A Novel Computational Approach for Global Alignment of Multiple Biological Networks

Engelbert MEPHU NGUIFO
LIMOS, UMR CNRS, Université Clermont Auvergne, France
engelbert.mephu_nguifo@uca.fr

Séminaire Bioinformatique
Département Informatique, UQAM, Montréal
26 mars 2019, 17h30

Co-auteurs : Warith Eddine DJEDDI*, Sadok BEN YAHIA,
* Thanks to French Embassy of Tunisia.

Bioinformatique au LIMOS – UMR CNRS

- Thème : Données-Services-Intelligence (DSI)
- Axe transversal : STIC pour SVE

- Collaborations avec les biologistes :
  - LMGE, GRED, INRA, …

- Fédération de recherches CNRS : Environnement

- Projets:
  - ✓ **Génomique, Protéomique, Métagénomique, …**

Bioinformatique au LIMOS :

- Projets de recherche :
  - ✓ **Génomique :**
    - ✓ Indexation de séquences d'ADN par hachage perceptuel (-INRA),
    - ✓ …
  - ✓ **Métagénomique :**
    - ✓ Etude de la biosphère rare microbienne (-LMGE). ---- JCB 2019, online version
    - ✓ Reconstruction de génomes microbiens à partir de données de séquençages de métagénomes (-LMGE-INRA)
    - ✓ …
  - ✓ **Protéomique :**
    - ✓ Etude de structures tridimensionnelles de protéines. ---- JCB 2014
    - ✓ Etude de la résistance aux radiations chez des Bactéries (-CNSTN)
    - ✓ …
  - ✓ **Interatomique :**
    - ✓ DropNet : a web portal for integrated analysis of Drosophila protein–protein interaction networks (-GRED). --- NAR 2012
    - ✓ Alignement des PPI (-LIPAH), ---- TCBB 2018
    - ✓ …

Bioinformatique au LIMOS :

Smoothing 3D protein structure motifs through graph mining and amino-acids similarities

Wajdi Dhifli[*,**], Rabie Saidi[***] and Engelbert Mephu Nguifo[*,**]

[*]LIMOS - Blaise Pascal University - Clermont University, Clermont-Ferrand 63000, France.
[**]LIMOS - CNRS UMR 6158, Aubière 63173, France.
[***]European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, United Kingdom

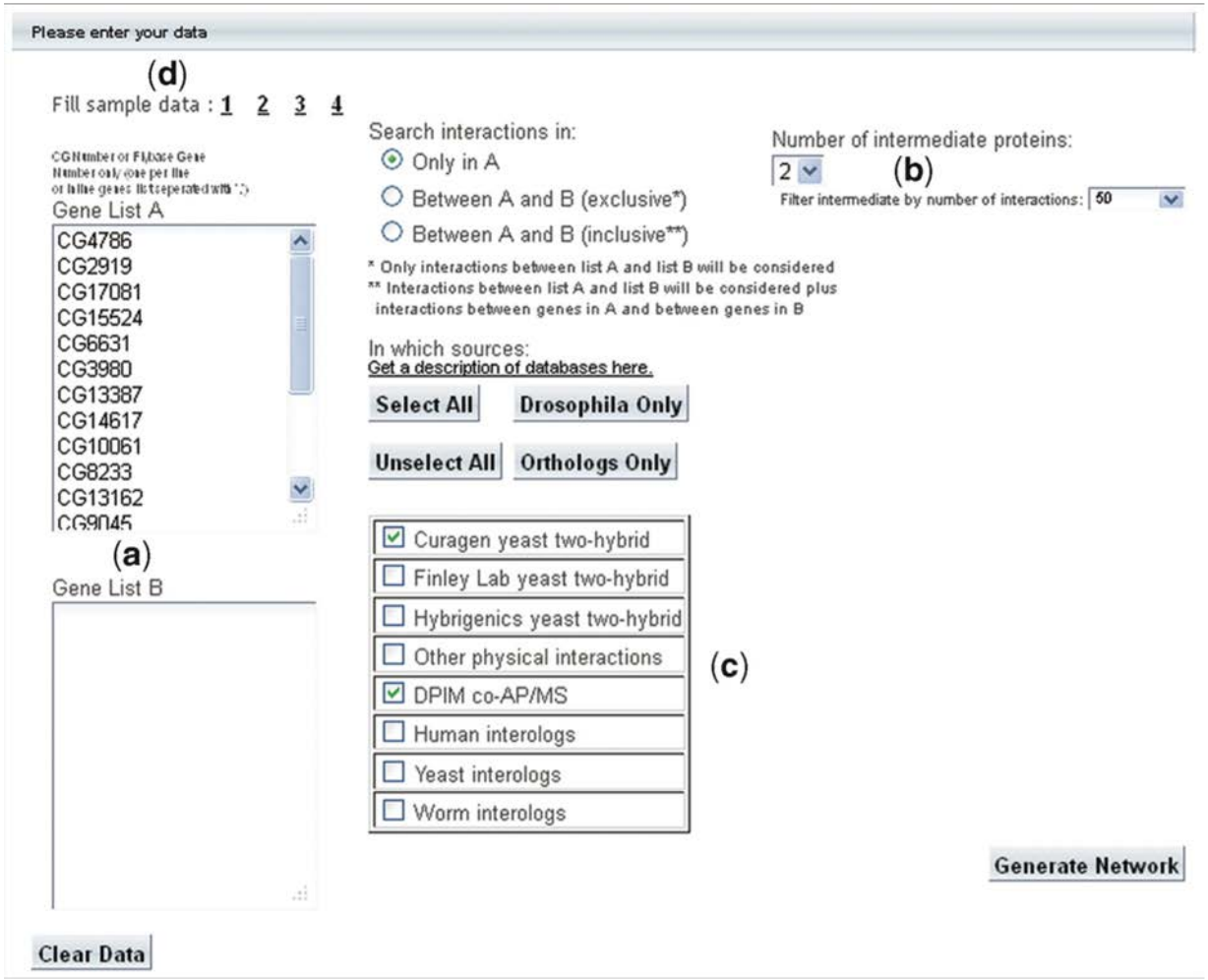
June 28, 2013

JCB, 2014

Abstract

One of the most powerful techniques to study proteins is to look for recurrent fragments (also called substructures), then use them as patterns to characterize the proteins under study. Although protein sequences have been extensively studied in the literature, studying protein three-dimensional (3D) structures can reveal relevant structural and functional information which may not be derived from protein sequences alone. An emergent trend consists in parsing proteins 3D structures into graphs of amino acids. Hence, the search of recurrent substructures is formulated as a process of frequent subgraph discovery where each subgraph represents

Bioinformatique au LIMOS :



From: DroPNet: a web portal for integrated analysis of Drosophila protein–protein interaction networks
Nucleic Acids Res. 2012;40(W1):W134-W139. doi:10.1093/nar/gks434

Intl. Workshop on Bioinformatics and Artificial Intelligence (BAI)

- BAI @ IJCAI : http://bioinfo.uqam.ca/IJCAI_BAIyyyy/
  - 2015 (Bueno Aires, Argentina),
  - 2016 (New York, USA),
  - 2017 (Melbourne, Australia)
- WCB & BAI @ ICML & IJCAI 2018 (Stockholm, Sweden)
- *WCB & BAI @ ICML 2019* (Long Beach, CA, USA)

- Special issues : **Journal of Computational Biology** (JCB)
  - **Vol. 24(8): 733, 2017** : selected papers BAI 2015 et BAI 2016
  - **Vol. 26(x): xxx, 2019** : sel. papers BAI 2017 et WCB&BAI 2018

- Diallo A.B., Mephu Nguifo E., Zaki M., Dhifli W., …

Outline

Introduction

Background / Related works

MAPPIN

Experimental results

Conclusion

Introduction

Why Networks?

