Zhong *et al. BMC Bioinformatics* 2020, **21**(Suppl 13):385
https://doi.org/10.1186/s12859-020-03672-6

BMC Bioinformatics

RESEARCH Open Access

Twadn: an efficient alignment algorithm based on time warping for pairwise dynamic networks

Yuanke Zhong[1†], Jing Li[2†], Junhao He[1], Yiqun Gao[1], Jie Liu[1], Jingru Wang[1], Xuequn Shang[1] and Jialu Hu[1,3*]

From The 18th Asia Pacific Bioinformatics Conference
Seoul, Korea. 18-20 August 2020

*Correspondence:
jhu@nwpu.edu.cn
†Yuanke Zhong and Jing Li contributed equally to this work.
[1]School of Computer Science, Northwestern Polytechnical University, West Youyi Road 127, 710072 Xi'an, China
[3]Centre of Multidisciplinary Convergence Computing, School of Computer Science, Northwestern Polytechnical University, 1 Dong Xiang Road, 710129 Xi'an, China
Full list of author information is available at the end of the article

Abstract

Background: Network alignment is an efficient computational framework in the prediction of protein function and phylogenetic relationships in systems biology. However, most of existing alignment methods focus on aligning PPIs based on static network model, which are actually dynamic in real-world systems. The dynamic characteristic of PPI networks is essential for understanding the evolution and regulation mechanism at the molecular level and there is still much room to improve the alignment quality in dynamic networks.

Results: In this paper, we proposed a novel alignment algorithm, Twadn, to align dynamic PPI networks based on a strategy of time warping. We compare Twadn with the existing dynamic network alignment algorithm DynaMAGNA++ and DynaWAVE and use area under the receiver operating characteristic curve and area under the precision-recall curve as evaluation indicators. The experimental results show that Twadn is superior to DynaMAGNA++ and DynaWAVE. In addition, we use protein interaction network of Drosophila to compare Twadn and the static network alignment algorithm NetCoffee2 and experimental results show that Twadn is able to capture timing information compared to NetCoffee2.

Conclusions: Twadn is a versatile and efficient alignment tool that can be applied to dynamic network. Hopefully, its application can benefit the research community in the fields of molecular function and evolution.

Keywords: PPI network, Dynamic network, Network alignment, Dynamic time warping

Background

In recent years, due to the rapid development of biotechnology, we can obtain a large amount of biological data, such as: gene expression data, methylation data, protein interaction network data and so on [1]. Protein is a substance closely related to life and various forms of life activities. It plays a vital role in almost all life activities. Therefore, research

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

on proteins plays a crucial role in our biological research. Protein is not a single biological function. It usually interacts with other proteins to perform certain biological functions [2–5]. All protein interactions form a protein-protein interaction (PPI) network. Most of networks are dynamic in real-world systems. For instance, PPI could change over time, and online professional network will also evolve over time [6]. A large number of PPIs are transient interactions, which briefly exists in only certain cellular context related with cell types, cell cycle stages etc. However, most of network alignment (NA) methods are designed for static networks [7], since static networks were used to model complex real-world systems. The aim of NA on PPI networks is to find an optimal node mapping that can indicate similar biological meanings between matched proteins. However, these networks actually change over time. The dynamic characteristic of PPI networks is essential for the understanding of evolution and regulation mechanism at the molecular level. Some pioneer works [8] attempt to improve NA quality using dynamic network model on evolving systems. This new computational framework can use dynamic characteristic as a supplementary information in the measure of node similarity, whereas it also suffers from the lack of high-confidence dynamic networks of real-world systems. Network study consists of a lot of parts, such as KF-finder [9], which can identify key factors from host-microbial networks in cervical cancer, besides, detection of network motif [10] is also a major search of network. In this paper, we focus on NA, which can be used to predict protein function by transferring functional knowledge from a well-studied species to a poorly-studied species.

There are two categories of alignment methods according to the target regions of networks: global alignment and local alignments. Global alignment is to find one global node mapping for compared networks [11], while local alignment aims to identify multiple conserved subregions which reflect putative functional modules of biological systems [12]. Alignments of two networks are called pairwise network alignments, those of three or more are termed as multiple network alignments. In this paper, we aim to address the global alignment problem of two dynamic networks. IsoRank was originally proposed to solve pairwise global alignment. It was intuitively guided by the assumption that one protein is a good match for another protein in the other compared network if their neighborhood topologies and sequences are similar. Many more alignment tools were developed to improve the algorithm performance of existing methods over the past decade. Among these, there are NETAL [13], H-GRAAL [14], MAGNA [15], MAGNA++ [16], which can provide one-to-one global node mapping for two compared networks. To find protein match sets for multiple species, IsoRankN [17], NetCoffee [11], SMETANA [18] and multiMAGNA++ [19] were used to find one global node mapping for multiple PPI networks. All these algorithms focused on aligning protein pairs based on static networks, although these networks evolve over time. DynaMAGNA++ [8] and DynaWAVE [20] were recently proposed to make up this deficiency. DynaMAGNA++ is the first dynamic NA algorithm, which was adapted from the MAGNA++ method. DynaMAGNA++ takes two measures (node conservation and edge conservation) to capture functionally conserved proteins. However, there is still much room to improve the alignment quality in dynamic networks. It is still a challenge to solve the alignment problem for dynamic networks.

