergraph a local well-defined alignment between G and G' as a solution of the corresponding weighted bipartite hypergraph assignment problem.

Running Example: Local alignment of PPI networks

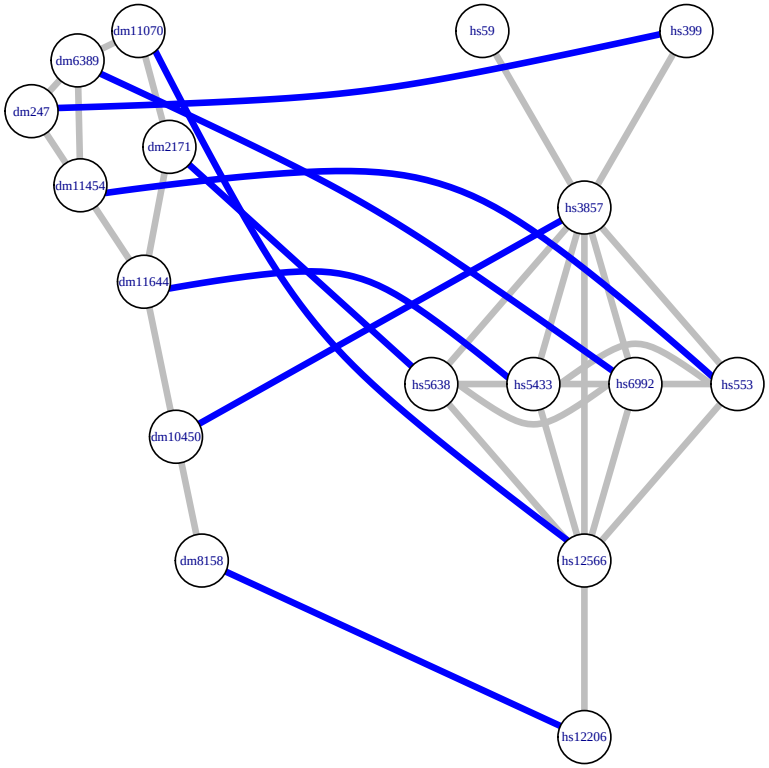


5. Global meaningful alignment of PPI networks

To end, **AligNet**:

- Removes the nodes in G and G' that have already been aligned.
- Repeat step $2-4$ while there exist nodes not aligned belonging to the domain or the codomain of some alignment $\eta_{u,u'}$ with (updated) score > 0.

Running Example: Global meaningful alignment of PPI networks



AligNet Testing

Following the tests and results presented in

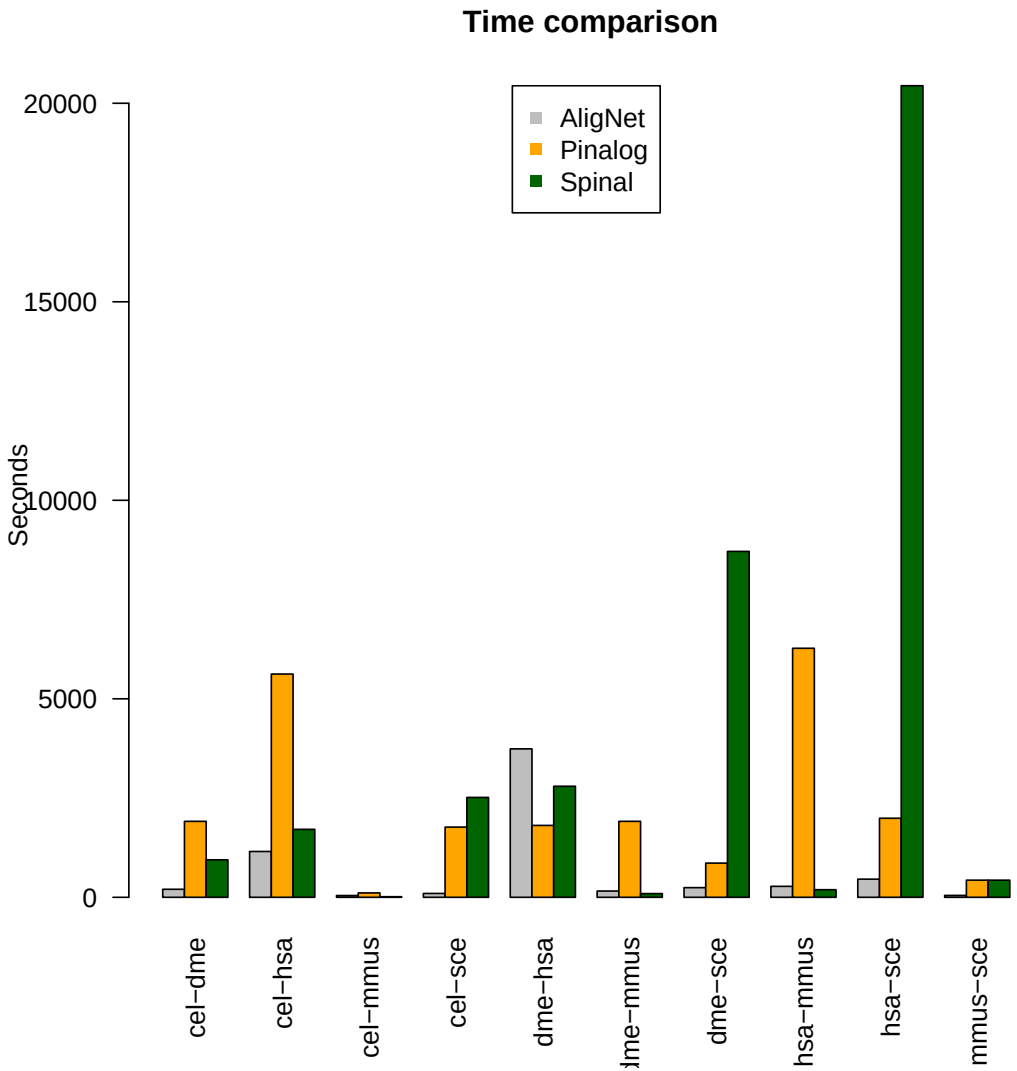
Connor C., Jugal K. *A comparison of algorithms for the pairwise alignment of biological networks*. Bioinformatics 30.16:2351–2359 (2014)