Networks are everywhere... *especially in Biology!*	*Networks are powerful tools... especially in Biology!*
• Molecular networks • Cell-cell communication • Nervous systems	• Reduce complexity • More efficient than tables • Great for data integration • Intuitive visualization

Protein-Protein Interactions (PPI)

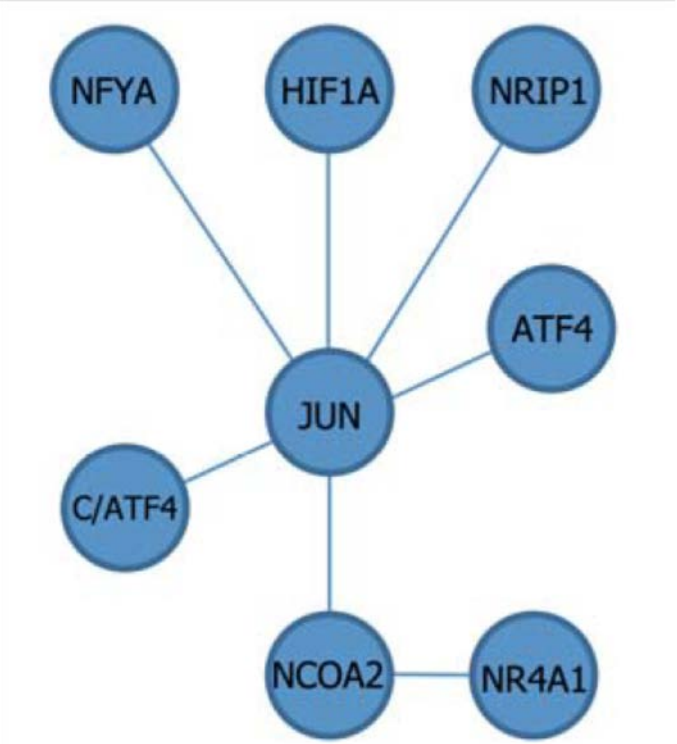
- Interaction between two proteins is carried out by several biochemical events

- The forces responsible for these interactions include:
  - ✓ **Electrostatic forces:** Forces interacting between static electrically charged particles
    - ✓ **Hydrogen bonds:** electrostatic attraction between hydrogen (H) and highly electronegative atom (e,g. O, N)
    - ✓ **Van der waals forces:** residual attractive or repulsive forces between molecules or atomic groups,
  - ✓ **Hydrophobic interactions:** Maximize hydrogen bond …

- Play an essential role in the proper functioning of living cells

A protein-protein interaction network

- PPI is represented as undirected edges (the physical relationships) between proteins.
- Proteins are represented as nodes that are linked by undirected edges.



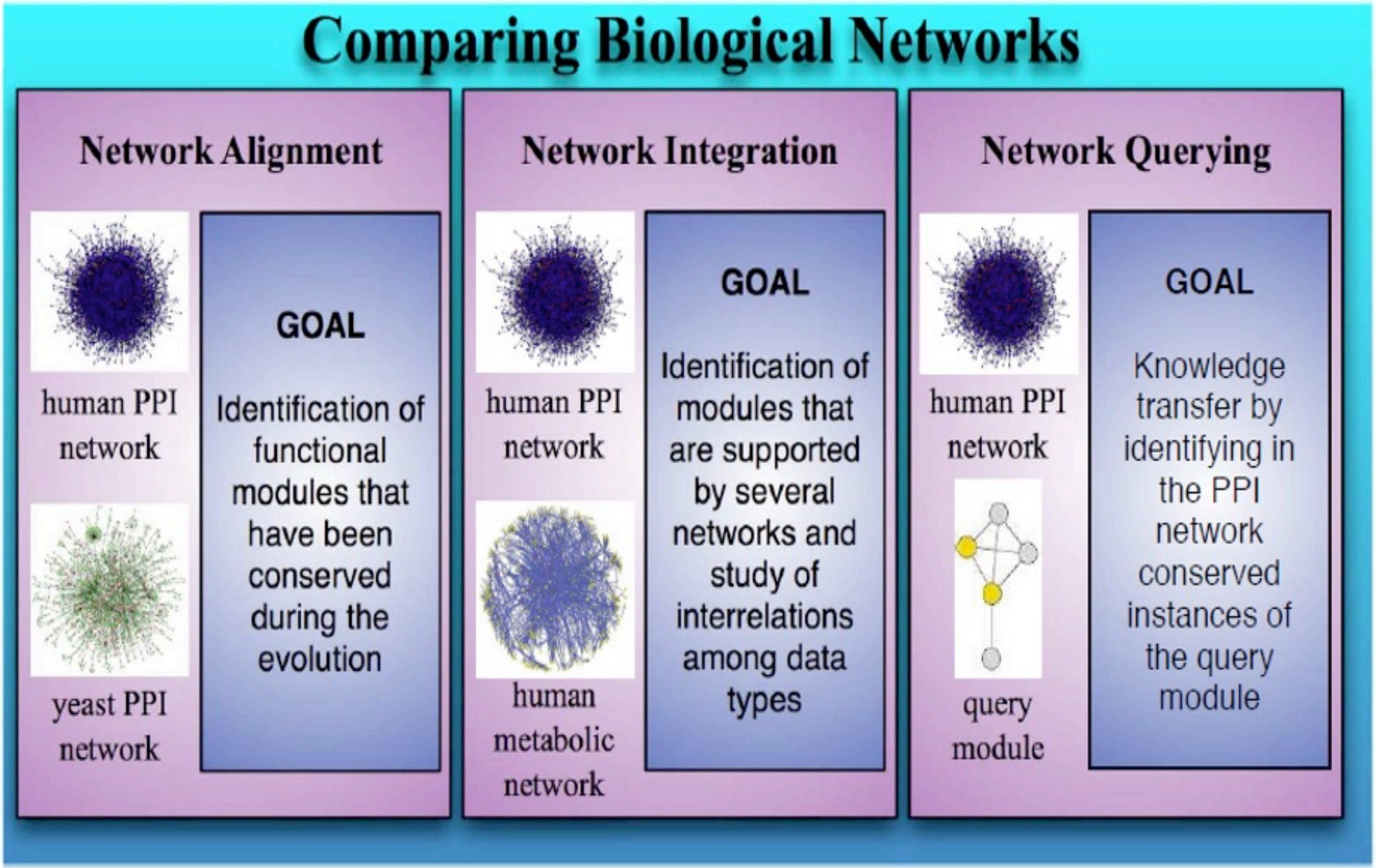
PPI network for nucleic acid metabolism pathway :

NFYA - Nuclear transcription factor Y subunit alpha,
HIF1A - Hypoxia inducible factor 1 alpha,
NRIP1 - Nuclear receptor interacting protein 1,
NCOA2 - Nuclear receptor co-activator 2,
NR4A1 – Nuclear receptor sub-family 4 group A member 1;
ATF4 – Activating transcription factor 4 (Cyt),
JUN – Transcription factor activator protein 1 (Nuc),
C/ATF4 - Cyclic AMP-dependent transcription factor ATF-4 (Cyt) (Nuc).