To overcome these issues, we proposed a novel NA algorithm based on a technique termed as dynamic time warping (DTW) to align dynamic PPI networks across species. A

5-tuple-feature vector was calculated on each node of each time snapshot. A target scoring function was used to evaluate the quality of NA, which integrates both topology and sequence information. Then, the alignment problem is transformed into an optimization problem. Simulated annealing was applied to iteratively search for a near-optimal global node mapping between two compared networks.

Methods

The Twadn algorithm returns the optimal alignment results over two given dynamic networks. A dynamic network can be seen as a series of static networks based on a time sequence. So the structure feature of each static network can be extracted by a traditional static NA. In our program, one of our previous work NetCoffee2 [21] was applied to extract the topological feature of each node in the network. Then we can get a sequence of features of each vertex in the dynamic network. Simulated annealing algorithm was used to search for a near-optimal solution. Twadn's algorithm framework is shown in Fig 1. and it has four major steps: 1) perform pair-wise sequence alignment for all pairs of proteins, and select the similar pairs which are statistically significant; 2) extract the topological feature of each vertex in each of the static network using NetCoffee2; 3) calculate the dynamic time warping similarity of all pairs of proteins; 4) use simulated annealing algorithm to find an optimal NA.

Sequence-based similarity

We use the open source tool BLASTP [22] to sequence alignment of all proteins in the network, and obtain the sequence similarity scores e-value and bit-score for each protein pair. Considering that the amino acid sequence that affects protein function may be just a functional region of the sequence, we use the e-value parameter for preliminary filtering and select those protein pairs with e-value less than 1e-7 as the e-value can affect the coverage of predicted homologous proteins by the NetCoffee2 algorithm. Note that Ω denotes the candidates of homology proteins. Given a protein pair u and v, the sequence similarity $s_h(u, v)$ can be calculated in the following formula:

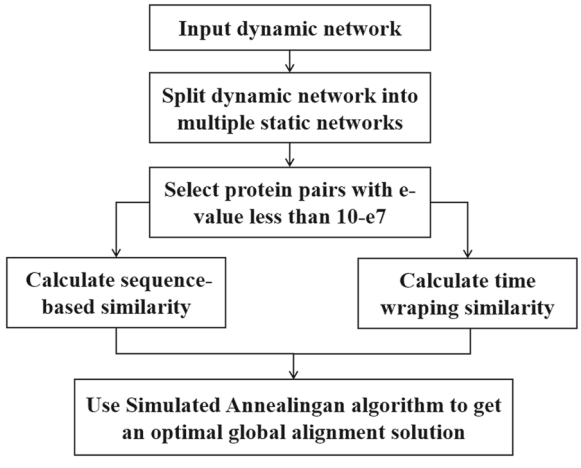

Fig. 1 Algorithm framework of Twadn

$$s_h(u,v) = \frac{\varepsilon(u,v) - \varepsilon_{min}(u,v)}{\Delta\varepsilon} \tag{1}$$

Here, $\varepsilon(u,v)$ can be $-log(e-value)$ or bit-score of the protein pair u and v, $\Delta\varepsilon$ is the largest difference between any two pairs of homolog in Ω, and $\Delta\varepsilon = \varepsilon_{max}(u,v) - \varepsilon_{min}(u,v)$, which servers as a normalization factor. The similarity values calculated by this method are in the interval [0, 1], where 0 represents the least similar protein pair and 1 represents the most similar protein pair.

5-tuple-feature vector of every vertex

Dynamic networks can be regarded as a series of static networks at many snapshots. Here, we attempt to construct a 5-tuple-feature $(\gamma, \sigma, \tau, \eta, \theta)$ for each node in the static network to represent local connectivity of its corresponding node. We denote the adjacent matrix of a netwrok G as $M_{n\times n}$. Since M is real and symmetric, it must has a major normalized eigenvector $K = (k_1, k_2, ..., k_n)$, which is the normalized eigenvector of the largest eigenvalue. Then, we use k_i, $1 \le i \le n$ as the reputation of the node v_i while the greater the reputation is, the more important the node is. Therefore, we use k_i as the first element of the 5-tuple-feature vector (i.e. γ) for node v_i. The set of neighbors of v is denoted as N_v. Then, we use $|N_v|$ as the second element of the 5-tuple-feature vector (i.e. σ), the sum of the reputation of these nodes $\sum_{x\in N_v} k_x$ as the third element(i.e. τ). Let us denote nodes that are 2-step away from v as N_v^2 and all nodes in N_v^2 are not directly connected to v. Then, we use $\left|N_v^2\right|$ as the fourth element (i.e. η). The last element η is calculated by the following formula:

$$\frac{1}{2}\sum_{x\in N_v^2} k_x p_{xv} \tag{2}$$

where p_{xv} presents the number of the shortest paths from x to v.

Dynamic time warping similarity

The input for calculating the dynamic time warping similarity between networks is two dynamic network $DN1$ and $DN2$, and output is a matrix S in which each element s_{ij} represents the time warping similarity between protein i from $DN1$ and protein j from $DN2$. Therefore, for each protein in $DN1$, we need to calculate it's time warp similarity with each protein in $DN2$. Here we carefully explain the calculation method of the time warping similarity between the proteins P and Q of $DN1$ and $DN2$, then through apply this method to all other protein pairs, we can obtain the result matrix S.