we have downloaded from the IsoBase database the PPI networks of the organisms:

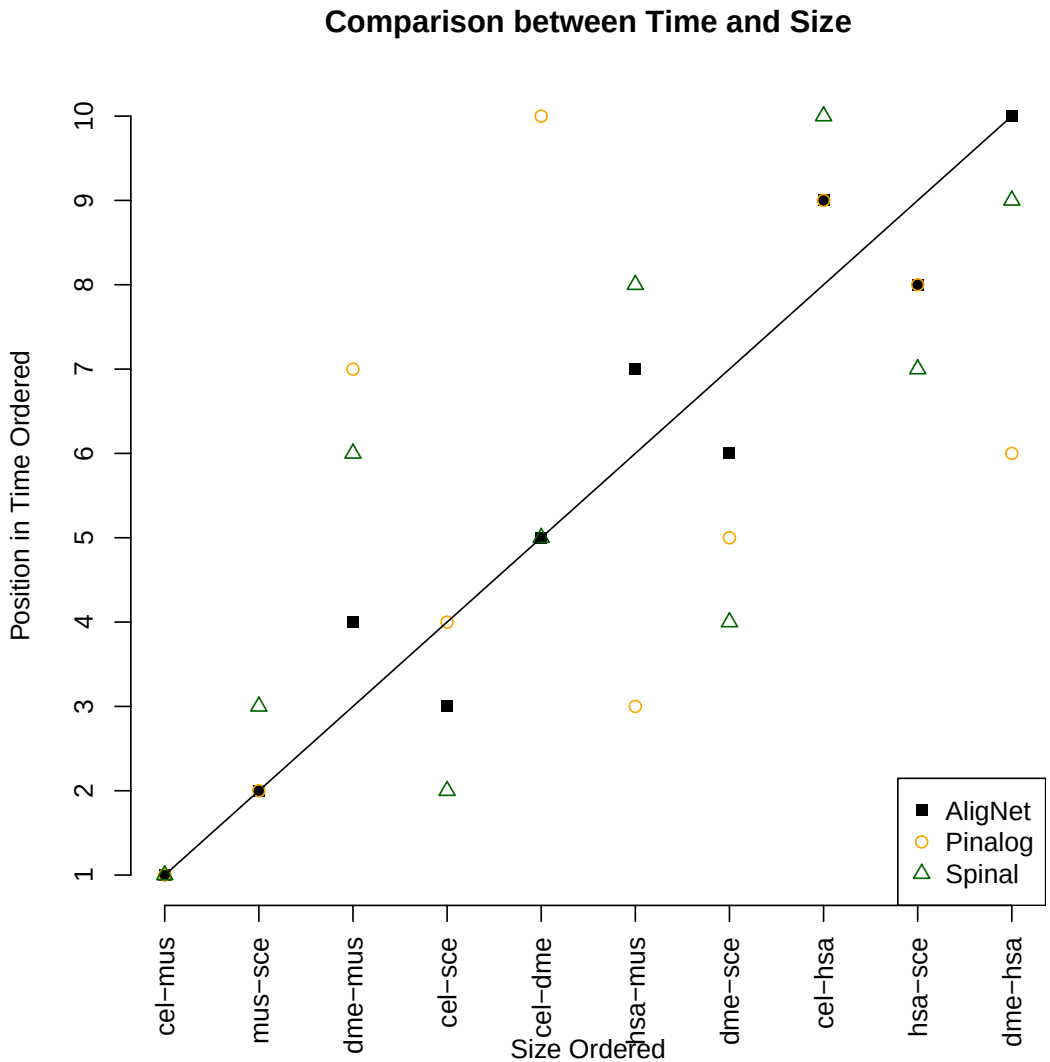
- *C. elegans*
- *D. melanogaster*,
- *S. cerevisiae*,
- *M. Musculus*,
- *H. sapiens*