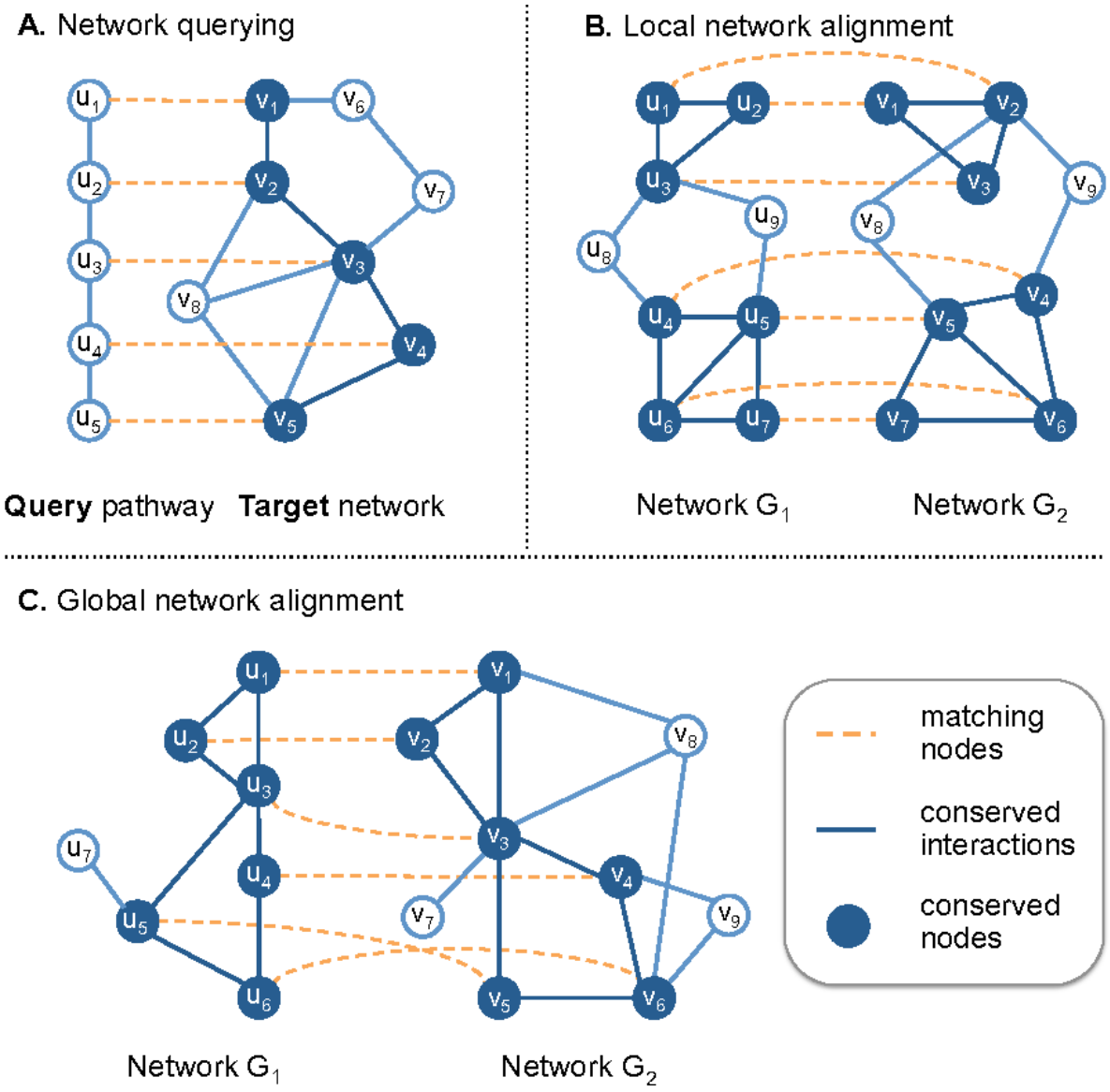
Source : IEEE/ACM TCBB 13(4): 689-705, 2016

Types of Protein-Protein interaction

- PPIs can be classified on the bases of
  - ✓ **Stability :**
    - Stable: Always stable and active (e.g., Hormones, Hemoglobin)
    - Transient: Control the majority of cellular processes, can be strong or weak, fast or slow
  - ✓ **Structural :**
    - Homo-oligomer: Same type of subunits (e.g., Enzymes)
    - Hetero-oligomer: Different types of subunits (e.g., G-proteins)
  - ✓ **Chemical bonding :**
    - Covalent bonding: Share electron pairs
    - Non Covalent Bonding: Rather sharing electrons, involves in some electromagnetic forces



Source: Fionda, Valeria. "Biological network analysis and comparison: mining new biological knowledge." *Open Computer Science* 1.2 (2011): 185-193.



Source : Yoon, Byung-Jun, Xiaoning Qian, and Sayed Mohammad Ebrahim Sahraeian. "Comparative analysis of biological networks using Markov chains and hidden Markov models." *IEEE Signal Processing Magazine* 29(1):22-34, (2012).

PPI Network Alignment

- PPI networks alignment enables us to **uncover the relationships** between different species

- Network alignment can be used to **transfer biological knowledge** between species

- A comparative analysis of PPI networks provides **insight into species evolution** and information about evolutionarily conserved biological interactions, such as pathways across multiple species

Outline

Introduction

Background / Related works

MAPPIN

Experimental results

Conclusion

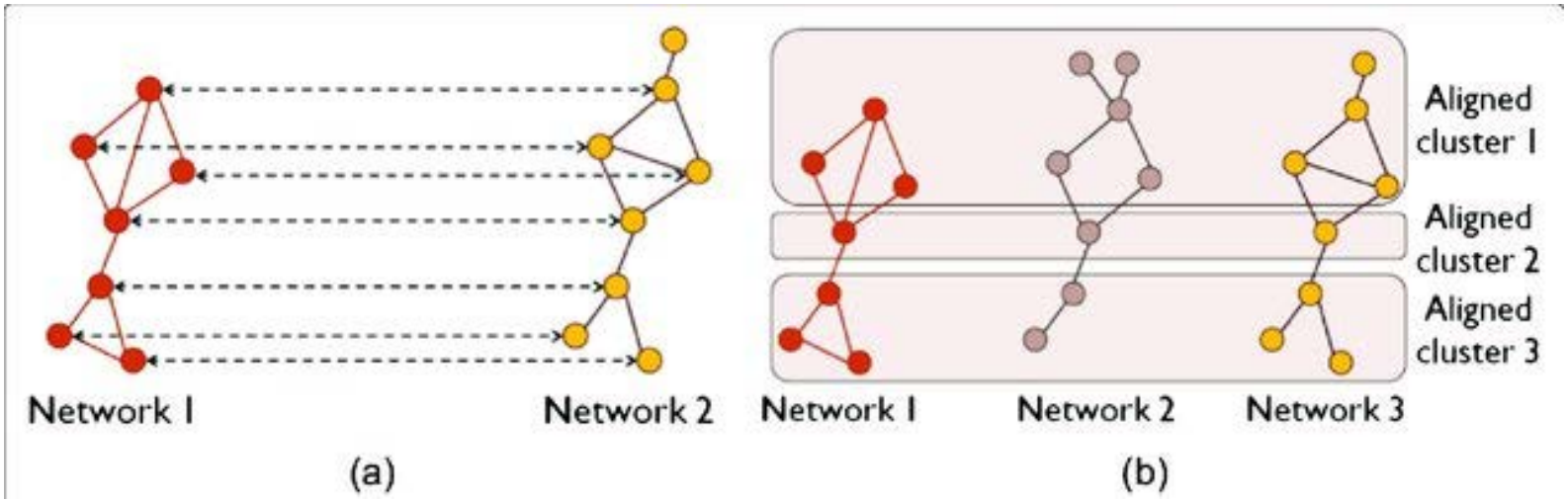
PPI Alignment

- Graph alignment problem
- Subgraph isomorphism
  - NP-complete
- Approximate solutions
  - Many existing approaches depending on :
    - Node similarities (scoring functions)
    - Search methodologies
  - Domain knowledge can help

Pairwise vs Multiple Network Alignment

- Network alignment (NA) can be pairwise (PNA) and multiple (MNA):
  - ✓ PNA produces aligned node pairs between two networks (Fig.a),
  - ✓ MNA produces aligned node clusters between more than 2 networks (Fig.b).