Since proteins from different dynamic network may have different snapshot number, the length of each sequence might be different. So, Euclidean distance is unable to be used in measuring the similarity of two given nodes as their differences in sequence length. Suppose P and Q appearing m and n times in their snapshots respectively, so their time sequences can be written as, $p = (p_1, p_2, ..., p_i, ..., p_m), q = (q_1, q_2, ..., q_j, ..., q_n)$. Here, p_i and q_j are the 5-tuple-feature vectors of time sequences p and q at its ith and j^{th} appearance in the snapshot.

DTW is one of the algorithms for measuring similarity between two temporal sequences, which may vary in length and it's proved that DTW is a robust distance measure for time series, allowing similar shapes to match even if they are out of phase in the

time index or different length of time series. So we apply DTW algorithm to calculate the time warping similarity between the sequences p and q.

To align two sequence using DTW, we construct a matrix $D_{m\times n}$ in which each matrix element value d_{ij} represents distance between p_i and q_j $\left(i.e.\ d_{ij} = \sum_{k=1}^{5}\left(p_{ik} - q_{jk}\right)^2\right)$, which capture the time feature of protein in the network. Here, we define a warping path to describe the time correspondence between p and q as follow:

$$W = w_1, w_2, ..., w_k, \max(m,n) \le k \le m + n - 1 \tag{3}$$

The form of w_k is (i,j), which represents that this path passes through the lattice corresponding to p_i and q_j. DTW is a typical optimization problem and its purpose is to find the warping path that minimizes the cumulative distance between two sequence as follow:

$$DTW(p,q) = \min_{W} \sum_{i=1}^{k} \delta\left(w_i\right) \tag{4}$$

Here, $\delta(w_k) = d_{ij}$ is the distance between two time series elements of w_k.

The selection of this path needs to meet the following constraints: 1)$w_1 = (1,1)$ and $w_k = (m,n)$; 2) If $w_{k-1} = (i,j)$ and $w_k = \left(i',j'\right)$, then $i <= i' <= i+1, j <= j' <= j+1$. Then, the search for an optimal path can be transformed into a dynamic programming problem. We define the recurrence formula of the cumulative distance of each pair of proteins as follow:

$$\lambda(i,j) = d(p_i, q_j) + min\{\lambda(i-1,j-1), \lambda(i-1,j), \lambda(i,j-1)\} \tag{5}$$

where $\lambda(i,j)$ is the dynamic time warping distance of p and q. With the warping path W, the node similarity of any two nodes p and q in the dynamic network can be calculated with the Gaussian function $s_t(p,q) = exp\left(-\frac{1}{2}[\,\lambda(m,n)]^2\right)$.

Figure 2 shows an example of calculate dynamic time warping similarity of each two nodes in dynamic network $DN1$ and $DN2$. The matrix (A) are the 5-tuple-feature sequence of node a at five snapshots. Red numbers in matrix (B) constitute the warping path, which can be used to calculate the dynamic time warping distance of node a and node B. The matrix (C) contains all dynamic time warping distance of each two nodes in $DN1$ and $DN2$. The matrix (D) contains all dynamic time warping similarity of each two nodes in $DN1$ and $DN2$.

Simulated annealing

To find an optimal NA, Twadn used the simulated annealing technique to search for an approximately optimal solution, which maximizes the sequence similarity and dynamic time warping similarity with an objective function. The objective function writes as $f(A) = \sum_{m\in A} s_m$, where A and m refer to all possible match sets and a match set, respectively. A match set is a putative functional orthologs that could be a group of functionally related proteins. Suppose there is a match set $m = (m_1, m_2, ..., m_v)$, the alignment score of m is:

$$s_m = \sum_{i,j,i\neq j} \alpha s_h(i,j) + \sum_{i,j,i\neq j} (1-\alpha) s_t(i,j), i,j \in \{m_1, m_2, ..., m_v\} \tag{6}$$

where $s_h(i,j)$ and $s_t(i,j)$ is sequence-based similarity and dynamic time warping similarity of protein i and j described above. With this definition, an optimal global alignment