and aligned each pair of them with **AligNet**, PINALOG and SPINAL.

AligNet Testing: 1. Running Times



AligNet Testing: 1. Running Times

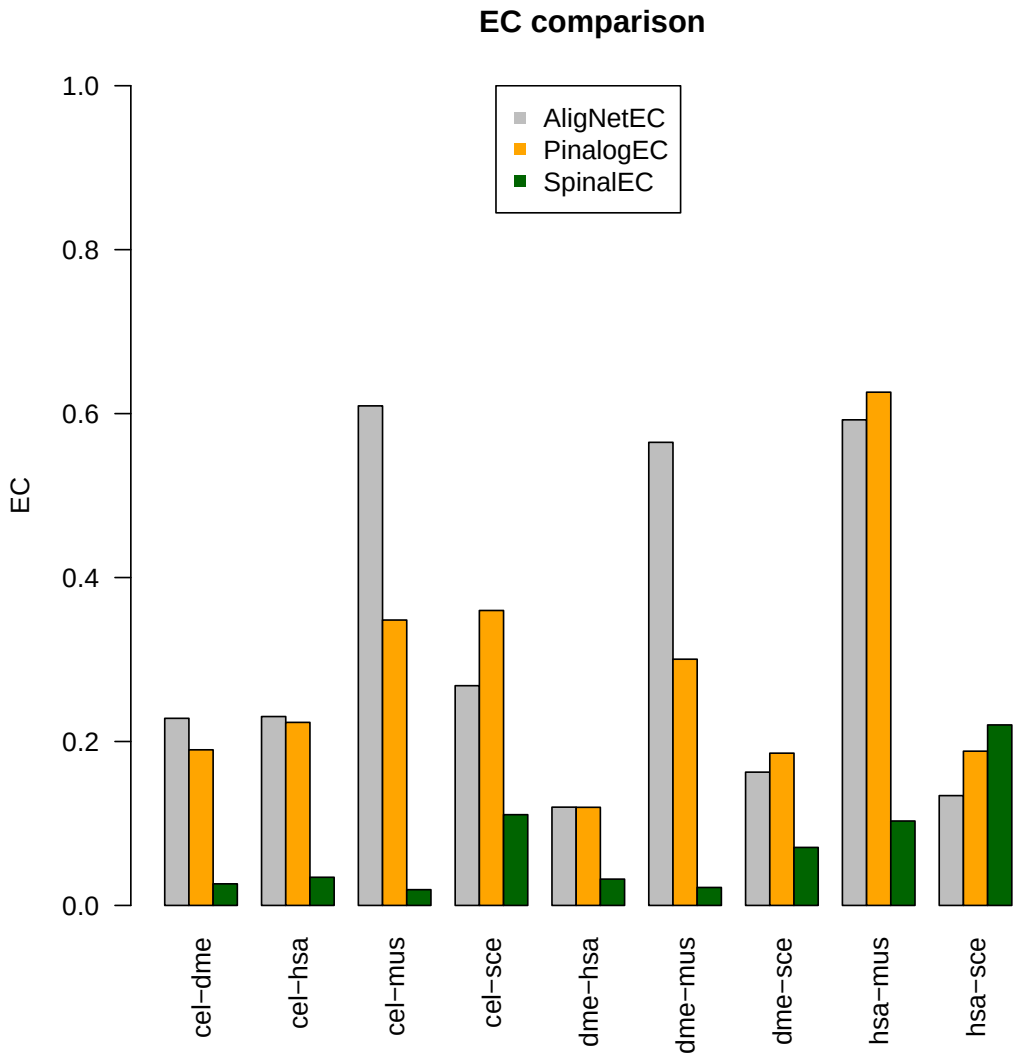


Alignet Testing: Edge correctness ratio

The *edge correctness ratio* of a mapping $\mu : G \to G'$ is the ratio of the edges that are preserved by μ, and it is defined by

$$EC(\mu) = \frac{\left|\left\{\{u, v\} \in E \mid \{\mu(u), \mu(v)\} \in E'\right\}\right|}{|E|}.$$