Note: Recently, the focus has shifted from PNA to MNA, because **MNA captures conserved regions** between more networks than PNA (and MNA is thus considered to be more insightful), though at higher computational complexity.



Source: Faisal, Fazle E., et al. "The post-genomic era of biological network alignment." *EURASIP Journal on Bioinformatics and Systems Biology* 2015.1 (2015): 3.

Pairwise PPI Alignment

- $G_1 = (V_1,E_1)$, $G_2 = (V_2,E_2)$, $|\boldsymbol{V_1}| = n$, $|\boldsymbol{V_2}| = m$, $(u,v) \in E_i$ s.t. $u,v \in V_i$

Problem : Find an *injective function* $f : V_1 \rightarrow V_2$ that aligns each node in V_1 to only one node in V_2

$$f(u) = \{v, \text{where } u \in V_1 \text{ and } v \in V_2\}$$

- Variant : f can be *partially* defined

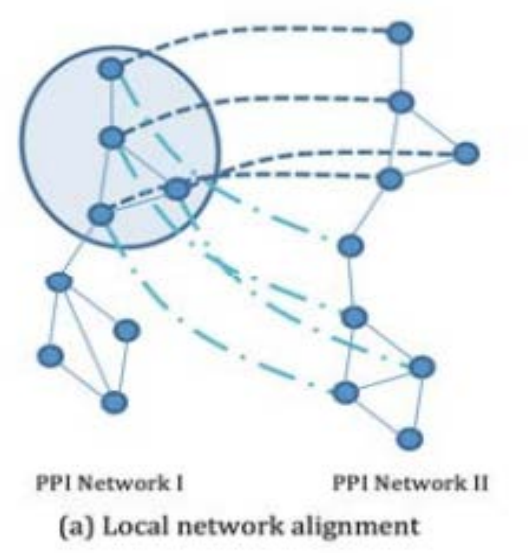
- Best alignment :
  - A : set of all possible alignments
  - One that has the maximum score using a scoring function S

$$a = \textit{argmax}_{\,ai \in A}\, S(a_i)$$

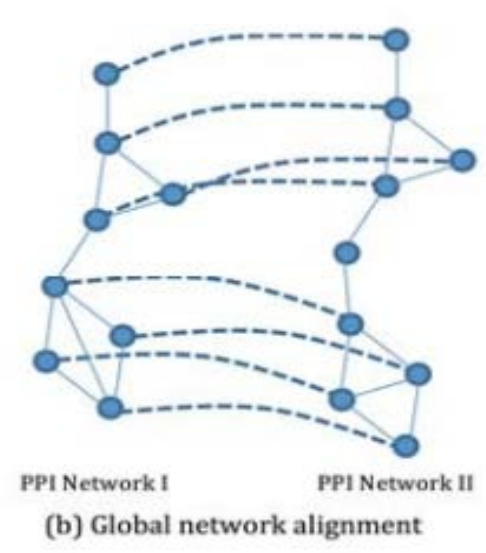
PPI Alignment : classification

❖ According to type of PPI network alignment:

- Local Area Network (LNA) : small similarity regions are independently adapted, and many of these regions may overlap in a contradictory manner.
- Global Network Alignment (GNA) : each node of the lower network is uniquely aligned to a single, better matching node in the large network.



(a) Local network alignment



(b) Global network alignment

Remark :
- LNA is more faithful to biological theory, but difficulty of interpreting LNA results
➔ GNA

Source :
Ahed Elmsallati, Connor Clark, Jugal Kalita
IEEE/ACM TCBB 13(4):689-705, 2016

PPI Alignment : Validation

- Topological Assessment :
  - Unsupervised
    - Edge Correctness
    - Induced Conserved Structure
    - Symmetric Substructure Score
  - Supervised
    - Node Correctness
    - Interaction Correctness

$$EC(G_1, G_2, f) = \frac{|f(E_1) \cap E_2|}{|E_1|}$$

$$ICS(G_1, G_2, f) = \frac{|f(E_1) \cap E_2|}{|E_{G_2[f(V_1)]}|}$$

$$S^3(G_1, G_2, f) = \frac{|f(E_1) \cap E_2|}{|E_1| + |E_{G_2[f(V_1)]}| - |f(E_1) \cap E_2|}$$

$$NC(G_1, G_2, f) = \frac{|u_i : f(u_i) = g(u_i)|}{|V|} \times 100$$

PPI Alignment : Validation

- Biological Assessment :
  - Use Gene Ontologoly (GO) annotations
    - Resnik ontological similarity
    - GO Consistency (GOC). --- similar to Jaccard index

$$GOC(G_1, G_2, f) = \sum_{(u_i, v_j) \in a} \frac{|GO(u_i) \cap GO(v_j)|}{|GO(u_i) \cup GO(v_j)|}$$

  - Consistency : Assess the functional coherence
    - Mean Entropy
    - Mean Normalised Entropy
- Other Assessment :
  - Coverage :

    amount of protein in the whole set of proteins that are covered by the alignment

Background / Related works

- SMETANA is a many-to-many global MNA algorithm, tries to find correspondences by using a semi-Markov random-walk model. Compute pairwise sequence scores and pairwise topological scores.

- BEAMS is a fast approach that constructs global many-to-many MNA from the pairwise sequence similarities of the nodes by using a backbone (seed) extraction and merge strategy.

Background / Related works

- IsoRankN (IsoRank-Nibble) is the first global MNA algorithm that uses both pairwise sequence similarities and network topology, to generate many-to-many alignments.
  - It applies IsoRank to derive pairwise alignment scores between every pair of networks, and then employs a PageRank-Nibble algorithm to cluster all the proteins by their alignment score.

Background / Related works

- NetCoffee aligns multiple PPI networks based only on sequence similarity and does not take into account the topology of the considered networks.
  1. Its alignment strategy constructs a weighted bipartite graph for each pair of networks, searches for candidate edges from each bipartite graph by solving maximum weight bipartite matching problem.
  2. NetCoffee applies a triplet approach similar to T-Coffee to compute the edge weights of the kpartite graph. Then, the algorithm finds candidate edges in the bipartite graphs and combines qualified edges through **simulated annealing**.

Background / Related works

- PINALOG is a global network alignment algorithm which combines information from protein sequence, function and network topology.
  - ✓PINALOG forms the alignment between two PPINs based on the similarities of protein sequence and the protein function between the two networks. Functional similarity is formalized using GO (gene ontology) annotations.

Background / Related works

- Although few methods have been developed for multiple PPI network alignment and thus, new network alignment methods are of a compelling need.
- Moreover, many alignment tools encounter limitations in introducing the functional similarities during the alignment process because it needs faster and more efficient alignment tool especially for the alignment of multiple PPI networks.

- Note : Most of them make use of the Gene Ontology (GO) at the final validation step of the quality of the final alignment and not during the alignment process.