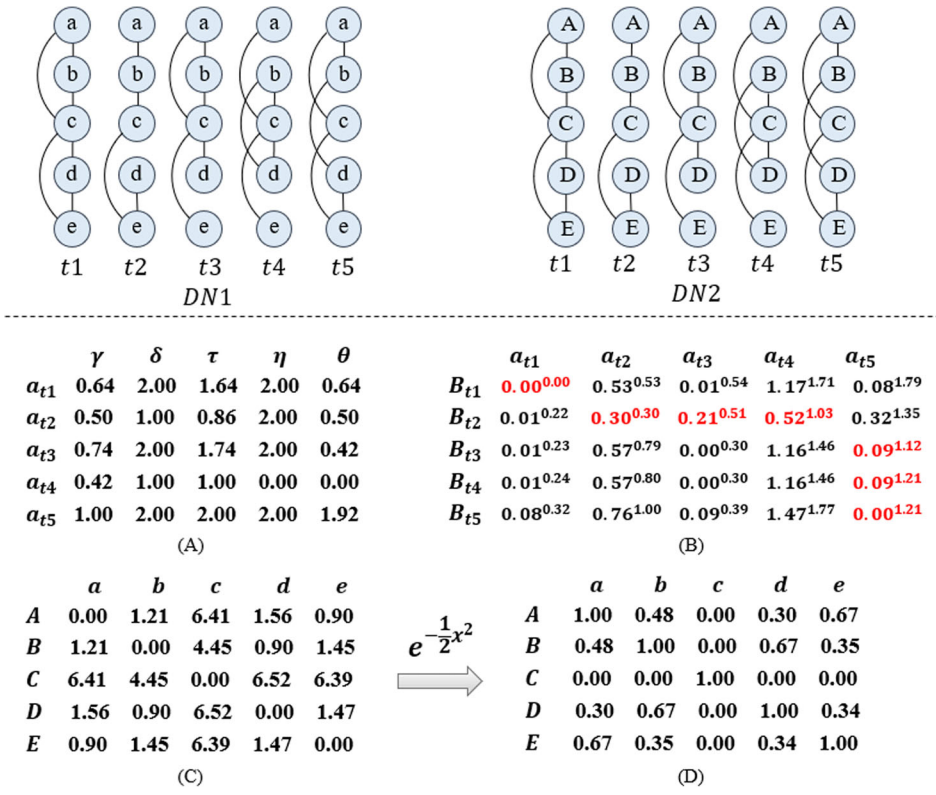

Fig. 2 Dynamic time warping similarity between dynamic network *DN*1 and *DN*2

solution could be solved by maximizing a target function:

$$A^* = arg \max_A f(A) = arg \max_A \sum_{m \in A} s_m \tag{7}$$

Result

To test our method, Twadn was evaluated on both simulated dynamic networks and real-world dynamic networks. We use a simulated dynamic network to compare Twadn, DynaMAGNA++ and DynaWAVE and evaluate the quality of alignment results in terms of area under the precision-recall curve(AUPR) and area under the ROC curve(AUROC). To show the ability of characterizing the time information in dynamic network, Twadn and NetCoffee2 were implemented on two real-world dynamic networks. We aggregated a dynamic network into a static network. In the static version, the network has the same set of nodes as the dynamic network, and a static edge will exist between two nodes if there is at least one edge between the same two nodes in the dynamic network. This kind of dynamic network aggregation was commonly applied in other time series network analyses, such as [23].

Evaluation using synthetic dynamic network

Model of synthetic dynamic network generation

To simulate a network that mimics the evolution of protein-protein interaction networks, we generated the simulated network using a scale free gene duplication network model

[20]. First, a small seed network was given with two connected nodes. Then, we add a node and some edges to network at each step, which simulates the gene duplication and divergence mechanism during the evolution of PPI networks:

Duplication: A node i is selected at random. A new node i', with a link to all the neighbors of i, is created. With probability p a link between i and i' is established.

Divergence: For each of the nodes j linked to i and i', we choose randomly one of the two links (i, j) or (i', j) and remove it with probability q.

Evaluation measures

A good NA approach should be able to produce high quality alignments between networks that are similar, and to produce low quality alignments between networks that are dissimilar [8]. It assumes that networks originating from a model with a same set of parameters should be more similar than these from different parameters [20]. We generate 20 dynamic networks using two sets of parameters $p = 0.3, q = 0.7$ and $p = 0.7, q = 0.6$. Each model generates 10 dynamic networks. We align all possible pairs of the synthetic networks by using Twadn, DynaMAGNA++ and DynaWAVE. Alignment quality of the $C_{20}^2 = 190$ pairs of synthetic dynamic networks can reflect the alignment power of the three algorithms. Alignment tools can classify all pairs of networks based on the alignment score into two categories, similar pairs and dissimilar pairs. With a given threshold ρ, these network pairs with a score $s > \rho$ will be categorized into similar pairs, others will go to dissimilar pairs.

Afterwards, we can compare the performance of all the alignment algorithms using precision-recall and receiver operating characteristic (ROC). The precision is the fraction of network pairs which are true positive among all pairs which have a score $s > \rho$. The recall is the fraction of network pairs which are true positive among all pairs which are true (similar) network pairs. A precision-recall curve can be plotted by adjusting the threshold ρ from 0 to the maximum observed alignment score. The area under the precision-recall curve (AUPR) and F-score are two commonly used measures for the performance of classification methods. The measure F-score is the harmonic mean of precision and recall which can be calculated by following function:

$$F = 2 \times \frac{precision \times recall}{precision + recall} \tag{8}$$

We used the F-score at the point that precision and recall are equal, termed as F-score$_{cross}$ and the score when F-score gets the maximum, termed as F-score$_{max}$. We also evaluated the binary classification according to the receiver operating characteristic (ROC) curve, which was created by plotting the true positive rate (TPR) against the false positive rate (FPR). Here, TPR is the same as recall. FPR is the fraction of false positive network pairs among all false pairs (i.e. dissimilar network pairs). The ROC curve can be plotted by adjusting the criterion ρ. The area under the ROC curve (AUROC) can be calculated after we got the plot.