Alignet Testing: Edge correctness ratio



Alignet Testing: Functional Coherence Value

The *functional coherence value*, or *GO consistency*, of a mapping $\mu : G \to G'$ is defined as

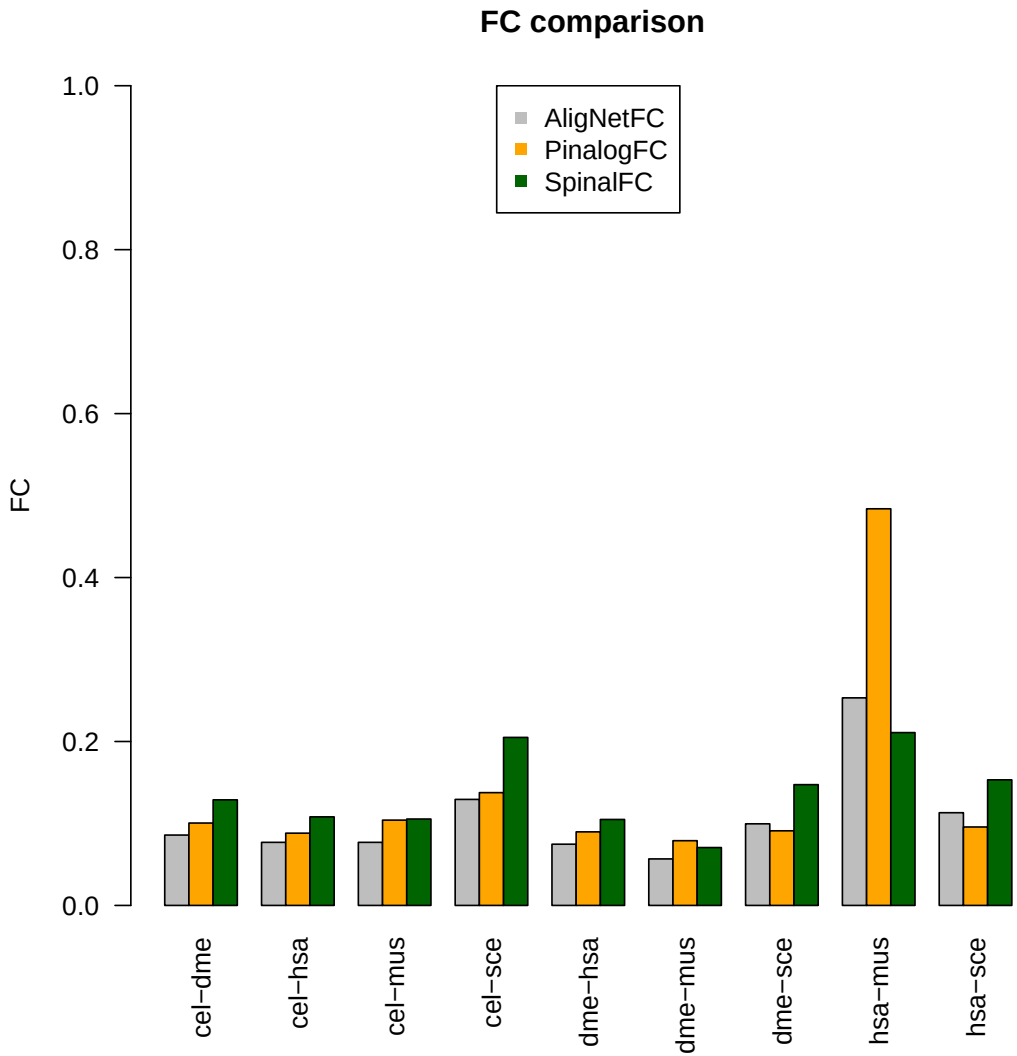
$$FC(\mu) = \frac{\sum_{u \in V} FS(u, \mu(u))}{|V|}$$

where, on its turn, the functional similarity score FS is defined by

$$FS(u, u') = \frac{|GO(u) \cap GO(u')|}{|GO(u) \cup GO(u')|},$$

with $GO(u)$ and $GO(u')$ the sets of GO annotations of the proteins u and u', respectively.

Alignet Testing: Functional Coherence Value



Conclusions and Further Work

- AligNet is faster than PINALOG and SPINAL
- The Edge correctness ratio obtained with AligNet is higher than other aligners
- The Functional coherence value obtained with AligNet is reasonable

Apply AligNet to

- Protein complexes prediction
- Protein-protein interaction prediction
- Orthology prediction