Gene Ontology / Goals

- Develop a set of controlled, structured vocabularies – gene ontology (GO) to describe aspects of molecular biology
- Describe gene products using vocabulary terms (annotation)
- Provide a public resource, allowing access to the GO, annotations and software tools developed for use with the GO data
- www.geneontology.org

Gene Ontology / The Three Ontologies

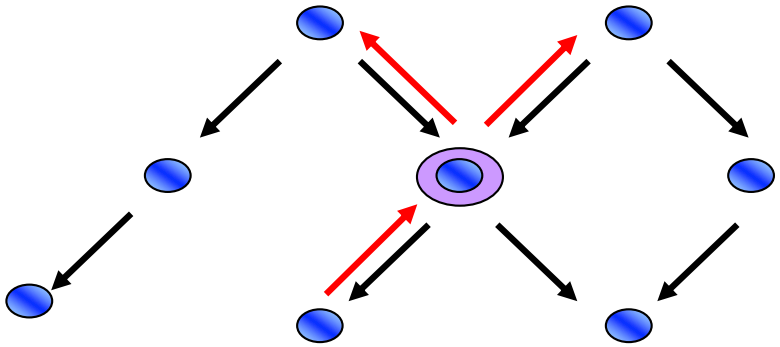
- **Molecular Function** — describes activities, or tasks, performed by individual or by assembled complexes of gene products (DNA binding, transcription factor)

- **Biological Process** — a series of events accomplished by one or more ordered assemblies of molecular functions. NOT a "pathway"! (mitosis, signal transduction, metabolism)

- **Cellular Component** — location or complex , a component of a cell, that also is part of some larger object (nucleus, ribosome, origin recognition complex)

Gene Ontology / Relationships between terms

Directed acyclic graph: each child may have one or more parents



Every path from a node back to the root must be biologically

accurate (the true path rule)

Relationship types:

- **is_a :** class-subclass relationship, meaning that **a** is a type of **b**

Exemple: **nuclear chromosome** is_a **chromosome**.

- **part_of** : physical part of (component) subprocess of (process)

part_of **c** part_ of **d**, meaning that whenever **c** is present, it is a part of **d**, but **c** doesn't always have to be present.

Example: **nucleus** part_of **cell** ; meaning that nucleus are always part of a cell, but not all cells have nucleus.

The Gene Ontology Annotation database (GOA)

- The Gene Ontology Annotation database (GOA) contains a list of associations between UniProtKB identifiers and GO terms.
- But, only 558,681 protein sequences in UniProtKB have an experimentally determined annotation.
- As these annotations come from various labs and genome annotation consortia, neither the proteins nor the GO terms are studied uniformly.
- Experimental annotations, which usually describe a protein function in part or at a high level, are expensive to obtain, rare, and collected with bias.

Outline

Introduction

Background / Related works

MAPPIN

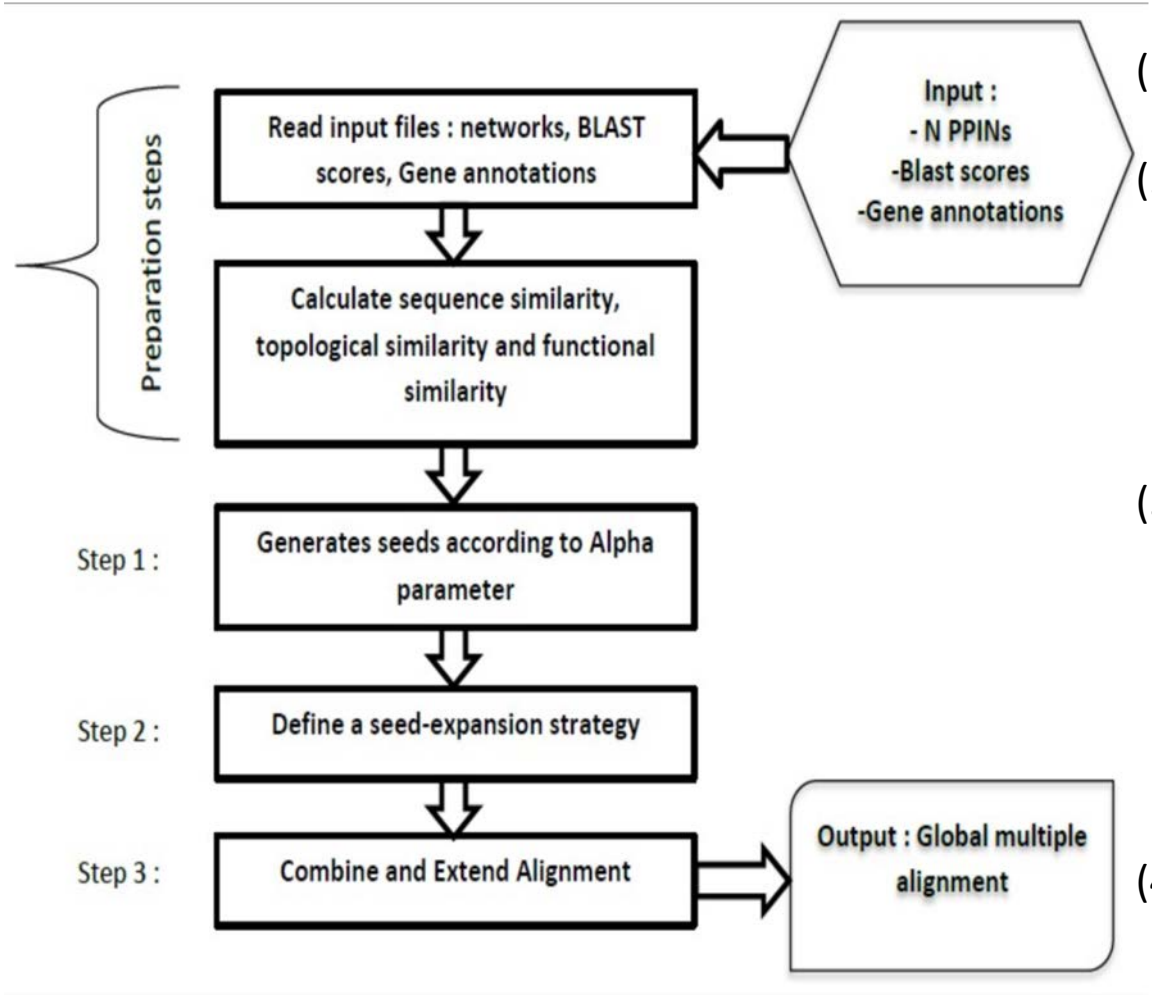
Experimental results

Conclusion

MAPPIN (GOA + PPI)

- MAPPIN uses sequence similarity together with the Gene Ontology Annotation (GOA) of proteins to incorporate functional similarity between the proteins and perform the matching among the proteins of different species.

Workflow of our approach



Our approach in four major steps:

(1) Parsing the n PPI networks;

(2) Giving a calculated weight to each edge in the bipartite graphs using the information in the GOA (Gene Ontology Annotation) and sequence level for each aligned protein;

(3) Collecting seed with high similarity scores from the bipartite graphs, each seed is expanded in an iterative fashion by exploring the local neighborhood for each compared protein;

(4) Finally, MAPPIN applies a simulated annealing (SA) function in order to find a global alignment.