Performance on synthetic dynamic networks

For synthetic dynamic networks, we aim to develop a NA tool to distinguish these similar network pairs from the other. So we use AUPR, F-score$_{cross}$, F-score$_{max}$, and AUROC to evaluate the quality of the alignment algorithm. As shown in Table 1, the performance

Table 1 Network discrimination performance of DynaMAGNA++, DynaWAVE and Twadn. For biological synthetic networks, with respect to the area under the precision-recall curve (AUPR), F-score at which precision and recall cross and are thus equal (F-score$_{cross}$), maximum F-score (F-score$_{max}$), and the area under the ROC curve (AUROC). In each column, the best score is bolded

algorithm	AUPR	F-score$_{cross}$	F-score$_{max}$	AUROC
DynaMAGNA++	0.467	0.489	0.642	0.507
DynaWAVE	0.600	0.556	0.642	0.594
Twadn	**0.653**	**0.589**	**0.735**	**0.718**

of Twadn outperforms all other alignment algorithms in terms of AUPR, F-score$_{cross}$, F-score$_{max}$, and AUROC, which are 0.653, 0.589, 0.735, and 0.718, respectively. DynaWAVE shows a better performance than DynaMAGNA++ in terms of AUPR, F-score$_{cross}$, and AUROC. From Fig. 3, it shows that the PR curve of DynaMAGNA++ starts from origin of the coordinate system. It means that DynaMAGNA++ failed to discriminate the one with the best alignment score. From Figs. 3 and 4, Twadn is the best aligner according to both the PR curve and the ROC curve over all.

Evaluation using real-world dynamic networks

Experimental design on real-world dynamic networks

To show the alignment capabilities in real-world networks, we evaluated our algorithm on PPI networks. In contrast to static NA tools, Twadn is able to capture dynamic features of a node in the time axis, which would benefit the alignment quality. It's difficult to tell whether two biological networks come from a same evolution model or not because of the ambiguity of evolution model in biology. So we use a randomized (noisy) version of the network (see below). The larger randomized noise level, the more dissimilar the two dynamic network are. We have two randomized versions of the network. One randomizes only the temporal aspect of the network and another randomized both temporal aspect and structure aspect.

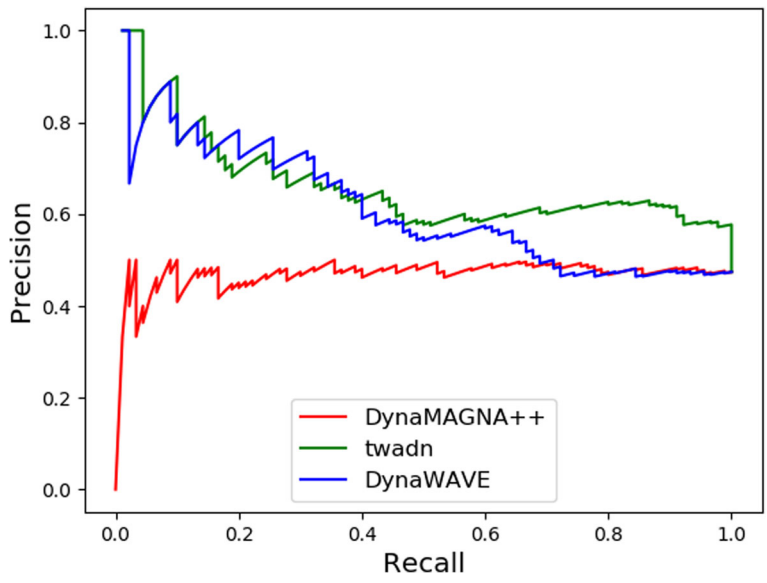

Fig. 3 Network discrimination performance of DynaMAGNA++ and Twadn for biological synthetic networks with respect to precision-recall curve

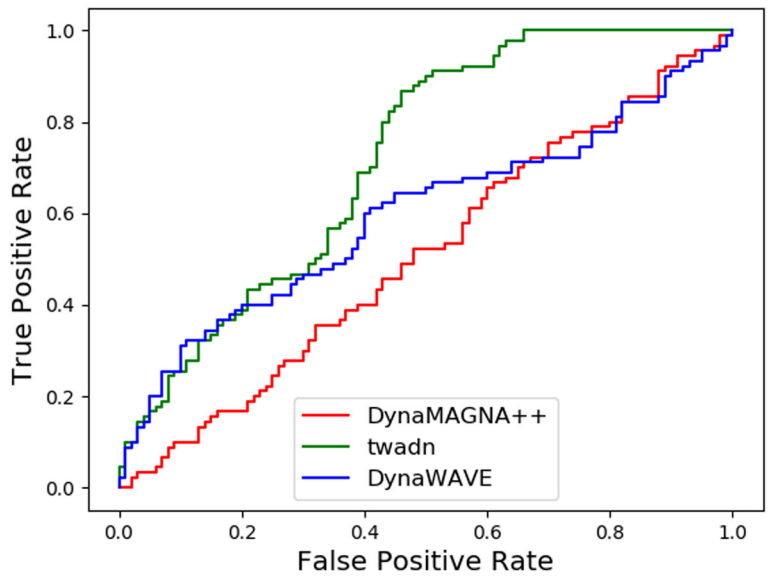

Fig. 4 Network discrimination performance of DynaMAGNA++ and Twadn for biological synthetic networks with respect to ROC curve

