Learning Heat Diffusion for Network Alignment

Sally Sisi Qu, Frost Mengmeng Xu, Bernard Ghanem, and Jesper Tegner
King Abdullah University of Science and Technology

The 2020 ICML Workshop on Computational Biology

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology
Living Systems Laboratory
VCC VISUAL COMPUTING CENTER

Contributions

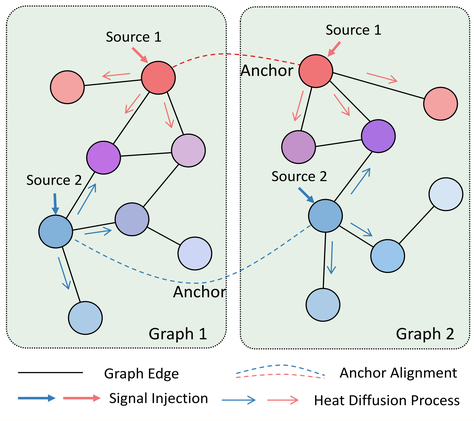
EDNA: Evolutionary heat Diffusion- based Network Alignment

- uses the diffusion signal as a proxy for computing node similarities between networks
  - ✓ the most **accurate** alignments
  - ✓ increased **robustness** against noise
  - ✓ superior **scalability**
- works as a wrapper around available network alignments/embeddings

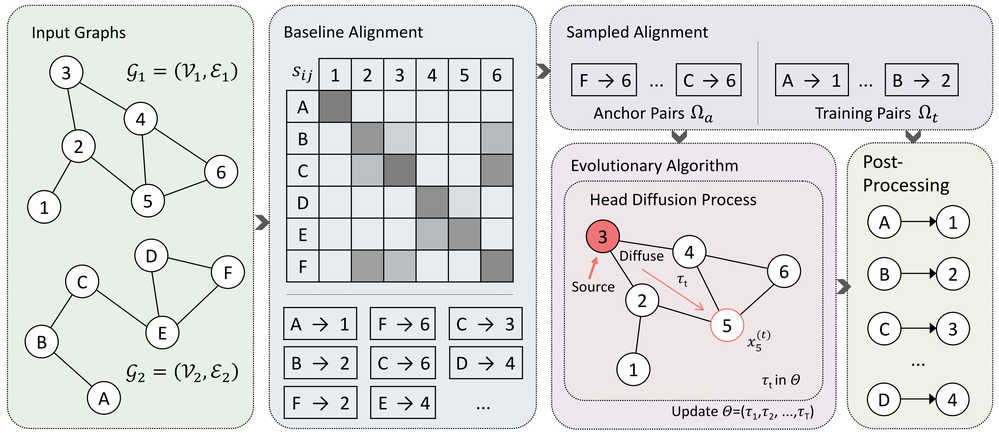
Overview

- **Protein protein interaction networks**: proteins as nodes, interactions as edges
- **Local and global structure**: signals of anchor nodes from baseline alignments diffuse to all nodes
- **Evolutionary algorithm**: fine-tune the diffusion parameter Θ to further improve performance

Graph Formulation



Architecture



$$s_{ij} = \exp(-\|\boldsymbol{e}_i - \boldsymbol{e}_j\|_2^2)$$

$$\forall \nu_i \in \mathcal{G}_1, \boldsymbol{x}_i^{(0)} = \begin{cases} \boldsymbol{f}_k, & \exists \Omega_a(k) = (i,j) \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

$$\forall \nu_j \in \mathcal{G}_2, \boldsymbol{x}_j^{(0)} = \begin{cases} \boldsymbol{f}_k, & \exists \Omega_a(k) = (i,j) \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

$$\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)} + \sigma(\tau_t(\mathbf{D} - \mathbf{A})\mathbf{X}^{(t)})$$

$$f(\Theta) = \frac{1}{\beta} \sum_{(i,j)\in\Omega_t} \mathbf{1}\{j = \underset{l}{\operatorname{argmin}}(\|\boldsymbol{x}_i^{(T)} - \boldsymbol{x}_l^{(T)}\|_2^2)\}$$

Performance

Method	Acc@1	Acc@5	*EC*	S^3
Baseline				
Struc2vec	63.9	88.7	78.6	75.7
REGAL	59.6	77.4	49.3	46.3
ndegree	65.0	80.9	63.1	60.3
After EDNA				
Struc2vec	73.4	95.4	55.4	55.6
REGAL	92.6	96.5	97.7	98.5
ndegree	92.6	96.7	98.1	98.8

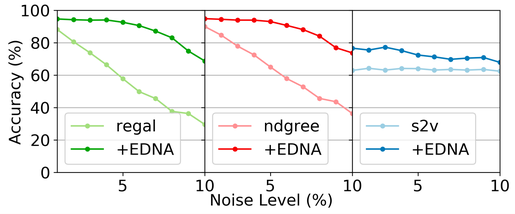
Accuracy of baseline alignments and EDNA with varying noise levels from 1% to 10%.

$$\mathbf{P}\text{: } \mathbf{A}_1^{perm} = \mathbf{P}\mathbf{A}_1\mathbf{P}^\top$$

The implementation with EDNA consistently outperforms baselines using different features from (struc2vec and ndegree) and NA method (REGAL).

Accuracy of baseline alignments and EDNA for fixed 5% noise level.

- **Accuracy**: the ratio of the number of correct alignments to the total number of groudtruth alignments
- **EC**: **edge correctness,** the percentage of edges in one network aligned to the edges in another
- **S³**: **symmetric substructure score**. Penalize the alignments that map sparse regions of the network to denser ones.



Ablation Study

Node Number	500	1000	2000	4000	8000
Top 1 Acc.	67.0	65.4	62.0	58.2	56.2
Top 5 Acc.	76.0	70.8	67.5	62.5	58.4
CPU (mean)	17.8	52.5	193	795	3041
CPU (std)	1.53	10.5	12.4	62.1	18.3
GPU (mean)	4.85	10.9	34.5	135	272
GPU (std)	0.46	0.17	0.59	3.7	1.4

Scalability of EDNA with network size.

- Erdos-Renyi networks a constant average degree of 10

Observation:

- With increasing network size the alignment accuracy decreases slightly, while the running time on CPU or on GPU increases.
- However, the simulation of the heat diffusion process on a GPU is much faster.

Method	Acc@1	Acc@5	*EC*	S^3
ndegree	35.63	57.09	27.77	23.53
Diffusion	52.75	73.54	56.96	56.26
EDNA	77.89	86.58	70.22	71.29

Ablation study of EDNA.
ndegree as a baseline alignment method to compare with the performance after applying diffusion and full EDNA. p_s is set to 10%.

Future Work