Workflow of our approach

Input: $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, alpha
Output: Biological score Matrix $\grave{BM}$
for all $p_i \in V_1$ **do**
  for all $p_j \in V_2$ **do**

$$s_{seq}(p_i, p_j) \longleftarrow \frac{BLAST(p_i, p_j)}{\sqrt{BLAST(p_i, p_i) \times BLAST(p_j, p_j)}};$$

$$s_{funct}(p_i, p_j) \longleftarrow s_{Schlicker}(p_i, p_j);$$

$$\grave{BM}_{ij} \longleftarrow \alpha s_{seq}(p_i, p_j) + (1 - \alpha) s_{funct}(p_i, p_j);$$

  end for
end for
return $\grave{BM}$

SimilarityScore (G_1, G_2, α)

Input: Set of network $G_1(V_1, E_1)$, $G_2(V_2, E_2)$ ... $G_k(V_k, E_k)$, α, τ, K, T_{min}, T_{max}, s

Output: A set of global Multiple match-sets

Initialize $V^* = \emptyset$

Initialize $E^* = \emptyset$

$\Omega \longleftarrow \emptyset$

$A \longleftarrow \emptyset$

for $i = 1$ to k **do**

  for all all remaining networks G_j **do**

    $GP_{(ij)} \longleftarrow pairwiseAlignment(G_i, G_j, \alpha, \tau)$ ▷ Create node alignment

    for each node of G_i, $v \in V_i$ **do**

      $VertexCluster(v) = \{v\}$

      for each each pairwise alignment $GP_{(ij)}$ **do**

        $VertexCluster(v) = VertexCluster(v) \cup VertexCluster_{ij}(v)$

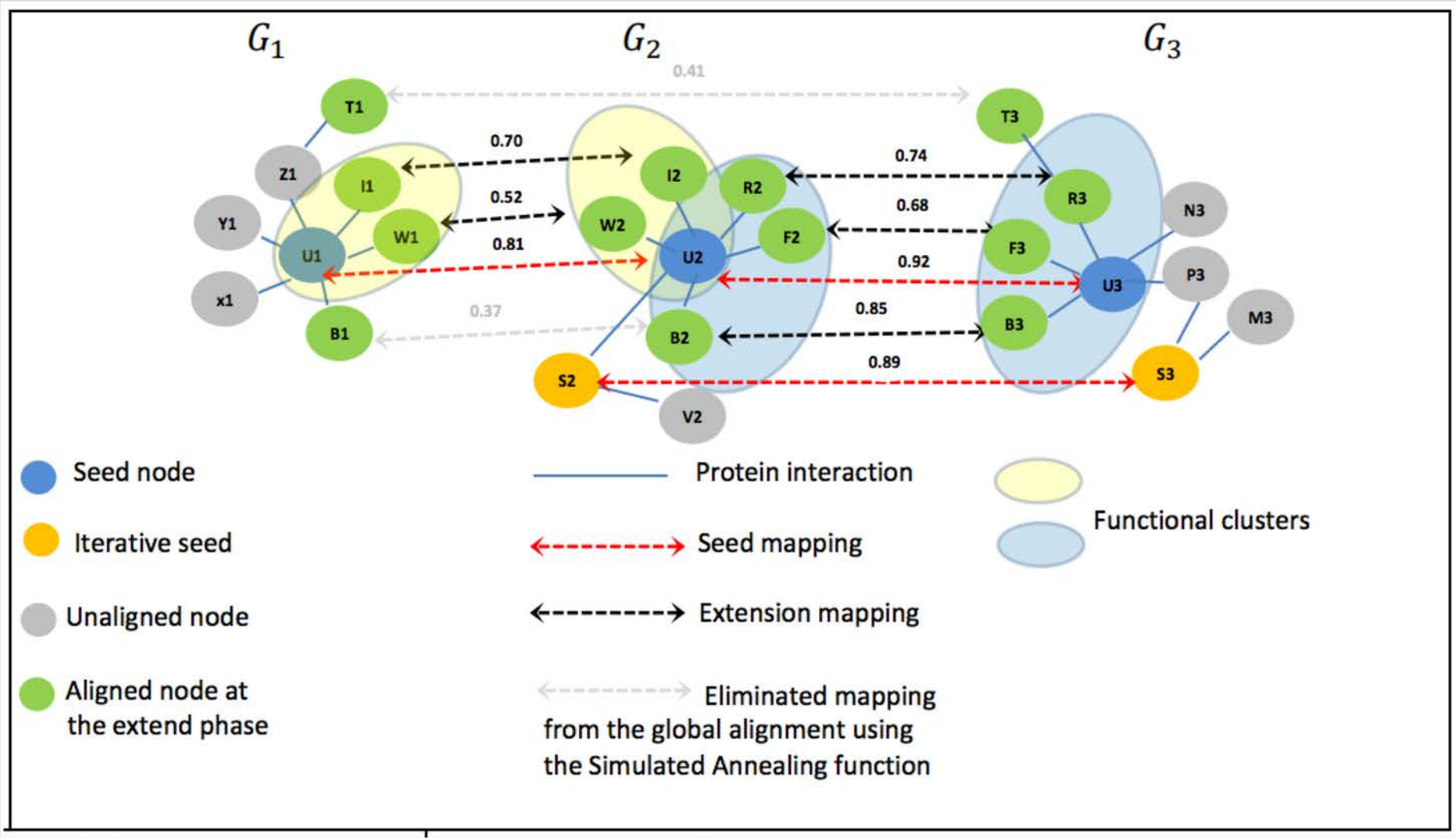
      end for

      ▷ Concatenate sets

      $V^* = V^* \cup V^*.VertexCluster(v)$

end for
for each edge of G_i, $(u,v) \in E_i$ **do**
$EdgeCluster(u,v) = \{(u,v)\}$
for each pair $(k,l) \in VertexCluster(u) \times VertexCluster(v)$, $(u,v) \in E_i$ **do**
if (k,l) form an edge **then**
$EdgeCluster(u,v) \longleftarrow EdgeCluster(u,v) \cup (k,l)$
end if
end for ▷ Concatenate sets
$E^* = E^* \cup E^*.EdgeCluster(u,v)$
end for
end for
end for
$\Omega \longleftarrow Seed - Expansion(E^*, V^*)$ ▷ Generation a feasible solution with a set of mutually disjoint match sets. The parameters K, T_{min}, T_{max} and s control the SA
$A \longleftarrow Simulated - annealing(\Omega, K, T_{min}, T_{max}, s)$
return A

MAPPIN : Example



MAPPIN : Example

1- Building the three PPI networks:

G1.net

$$\begin{Bmatrix} U1-W1 \\ U1-I1 \\ U1-B1 \\ U1-Z1 \\ U1-Y1 \\ U1-X1 \\ Z1-T1 \end{Bmatrix}$$

G2.net

$$\begin{Bmatrix} U2-F2 \\ U2-R2 \\ U2-I2 \\ U2-W2 \\ U2-B2 \\ U2-S2 \\ S2-V2 \end{Bmatrix}$$

G3.net

$$\begin{Bmatrix} U3-P3 \\ U3-N3 \\ U3-R3 \\ U3-F3 \\ U3-B3 \\ R3-T3 \\ P3-S3 \\ S3-M3 \end{Bmatrix}$$

2- Bipartite graphs: Assigning a weight for each interaction:

G1-G2.net

$$\begin{Bmatrix} U1-U2=0.81 \\ I1-I2=0.70 \\ W1-W2=0.52 \\ B1-B2=0.37 \\ T1-I2=0.12 \\ Z1-R2=0.09 \\ Y1-W1=0.04 \\ \dots \end{Bmatrix}$$