Since there is a lack of available experimental dynamic molecular networks, we create a dynamic Drosophila melanogaster PPI network from an artificial temporal sequence of static PPI networks. Here, the static PPI network that are used as snapshots of the dynamic PPI network are all real-world networks, it is just their temporal sequence that is artificial. The sequence consists of seven static PPI network snapshots: at the first snapshot, network have 70% high confidence interactions of original network. at second snapshot, we add 5% high confidence interactions. Now the network have 75% interactions of original which have greater confidence value than the rest interactions. We continue to add 5% interaction until the network has hole 100% interaction of original network. Then we get a real-world dynamic network with seven static PPI network snapshots. Then we generate two randomized versions of network, as follow:

Randomizes only temporal aspect of network: Since the difference between dynamic NA and static NA is that the former accounts for the temporal aspect of the data more explicitly than the latter, we first randomizes only temporal aspect of network, which means randomized network will preserve as much structure as possible of dynamic network. They are only different from each other in time information. This way, the only difference observed between Twadn's and NetCoffee2's performance will be the consequence of considering the temporal aspect of the data. We randomize network with a parameter the noisy level p, the larger the p value, the more noise is added. Given the noise level p and $G = \left(g_1, g_2, ..., g_7\right), g_i = (V_i, E_i)$, for each e_{ab} $(a \in V_i, b \in V_i)$ in E_i , with probability p, we arbitrarily select e_{cd} $\left(c \in V_j, d \in V_j\right)$ from another snapshot $g_j (j \neq i)$ and swap this two interactions. The specific method of swap is to delete the edge e_{ab} in the snapshot g_i and add an edge e_{cd}. At the same time, delete e_{cd} in the snapshot g_j and increase e_{ab}. By randomizing only temporal aspect of network, an dynamic network can be aggregated into the same static network as its noisy version's.

Randomizes both temporal aspect and structure aspect of network: To observe the performance of Twadn and NetCoffee2 in different sets of noisy versions of the

original network using a somewhat more flexible randomization scheme that does not conserve the structure of the flattened version of the original dynamic network, we randomize both temporal aspect and structure aspect of network. Given the noisy level p and $G = \left(g_1, g_2, ..., g_7\right), g_i = (V_i, E_i)$, for each e_{ab} $(a \in V_i, b \in V_i)$ in E_i, with probability p, we arbitrarily select e_{cd} $\left(c \in V_j, d \in V_j\right)$ from another snapshot $g_j (j \neq i)$, if there is no edge connection between node a and d, and there is no edge connection between node b and node c, then connect a and d, b and c. If the process of adding noise creates a loop (i.e., an edge from a node to itself) or a multiple link (i.e., duplicate edge between the same nodes), then we undo it and re-randomly select another edge to do the above process. By randomizing both temporal aspect and structure aspect of network, the resulting compressed version of the static network is different from the original dynamic network.

Evaluation measures

For NA of Drosophila dynamic networks with increased noise, we know the true alignment results (the same protein can be considered a homologous protein). So we use the Alignment Score, which is the algorithm's objective function value (Eq. 3) and node correctness (NC) to measure the network alignment results and the NC can be calculate as follow:

$$NC = \frac{N_{correct}}{N_{all}} \tag{9}$$

where $N_{correct}$ represents the number of correctly aligned protein pairs and N_{all} represents the count of all protein pairs. The greater the node-correctness is, the better the algorithm is. When noise is added to the dynamic network only on the timing information, the dynamic network will become more and more dissimilar to the original network as the level of noise increases, while the dynamic network will be compressed into the same static network. Then as the noise level increases, the comparison result of Twadn will become worse and worse, and the result of NetCoffee2 should not change much. When randomizes both temporal aspect and structure aspect of network, if the dynamic network is compressed to a static version, different static networks will be obtained. As noise level increase, the noise and original versions of dynamic and static networks become more and more dissimilar. Then if Twadn and NetCoffee2 are used to compare the dynamic network with the static network respectively, the comparison results of the two algorithms should be worse as the noise level increases.

Result

As Fig. 5 shows, the alignment score of Twadn decrease with the increase of noisy, while the alignment score of NetCoffee2 does not change when randomizing only temporal aspect. It is reasonable since the flattened version of noisy of dynamic network is the same as original network's. The alignment quality of NetCoffee2 and Twadn both decrease when randomizing both temporal aspect and structure aspect of network in Fig. 6. This illustrates that Twadn really capture the time information of dynamic network compared with NetCoffee2. In other hand, the node-correctness of Twadn is higher than NetCoffee2 in every noisy level, which means Twadn is superior to NetCoffee2. We think it is reasonable as Twadn capture the time information of dynamic network while NetCoffee2 does not.

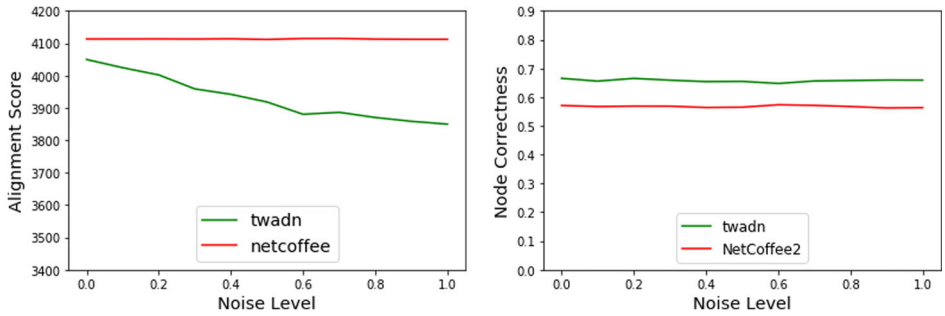

Fig. 5 The performance of node correctness and alignment score when randomizing only temporal aspect of network

Discussion

Although the experimental results have been able to achieve better results than existing algorithms. There are still many problems in the work done in this paper that can be further explored: 1) The algorithm temporarily does not support multiple dynamic networks for simultaneous alignment. However, according to the framework of the simulated annealing algorithm, it is hoped that multiple networks can be aligned at the same time in the future. 2) The algorithm is currently only used in PPI networks, and it is expected to be applied to other networks in the future, such as gene regulatory networks, metabolic networks, etc.

Conclusion

NA is a very important computational framework for understanding molecular function and phylogenetic relationships. Although many NA methods have been developed in the last decade, most of these focused on aligning proteins in static PPI networks. All species and PPIs evolve in different speed. Therefore, there is an urgent demand to develop efficient computational tools to deal with these dynamic networks. To supplement this shortcoming, a novel method Twadn based on the dynamic model of networks was proposed, which can include the time information of molecular interactions. We construct a 5-tupe-feature vector and an optimal warping path to extract topology structures and evolving patterns of all nodes in networks. Twadn was applied in both synthetic datasets and real biological datasets. The synthetic dataset was generated based on a scale-free gene duplication model. Results show that Twadn is superior to DynaMAGNA++ and

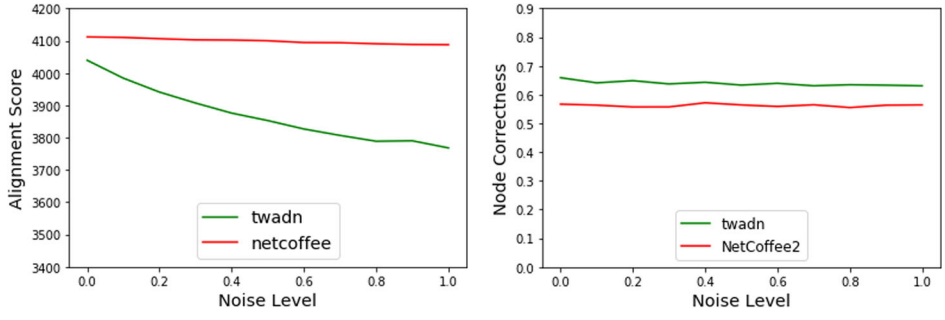

Fig. 6 The performance of node correctness and alignment score when randomizing both temporal aspect and structure aspect of network

DynaWAVE in synthetic network. At the same time, in order to show that the Twadn can capture timing information compared to the static NA, we add timing information and noise to the Drosophila protein interaction network, and then run the Twadn and NetCoffee2, experimental results in line with our expectations. Dynamic NA algorithm do capture timing information. It suggests that Twadn is a versatile and efficient alignment tool that can be applied to dynamic network. Hopefully, its application can benefit the research community in the fields of molecular function and evolution.

Abbreviations

AUPR: Area under the precision-recall curve; AUROC: Area under the receiver operating characteristic curve; DN: Dynamic network; DTW: Dynamic time warping; FPR: False positive rate; NA: Network alignment; NC: Node correctness; PPI: Protein-protein interaction; ROC: Receiver operating characteristic; TPR: True positive rate; Twadn: Time warping based alignment algorithm for pairwise dynamic networks

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 21 Supplement 13, 2020: Selected articles from the 18th Asia Pacific Bioinformatics Conference (APBC 2020): bioinformatics*. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-13.

Authors' contributions

YKZ and JL designed the computational framework, JH and YQG implemented the program. JL, XS and JW performed all the analyses of the data; YKZ and JLiu wrote the manuscript; JLH is the major coordinator, who contributed a lot of time and efforts in the discussion of this project. All author(s) read and approved the final manuscript.

Funding

Publication costs were funded by the National Natural Science Foundation of China (Grant No. 61702420); This project has also been funded by the National Natural Science Foundation of China (Grant No. 61332014, 61702420 and 61772426); the China Postdoctoral Science Foundation (Grant No. 2017M613203); the Natural Science Foundation of Shaanxi Province (Grant No. 2017JQ6037); the Fundamental Research Funds for the Central Universities (Grant No. 3102018zy032); the Top International University Visiting Program for Outstanding Young Scholars of Northwestern Polytechnical University.

Availability of data and materials

The source code of Twadn is freely available at: https://github.com/screamer/twadn.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Computer Science, Northwestern Polytechnical University, West Youyi Road 127, 710072 Xi'an, China. [2]Xi'an Mingde Institute of Technology, Fenghe Campus, Fenghe Campus, 710124 Xi'an, China. [3]Centre of Multidisciplinary Convergence Computing, School of Computer Science, Northwestern Polytechnical University, 1 Dong Xiang Road, 710129 Xi'an, China.