G1-G3.net

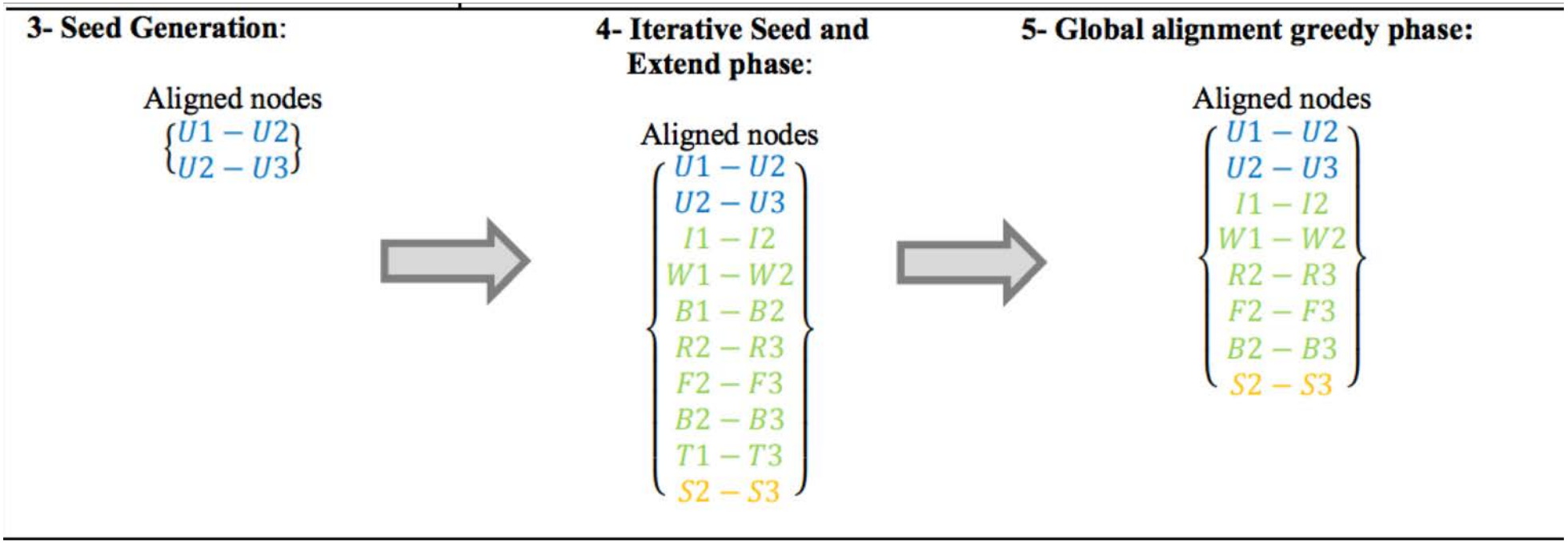
$$\begin{Bmatrix} T1-T3=0.41 \\ Y1-P3=0.14 \\ X1-M3=0.06 \\ \dots \end{Bmatrix}$$

G2-G3.net

$$\begin{Bmatrix} U2-U3=0.92 \\ R2-R3=0.74 \\ F2-F3=0.68 \\ B2-B3=0.85 \\ S2-S3=0.89 \\ S2-N3=0.17 \\ V2-N3=0.02 \\ \dots \end{Bmatrix}$$

=> MAPPIN takes into consideration mapping pairs greater than the threshold fixed at 0.3 for example

MAPPIN : Example



Theoretical Time Study

- Suppose we have k networks, where :
  - the maximum network size is $\boldsymbol{n = max_i|V_i|}$,
  - the maximum number of interactions in a network is $\mathbf{m} = \boldsymbol{max_i|E_i|}$.

- **Suppose there is a bipartite graph**, $B_s = (V_{s1} \cup V_{s2},\ \ E_s)$
  - the running time complexity on B_s is about $O(|V_{s1} \cup V_{s2}|.log\lceil E_s \rceil)$.

- So, the collection of candidate edge costs $\binom{\boldsymbol{k}}{\boldsymbol{2}} \boldsymbol{O(nlog(n))}$ time.

- Running the Simulated Annealing only depends of two parameters of the cooling scheme, K and N, which are independent of the number of compared species k.

Summary

Aligner	Time	Pairwise	Multiple	GNA	LNA	Year
PROPER	?	×	-	×	-	2016
PINALOG	?	×	-	×	-	2012
IsoRank	$O(n^4)$	×	-	×	-	2007
IsoRankN	$O(n^k)$	×	×	×	-	2009
SMETANA	$O(nk^3)$	×	×	×	-	2013
NetCoffee	$O(knlog(n))$	-	×	×	-	2014
MAPPIN	$\binom{k}{2} O(nlog(n))$	×	×	×	-	2018
BEAMS	?	×	×	×	-	2013

Outline

Introduction

Background / Related works

MAPPIN

Experimental results

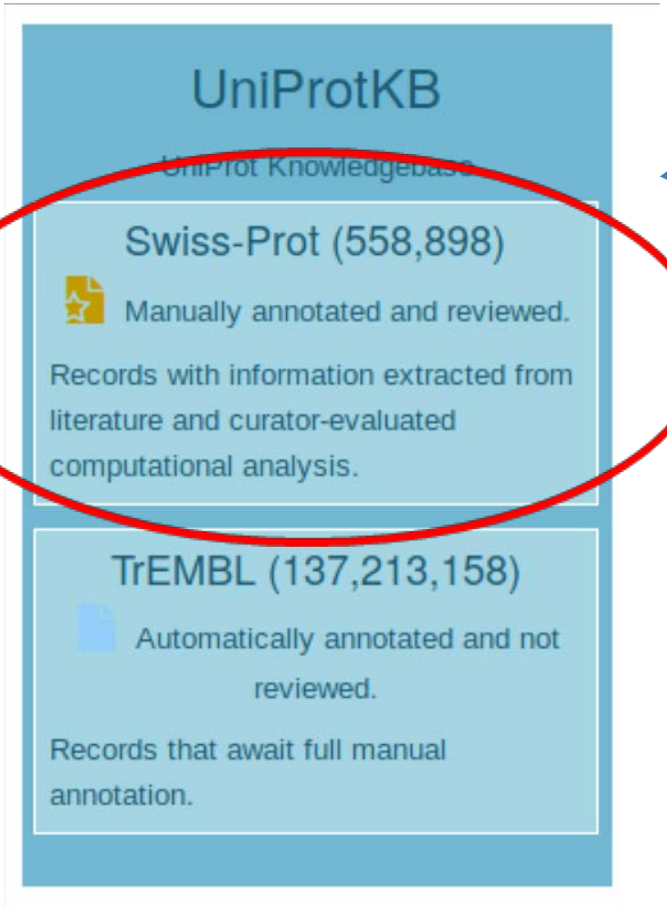
Conclusion

Data sets

Characteristics of the PPI Networks and Datasets from Eight Species

Species	Proteins	Interactions	D1	D2	D3	D4	D5
Arabidopsis	2651	5235					×
C.elegans	4305	7746	×	×	×	×	×
D.melanogaster	8374	25610	×	×	×	×	×
E.coli	2818	13841				×	×
H.sapiens	9003	34935	×	×	×	×	×
M.musculus	2897	4372		×	×	×	×
Rat	1150	1305		×	×	×	×
S.cerevisiae	5674	49830			×	×	×

Quality Validation



Validation on Swiss-Prot Database

Measures :

- Coverage : percentage of proteins in the whole set of proteins that are covered

- Mean Entropy

$$H(A_1) = -\sum_{i=1}^{d} p_i \log p_i$$

p_i = fraction of A_1 with GO term i

d = number of GO terms in each cluster

- Mean Normalized Entropy

$$\bar{H}(A_1) = \frac{1}{\log d} H(A_1).$$

- Runtime

Evaluation |Results|

Default parameters

Measure	MAPPIN	NetCoffee	SMETANA	BEAMS
D1 (Multiple Alignment)				
CV(%)	**73.8**	59.6	72.9	71.8
ME	0.324	**0.128**	0.274	0.231
MNE	0.233	**0.13**	0.256	0.231
Time	39mn	**15s**	50s	20s
D2 (Multiple Alignment)				
CV(%)	73.6	60.8	**74.6**	71.8
ME	0.368	**0.196**	0.312	0.283
MNE	0.256	**0.195**	0.276	0.253
Time	42mn	**45s**	225s	1h30
D3 (Multiple Alignment)				
CV(%)	63.4	53.4	**64.5**	63.4
ME	0.411	**0.264**	0.381	0.326
MNE	0.283	**0.251**	0.294	0.286
Time	42mn	**57s**	321s	3h
D4 (Multiple Alignment)				
CV(%)	60.8	52.7	63	**61.4**
ME	0.393	**0.246**	0.351	0.526
MNE	0.273	**0.241**	0.297	0.392
Time	44mn	**2.45s**	521s	≈8h
D5 (Multiple Alignment)				
CV(%)	**59.8**	53.2	-	58.3
ME	0.384	**0.248**	-	0.264
MNE	0.27	**0.242**	-	0.27
Time	44mn	**3.41s**	-	≈13h

Discussion |Results|

- MAPPIN algorithm can occasionally be efficient in terms of CV, ME and MNE across all cases, showing that it can accurately align real PPI networks.
- For D1 and D5 datasets, MAPPIN outperforms its competitors in terms of CV. On average, our approach provides an acceptable lower entropy values.
- NetCoffee also shows good performance on the all datasets, with a slightly lower CV and achieves entropy scores lower than all the compared approach.
- In addition, SMETANA gives a good coverage for all the five datasets, but it couldn’t align the dataset D5.
- For D4 and D5 datasets, BEAMS struggles to provide a coherent alignment in a reasonable time.

Discussion |Results|

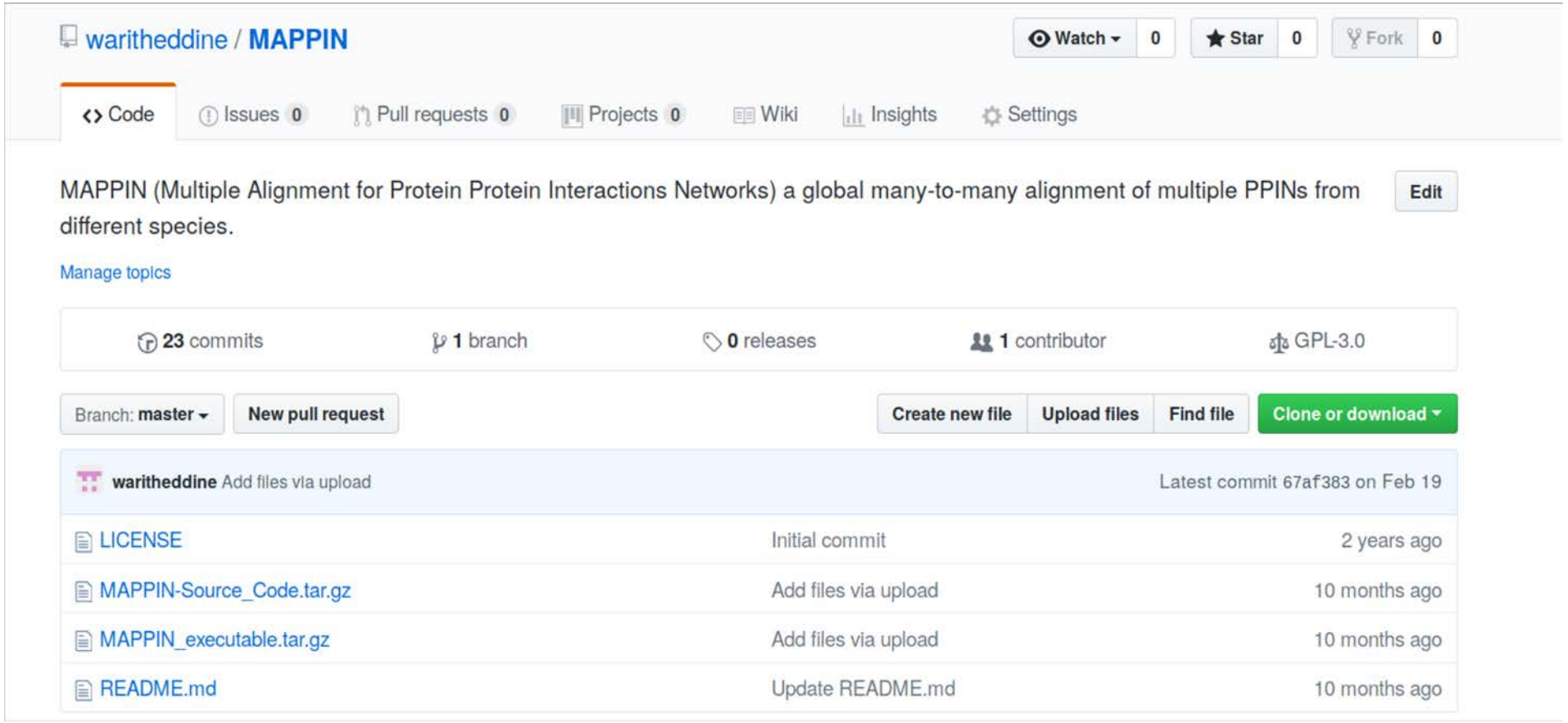
MAPPIN gives encouraging results in terms of coverage and consistency compared to its competitors.

Indeed, these results stand on the incompleteness of the GO annotation of proteins. In addition, the assignment of more and less specific annotation terms, for each protein, also has a negative impact on the accuracy of the produced alignments.

Moreover, the high number of unannotated protein isoforms, that have considerably different functions, often play radically different roles within tissues and cells, leads to worse biological alignment quality.

Availability

- https://github.com/waritheddine/MAPPIN



Outline

Introduction

Background / Related works

MAPPIN

Experimental results

Conclusion

Conclusion

MAPPIN	NetCoffee
It aligns two or more PPI networks	It aligns 3 networks or more, so it can not align two networks.
The topological similarity is used for the detection of hubs and in phase of Seed Expansion	Topological similarity is based on the T-Coffee approach.
It includes the functional similarity during the alignment process from the Gene Ontology Annotation (GOA) collected from UniProt-GOA	It doesn't rely on functional similarity. The Gene Ontology, used after the process of the alignment in order to test the coherence of the alignments.
It rigorously combines protein sequence similarity, network topology similarity and functional similarity (using GO) into a suitable scoring scheme for aligning *k* multiple networks.	It rigorously combines protein sequence similarity and network topology similarity for aligning *k* multiple networks.

Conclusion

- ✓ MAPPIN : an effective method for PPI network alignment.
  - ✓ Test on the five eukaryotic species.
  - ✓ Results consistent with existing approaches,
    - ✓ lead to better functional predictions.
- ✓ Shortcomings :
  - ✓ Runtime with GO Annotations
  - ✓ Changes (temporal, …) on alignment
    - ✓ Evolving alignment, Dynamics
  - ✓ …

A Novel Computational Approach for Global Alignment for Multiple Biological Networks

3 Author(s) Warith Eddine Djeddi ; Sadok Ben Yahia ; Engelbert Mephu Nguifo View All Authors

Abstract:
Due to the rapid progress of biological networks for modeling biological systems, a lot of biomolecular networks have been producing more and more protein-protein interaction (PPI) data. Analyzing protein-protein interaction networks aims to find regions of topological and functional (dis)similarities between molecular networks of different species. The study of PPI networks has the potential to teach us as much about life process and diseases at the molecular level. Although few methods have been developed for multiple PPI network alignment and thus, new network alignment methods are of a compelling need. In this paper, we propose a novel

Related Articles

Detecting Essential Proteins Based on Network Topology, Gene Expression Data, and Gene Ontology Information
IEEE/ACM Transactions on Computational Biology and Bioinformatics
Published: 2018