Published: 17 September 2020

References

1. Zhong Y, Li J, Liu J, Zheng Y, Shang X, Hu J. Deep learning enables accurate alignment of single cell rna-seq data. In: Yoo I, Bi J, Hu J, editors. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). New York: IEEE. p. 778–81. https://doi.org/10.1109/BIBM47256.2019.8982969.
2. Hu JL, Wang JR, Lin JA, Liu TW, Zhong YK, Liu J, Zheng Y, Gao YQ, He JH, Shang XQ. Md-svm: a novel svm-based algorithm for the motif discovery of transcription factor binding sites. BMC Bioinformatics. 2019;20:8. https://doi.org/10.1186/s12859-019-2735-3.
3. Hu JL, Gao YQ, Li J, Shang XQ. Deep learning enables accurate prediction of interplay between lncrna and disease. Front Genet. 2019;10:7. https://doi.org/10.3389/fgene.2019.00937.
4. Hu JL, Gao YQ, Li J, Zheng Y, Wang JR, Shang XQ. A novel algorithm based on bi-random walks to identify disease-related lncrnas. BMC Bioinformatics. 2019;20:569. https://doi.org/10.1186/s12859-019-3128-3.
5. Hu J, He J, Li J, Gao Y, Zheng Y, Shang X. A novel algorithm for alignment of multiple ppi networks based on simulated annealing. BMC Genomics. 2019;20(Suppl. 13):932. https://doi.org/10.1186/s12864-019-6302-0.
6. Oentaryo LEAXJSea RJ. Talent flow analytics in online professional network. Data Sci Eng. 2018;3(3):199–220. https://doi.org/10.1007/s41019-018-0070-8.

7. Hu JL, Reinert K. Localali: an evolutionary-based local alignment approach to identify functionally conserved modules in multiple networks. Bioinformatics. 2015;31(3):363–72. https://doi.org/10.1093/bioinformatics/btu652.
8. Vijayan V, Critchlow D, Milenkovic T. Alignment of dynamic networks. Bioinformatics. 2017;33(14):180–9. https://doi.org/10.1093/bioinformatics/btx246.
9. Hu JL, Gao YQ, Zheng Y, Shang XQ. Kf-finder: identification of key factors from host-microbial networks in cervical cancer. BMC Syst Biol. 2018;12(4):54. https://doi.org/10.1186/s12918-018-0566-x.
10. Hu JL, Shang XQ. Detection of network motif based on a novel graph canonization algorithm from transcriptional regulation networks. Molecules. 2017;22(12):9. https://doi.org/10.3390/molecules22122194.
11. Hu JL, Kehr B, Reinert K. Netcoffee: a fast and accurate global alignment approach to identify functionally conserved proteins in multiple networks. Bioinformatics. 2014;30(4):540–8. https://doi.org/10.1093/bioinformatics/btt715.
12. Kalaev M, Smoot M, Ideker T, Sharan R. Networkblast: comparative analysis of protein networks. Bioinformatics. 2008;24(4):594–6. https://doi.org/10.1093/bioinformatics/btm630.
13. Behnam N, Ahmadreza K, Somaye H, Seyed Shahriar A. Netal: a new graph-based method for global alignment of protein-protein interaction networks. Bioinformatics. 2013;29(13):1654–62. https://doi.org/10.1093/bioinformatics/btt202.
14. Paufique J, Madec PY, Kolb J, Kiekebusch MJ, Arsenault R, Siebenmorgen R, Downing M, Hibon P, Valenzuela JJ, Haguenauer P. Graal on the mountaintop. In: Adaptive Optics Systems V. vol. 9909. Bellingham: SPIE; 2016. p. 806–20. https://doi.org/10.1117/12.2232826.
15. Saraph V, Milenkovic T. Magna: Maximizing accuracy in global network alignment. Bioinformatics. 2014;30(20):2931–40. https://doi.org/10.1093/bioinformatics/btu409.
16. Vijayan V, Saraph V, Milenkovic T. Magna++: Maximizing accuracy in global network alignment via both node and edge conservation. Bioinformatics. 2015;31(14):2409–11. https://doi.org/10.1093/bioinformatics/btv161.
17. Chung-Shou L, Kanghao L, Michael B, Rohit S, Bonnie B. Isorankn: spectral methods for global alignment of multiple protein networks. Bioinformatics. 2009;25(12):253–8. https://doi.org/10.1093/bioinformatics/btp203.
18. Sahraeian SM, Yoon BJ. Smetana: accurate and scalable algorithm for probabilistic alignment of large-scale biological networks. Plos ONE. 2013;8(7):67995. https://doi.org/10.1371/journal.pone.0067995.
19. Vijayan V, Milenković T. Multiple network alignment via multimagna++. IEEE/ACM Trans Comput Biol Bioinforma. 2018;15(5):1669–82.
20. Vijayan V, Milenkovic T. Aligning dynamic networks with dynawave. Bioinformatics. 2017;34(10):1795–8. https://doi.org/10.1093/bioinformatics/btx841.
21. Hu JL, He JH, Gao YQ, Zheng Y, Shang XQ. Netcoffee2: A novel global alignment algorithm for multiple ppi networks based on graph feature vectors. In: Huang DS, Jo KH, Zhang XL, editors. Intelligent Computing Theories and Application, Pt Ii, Lecture Notes in Computer Science, vol. 10955. Cham: Springer; 2018. p. 241–6.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.
23. Holme P. Modern temporal network theory: a colloquium. Eur Phys J B. 2015;88(9):234. https://doi.org/10.1140/epjb/e2015-60657-4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC